



**UNIVERSIDADE FEDERAL DO TOCANTINS  
CAMPUS UNIVERSITÁRIO DE PALMAS  
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

**MINERAÇÃO DE DADOS UTILIZANDO A BASE DE DADOS DOS  
ATENDIMENTOS DO SISTEMA NACIONAL DE EMPREGO - SINE DE  
PALMAS/TO**

**LUCAS GUILHERME PONTES LIMA**

**PALMAS (TO)**

**2019**

LUCAS GUILHERME PONTES LIMA

MINERAÇÃO DE DADOS UTILIZANDO A BASE DE DADOS DOS  
ATENDIMENTOS DO SISTEMA NACIONAL DE EMPREGO - SINE DE  
PALMAS/TO

Trabalho de Conclusão de Curso II apresentado  
à Universidade Federal do Tocantins para  
obtenção do título de Bacharel em Ciência da  
Computação, sob a orientação do(a) Prof.(a) Dr.  
George Lauro Ribeiro de Brito.

Orientador: Dr. George Lauro Ribeiro de Brito

PALMAS (TO)

2019

LUCAS GUILHERME PONTES LIMA

MINERAÇÃO DE DADOS UTILIZANDO A BASE DE DADOS DOS  
ATENDIMENTOS DO SISTEMA NACIONAL DE EMPREGO - SINE DE  
PALMAS/TO

Trabalho de Conclusão de Curso II apresentado à UFT – Universidade Federal do Tocantins – Campus Universitário de Palmas, Curso de Ciência da Computação foi avaliado para a obtenção do título de Bacharel e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Data de aprovação: 26 / 3 / 2019

Banca Examinadora:

---

Prof. George Lauro Ribeiro de Brito Dr.

---

Prof. Gentil Veloso Barbosa Dr.

---

Prof. Wesley da Costa Arruda Dr.

**Dados Internacionais de Catalogação na Publicação (CIP)**  
**Sistema de Bibliotecas da Universidade Federal do Tocantins**

---

L732m Lima, Lucas Guilherme Pontes.  
Mineração de dados utilizando a base de dados dos atendimentos do Sistema Nacional de Emprego - SINE de Palmas/To. / Lucas Guilherme Pontes Lima. – Palmas, TO, 2019.  
53 f.  
  
Monografia Graduação - Universidade Federal do Tocantins – Câmpus Universitário de Palmas - Curso de Ciências da Computação, 2019.  
Orientador: George Lauro Ribeiro de Brito  
  
1. Mineração de Dados. 2. Atendimento. 3. SINE. 4. Algoritmo Apriori. I. Título

**CDD 004**

---

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer forma ou por qualquer meio deste documento é autorizado desde que citada a fonte. A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184 do Código Penal.

**Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os dados fornecidos pelo(a) autor(a).**

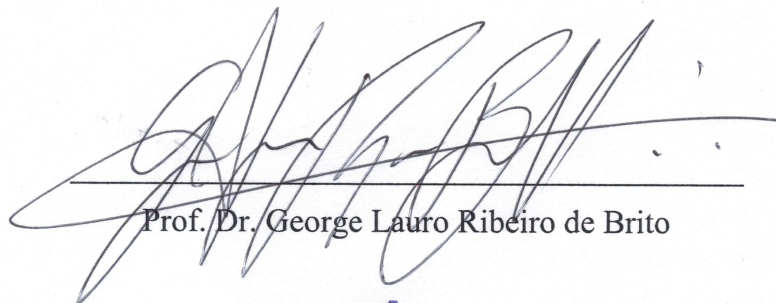
UNIVERSIDADE FEDERAL DO TOCANTINS  
CÂMPUS DE PALMAS  
CURSO DE CIÊNCIA DA COMPUTAÇÃO

109 Norte Av. NS-15, ALCNO- 14, bloco II, Sal. 20 | 77001-090 | Palmas/TO  
(63) 3232-8027 | www.uft.edu.br | compalmas@uft.edu.br



**ATA DE DEFESA DA DISCIPLINA DE PROJETO DE GRADUAÇÃO II**

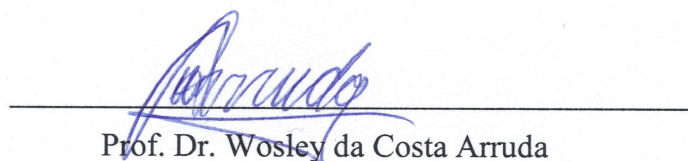
Ao Vigésimo Sexto dia do mês de Março de 2019 realizou-se a defesa de Projeto de Graduação, da disciplina de Projeto de Graduação II do discente **Lucas Guilherme Pontes Lima** do curso de Ciência da Computação do Campus Universitário de Palmas da Universidade Federal do Tocantins (UFT), intitulado “Mineração de dados utilizando a base de dados dos Atendimentos do Sistema Nacional de Emprego - SINE de Palmas/TO”, realizado sob a responsabilidade do Orientador Prof. Dr. George Lauro Ribeiro de Brito. Tendo como Comissão Avaliadora, os professores: Prof. Dr. George Lauro Ribeiro de Brito, Prof. Dr. Gentil Veloso Barbosa e Prof. Dr. Wosley da Costa Arruda, os quais, após avaliação, consideraram o discente APROVADO. Nada mais tendo a constar, assinaram esta Ata os presentes:



Prof. Dr. George Lauro Ribeiro de Brito



Prof. Dr. Gentil Veloso Barbosa



Prof. Dr. Wosley da Costa Arruda

*Aos meus pais, Maria e Lima*

## AGRADECIMENTOS

Agradeço principalmente aos meus pais, por aceitarem quando eu decidi fazer este curso e que sempre me apoiaram durante todos os anos que estive cursando, ansiosos para me ver formado, graças a eles tive forças para chegar aonde estou.

Aos Amigos que conheci durante a graduação e aqueles que sempre estiveram junto a mim tentando vencer os obstáculos.

A minha namorada que esteve comigo a maior parte do tempo durante a graduação.

Ao meu Orientador, professor Doutor George Lauro Ribeiro de Brito pelo apoio nesse projeto e aos professores do curso de Ciência da Computação.

A meus colegas de trabalho e a todos que direta e indiretamente me ajudaram a adquirir conhecimento durante o curso.

## RESUMO

Com a grande procura por serviços relacionados ao trabalho, como procura de vagas de emprego, requerimento do Seguro Desemprego ou mesmo emissão de carteira de trabalho, houve uma necessidade de um Sistema para o atendimento às pessoas que procuram o Sistema Nacional de Emprego (SINE). O Sistema de Gerenciamento de Atendimento (SGA), assim chamado, foi implantado para suprir essa necessidade de organização, melhorar no atendimento e mesmo coleta de dados dos atendimentos prestados pelo SINE, possibilitando atender à população com mais efetividade e rapidez. Com base nisso, foi feito um estudo de caso, através da mineração de dados da referida base de dados do Sistema de Gerenciamento de Atendimento (SGA), para que se possa demonstrar a partir dos resultados possíveis problemas a serem solucionados, ou mesmo melhorias que podem ser feitas a partir disso.

**Palavra-chave:** Mineração de Dados. Atendimento. SINE. Algoritmo Apriori.



## ABSTRACT

With the great demand for services related to work, such as job search, unemployment insurance application or even issuance of a work permit, there was a need for a System for the care of people seeking the National Employment System (SINE) . The Service Management System (SGA), so called, was implemented to meet this need for organization, improvement in the service and even data collection of the services provided by SINE, making it possible to serve the population more effectively and quickly. Based on this, a case study will be done, through the data mining of said database of the Service Management System (SGA), so that it can be demonstrated from the possible results problems to be solved, or even improvements that can be made from it.

**Keywords:** Mineração de Dados. Atendimento. SINE.

## ABREVIATÓES

<b>SINE</b>	Sistema Nacional de Emprego
<b>SGA</b>	Sistema de Gerenciamento de Atendimento
<b>IMO</b>	Intermediação da Mão de Obra
<b>TI</b>	Tecnologia da Informação
<b>IA</b>	Inteligência Artificial
<b>IP</b>	Internet Protocol
<b>FAT</b>	Fundo de Amparo ao Trabalhador

## LISTA DE FIGURAS

Figura 2.1 – Visão Pragmática das etapas do processo de KDD. Retirada de Goldschmidt e Passos (2015). . . . .	19
Figura 2.2 – Visão da etapa seleção/redução do processo de KDD. Retirada de Goldschmidt e Passos (2015). . . . .	20
Figura 2.3 – Utilização do método de Preencher com Mineração. Retirada de Goldschmidt e Passos (2015). . . . .	22
Figura 2.4 – Enriquecimento de Dados no processo de KDD. Retirada de Goldschmidt e Passos (2015). . . . .	23
Figura 3.1 – Fluxo de funcionamento dos atendimentos do SINE. Fonte: SGA (2019). . . . .	27
Figura 3.2 – Tela Inicial de Login SGA. Fonte: SINE-TO (2016). . . . .	28
Figura 3.3 – Tela Gerenciamento do Administrador do SGA. Fonte: SINE-TO (2016). . . . .	28
Figura 3.4 – Tela do Modulo de Atendimento do SGA. Fonte: SINE-TO (2016). . . . .	29
Figura 3.5 – Tela de Triagem para geração de senhas do SGA. Fonte: SINE-TO (2016). . . . .	29
Figura 3.6 – Tela de Estatísticas com Gráfico - SGA. Fonte: SINE-TO (2016). . . . .	30
Figura 3.7 – Tela do Módulo de Prioridades do SGA. Fonte: SINE-TO (2016). . . . .	30
Figura 4.1 – Diagrama do Banco de Dados - SGA. Fonte: SINE-TO (2016). . . . .	32
Figura 4.2 – Gráfico - Total de Registros do atributo Serviço. Fonte: SINE-TO (2016). . . . .	36
Figura 4.3 – Gráfico - Total de Registros do atributo Prioridade. Fonte: SINE-TO (2016). . . . .	37
Figura 4.4 – Gráfico - Total de Registros do atributo Status do Atendimento. Fonte: SINE-TO (2016). . . . .	37
Figura 4.5 – Janela Inicial da Ferramenta Weka. Fonte: (WEKA, 2019). . . . .	38

Figura 4.6 – Janela Principal - Pré processamento da Ferramenta Weka. Fonte: (WEKA, 2019). . . . .	38
Figura 4.7 – Sub-janela de Parâmetros do Algoritmo Apriori na Ferramenta Weka. Fonte: (WEKA, 2019). . . . .	39
Figura 4.8 – Pseudocódigo do algoritmo Apriori. Fonte: Araujo (2016). . . . .	40
Figura 5.1 – Tabela Análise de Resultados 1.1. . . . .	43
Figura 5.2 – Tabela Análise de Resultados 1.2. . . . .	44
Figura 5.3 – Tabela Análise de Resultados 1.3. . . . .	44
Figura 5.4 – Tabela Análise de Resultados 1.4. . . . .	45
Figura 5.5 – Tabela Análise de Resultados 1.5. . . . .	45
Figura 5.6 – Tabela Análise de Resultados 1.6. . . . .	46
Figura 5.7 – Tabela Análise de Resultados 1.7. . . . .	47
Figura 5.8 – Tabela Análise de Resultados 1.8. . . . .	48
Figura 5.9 – Tabela Análise de Resultados 1.9. . . . .	48
Figura 5.10 – Tabela Análise de Resultados 1.10. . . . .	49

## LISTA DE TABELAS

Tabela 2.1 – Tarefas e Técnicas de Mineração de Dados. Retirada de Correa e Machado (2007). . . . .	25
Tabela 3.1 – Tabela de Serviços fornecidos no SGA do <i>SINE</i> . Fonte: SINE-TO (2016). . . . .	31
Tabela 4.1 – Tabela de Serviços do SGA. Fonte: SINE-TO (2016). . . . .	34
Tabela 4.2 – Tabela de Prioridades do SGA. Fonte: SINE-TO (2016). . . . .	34
Tabela 4.3 – Tabela do Status do Atendimento do SGA. Fonte: SINE-TO (2016). . . . .	35
Tabela 5.1 – Tabela Resultado Apriori - Serviço x prioridade. . . . .	41
Tabela 5.2 – Tabela Resultado Apriori - Serviço x Status do Atendimento. . . . .	41
Tabela 5.3 – Tabela Resultado Apriori - Prioridade X Status do Atendimento. . . . .	42
Tabela 5.4 – Tabela Resultado Apriori - Serviço x prioridade x status do atendimento. . . . .	42
Tabela 5.5 – Tabela Resultado Apriori - Serviço x prioridade x status do atendimento. . . . .	43
Tabela 5.6 – Tabela Resultados 5.1. . . . .	47

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>16</b>
<b>1.1</b>	<b>Justificativa</b> . . . . .	<b>16</b>
<b>1.2</b>	<b>Objetivos</b> . . . . .	<b>17</b>
1.2.1	Objetivo Geral . . . . .	17
1.2.2	Objetivos Específicos . . . . .	17
<b>1.3</b>	<b>Organização do Projeto</b> . . . . .	<b>17</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	<b>18</b>
<b>2.1</b>	<b>O Processo KDD (Knowledge Discovery in Databases)</b> . . . . .	<b>18</b>
<b>2.2</b>	<b>As Etapas do Processo KDD</b> . . . . .	<b>19</b>
2.2.1	O Pré-processamento dos Dados . . . . .	20
2.2.1.1	Seleção e Redução dos Dados . . . . .	20
2.2.1.2	Limpeza dos Dados . . . . .	21
2.2.1.3	Codificação dos Dados . . . . .	22
2.2.1.4	Enriquecimento dos Dados . . . . .	23
2.2.2	Mineração de Dados . . . . .	23
2.2.2.1	As Técnicas e Tarefas de Mineração de Dados . . . . .	24
2.2.2.2	Análise de Afinidades . . . . .	25
2.2.2.3	Regras de Associação . . . . .	25
2.2.3	O Pós-processamento dos Dados . . . . .	26
<b>3</b>	<b>O SISTEMA DE GERENCIAMENTO DE ATENDIMENTO (SGA)</b>	
	<b>27</b>	
<b>3.1</b>	<b>Módulo de Atendimento</b> . . . . .	<b>28</b>
<b>3.2</b>	<b>Módulo Triagem</b> . . . . .	<b>29</b>
<b>3.3</b>	<b>Módulo de Estatísticas</b> . . . . .	<b>29</b>

3.4	Módulo de Prioridades . . . . .	30
3.5	Módulo de Serviços . . . . .	31
4	MÉTODOS E O PROCESSO DE KDD NA BASE DE DADOS DO SGA . . . . .	32
4.1	Pré-Processamento dos Dados do SGA . . . . .	33
4.1.1	Fase de Seleção dos Atributos . . . . .	33
4.1.2	Fase de Limpeza e Seleção dos Dados . . . . .	35
4.1.3	Transformação dos Dados - Enriquecimento e Codificação . . . . .	35
4.2	Análise Exploratória de Dados . . . . .	36
4.3	A Ferramenta Utilizada . . . . .	37
5	RESULTADOS ENCONTRADOS . . . . .	41
5.1	Análise dos Resultados . . . . .	43
6	CONCLUSÃO . . . . .	50
6.1	Trabalhos Futuros . . . . .	50
	REFERÊNCIAS . . . . .	52

## 1 INTRODUÇÃO

O Sistema Nacional de Emprego (SINE) é um programa do Ministério do Trabalho e Emprego, viabilizado com recursos do Fundo de Amparo ao Trabalhador (FAT). Tem por objetivo oferecer serviços de: Pagamentos do Seguro Desemprego, emissão de carteira de trabalho e inserção do trabalhador no mercado de trabalho através da Intermediação da Mão de Obra (IMO) (FGTAS, 2019).

