# Classificação de ressarcimento de verbas em processos de judicialização da saúde

*Manuscript template for the REPA Journal*

**Marlon Castro 1**  [orcid.org/0000-0002-5830-2090](https://orcid.org/0000-0002-5830-2090)

**Elisson 1**  [orcid.org/0000-0001-5727-2427](https://orcid.org/0000-0001-5727-2427)

**Isac 1**  [orcid.org/0000-0001-5727-2427](https://orcid.org/0000-0001-5727-2427)

**Tiago Cabral 1**  [orcid.org/0000-0001-5727-2427](https://orcid.org/0000-0001-5727-2427)

1 Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil,

**E-mail do autor principal: Marlon Castro** [msc@ecomp.poli.br](mailto:msc@ecomp.poli.br)

**Resumo**

Conforme o site *agenciabrasil.ebc.com.br,* só no ultimo ano houve um acréscimo de 130% nos casos de judicialização da saúde em todo o pais, o direito à saúde garantido na constituição dá respaldo às diversas ações impetradas por diversas pessoas que quase sempre ganham na justiça o direito de receber medicamentos caríssimos ou algum tratamento especifico de saúde o qual seria obrigação do governo arcar.

Em causas onde envolvem o Estado de Pernambuco, por vezes na decisão judicial, fica explicita a obrigação conjunta entre o Governo Estadual e o Federal no sentido de arcar com a compra do medicamento ou tratamento. Como os dados são desconexos, as informações acerca dos medicamentos comprados existem em diversas bases distintas, fazendo-se necessário a junção desses dados a fim de agrupá-los para executar processos de mineração.

**Palavras-Chave*:***Restituições de verba; Arvores de decisão; Judicialização da saúde.

***Abstract***

*According to the website agenciabrasil.ebc.com.br, in the last year alone there was an increase of 130% in cases of judicialization of health throughout the country, the right to health guaranteed in the constitution supports the various actions brought by several people who almost always win in court the right to receive expensive drugs or some specific health treatment which would be the government's obligation to bear.*

*In cases involving the State of Pernambuco, sometimes in the court decision, the joint obligation between the State Government and the Federal Government to pay for the purchase of the drug or treatment is explicit. As data are disconnected, information about purchased drugs exists on several different bases, and it is necessary to merge these data in order to group them together to perform mining processes.*

***Key-words:*** Refunds of funds; Decision trees; Health Judicialization.

## Introdução

Recentemente houve um considerável crescimento em Judicialização da Saúde, apresentando um aumento de 130% entre os períodos de 2017 e 2018 conforme relatório encomendado pelo CNJ em março deste ano. Dentre as ações que mais se destacam, estão as relacionadas à convênios, cerca de 30%, os pedidos de seguros em saúde apontam em 21%, enquanto que o fornecimento de medicamentos e tratamentos hospitalares ficam com 12% dos casos.

Isso em relação à todos os processos no País, no estado de Pernambuco, no entanto, não há pesquisas ou relatórios que apontem números a serem discutidos neste documento.

Mas o que é Judicialização da saúde ?

Quando um paciente, por exemplo, uma individua com câncer de mama, tem indicação para utilizar um tratamento específico, considerado o mais adequado para o seu tipo de tumor, e ele não está disponível no SUS, a única forma dessa paciente obtê-lo pode ser entrando na justiça contra o governo para reivindicar as doses necessárias, já que o acesso universal e igualitário à saúde é um direito garantido pela Constituição. Muitas vezes essa acaba sendo a única alternativa para que a paciente tenha acesso ao tratamento de que precisa. Esse processo se chama judicialização em saúde.

Infelizmente essa alternativa não produz igualdade no acesso ao tratamento, uma vez que nem todas as pacientes sabem que podem ou não fazer para recorrer a esse recurso quando não há o tratamento disponível na rede pública. De modo geral enfrentar um processo judicial enquanto luta contra uma doença é uma tarefa difícil.

Além de se envolver com questões burocráticas exigidas, a paciente precisa lidar com a incerteza sobre o resultado da ação e o medo do câncer evoluir enquanto aguarda pelo tratamento que já deveria estar utilizando para controlá-lo. Uma ação judicial pode demorar muito tempo e é desgastante para uma pessoa fragilizada. Algumas pacientes acabam morrendo esperando por uma decisão na Justiça.