Por não ser centralizado, fica a cargo do governo de cada estado o modo como a população passa a ser atendida, sendo por meio da internet ou comparecendo às agências físicas do SINE, como é o caso do SINE de Palmas/TO.

Para melhorar a organização de registro e facilitar o processo de atendimento às pessoas que procuram o SINE, foi implantado o Sistema de Gerenciamento de Atendimento (SGA), que é um sistema open-source disponibilizado para a comunidade geral que deseja utilizar, o qual faz a gerência dos atendimentos através de um painel de senhas e armazena tais informações em um banco de dados interno.

Por apresentar uma quantidade enorme de atendimentos diariamente, é gerado assim um volume enorme de informações. Com isso se torna quase impossível analisar manualmente de forma efetiva tal conjunto de dados, ainda que utilizando ferramentas para tal tarefa. Então com o surgimento da Mineração de Dados isso se tornou possível de realizar. Com isso apareceu também ferramentas para facilitar a análise desse conglomerado de dados, como é o caso da ferramenta Weka, que faz esse trabalho de forma automática utilizando algumas configurações básicas.

A Weka é uma ferramenta/aplicativo que serve para realizar todo o processo de mineração de dados e para isto utiliza por padrão um algoritmo muito conhecido na área de mineração, o Apriori. Entretanto não impossibilita a ferramenta de utilizar outros algoritmos para tal fim. Será ela a escolha para esse Projeto e melhor explicada mais a frente.

### 1.1 Justificativa

Hoje muitas das empresas e órgãos utilizam do setor de Tecnologia da Informação (TI) para otimizar, automatizar e agilizar o serviço ou trabalho. Os sistemas utilizados pelas mesmas gera no decorrer do tempo de utilização um número crescente de informações nos bancos de dados, assim observa-se que surge uma necessidade de explorar o conhecimento escondido dessas informações.

Como já informado, o Apriori utilizado no Weka é um importante algoritmo para essa finalidade, o que se verificou que seria ideal para identificar se possível, quais são os serviços do SINE mais procurados pela população, quanto tempo em média as pessoas



são atendidas, desde a chegada até a finalização do atendimento, ou ainda podendo até mesmo descobrir a relação entre a oferta e demanda de trabalho por meio do serviço de Intermediação da Mão de Obra (IMO) prestado pelo SINE. Tendo em conta todo o processo feito no Weka, no final se possa analisar as informações reveladas, entender a demanda e possíveis problemas a serem solucionados.

## **1.2 Objetivos**

### **1.2.1 Objetivo Geral**

O presente projeto tem o objetivo de fazer a análise de um grande conglomerado de registros oriundos dos atendimentos do SGA. Visando com conhecimento oculto gerado dos resultados obtidos, realizar as devidas melhorias no órgão, como na demanda dos atendimentos ou no próprio SGA se necessário.

### **1.2.2 Objetivos Específicos**

- Aplicar o Processo de Descoberta de Conhecimento através da mineração de dados na base de dados do SGA.
- Identificar possíveis relacionamentos entre atributos da base de dados utilizando ferramenta Weka.
- Analisar e Comparar resultados da pesquisa obtida.
- A partir da análise dos resultados, repassá-los à equipe responsável para que ela possa utilizá-los na melhoria do atendimento do órgão e do Sistema SGA.

## **1.3 Organização do Projeto**

No Capítulo 2, toda a fundamentação desde as etapas do KDD, pré-processamento, seleção, limpeza, codificação, e enriquecimento até o processo central de mineração dos dados e pós-processamento.

No Capítulo 3, se refere ao funcionamento e principais módulos do Sistema de Gerenciamento de Atendimento (SGA), contando ainda com um breve conhecimento de seus atributos e variáveis.

Já no Capítulo 4 se dá início a aplicação do processo de KDD na base de dados do SGA, passando por todo o processo até a chegar na fase de mineração.

No Capítulo 5 apresenta os resultados obtidos durante todo esse projeto, ampliando o conhecimento.

E por último no capítulo 6, se conclui o trabalho, mostrando a importância dele para geração de um novo conhecimento a partir do conhecimento inicialmente oculto.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo aborda os conceitos básicos e necessários para a compreensão deste estudo. Primeiramente será abordado o processo de descoberta de conhecimento, que ao longo desse projeto será chamado de KDD (*Knowledge Discovery in Databases*). Esse processo de KDD é de fundamental importância para se entender o que será feito ao longo do trabalho é ele que seleciona e modela os dados, fazendo verificação dos atributos que podem ser relevantes para a análise.

Subsequentemente virá o tema mineração de dados, já que é nessa etapa que se aplica a descoberta, onde os dados terão maior importância pois serão através deles extraído as informações, classificações dos campos e estudo dos registros .

### 2.1 O Processo KDD (Knowledge Discovery in Databases)

“Aprendizado de máquina é uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas através da solução bem sucedida de problemas anteriores. Os diversos sistemas de aprendizado de máquina possuem características particulares e comuns que possibilitam sua classificação quanto à linguagem de descrição, modo, paradigma e forma de aprendizado utilizado” Monard e Baranauskas (2003).

Existem dois tipos de tarefas de aprendizado de Máquina, o supervisionado e o não supervisionado; este primeiro se subdivide em: Regressão, onde buscam encontrar como uma variável se comporta ou evolui na medida em que outra variável sofre oscilações; modelo muito utilizado na área de estatística básica. Já no aprendizado supervisionado por Classificação a ideia é prever variáveis categóricas, e não numéricas, como é o caso da regressão. Como exemplo de prever variáveis binárias ou classificadas em faixas, como níveis de renda, faixas de idade, etc, Prates (2018).

As tarefas de Aprendizado não supervisionado que serão utilizadas nesse processo de KDD, se referem ao reconhecimento de informações relevantes nos dados sem qualquer elemento externo que possa guiar o aprendizado. Tais tarefas tem por finalidade encontrar uma descrição compacta e simples dos dados, realizando uma busca por padrões que se repetem de maneira contínua na associação entre atributos e agrupamento de um coleção de dados, lidando com a identificação de grupos de acordo com a similaridade entre os dados Feuser (2017).

“As regras de associação devem evidenciar não apenas os conjuntos triviais de dados, mas também algumas relações que não estão aparentes, facilitando a tomada de

decisão. A técnica de regras por associação procura itens que ocorrem de forma frequente no conjunto de dados” Vasconcelos e Carvalho (2004).

Tendo o KDD como objetivo fundamental de trazer algo relevante, valido, útil e novo sobre uma determinada base de dados através de algoritmos, deve esse processo ser feito de forma detalhista e cuidadosa, já que é necessário ter bastante paciência e atenção ao que se está fazendo, pois é o passo inicial para se obter os dados.

Por existir um grande conjunto de obras sobre tal tema, os termos KDD e mineração de dados passam a ser confundidos entre si Rezende (2003), ele afirma ainda em sua obra que para se extrair conhecimento o processo é aplicado em grandes volume de dados.

Não somente ele sustenta isso, Fayyad, Piatetsky-Shapiro e Smyth (1996) também afirma: “o KDD (*Knowledge Discovery in Databases*) é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados”.

O termo mineração de dados não é muito usado pois é muito longo, poderia muito bem ter um nome mais conveniente Han, Kamber e Pei (2011). Ainda para Han, Kamber e Pei (2011), “A descoberta de conhecimento em bases de dados é um processo mais amplo que a mineração de dados. Envolve várias etapas onde a mineração de dados é parte deste processo”.

## 2.2 As Etapas do Processo KDD

O KDD possui uma série de etapas Operacionais a serem seguidas:

1. O Pré-processamento dos Dados - contempla a limpeza e preparação dos dados
2. A Mineração de Dados - compreende a mineração para descoberta de padrões
3. O Pós-processamento dos Dados - compreende a avaliação e interpretação do resultado obtido



Figura 2.1 – Visão Pragmática das etapas do processo de KDD. Retirada de Goldschmidt e Passos (2015).

### 2.2.1 O Pré-processamento dos Dados

Nesta etapa é onde ocorre a preparação dos dados a serem processados pelas ferramentas de mineração de dados na etapa seguinte do processo de KDD de modo a otimizarem os resultados, o que normalmente acontece em uma grande quantidade de registros, segundo Correa e Machado (2007).

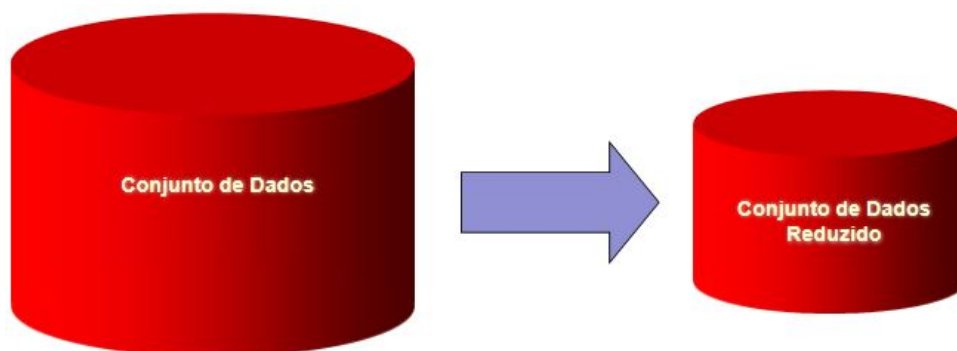
Como esta parte é a que gasta a maior parte do tempo com o processo de descoberta de conhecimento, aplicando técnicas de limpeza de dados, eliminando valores inutilizáveis ou substituindo valores. Tanto o tempo gasto, quanto o método empregado para a preparação de dados influenciam diretamente no resultado sendo ele favorável ou desfavorável.

Na maioria dos casos, as bases já estão organizadas em um *Data Warehouse*, que nada mais é do que um depósito de dados digitais que serve para armazenar informações detalhadas, orientadas e organizadas, que mantém históricos de informações, corrigindo ainda erros e reestruturando dados sem que afete o sistema de operação, para que facilite e melhore os dados analisados DATE (2004). Com eles, essa etapa de pré-processamento se torna bastante facilitada.

O fato de uma base de dados não se encontrar em um *Data Warehouse*, não impede que os dados sejam analisado pelos algoritmos de mineração, entretanto este fato implicará em um gasto de tempo maior nas etapas de preparação, seleção e limpeza dos dados.

#### 2.2.1.1 Seleção e Redução dos Dados

Fase esta, onde seleciona-se entre os dados de um conjunto de dados, quais deverão ser efetivamente considerados para a análise, ou seja, seleciona-se e limita-se o que para o processo de KDD será importante, fazendo a escolha e modelagem da base de dados, atributos e registros, obtendo como consequência uma base de dados reduzida em tamanho total.



**Figura 2.2 – Visão da etapa seleção/redução do processo de KDD. Retirada de Goldschmidt e Passos (2015).**

Segundo Correa e Machado (2007), para que o processo de KDD seja proveitoso, é

necessário identificar e escolher na base de dados os seus respectivos registros e atributos corretos. Dessa forma, algo que deve-se fazer primeiro é identificar a fonte de dados primária que será alvo da pesquisa, ou seja, o “C de Dados” inicial, sem nenhuma alteração. Para começar essa etapa, as fontes deverão estarem todas reunidas em uma única base de dados ou arquivo, deve-se atentar a isso pois, muitas vezes as fontes de dados estão em formatos diferentes como: banco de dados, planilhas eletrônicas ou mesmo em um *Data Warehouse*. Isso implica em uma maior dificuldade de reuni-las, se todas as fontes de dados tiver essa multiplicidade.

É parte também dessa tarefa de seleção determinar quais atributos serão examinados na mineração. Importante frisar que deve-se tomar muito cuidado, pois no final se obtém uma representação reduzida em volume mas ainda sim deve-se produzir resultados de análise idênticos ou similares aos obtidos com o conjunto completo de atributos. Como exemplo pode-se verificar em relação a base de dados do SGA, onde apresenta alguns atributos como: *id do usuário*, *numero da senha*, *serviço*, *prioridade*. Analisando esses atributos, podemos desprezar os atributos não importantes como *id do usuário* e *numero da senha* pois são atributos que apresentam baixa importancia de informações para nesse processo de mineração.

#### 2.2.1.2 Limpeza dos Dados

Conforme disse Goldschmidt e Passos (2015), tal tarefa de limpeza é composto de ações que acontecem nos dados selecionados para garantir tanto a qualidade quanto a veracidade, integridade e completude. Provavelmente os dados apresentem registros duplicados, erros, ou apresente dados ausentes após a união dos mesmos. Isso pode também ocorrer em dados que foram coletados de uma fonte única.

A limpeza dos dados tem por finalidade de excluir três tipos de situações: dados incompletos que possuem informações ausentes para determinados atributos ou dados poucos detalhados; dados ruidosos que estão errados ou contenham valores considerados divergentes; ou por último dados inconsistentes que contenham algum tipo de discrepância semântica entre si. Assim será utilizado nessa etapa de Limpeza o processo de exclusão de dados incompletos, com informações ausentes, pois alguns registros da base de dados se encontram em falta.

Han, Kamber e Pei (2011) diz que existem alguns métodos para tratar dados incompletos e que a natureza do atributo, a quantidade de registros e o número de dados ausentes são fundamentais para a escolha do método mais adequado e claro, todos eles apresentam vantagens e desvantagens. Métodos enumerados a seguir:

1. Ignorar o registro - que consiste em analisar todos os casos, removendo do processo de KDD os que apresentam valores ausentes.
2. Preencher manualmente o valor em falta - Requer uma pessoa especializada do

domínio, examinando cada caso unicamente. Porém pode ser tornar inviável devido a quantidade de registros faltantes

3. Preenchimento com valores de Globais Constantes - Método rápido e fácil para implementar, porém pode adulterar o aspecto geral da base, tornando tendencioso.
4. Preencher com a média estatística dos valores do Atributo - Utiliza-se técnicas estatísticas práticas e eficientes, porém a média pode tornar um processo de mineração também tendencioso.
5. Preencher com Método de Mineração - usar o valor mais provável, usando abordagens aprimoradas capazes de aprender o relacionamento entre atributos, no entanto exige uma complexidade de implementação e execução. Veja na figura 2.3.

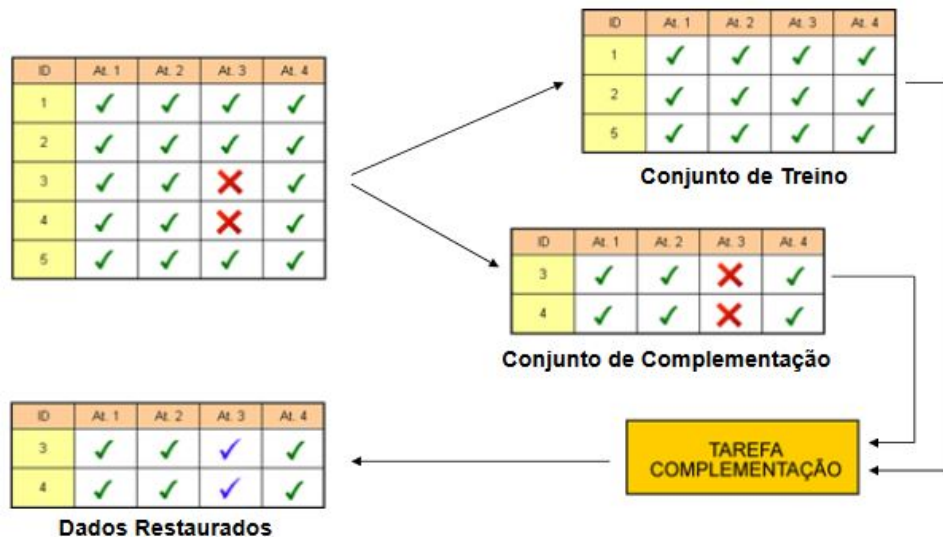


Figura 2.3 – Utilização do método de Preencher com Mineração. Retirada de Goldschmidt e Passos (2015).

### 2.2.1.3 Codificação dos Dados

Nesta etapa, os dados são colocados no mesmo formato se necessário, para que fiquem compatíveis com as ferramentas que fazem a mineração dos dados. Conforme Goldschmidt e Passos (2015), alguns algoritmos precisam que os dados tenham o formato nominal para que se obtenham os melhores resultados, enquanto outros necessitam estar em numérico. Isso consiste na simples substituição dos valores numéricos por nominais. Esse é o caso do algoritmo Apriori, que necessita que estejam em nominal, pois somente assim ele consegue gerar conjuntos de padrões, onde os mais frequentes são mantidos e os outros são eliminados .

Exemplo:

Sexo:

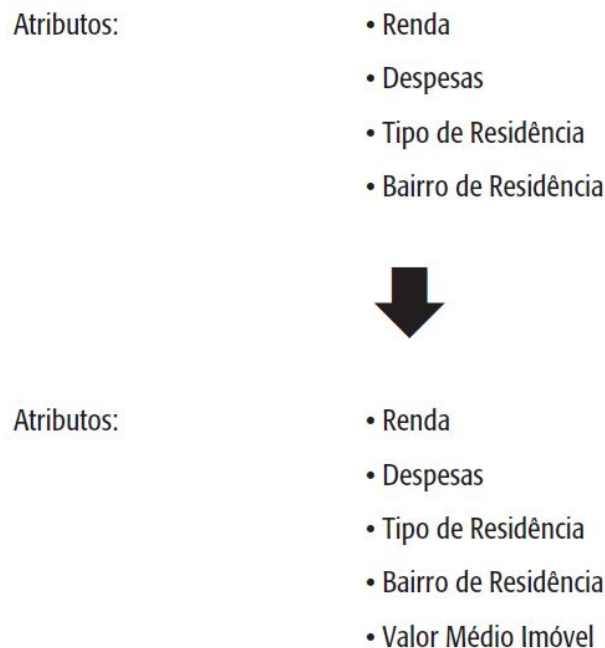
1' → F'

2' → M'

#### 2.2.1.4 Enriquecimento dos Dados

Segundo Goldschmidt e Passos (2015) “A fase de enriquecimento consiste em conseguir agregar mais informações aos registros existentes para que estes forneçam mais elementos para o processo de descoberta de conhecimento”. Para enriquecer se cria novos atributos tendo como base outros já existentes.

Normalmente enriquecer envolve a criação de novos atributos ou mesmo novas tabelas inteiras na base de dados, elas não ficam restritas a preencher informações ausente e assim se diferenciam das operações de limpeza. Esta agregação de novas informações aos dados devem contribuir diretamente para o KDD, porém muitas vezes devido ao alto custo de implementação esse processo fica inviável. Lembrando ainda que por ser informações que não constam na base de dados originalmente, devem ser conhecidas e validadas por especialistas na área.



**Figura 2.4 – Enriquecimento de Dados no processo de KDD. Retirada de Goldschmidt e Passos (2015).**

#### 2.2.2 Mineração de Dados

Definitivamente essa etapa é a mais importante no processo de KDD, pois é nela de fato que acontece toda a busca por novos conhecimentos e dados úteis da Base. Portanto é por isso que muitos autores tratam o processo de KDD e a mineração de dados como dois processos iguais, já para outros a mineração de dados é tratada como um processo

completo. Importante salientar que é nessa parte que o Algoritmo Apriori será executado gerando as regras de associação.

De acordo com Carvalho (2005), “Datamining ou Mineração de Dados é um conjunto de técnicas de Inteligência Artificial para a descoberta de conhecimento novo, não facilmente visível a olho nu, que possa estar presente em grandes massas de dados, normalmente armazenadas de forma digital em bancos de dados empresariais”.

Groth (1998) afirma que “é o processo de descoberta automática de informações”.

Por último mas não menos importante KREMER (1999) ressalta que Mineração de Dados é uma tecnologia que se usa para conhecer novas informações de forma estratégica que estavam ocultas em meio a grande massa de dados.

A origem da abordagem, a Mineração de dados tem maior destaque na tecnologia, na pesquisa da estatística e Inteligencia Artificial (IA) ou mesmo no processo de descoberta de conhecimento.

Logo, a mineração de dados utiliza algoritmos e técnicas de diversas áreas de conhecimento para que seja analisado grande volume de dados visando a extração de padrões sobre estes dados. Porém um fator que influência na escolha desses algoritmos tem relação com os tipos de variáveis envolvidas em cada problema específico e a definição da abordagem que se utilize para a solução.

#### 2.2.2.1 As Técnicas e Tarefas de Mineração de Dados

Conforme Correa e Machado (2007) “As tarefas definem a finalidade a que se propõe a análise dos dados. Para execução destas tarefas são aplicadas algumas técnicas que podem ser específicas ou não para uma determinada tarefa”. Lembrando o que foi falado mais acima, tais técnicas não envolve somente a área da tecnologia computacional, mas também IA, modelos matemáticos, e outros campos de pesquisa.

Veja a seguir as tarefas e técnicas de mineração de dados na tabela 2.1.



TAREFAS	TÉCNICAS
Classificação	Árvores de decisão, regressão, redes neurais
Estimação	Regressão e redes neurais
Previsão	Redes neurais, séries temporais, regressão, árvore de decisão, raciocínio baseado em casos
Análise de afinidades	Regras de associação
Análise de agrupamentos	Redes neurais, algoritmos genéticos e específicos

**Tabela 2.1 – Tarefas e Técnicas de Mineração de Dados. Retirada de Correa e Machado (2007).**

Dentre as Técnicas mostradas anteriormente na tabela 2.1, a utilizada pelo algoritmo apriori nesse trabalho é a Técnica de Regras de Associação através das tarefas de Análise de afinidades.

#### 2.2.2.2 Análise de Afinidades

Afinidades são unidades de informação que se repetem, na sequência em que aparece quanto na sua estrutura de formação. Na Mineração de dados tem como tarefa mais comum a busca desses padrões, segundo Correa e Machado (2007).

#### 2.2.2.3 Regras de Associação

Em grandes bancos de dados as regras de associação procuram por associações relevantes ou pelas relações de correlação entre os atributos Han, Kamber e Pei (2011). Tais regras de associação foram introduzidas por Agrawal, Imielinsk e Swami (1993), por meio do algoritmo Apriori, que utiliza na busca de itens frequentes uma forma interativa.

#### Definições

Correa e Machado (2007) diz que:

Considere um conjunto de itens  $I = (I_1, I_2, I_3, \dots, I_n)$ .

Uma regra de associação é uma expressão na forma  $A \rightarrow B$ , tal que A e B são chamados de itensets frequentes, sendo que  $A \neq \emptyset$  e  $B \neq \emptyset$ , onde A é denominado antecedente e o B conseqüente da regra.

### Suporte de uma Regra

Segundo Correa e Machado (2007) “O suporte de uma regra é o número de ocorrências da regra em relação ao número total de transações”. Ou seja, o suporte é quantidade de registros em que um conjunto de itens (A) aparece, dividido pela quantidade total de registros presentes na base de dados

$$Suporte(A \rightarrow B) = \frac{(TotalregA \cup B)}{Totalreg}$$

### Confiança de uma Regra

Tal confiança é o número de ocorrências da regra em relação ao número de ocorrência do itenset anterior a regra, Correa e Machado (2007).

$$Confiana(A \rightarrow B) = \frac{Suporte(A \cup B)}{SuporteA}$$

### 2.2.3 O Pós-processamento dos Dados

É nesta etapa que se consolida todo o conhecimento gerado pelo processo de KDD, criando modelos que podem ser de diversas formas representadas dependendo é claro dos dados resultantes, como normalmente mais se utiliza em tabelas, planilhas, gráficos em 2D e 3D, regras ou mesmo na forma de arvores.

Essas técnicas que são utilizadas para representar esses dados novos visualmente, facilitam a compreensão humana, pois desse modo se associa melhor as informações adquiridas. Porém, chegar nessa etapa final não acaba com a tomada de conhecimento, já que deve ser tudo analisado por especialistas da área em relação a validade dos resultados adquiridos e na qualidade dos dados gerados.

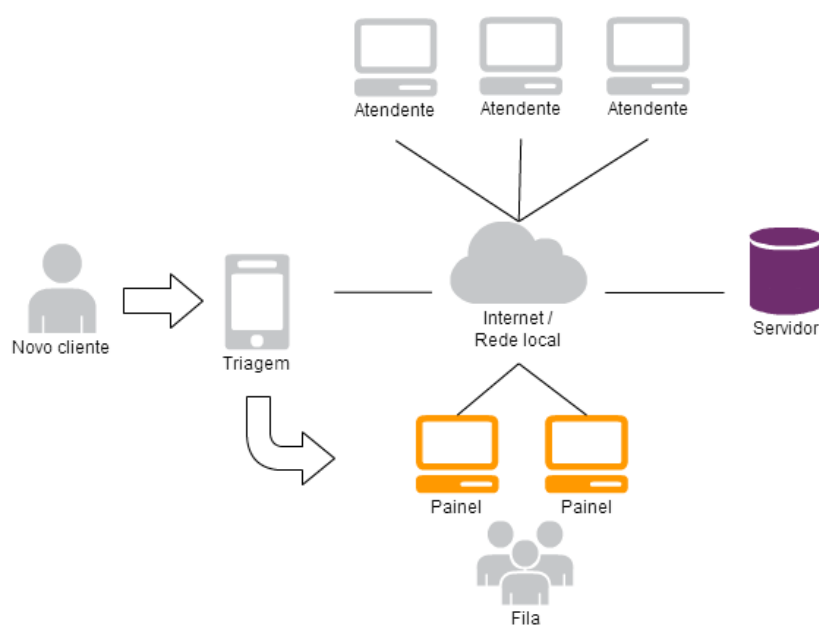
Em ambos os casos pode-se entender que os resultados obtidos sejam insatisfatórios, assim necessita a retomada de toda a etapa de mineração, até mesmo retornar a etapa de pré-processamento, verificando as técnicas utilizadas e parâmetros, para que possam ser mudados para refazer todo o processo novamente aplicando outras propriedades. É bem comum esse processo ser refeito várias vezes para chegar a um resultado satisfatório.

### 3 O SISTEMA DE GERENCIAMENTO DE ATENDIMENTO (SGA)

O SGA é um sistema de atendimentos de código aberto SGA (2019), disponível para a comunidade em geral que deseje utiliza-lo, ele foi implantado no SINE pois havia uma grande necessidade de algum sistema para armazenar as informações, agilizar e melhorar o atendimento dos serviços prestados à população.

A partir de agora apresenta-se o funcionamento do sistema SGA, para que posteriormente entenda-se melhor como o banco de dados funciona e quais atributos e registros se encontram armazenados nele. Essa etapa é fundamental para o entendimento geral do SINE e do SGA, também os serviços prestados e qual perfil das pessoas que são atendidas.

A figura 3.1 mostra de maneira geral um fluxograma de como funciona o atendimento no SINE gerenciado pelo SGA, desde a chegada do cliente até o fim do atendimento. Primeiramente a pessoa chega ao SINE e através dos atendentes responsáveis pela triagem, retiram-se a senha referente ao atendimento desejado. Logo após, os clientes aguardam atendimento sentados de frente para o Painel de senhas, onde as senhas são chamadas e indicadas em qual guichê ou mesa o cliente será atendido.