Em ações proferidas pelas diversas comarcas espalhadas pelo estado de Pernambuco, muitas vezes a obrigação do cumprimento da determinação judicial, que deve ser respeitada, cabe ao Estado e à União, na qual muitas vezes apenas o governo de Pernambuco arca com o ônus da determinação. À Longo prazo, estes valores desembolsados pelo Estado vão se acumulando e chegando à casa dos milhões facilmente, já que existem remédios que custam até R$ 500.000,00 apenas uma caixa.

A Secretaria de Saúde por sua vez, por mais registros dos processos que tenha, não dispõe de métodos para identificar as comarcas com mais chances de ressarcimento, o que torna o problema ainda maior, pois o Estado sabe que tem verbas a ressarcir da União, mas não sabe quais e onde estão as mais facilmente acessíveis.

## Fundamentação teórica

### Mineração de dados

### 

Entende-se por Mineração de dados, um conjunto de técnicas e procedimentos computacionais visando extrair conhecimento de bases de dados geralmente grandes. Quase sempre as empresas e órgãos públicos armazenam seus dados em bancos de dados relacionais, que por muitas vezes se interligam com outras bases às vezes até com paradigmas e funcionalidades diferentes. Também é bastante comum, essas organizações que embora tenha sistemas informatizados, utilizem também planilhas de dados para armazenar informações referentes a algum processo ou setor específico.

        Durante anos trabalhando dessa forma, por vezes pode-se chegar a um momento em que a instituição deseja obter determinada informação dessas, só que as mesmas não estão interligadas, possuem dezenas de atributos ou contém informações desorganizadas.

        A partir daí, é feita a coleta dos dados, seleção de atributos, eliminação de atributos inválidos e desnecessários e junção de dados de todas essas bases. Os dados são então tratados de forma a servirem de entrada para algoritmos classificadores, clusterizadores, preditores e diversos outros tipos de algoritmos utilizados para se extrair conhecimento de um montante de dados.

**Área de conhecimento**

Durante os encontros com a Secretaria de Saúde e Professor, a equipe verificou a possibilidade de utilizar um classificador. Com o objetivo de tentar classificar as comarcas que mais tenham possibilidade de ressarcimento.

Para tal feito, tornou-se necessária a disponibilização da base de dados contendo o histórico de recursos com ressarcimento bem sucedidos para que servisse de entrada para o classificador.

Um classificador é um algoritmo que entende as regras de associação entre um conjunto de atributos de forma a prover uma Classe específica.

### Materiais e métodos

Para efeito desta publicação serão utilizados os seguintes materiais:

* **Orange** – Utilizado para toda a modelagem dos dados, este aplicativo é responsável por mais de 80% do trabalho efetuado uma vez que o mesmo possui ferramentas para praticamente todo tipo de modelagem.
* **Weka** – Software utilizado para geração de dados estatísticos sobre os modelos gerados, com fim de confrontá-los com as análises vindas do Orange.
* **Csvkit** – Ferramenta utilizada em linha de comando para Linux que consiste em execução de queries SQL em arquivos CSV. Com este software foi possível extrair algumas informações básicas da distribuição dos dados.

Os procedimentos específicos a seguir definem em ordem cronológica os levantamentos dos dados para a elaboração deste trabalho.

* Recebimento da base de dados por parte da Secretaria de Saúde;
* Análise inicial dos dados, buscando efetuar uma junção e posterior condensação destes numa só tabela;
* Caracterização e eliminação dos atributos (seção 3.3);
* Transformação dos atributos;
* Padronização dos dados;
* Normalização dos dados;
* Criação do modelo;
* Testes e geração de gráficos;

### Descrição da base

A base repassada possui a estrutura conforme figura 1.

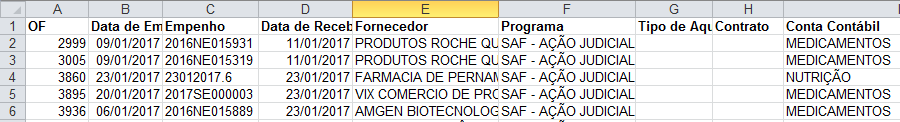


Figura 1 - Histórico de compras 2017 a 2019

Estes dados por si só não retratavam muita coisa, afinal, são apenas dados históricos de compras efetuadas pelo Estado para cumprimento das ordens judiciais. Mesmo assim, não se tinham nesta base dos dados das comarcas, fóruns e cidades nas quais se deram entradas nos processos. E isso é necessário uma vez que o objetivo é “mapear” as comarcas com maior possibilidade de ressarcimento por parte do Governo Federal.