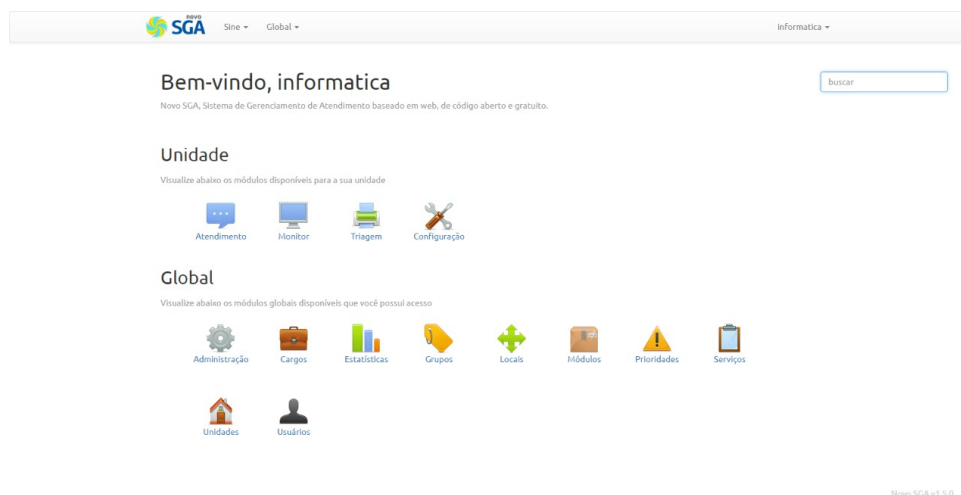
**Figura 3.1 – Fluxo de funcionamento dos atendimentos do SINE. Fonte: SGA (2019).**

A figura 3.2 representa a tela de login do SGA, o qual é acessado via endereço web local por IP. Infelizmente por motivos de infraestrutura na rede como internet de baixa velocidade ou cabeamento ruim, ele não se encontra disponível online, para acesso externo.



**Figura 3.2 – Tela Inicial de Login SGA. Fonte: SINE-TO (2016).**

Na próxima figura 3.3, tem-se o painel de gerenciamento do administrador do sistema, onde será tratado em seguida algumas funções relevantes a serem informadas sobre o atendimento feito através do SGA.



**Figura 3.3 – Tela Gerenciamento do Administrador do SGA. Fonte: SINE-TO (2016).**

### 3.1 Módulo de Atendimento

É o módulo responsável por prestar todo o serviço para o cliente que aguarda atendimento. Quem o opera são os funcionários do SINE que trabalham em guichês ou mesas de atendimento. Por meio desse módulo que é feito o chamado da próxima senha a ser atendida, alteram-se também se necessário, o número do guichê e claro realizam-se todos os serviços disponíveis no Órgão. Na figura 3.4 observa-se bem toda a tela utilizada por esses operadores.



Figura 3.4 – Tela do Módulo de Atendimento do SGA. Fonte: SINE-TO (2016).

### 3.2 Módulo Triagem

Esse módulo é responsável por gerar as senhas de atendimento. O atendente responsável por esse serviço é na verdade o primeiro que atende o cliente em busca de serviços, pois é o primeiro passo a se fazer ao chegar no órgão, ou seja, emitir a senha. Veja na figura 3.5 que ele tem ativo 4 serviços que serão falados mais a frente nesse capítulo.

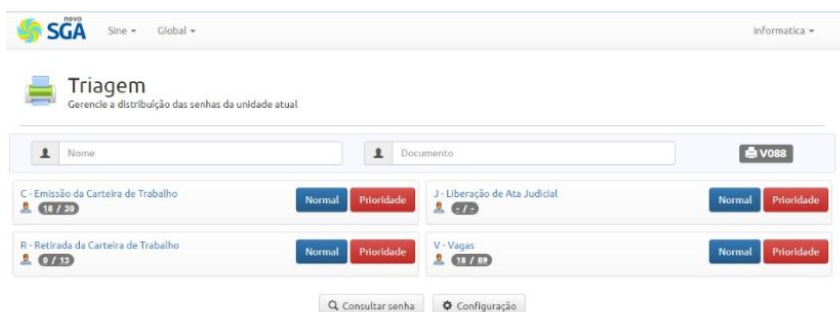


Figura 3.5 – Tela de Triagem para geração de senhas do SGA. Fonte: SINE-TO (2016).

Alguns módulos não serão comentados, pois não possuem uma relevância para este projeto, são módulos de configuração ou administração do sistema que não implica na finalidade que este projeto terá. Então passa-se ao próximo módulo, o de Estatísticas.

### 3.3 Módulo de Estatísticas

Aqui é somente para nível de informação, pois é um módulo que gera gráficos de informações do dia que se encontra armazenada na base de dados, porém esse gráfico ao final de todos os dias é zerado e reinicia a contagem no dia seguinte. Infelizmente não há possibilidade de utilizá-lo para processo de Descoberta de Conhecimento, mas é uma boa informação rápida para quem deseja-se conhecer sobre os atendimentos do dia.

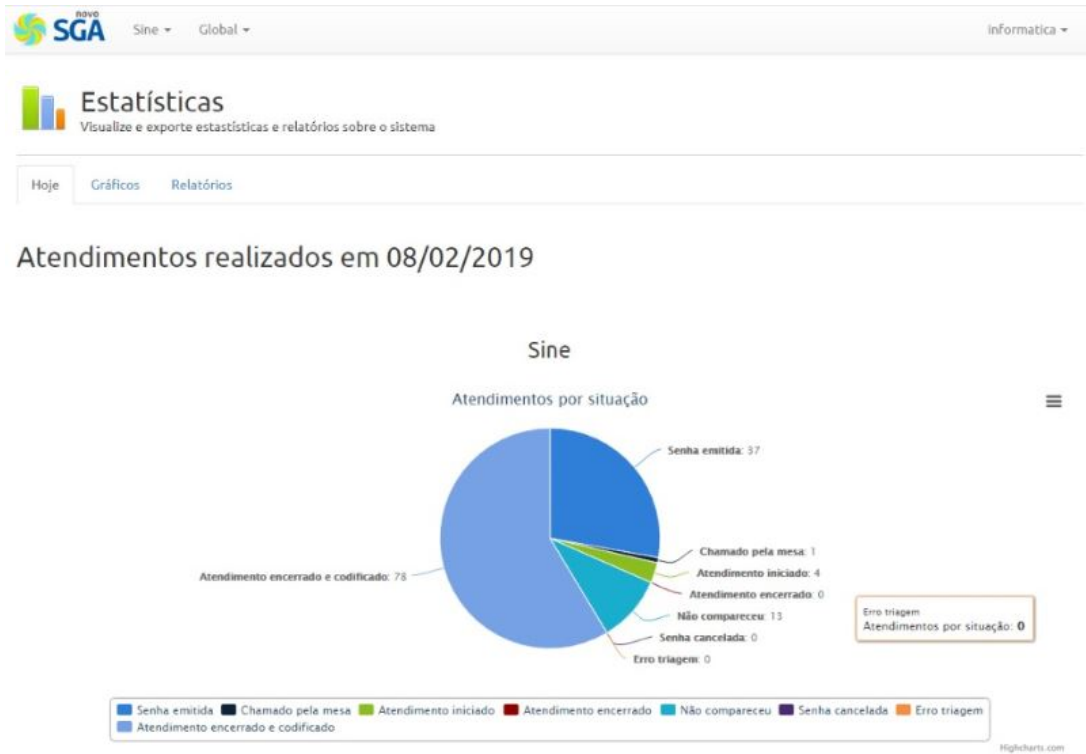


Figura 3.6 – Tela de Estatísticas com Gráfico - SGA. Fonte: SINE-TO (2016).

### 3.4 Módulo de Prioridades

Seguindo adiante, na figura 3.7 encontra-se o módulo de prioridades. Nele é possível a criação de novos tipos de Prioridades para atendimentos especiais, como ainda, definir para as prioridades já existentes, o peso que elas possuem. Podemos ver que todas as prioridades possuem peso “20”, com exceção dos chamados “Sem prioridade” ou seja, prioridade normal. Esse atributo é importante para o processo de descoberta de conhecimento.

The screenshot shows the 'Prioridades' (Priorities) module of the SGA system. It features a search bar and a 'Novo' (New) button. Below is a table listing the priority types:

#	Nome	Peso	Status
1	Sem prioridade	0	Ativo
2	Portador de Deficiência	20	Ativo
3	Gestante	20	Ativo
4	Idoso	20	Ativo
5	Outros	20	Ativo

Figura 3.7 – Tela do Módulo de Prioridades do SGA. Fonte: SINE-TO (2016).

### 3.5 Módulo de Serviços

Esse de fato é o módulo mais importante do SGA, pois é responsável pela principal função do órgão, a de execução de serviços, é definitivamente o essencial motivo pelo qual as pessoas procuram o SINE. A seguir consta uma tabela 3.1 com os serviços disponíveis:

Serviço
Consulta de Vagas
Seguro Desemprego
Emissão de Carteira de Trabalho
Retirada de Carteira de Trabalho
Triagem
Qualificação
Convocação
Curriculum
Liberação de Ata Judicial

**Tabela 3.1 – Tabela de Serviços fornecidos no SGA do SINE. Fonte: SINE-TO (2016).**

#### 4 MÉTODOS E O PROCESSO DE KDD NA BASE DE DADOS DO SGA

Independente da ferramenta ou algoritmo escolhido, faz-se necessário primeiramente escolher a base de dados. Como já mencionado nesse Projeto, o banco de dados utilizado foi do SGA, que foi desenvolvida em MySQL.

Para acessar o banco de dados utilizou-se o Sistema de Gerenciamento de banco de dados Mysql Workbench, a qual já se utilizava pelo administrador do SGA e portanto também já havia um conhecimento prévio de seu funcionamento.

Para passar por todo o processo de KDD principalmente a etapa de seleção dos dados, necessitou-se conhecer um pouco do funcionamento da base de dados do SGA e seus relacionamentos. A Figura 4.1 mostra o Diagrama de Entidade Relacionamento da Base de Dados do SGA.

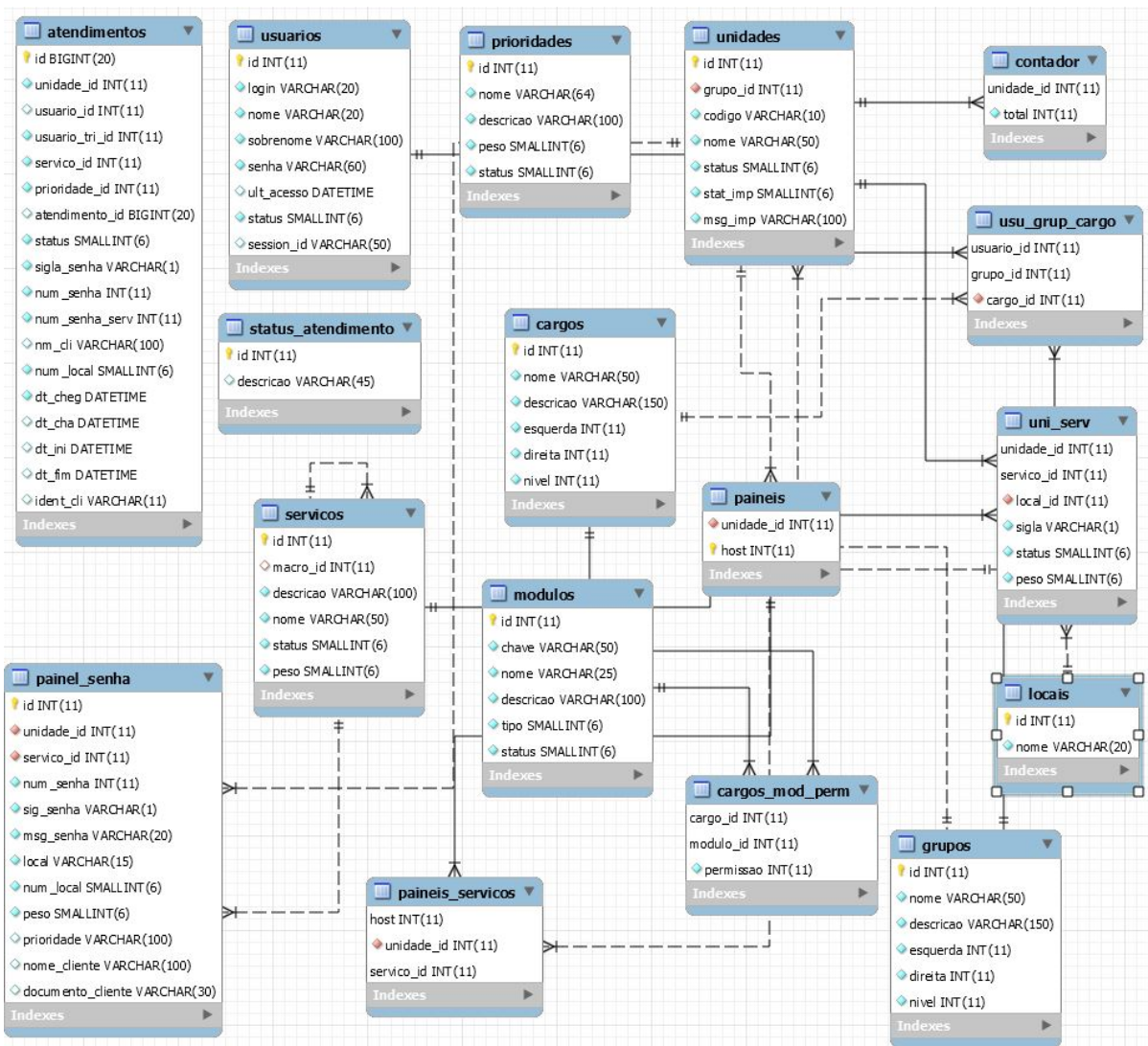


Figura 4.1 – Diagrama do Banco de Dados - SGA. Fonte: SINE-TO (2016).



A Base de Dados possui uma quantidade significativa de tabelas, porém somente três delas foram utilizadas neste trabalho que são: “*atendimentos*” que é responsável por armazenar as informações dos atendimentos em geral, com todos os atributos, porém com valores numéricos; e as tabelas “*servicos*” e “*prioridades*”, onde armazena as informações nominais dos serviços e prioridades respectivamente. O restante das tabelas não foram utilizadas diretamente, pois as mesmas tem importância e funcionalidades somente para funcionamento do SGA e não para esse trabalho em especial.

## 4.1 Pré-Processamento dos Dados do SGA

Aqui, onde é feita a preparação dos dados do SGA para serem posteriormente processados pela ferramenta de mineração escolhida, otimizando assim os resultados obtidos. Logo que foi analisado alguns registros, já houve a necessidade de eliminar alguns atributos que não seriam úteis obter conhecimento sobre eles ou que não encaixariam no processo de geração de regras de associação.

### 4.1.1 Fase de Seleção dos Atributos

Independente da base de dados, praticamente todas elas existem atributos que podem ser descartados ou eliminados para que reduza o número de regras que serão geradas, uma vez que este é um dos maiores problemas para análise dos resultados de um processo de mineração de dados, a grande quantidade de regras que são geradas.

Como mencionado anteriormente alguns atributos foram descartados como *id da unidade, número da senha, número do local, nome do cliente*.

Dentre os Atributos escolhidos ficaram:

- Serviço
- Prioridade
- Status do Atendimento

#### **Serviço**

O atributo Serviço são os serviços disponíveis para o cliente ao visitar o SINE, eles podem ter os seguintes valores, veja na tabela 4.1 a seguir:

Serviço
Consulta de Vagas
Seguro Desemprego
Emissão de Carteira de Trabalho
Retirada de Carteira de Trabalho
Triagem
Qualificação
Convocação
Curriculum
Liberação de Ata Judicial

**Tabela 4.1 – Tabela de Serviços do SGA. Fonte: SINE-TO (2016).**

### **Prioridade**

Nesse atributo fica registrado as prioridades, que podem ser vistos na tabela 4.2.

Prioridade
Atendimento Normal
Atendimento Prioritário Idoso
Atendimento Prioritário Gestante
Atendimento Prioritário Portadores de Deficiencia
Qualquer Outra Prioridade

**Tabela 4.2 – Tabela de Prioridades do SGA. Fonte: SINE-TO (2016).**

### **Status do Atendimento**

Por último temos o atributo Status do Atendimento que registra qual a situação se encontra o atendimento feito ao cliente. Segue na tabela 4.3.

Status do Atendimento
Atendimento Encerrado
Não Compareceu
Senha Emitida
Erro Triagem
Atendimento Encerrado
Chamado pela Mesa
Atendimento Iniciado
Senha Cancelada

**Tabela 4.3 – Tabela do Status do Atendimento do SGA. Fonte: SINE-TO (2016).**

#### 4.1.2 Fase de Limpeza e Seleção dos Dados

Não foi necessário muito processo nesta fase, pois a maioria dos dados já estavam prontos para passar em diante à etapa seguinte. A única coisa feita na base de dados no SGA foi:

- Eliminação de todos os registros que continham algum erro ou informações incompletas.

Como exemplo alguns registros no atributo status de atendimento se encontravam ausentes, pois alguns atendimentos não foram finalizados e nem mesmo continham informação a respeito.

#### 4.1.3 Transformação dos Dados - Enriquecimento e Codificação

A fase que será feita a preparação dos dados, alguns campos dos atributos serão transformados ou codificados para tornar possível ou facilitar a extração dos padrões e ainda auxiliar a visualização e análise final dos resultados. Está listado a seguir as ações efetuadas nessa transformação dos dados do SGA:

- Os valores do atributo *serviço* originalmente continha valores de *id de serviço* e portanto eram numéricos na base original do SGA e foram convertidos para valores nominais, pois a maioria das ferramentas de mineração trabalham com campos nominais para geração de regras de associação.
- Assim como no anterior, o atributo *prioridade* também continha valores numéricos na base original e também foram convertidos para valores nominais pelo mesmo motivo.

- Foi criado uma nova tabela para associar os valores do atributo *status do atendimento*, pois a tabela original continham valores inteiro, e como as regras de associação do algoritmo escolhido utiliza valores nominais.

## 4.2 Análise Exploratória de Dados

Essa análise parece ser algo simples, mas é importante para as estatísticas e faz parte para melhorar eficiência da interpretação e análise de regras de associação que serão geradas no processo de mineração de dados. É onde identifica e compara comportamentos; estuda a correlação entre as variáveis; encontra tendências.

Após passar por todas as etapas de Pre-processamento a base de dados do SGA possui um total de 115.054 registros de atendimentos. A seguir será apresentado a quantidade de registros em cada categoria de atributos escolhidos na etapa de seleção de dados.

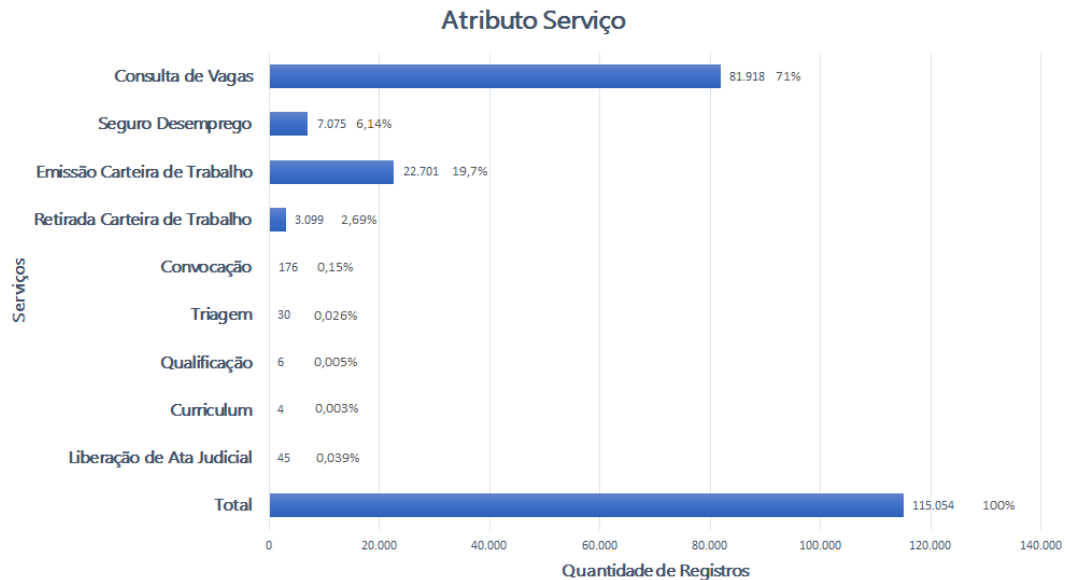


Figura 4.2 – Gráfico - Total de Registros do atributo Serviço. Fonte: SINE-TO (2016).

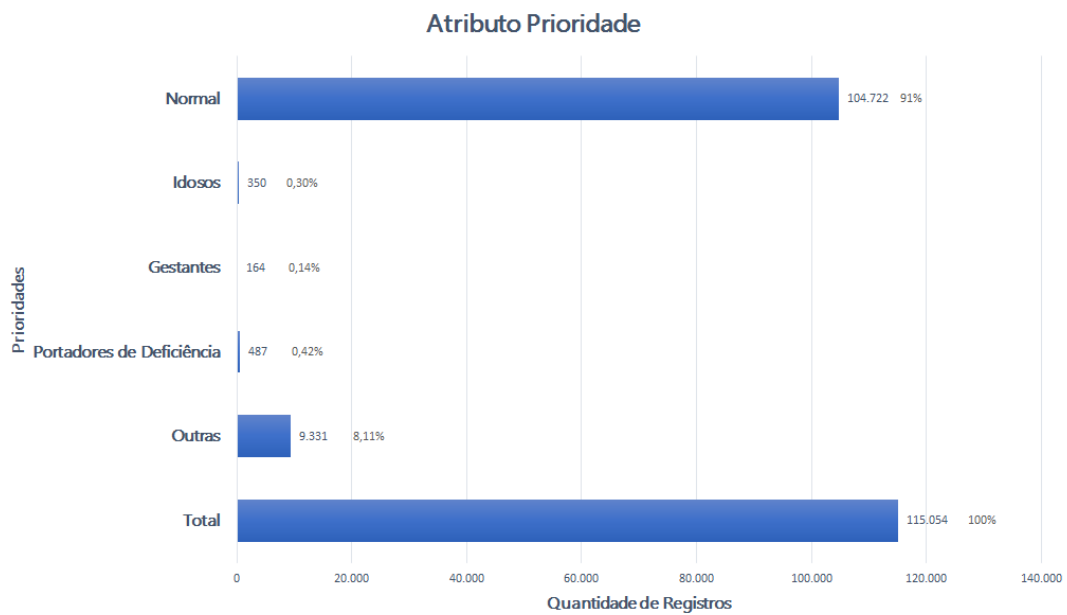


Figura 4.3 – Gráfico - Total de Registros do atributo Prioridade. Fonte: SINE-TO (2016).

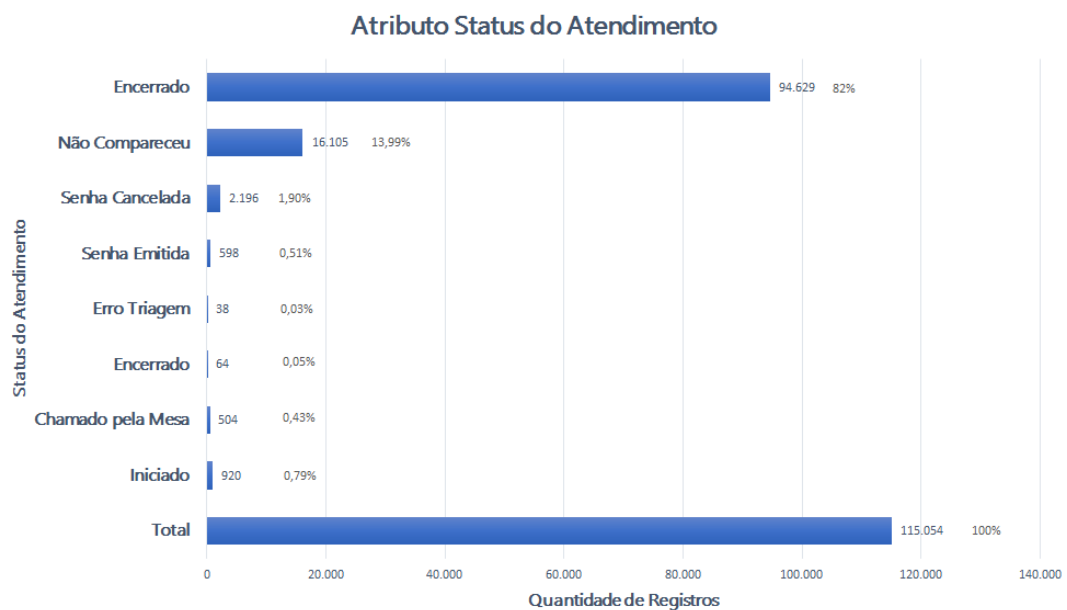


Figura 4.4 – Gráfico - Total de Registros do atributo Status do Atendimento. Fonte: SINE-TO (2016).

### 4.3 A Ferramenta Utilizada

No mercado encontra-se bastante ferramentas para mineração de dados com o único objetivo de gerar regras de associação, tanto de software livre como software proprietário.

A Weka (WEKA, 2019), é a ferramenta mais utilizada para mineração de dados e aprendizagem de máquina no meio acadêmico, desenvolvida na linguagem Java, pode ser utilizada desde sua interface gráfica, para interagir com arquivos de dados e produzir

resultados visuais; até mesmo através de linha de comando e também é um *software livre*. Essa foi a Ferramenta escolhida para esse trabalho porque é simples e possui um fácil manuseio.

Na maioria dos casos, os utilizadores da ferramenta convertem seus respectivos banco de dados que normalmente se encontram no formato de arquivo .sql em arquivo padrão da ferramenta, chamado de ARFF(.arff) ou ainda em formato .csv que também é aceito pela Weka. Veja a a janela principal da Weka na figura 4.5

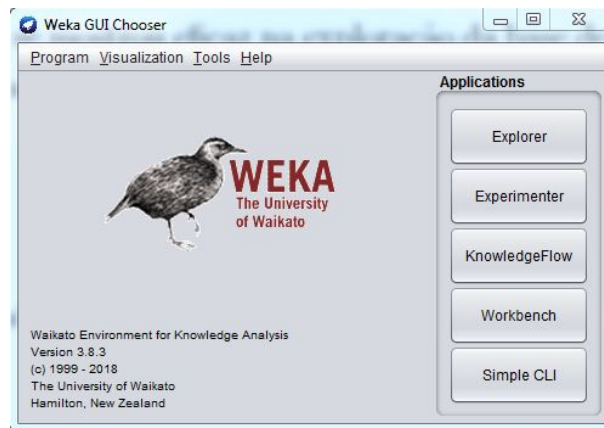


Figura 4.5 – Janela Inicial da Ferramenta Weka. Fonte: (WEKA, 2019).

A ferramenta também é capaz de se conectar diretamente com o banco de dados onde se encontra as informações, e foi desta maneira que obteve-se os dados da base, foi feito também todos os procedimentos de configuração para a conexão com a base de dados que encontrava-se sendo administrada no mysql Workbench, eliminando assim a necessidade de geração do arquivo em ARFF ou CSV.

Na figura 4.6 mostra a janela Principal do Weka já com a base de dados importada depois de todo o pré-processamento, onde será mostrado algumas informações.

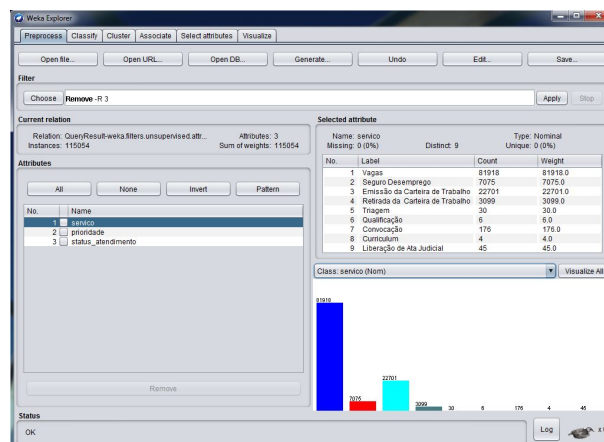
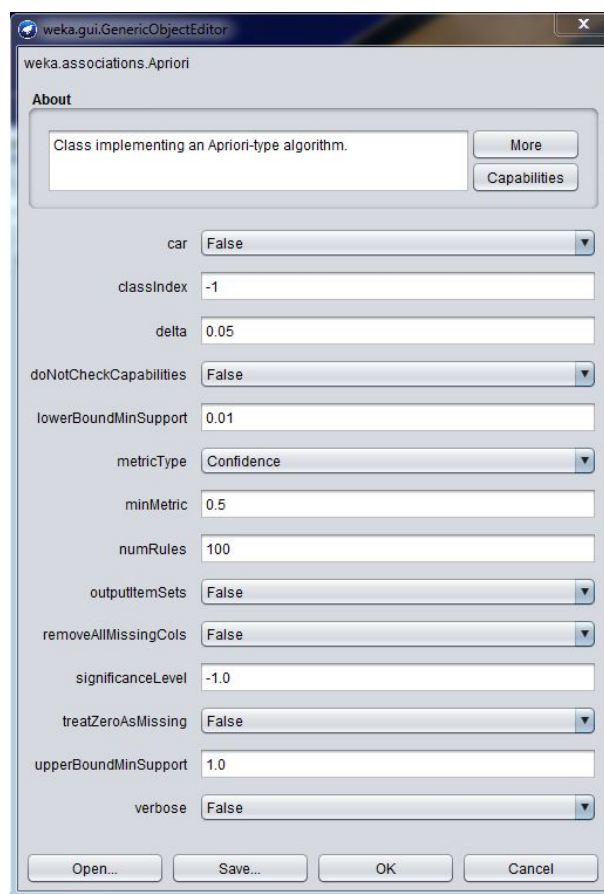


Figura 4.6 – Janela Principal - Pré processamento da Ferramenta Weka. Fonte: (WEKA, 2019).