Sendo assim, a figura 2 mostra parte da base principal e tem o que seria necessário ao estudo. No entanto, as bases da figura 1 e 2 são **desconexas** – não possuem chaves ou dados sob os quais podemos juntar essas tabelas – e não foi possível uni-las.

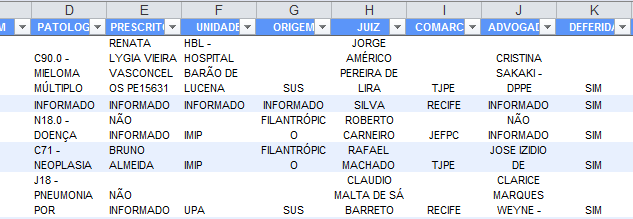


Figura 2 - Base com dados das ações

As bases juntas possuem 37 atributos – 19 da base da figura 1 e 18 da base da figura 2 – e sua maioria são categóricos e só possuem apenas 9 atributos numéricos, em meio a estes alguns serão descartados conforme **seção 3.5** por se tratarem apenas de nº sequencial e outros não contém variações do atributo (todos com o valor “0”).

### Análise descritiva

Dada importância desta analise, vale se destacar que consiste numa etapa importante no processo de mineração uma vez que através dela, antes de efetuar a mineração em si, podemos extrair informações úteis sobre os dados com auxilio de algumas ferramentas, tabelas e gráficos.

A figura 3 mostra a distribuição das ações judiciais separadas por origem. O gráfico permite observar que a maior parte dos processos vem do SUS seguindo de Entidades filantrópicas. Existe também uma parcela dos dados que possui o valor “não informado” presente nos atributos.

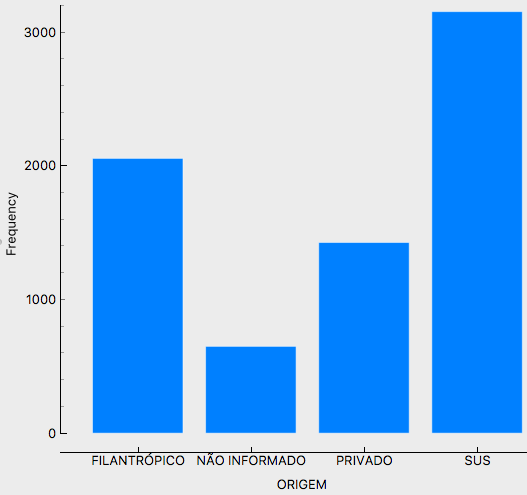


Figura 3 - Distribuição de ações por Origem

A figura 4 permite observar que a maior quantidade de ações julgadas provém da Cidade de Recife, seguindo por Jaboatão dos Guararapes, Olinda, Petrolina e Paulista.

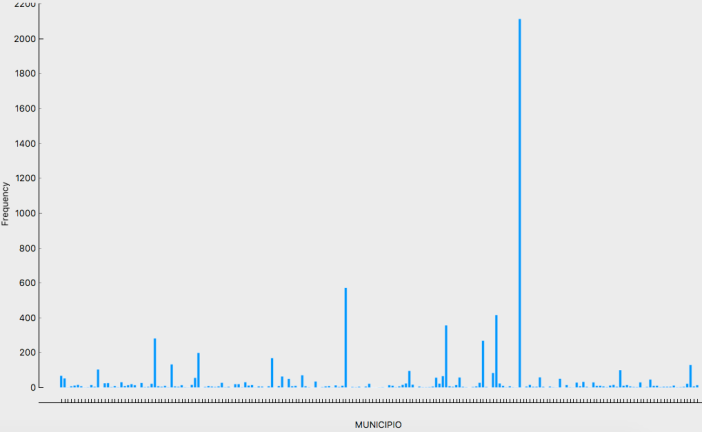


Figura 4 - Distribuição de Ações por Municipios  
Fonte: Autores

Dentre a grande quantidade de municípios, e levando em consideração que essas análises foram feitas em dados brutos, a figura 5 mostra de forma mais detalhada os 9 municípios com mais índice de ações.

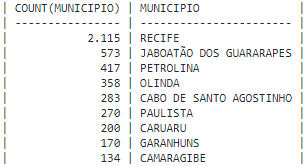


Figura 5 - Distribuição de Municípios detalhado

Os processos de judicialização da saúde se embasam em direitos previstos na Constituição Federal, sendo assim a maioria destas ações são deferidas conforme figura 6. O Direito à saúde é pétreo e indiretamente obriga aos estados, municípios e união a cumprirem o que determinam as sentenças.