O Algoritmo utilizado por trás do Weka nesse Projeto foi o algoritmo Apriori, escolhido após pesquisas constatarem que ele é o algoritmo mais utilizado para geração de regras de associação na mineração de dados, além de ser de simples utilização.

O Apriori faz uma busca aprofundada nos dados, identificando os itens de maior frequência e gerando um conjunto de regras de associação entre eles.

Na figura 4.7 abaixo observa-se a sub-janela com os parâmetros do algoritmo Apriori, onde podem ser alterados de acordo com a necessidade, algo que acontece todas as vezes é fazer a mudança desses valores e executar o algoritmo afim de encontrar a melhor combinação de configurações para um resultado de geração de regras de associação satisfatório.



**Figura 4.7 – Sub-janela de Parâmetros do Algoritmo Apriori na Ferramenta Weka. Fonte: (WEKA, 2019).**

A maioria dos parâmetros utilizados foram deixados como padrão, os únicos alterados em algumas gerações de regras foram: o valor de *upperBoundMinSupport*, *lowerBoundMinSupport*, *MinMetric* e *numRules*.

- *upperBoundMinSupport*: Valor do “support” responsável por definir em qual porcentagem da base de dados iniciará a geração de regras, ou seja, regras de atributos que aparecem mais vezes na base.

- *lowerBoundMinSupport*: valor que define o mínimo do “support”, até qual porcentagem gerará as regras dos atributos.
- *MinMetric*: É o valor referente a confiança, o valor utilizado nesse trabalho foi de no mínimo 50% de confiança para geração de regras
- *numRules*: Valor máximo de regras geradas pelo Apriori
- *delta*: Parâmetro responsável por decrementar o “support”.

No Geral o processo iniciará com “support” de 100% e será decrementado de 5% em 5% até chegar a 1% da base ou alcançar o número máximo de regras geradas (definido por *numRules*), tudo isso com valor de confiança mínimo de 50%.

O Algoritmo Apriori é dividido em duas etapas. A primeira etapa gera os conjuntos de itens (itemsets) frequentes, que são aqueles que mais aparecem no banco de dados, e a segunda etapa gera as regras de associação. A Figura abaixo mostra o pseudocódigo para realizar a primeira etapa do algoritmo Araujo (2016).

```

L1 = {large 1-itemsets};
for ( k = 2; Lk-1 ≠ ∅; k++ ) do begin
  Ck = apriori-gen(Lk-1); // New candidates
  forall transactions t ∈ D do begin
    Ct = subset(Ck, t); // Candidates contained in t
    forall candidates c ∈ Ct do
      c.count++;
  end
  Lk = {c ∈ Ck | c.count ≥ minsup}
end
Answer = ∪k Lk;

```

Figura 4.8 – Pseudocódigo do algoritmo Apriori. Fonte: Araujo (2016).

De acordo com o pseudocódigo apresentado na Figura 4.8, “ $D$  é o conjunto de transações  $t$  (registros) do *dataset* (conjunto de dados).  $\mathcal{L}_k$  é a lista de conjunto-de- $k$ -itens frequentes.  $\mathcal{C}_k$  é a lista de conjunto-de- $k$ -itens candidatos. A geração de  $\mathcal{C}_k$  é realizada fazendo combinação entre os elementos de  $\mathcal{L}_{k-1}$ , utilizando a função *apriori-gen()*. A função *apriori-gen()* recebe como parâmetro  $\mathcal{L}_{k-1}$  e retorna para  $\mathcal{C}_k$  a lista de novos candidatos” Araujo (2016).



## 5 RESULTADOS ENCONTRADOS

Foi feita várias combinações através dos atributos escolhidos. Primeiramente, comparações de pares de atributos e por último com todos os atributos. Os resultados encontram-se a seguir nas tabelas 5.1, 5.2, 5.3, 5.4. E logo em seguida será feita a análise detalhada dos resultados.

### Serviço X Prioridade

Relação de Regras e Atributos	Confiança
servico=Vagas 81918 ==>prioridade=Atendimento normal 79217	<conf:(0.97)>
servico=Seguro Desemprego 7075 ==>prioridade=Atendimento normal 6624	<conf:(0.94)>
servico=Emissão da Carteira de Trabalho 22701 ==>prioridade=Atendimento normal 17582	<conf:(0.77)>
prioridade=Atendimento normal 104722 ==>servico=Vagas 79217	<conf:(0.76)>
servico=Retirada da Carteira de Trabalho 3099 ==>prioridade=Qualquer outra prioridade 2005	<conf:(0.65)>
prioridade=Qualquer outra prioridade 9331 ==>servico=Emissão da Carteira de Trabalho 4867	<conf:(0.52)>

**Tabela 5.1 – Tabela Resultado Apriori - Serviço x prioridade.**

### Serviço X Status do Atendimento

Relação de Regras e Atributos	Confiança
servico=Seguro Desemprego 7075 ==>status_atendimento=Atendimento Encerrado 6228	<conf:(0.88)>
servico=Retirada da Carteira de Trabalho 3099 ==>status_atendimento=Atendimento Encerrado 2723	<conf:(0.88)>
servico=Emissão da Carteira de Trabalho 22701 ==>status_atendimento=Atendimento Encerrado 18908	<conf:(0.83)>
servico=Vagas 81918 ==>status_atendimento=Atendimento Encerrado 66676	<conf:(0.81)>
status_atendimento=Nao Compareceu 16105 ==>servico=Vagas 12652	<conf:(0.79)>
status_atendimento=Atendimento Encerrado 94629 ==>servico=Vagas 66676	<conf:(0.7)>
status_atendimento=Senha Cancelada 2196 ==>servico=Vagas 1376	<conf:(0.63)>

**Tabela 5.2 – Tabela Resultado Apriori - Serviço x Status do Atendimento.**

## Prioridade X Status do Atendimento

Relação de Regras e Atributos	Confiança
status_atendimento=Nao Compareceu 16105 ==>prioridade=Atendimento normal 14902	<conf:(0.93)>
status_atendimento=Atendimento Encerrado 94629 ==>prioridade=Atendimento normal 85937	<conf:(0.91)>
status_atendimento=Senha Cancelada 2196 ==>prioridade=Atendimento normal 1925	<conf:(0.88)>
prioridade=Qualquer outra prioridade 9331 ==>status_atendimento=Atendimento Encerrado 7847	<conf:(0.84)>
prioridade=Atendimento normal 104722 ==>status_atendimento=Atendimento Encerrado 85937	<conf:(0.82)>

**Tabela 5.3 – Tabela Resultado Apriori - Prioridade X Status do Atendimento.**

## Serviço X Prioridade X Status do Atendimento

Relação de Regras e Atributos	Confiança
servico=Vagas status_atendimento=Nao Compareceu 12652 ==>prioridade=Atendimento normal 12299	<conf:(0.97)>
servico=Vagas 81918 ==>prioridade=Atendimento normal 79217	<conf:(0.97)>
servico=Vagas status_atendimento=Atendimento Encerrado 66676 ==>prioridade=Atendimento normal 64402	<conf:(0.97)>
servico=Vagas status_atendimento=Senha Cancelada 1376 ==>prioridade=Atendimento normal 1323	<conf:(0.96)>
servico=Seguro Desemprego 7075 ==>prioridade=Atendimento normal 6624	<conf:(0.94)>
servico=Seguro Desemprego status_atendimento=Atendimento Encerrado 6228 ==>prioridade=Atendimento normal 5826	<conf:(0.94)>
servico=Retirada da Carteira de Trabalho prioridade=Qualquer outra prioridade 2005 ==>status_atendimento=Atendimento Encerrado 1865	<conf:(0.93)>
status_atendimento=Nao Compareceu 16105 ==>prioridade=Atendimento normal 14902	<conf:(0.93)>
status_atendimento=Atendimento Encerrado 94629 ==>prioridade=Atendimento normal 85937	<conf:(0.91)>
servico=Seguro Desemprego 7075 ==>status_atendimento=Atendimento Encerrado 6228	<conf:(0.88)>
servico=Seguro Desemprego prioridade=Atendimento normal 6624 ==>status_atendimento=Atendimento Encerrado 5826	<conf:(0.88)>
servico=Retirada da Carteira de Trabalho 3099 ==>status_atendimento=Atendimento Encerrado 2723	<conf:(0.88)>
status_atendimento=Senha Cancelada 2196 ==>prioridade=Atendimento normal 1925	<conf:(0.88)>
servico=Emissão da Carteira de Trabalho prioridade=Atendimento normal 17582 ==>status_atendimento=Atendimento Encerrado 14789	<conf:(0.84)>
prioridade=Qualquer outra prioridade 9331 ==>status_atendimento=Atendimento Encerrado 7847	<conf:(0.84)>
servico=Vagas prioridade=Qualquer outra prioridade 2047 ==>status_atendimento=Atendimento Encerrado 1709	<conf:(0.83)>
servico=Emissão da Carteira de Trabalho 22701 ==>status_atendimento=Atendimento Encerrado 18908	<conf:(0.83)>
prioridade=Atendimento normal status_atendimento=Nao Compareceu 14902 ==>servico=Vagas 12299	<conf:(0.83)>
servico=Seguro Desemprego 7075 ==>prioridade=Atendimento normal status_atendimento=Atendimento Encerrado 5826	<conf:(0.82)>
prioridade=Atendimento normal 104722 ==>status_atendimento=Atendimento Encerrado 85937	<conf:(0.82)>
servico=Vagas 81918 ==>status_atendimento=Atendimento Encerrado 66676	<conf:(0.81)>
servico=Vagas prioridade=Atendimento normal 79217 ==>status_atendimento=Atendimento Encerrado 64402	<conf:(0.81)>
servico=Emissão da Carteira de Trabalho prioridade=Qualquer outra prioridade 4867 ==>status_atendimento=Atendimento Encerrado 3920	<conf:(0.81)>
servico=Vagas 81918 ==>prioridade=Atendimento normal status_atendimento=Atendimento Encerrado 64402	<conf:(0.79)>
status_atendimento=Nao Compareceu 16105 ==>servico=Vagas 12652	<conf:(0.79)>
servico=Emissão da Carteira de Trabalho status_atendimento=Atendimento Encerrado 18908 ==>prioridade=Atendimento normal 14789	<conf:(0.78)>
servico=Emissão da Carteira de Trabalho 22701 ==>prioridade=Atendimento normal 17582	<conf:(0.77)>
status_atendimento=Nao Compareceu 16105 ==>servico=Vagas prioridade=Atendimento normal 12299	<conf:(0.76)>
prioridade=Atendimento normal 104722 ==>servico=Vagas 79217	<conf:(0.76)>
prioridade=Atendimento normal status_atendimento=Atendimento Encerrado 85937 ==>servico=Vagas 64402	<conf:(0.75)>
servico=Emissão da Carteira de Trabalho status_atendimento=Nao Compareceu 2561 ==>prioridade=Atendimento normal 1835	<conf:(0.72)>
status_atendimento=Atendimento Encerrado 94629 ==>servico=Vagas 66676	<conf:(0.7)>
prioridade=Atendimento normal status_atendimento=Senha Cancelada 1925 ==>servico=Vagas 1323	<conf:(0.69)>
servico=Retirada da Carteira de Trabalho status_atendimento=Atendimento Encerrado 2723 ==>prioridade=Qualquer outra prioridade 1865	<conf:(0.68)>
status_atendimento=Atendimento Encerrado 94629 ==>servico=Vagas prioridade=Atendimento normal 64402	<conf:(0.68)>
servico=Emissão da Carteira de Trabalho 22701 ==>prioridade=Atendimento normal status_atendimento=Atendimento Encerrado 14789	<conf:(0.65)>
servico=Retirada da Carteira de Trabalho 3099 ==>prioridade=Qualquer outra prioridade 2005	<conf:(0.65)>
status_atendimento=Senha Cancelada 2196 ==>servico=Vagas 1376	<conf:(0.63)>
prioridade=Atendimento normal 104722 ==>servico=Vagas status_atendimento=Atendimento Encerrado 64402	<conf:(0.61)>
status_atendimento=Senha Cancelada 2196 ==>servico=Vagas prioridade=Atendimento normal 1323	<conf:(0.6)>
servico=Retirada da Carteira de Trabalho 3099 ==>prioridade=Qualquer outra prioridade status_atendimento=Atendimento Encerrado 1865	<conf:(0.6)>
prioridade=Qualquer outra prioridade 9331 ==>servico=Emissão da Carteira de Trabalho 4867	<conf:(0.52)>

**Tabela 5.4 – Tabela Resultado Apriori - Serviço x prioridade x status do atendimento.**

## Serviço X Prioridade X Status do Atendimento - *Support* baixo

Relação de Regras e Atributos	Confiança
servico=Seguro Desemprego prioridade=Qualquer outra prioridade 386 ==>status_atendimento=Atendimento Encerrado 343	<conf:(0.89)>
servico=Emissão da Carteira de Trabalho status_atendimento=Atendimento Iniciado 206 ==>prioridade=Atendimento normal 183	<conf:(0.89)>
servico=Vagas prioridade=Atendimento prioritário para idosos 174 ==>status_atendimento=Atendimento Encerrado 151	<conf:(0.87)>
servico=Convocação 176 ==>prioridade=Atendimento normal 151	<conf:(0.86)>
servico=Vagas prioridade=Atendimento prioritário para portadores de deficiência 404 ==>status_atendimento=Atendimento Encerrado 345	<conf:(0.85)>
prioridade=Atendimento prioritário para idosos 350 ==>status_atendimento=Atendimento Encerrado 295	<conf:(0.84)>
prioridade=Atendimento prioritário para gestantes 164 ==>status_atendimento=Atendimento Encerrado 138	<conf:(0.84)>
prioridade=Atendimento prioritário para portadores de deficiência status_atendimento=Atendimento Encerrado 412 ==>servico=Vagas 345	<conf:(0.84)>
servico=Emissão da Carteira de Trabalho status_atendimento=Senha Emitida 220 ==>prioridade=Atendimento normal 184	<conf:(0.84)>
prioridade=Atendimento prioritário para portadores de deficiência 487 ==>servico=Vagas status_atendimento=Atendimento Encerrado 345	<conf:(0.71)>
prioridade=Qualquer outra prioridade status_atendimento=Senha Cancelada 248 ==>servico=Emissão da Carteira de Trabalho 162	<conf:(0.65)>
prioridade=Atendimento normal status_atendimento=Chamado pela Mesa 444 ==>servico=Vagas 263	<conf:(0.59)>
prioridade=Atendimento prioritário para idosos status_atendimento=Atendimento Encerrado 295 ==>servico=Vagas 151	<conf:(0.51)>

Tabela 5.5 – Tabela Resultado Apriori - Serviço x prioridade x status do atendimento.