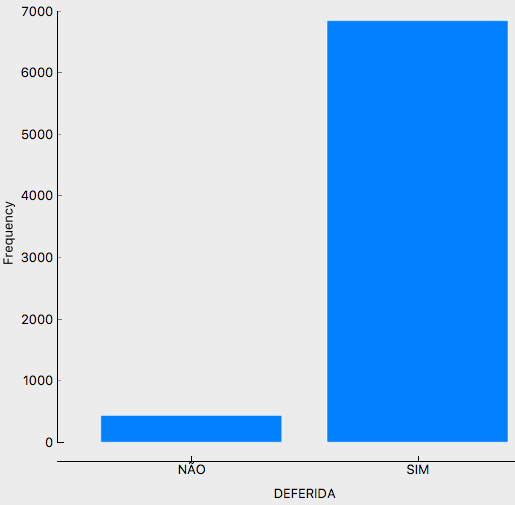


Figura 6 - Proporção de Ações Judiciais

Outra distribuição importante de comentar é a de UNIDADES, ou seja, de quais unidades de saúde mais saíram prescrições que culminaram em sentenças. A figura 7 e 8 nos mostra que o IMIP e Consultórios particulares são os principais meios que originaram esses processos.

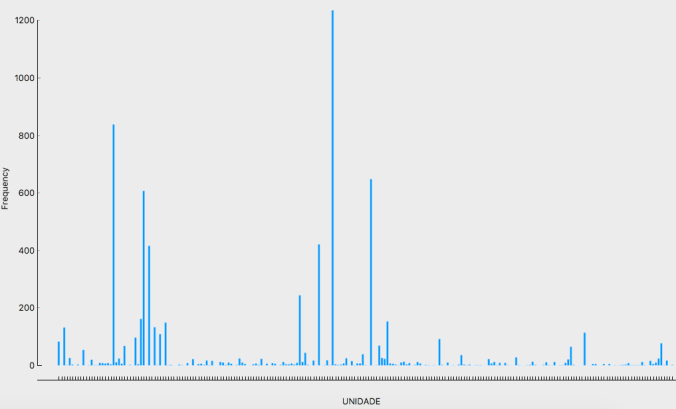


Figura 7 - Distribuição de unidades  
Fonte: Autores

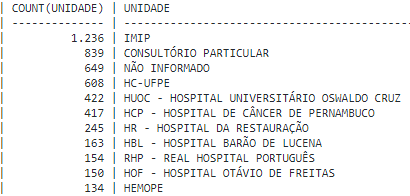


Figura - Detalhes das Unidades

### Sobre o KDD

O KDD – Descoberta de conhecimento em bases de dados foi amplamente difundido por consequência de um trabalho acadêmico elaborado por Fayyad et al., em 1996: “*From Data Mining to Knowledge Discovery in Databases*” onde os autores batizaram o processo.

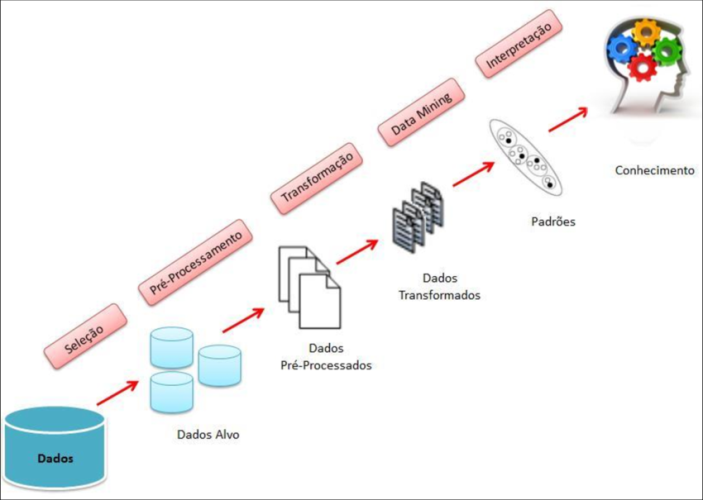


Figura 9 - Etapas do KDD  
Fonte: <https://www.researchgate.net/figure/Figura-1-Etapas-que-constituem-o-processo-de-KDD_fig1_303864815>