### 5.1 Análise dos Resultados

Com o objetivo de ampliar e melhorar o entendimento dos resultados, ao invés de gerar regras para somente uma comparação entre as 3 variáveis de atributos, foram feitas várias comparações. Inicialmente entre *serviço* e *prioridade* observou-se que quase na totalidade do “serviço de Consulta de vagas”, foram com “prioridade Normal”, com uma confiança de 97% (79217/81918); ou seja, de 81.918 consultas de vagas 79.217 são referentes a prioridade normal.



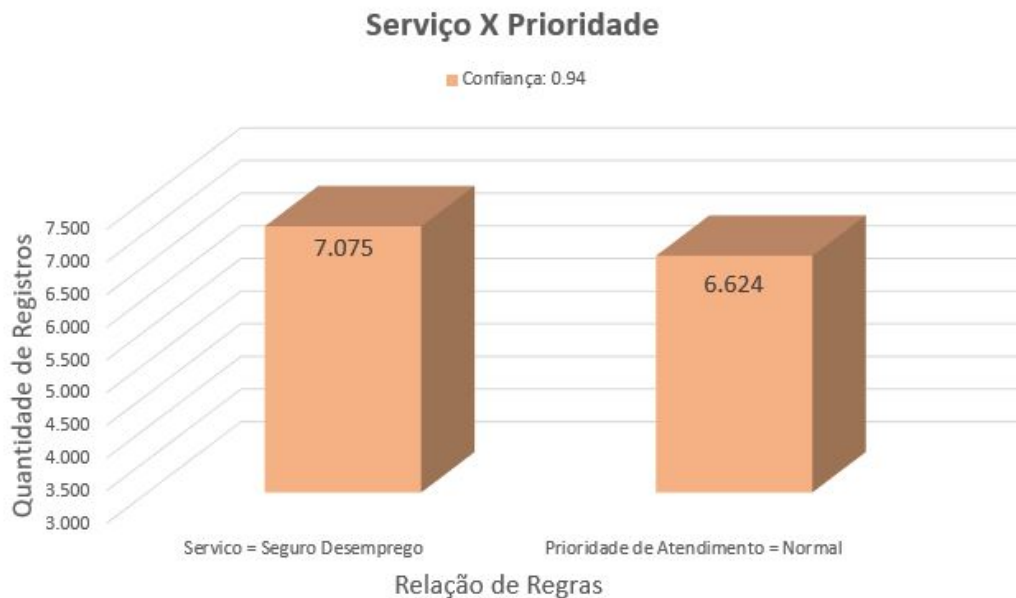
Figura 5.1 – Tabela Análise de Resultados 1.1.

Ainda referente a primeira geração de regras, podemos analisar no gráfico abaixo que de 9.331 registros de *Qualquer outra prioridade*, 4.867 são para Emitir Carteira de Trabalho, com uma confiança de 52%.



**Figura 5.2 – Tabela Análise de Resultados 1.2.**

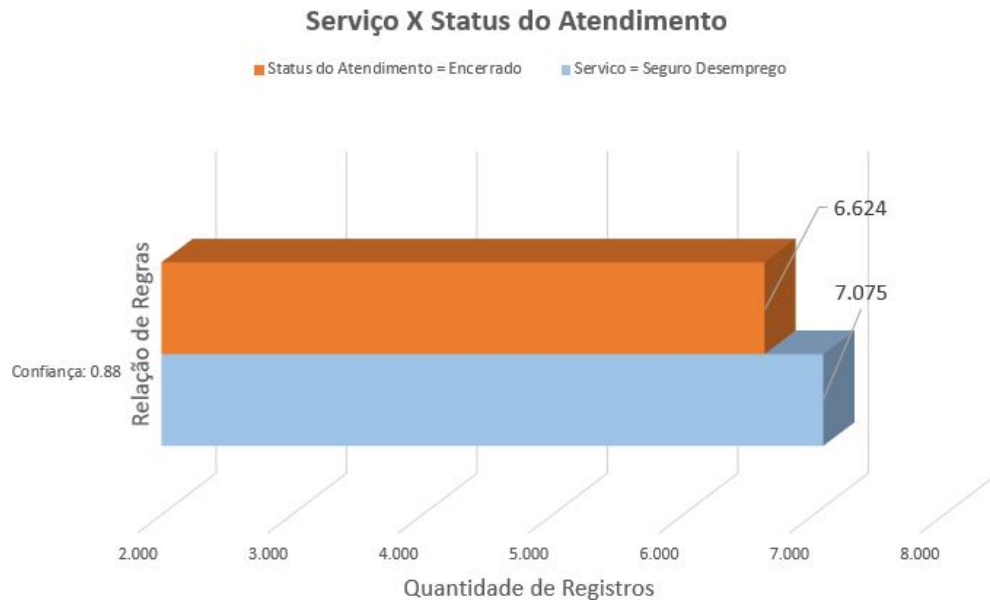
Outra regra gerada da primeira comparação é: Dos 7.075 registros de atendimentos para seguro desemprego, 6.624 são para atendimentos com prioridade normal, além de obter uma confiança de 94% como pode observar no gráfico da figura 5.3.



**Figura 5.3 – Tabela Análise de Resultados 1.3.**

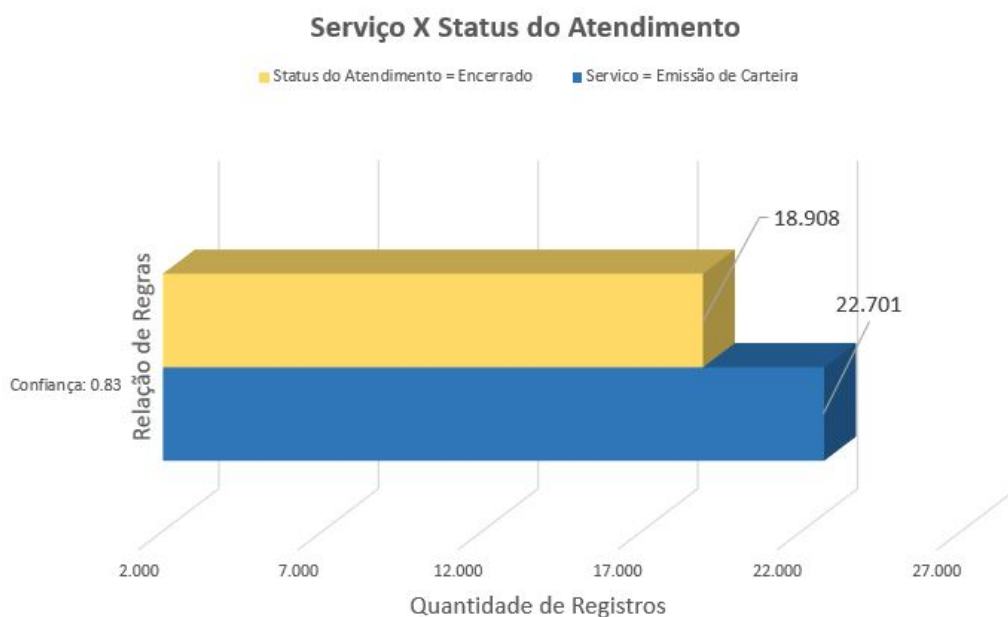
Já na segunda comparação, entre *Serviço* e *Status do Atendimento*, foi constatado que ainda existem muitos atendimentos sem encerramento, seja desistência por motivos pessoais, demora para ser atendido, ou quaisquer outros motivos que devem ser investigados pela equipe responsável; observou-se isto devido aos resultados de algumas regras

geradas, como por exemplo em: De 7.075 registros referentes ao serviço de seguro Desemprego, somente 6.228 foram encerrados; lembrando ainda que a confiança alcançada foi de 88%. Abaixo no gráfico da figura 5.4.



**Figura 5.4 – Tabela Análise de Resultados 1.4.**

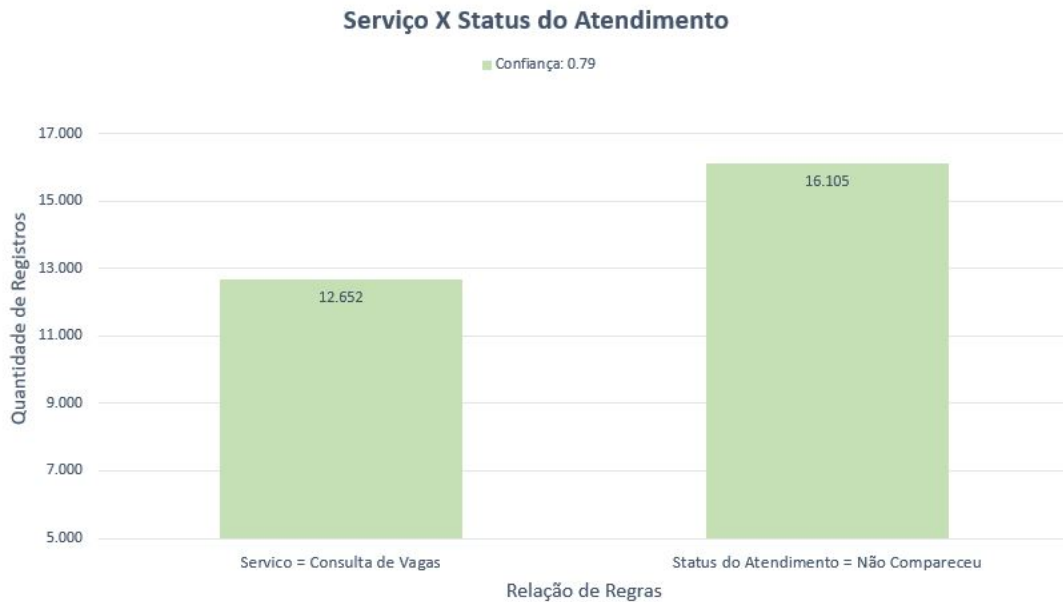
Seguindo adiante no gráfico da figura 5.5 em que a regra gerada, tem-se 22.701 atendimentos para *Emissão de Carteira de Trabalho* onde apenas 18.908 foram devidamente Encerrados e com uma confiança de 83%.



**Figura 5.5 – Tabela Análise de Resultados 1.5.**

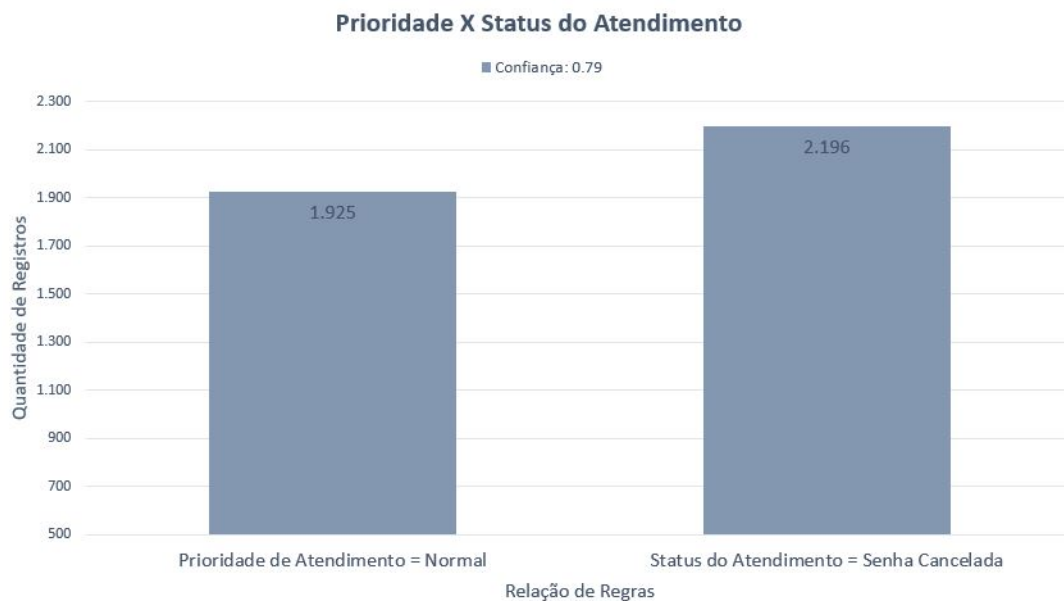
Seguindo a ordem da geração de regras, obtém-se a seguinte: Em 16.105 atendimentos onde os clientes não compareceram, 12.652 são relacionados a busca de vagas de

emprego. O que deve ser analisado é o por que dessa quantidade de clientes que não compareceram ao atendimento, o que aconteceu para não prosseguirem adiante. Lembrando também que a confiança continua alta com 79%. Abaixo o gráfico referente a essa associação.



**Figura 5.6 – Tabela Análise de Resultados 1.6.**

Na Associação entre *Prioridade* e *Status do Atendimento* pode-se evidenciar a regra gerada em que 2.196 atendimentos estavam com status de senha cancelada e dentre esses, 1.925 são relativos a prioridade de atendimento normal. Mais uma vez algo relacionado a atendimentos não finalizados e com confiança acima de 70%, ou melhor, exatamente com 79%; segue abaixo no gráfico 5.7.



**Figura 5.7 – Tabela Análise de Resultados 1.7.**

Após verificar todas as regras geradas pelo Apriori, ficou evidente que a maioria das regras envolvem a prioridade normal, pois o valor atribuído de suporte é muito alto e assim não gerou regras com a quantidade de registros mais baixos, como exemplo as prioridades de idosos, gestantes ou deficientes.

Devido a este motivo, foi necessário refazer o processo e assim alguns reajustes nos parâmetros do algoritmo Apriori. Foi feito alguns cálculos de suporte e resolveu-se a partir daí iniciar o suporte em 0,3% e para isso foi alterado o valor do parâmetro *upperBoundMinSupport* para 0.003 e o valor de *lowerBoundMinSupport* para 0.001. Isso significa que o suporte iniciará em 0,3% de toda a base de dados e irá percorrer até chegar em 0,1% ou alcançar o total de regras limitadas no parâmetro *numRules*.

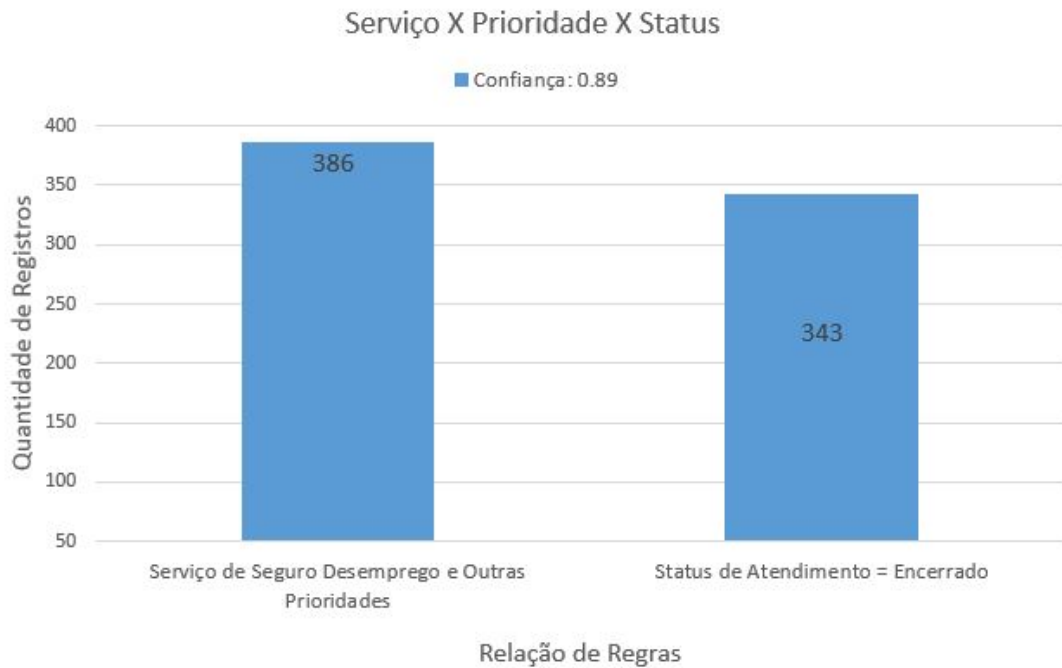
Somente depois dessas mudanças, que foi possível gerar regras referentes a atributos que continham uma pequena quantidade de registros, tais atributos se encontram na tabela 5.5 apresentada anteriormente. Algumas delas estão evidenciadas a seguir:

Relação de Regras e Atributos	Confiança
servico=Seguro Desemprego prioridade=Outras 386 ==>status atendimento=Atendimento Encerrado 343	<conf:(0.89)>
servico=Vagas prioridade=Prioritario para idosos 174 ==>status atendimento=Atendimento Encerrado 151	<conf:(0.87)>
servico=Vagas prioridade=Portadores de deficiência 404 ==>status atendimento=Atendimento Encerrado 345	<conf:(0.85)>
prioridade=Gestantes 164 ==>status atendimento=Atendimento Encerrado 138	<conf:(0.84)>
prioridade=Outras status atendimento=Senha Cancelada 248 ==>servico=Emissão da Carteira de Trabalho 162	<conf:(0.65)>

**Tabela 5.6 – Tabela Resultados 5.1.**

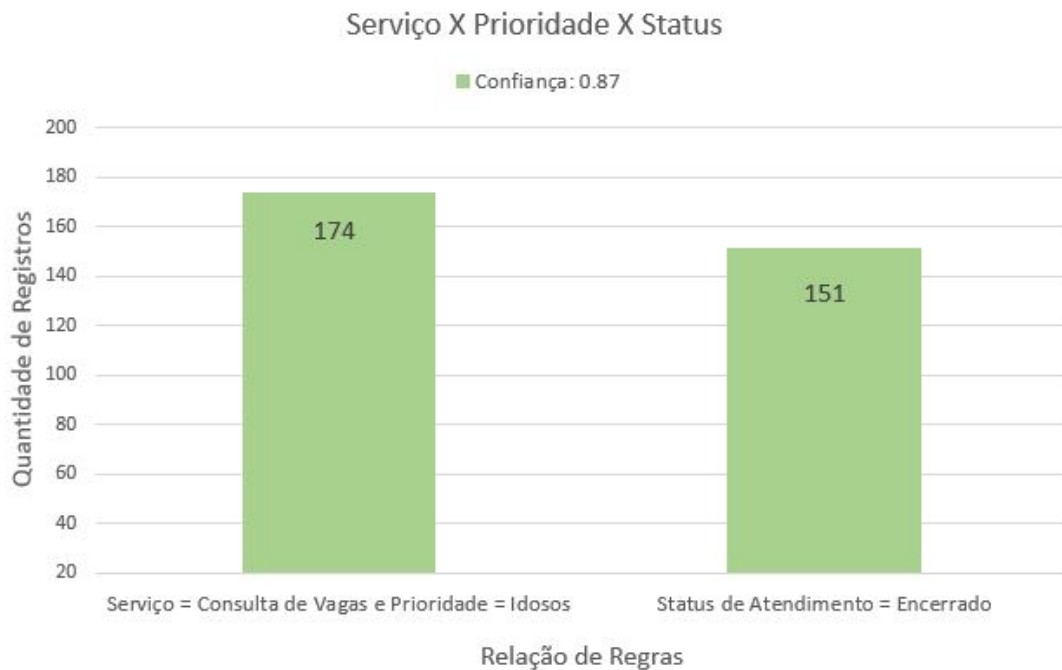
Para uma melhor visualização e comparação dos resultados obtidos da tabela 5.5 exposta anteriormente, se encontra abaixo alguns gráficos. No gráfico da Figura 5.8 a seguir, onde 386 atendimentos de seguro desemprego e outras prioridades, um total de

343 foram encerrados, ou seja, alguns ainda se encontram sem encerramento e com uma ótima confiança de 89%.



**Figura 5.8 – Tabela Análise de Resultados 1.8.**

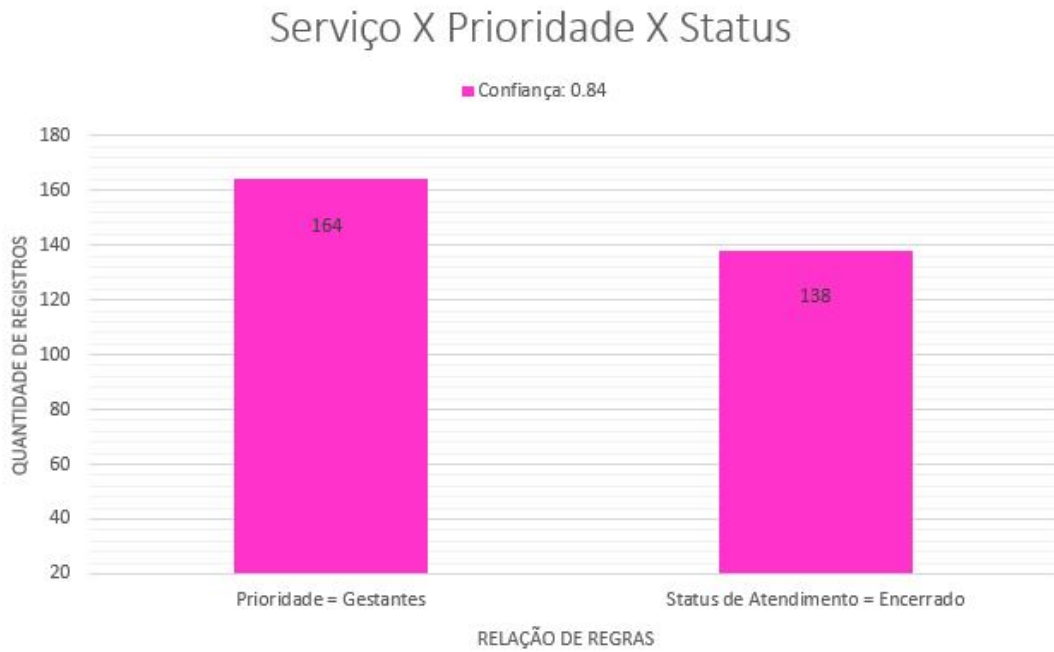
Adiante na terceira regra gerada após a mudança do valor de *support*, foram encontradas 404 registros de consulta de vagas para portadores de deficiência, onde desses registros, 345 foram encerrados, também com a confiança mantida nos 87% como mostra no gráfico da figura abaixo:



**Figura 5.9 – Tabela Análise de Resultados 1.9.**



A próxima associação de regra é a respeito de gestantes, uma informação também importante, onde se encontram 138 registros com status encerrados de 164 atendimentos à gestantes e mantendo a confiança alta.



**Figura 5.10 – Tabela Análise de Resultados 1.10.**

## 6 CONCLUSÃO

Desde a implantação do SGA, essa foi a primeira vez que os dados de registros da base de dados dele foi analisada totalmente. O processo poderia ter sido reduzido se dados já estivessem pré-processados e limpos, assim poderia ter sido usado alguns dos atributos que ficaram de fora como datas e horas dos registros.

O método utilizado no processo de KDD, se mostrou importante para os resultados finais do trabalho e sendo um passo fundamental, pois a mineração exige que os dados estejam padronizados, limpos e codificados, para que os resultados obtidos das regras geradas fossem confiáveis, e foi justamente isso que o KDD ajudou a realizar.

A ferramenta Weka utilizada nesse trabalho, atendeu as necessidades iniciais, que seriam de aplicar através de uma ferramenta viável, todo o processo de KDD desde o pré processamento até o pós processamento dos dados, juntamente com o algoritmo escolhido inicialmente, o Apriori. Esse trabalho teve sua importância, pois a partir da descoberta de conhecimento oculto dos resultados obtidos foi possível repassar à equipe de gerencia e de TI do SINE as informações geradas, para que assim eles realizem logo que possível melhoramentos futuros, tanto no atendimento direto aos clientes, quanto no SGA, isso a partir do que gestores e técnicos acharem necessário.

### 6.1 Trabalhos Futuros

Por motivos maiores ou problemas no pré-processo, alguns atributos foram excluídos desse trabalho, como por exemplo:

- o id do usuário atendente
- Data e hora de chegada do cliente
- Data e hora de senha chamada do cliente
- Data e hora de inicio do atendimento
- Data e hora de finalização do atendimento

Para que esses atributos sirvam no processo de mineração, será necessário um longo tempo de pré-processamento, removendo registros que se encontram vazios, padronizando os valores, convertendo e unindo-os. Isso traria uma maior quantidade de novas informações, principalmente relacionadas a desempenho de atendimento. Se preocupar com o desempenho é fundamental para aumentar a quantidade de atendimentos diários e melhorar a qualidade dos atendimentos. Um melhor desempenho significa atender mais clientes em menos tempo.

Futuramente seria de grande valia a implantação do SGA em todas as unidades do SINE no estado do Tocantins, assim seria possível reunir todos os registros de dados em uma única base de dados e aplicar a mineração na totalidade dessa base, gerando muito conhecimento novo.

## REFERÊNCIAS

- AGRAWAL, R.; IMIELINSK, T.; SWAMI, A. **Mining association rules between sets of items in large databases**. Washington, EUA: Acm sigmod record, 1993.
- ARAUJO, L. G. d. A. **DESENVOLVIMENTO DO MÓDULO DE EXTRAÇÃO DE REGRAS DE ASSOCIAÇÃO PARA O APRIORI-SENTIMENTALL**. Palmas, TO: Centro Universitario Luterano de Palmas, 2016.
- CARVALHO, L. A. V. D. **A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. São Paulo, Brasil: CIENCIA MODERNA, 2005.
- CORREA, U.; MACHADO, M. **Mineração De Dados De Help Desk Usando Rattle - o caso Petrobras**. Rio de Janeiro, Brasil: IBMEC. Rio de Janeiro, 2007. <<http://livros01.livrosgratis.com.br/cp047676.pdf/>>.
- DATE, C. J. **Introdução a Sistemas de Bancos de Dados**. 8ª Ed. Rio de Janeiro, Brasil: CAMPUS, 2004.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The kdd process for extracting useful knowledge from volumes of data. **Communications of the ACM**, ACM, v. 39, n. 11, p. 27–34, 1996.
- FEUSER, R. J. **MINERAÇÃO DE DADOS COM REGRAS DE ASSOCIAÇÃO APLICADA EM DADOS DE UNIDADE DE SAÚDE DE PRONTO ATENDIMENTO**. Pato Branco, PR: UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, 2017.
- FGTAS. **Fundação Gaúcha do Trabalho e Ação Social**. 2019. <<https://fgtas.rs.gov.br/sistema-nacional-de-emprego>>. Acesso em 15/03/2019.
- GOLDSCHMIDT, R.; PASSOS, E. **Data Mining Conceitos Técnicas Algoritmos Orientações e Aplicações**. [S.l.]: Elsevier Trade/Profissional; Edição: 2, 2015.
- GROTH, R. **Data Mining: A hands on approach**. [S.l.]: Prentice hall, 1998.
- HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. [S.l.]: Elsevier, 2011.
- KREMER, R. Sistema de apoio à decisão para previsões genéricas utilizando técnicas de data mining. **Trabalho de Conclusão de Curso. Universidade Regional de Blumenau**, 1999.
- MONARD, M. C.; BARANAUSKAS, J. A. **Sistemas inteligentes-fundamentos e aplicações**. São Paulo, Brasil, 2003. Acesso em 02/04/2019.
- PRATES, W. R. **Aprendizado de máquina: supervisionado e não supervisionado**. 2018.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. [S.l.]: Editora Manole Ltda, 2003.

SGA. **Novo SGA - Support queue management system**. 2019. <<http://novosga.org/>>. Acesso em 08/02/2019.

SINE-TO. **Sistema de Gerenciamento de Atendimento (SGA) - SINE**. 2016. <<https://setas.to.gov.br/>>. Acesso em 02/03/2019.

VASCONSELOS, L. M. R.; CARVALHO, C. L. **Aplicação de Regras de Associação para Mineração de Dados na Web**. Dissertação (Mestrado) — Universidade Federal de Goiás, Goiás, Brasil, 2004.

WEKA. **Weka 3: Data Mining Software in Java**. 2019. <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em 20/02/2019.