As etapas do KDD são

* Seleção: os dados são selecionados de uma ou mais fontes de dados, a partir de um entendimento do negócio em alto nível.
* Pré-processamento: os dados são sanitizados, incluindo limpeza de dados, eliminação de inconsistências, tratamento de dados faltantes etc.
* Transformação: os dados são transformados de acordo com a fase posterior, onde os padrões escondidos são realmente extraídos. Por exemplo, essa etapa pode incluir redução de dimensionalidade e mudanças de codificação.
* Mineração de Dados: após pré-processamento e transformação dos dados, modelos preditivos ou de sumarização são construídos a partir da área da Inteligência Artificial conhecida como Aprendizagem de Máquinas ou mesmo da Estatística. É aqui onde os padrões escondidos e valiosos são encontrados.
* Interpretação: nessa etapa, os padrões escondidos e os modelos inteligentes são avaliados e interpretados, normalmente em parceria com os especialistas do domínio e com o cliente.

### Seleção dos dados

Nas análises iniciais ficou evidente que necessitava de apenas 8 atributos dos 37 existentes na base, esse procedimento foi efetuado com o auxilio do especialista da área.

A citar, o campo CADAST, que possuía em média 10 ocorrências distintas na base foi eliminado por se tratar apenas das iniciais do nome da pessoa que inseriu os dados na base. Portanto, atributo sem relevância alguma para o problemas em questão.

### Pré-processamento

Assim como a maioria das bases de dados, esta também teve muitas inconsistências, a começar da falta de chave para unir as duas tabelas mais importantes do conjunto de dados, até atributos cujos campos só continham o valor “0” em todas as ocorrências, campos com valores vazios e atributos numéricos cujos valores continham letras.

Conforme falado foram feitas diversas eliminações verticais excluindo atributos com distorção na base. Ainda, foi feito descarte da base da figura 2 por falta de chave para união com a base da figura 1. Nesta ultima, foram incluídos os atributos COMARCA, MUNICIPIO e ORIGEM de forma aleatória visando a manipulação da base e deixá-la pronta para o processo de Data Mining.

Estas questões foram abordadas no decorrer de reuniões com os especialistas nos dados e foram enumeradas dificuldades diversas em relação ao fornecimento dos dados. Uma vez que os bancos de dados existentes no órgão são distintos, o que dificulta a extração de posterior união das bases para serem usadas no procedimento.

Foi feita também a eliminação horizontal de 1135 linhas da base, pois continham muitos atributos com valores vazios.

A transformação dos atributos categóricos em numéricos foi conseguida através da ferramenta scikit que gera as variáveis dummy, no Orange 3º processo também é bem simples e sem nenhuma linha de programação.

A figura 10 mostra o estado das variáveis antes da aplicação do procedimento.

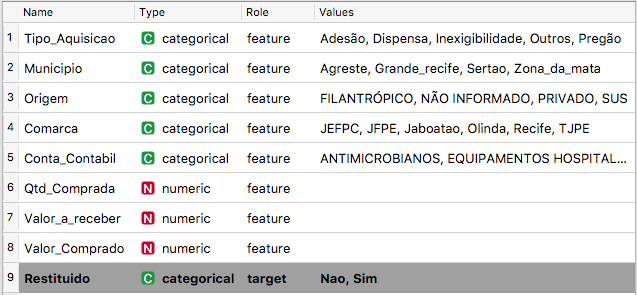


Figura 10 - variáveis de entrada

Na figura 11 percebe-se a mudança na quantidade de colunas da base e criação de diversos atributos binários decorrentes do processo de padronização e normalização.

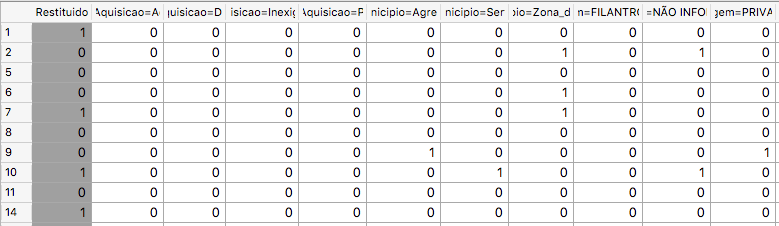


Figura 11 - Dados em formato dummy

A coluna RESTITUIU foi gerada para que pudesse se treinar o classificador, foram atribuídos aleatoriamente valores balanceando a base de treinamento em aproximadamente 50%.

Esses dados não devem contribuir para a análise final do problema elencado, pois são dados inexistentes que foram preenchidos automaticamente de forma a permitir o treinamento do modelo, já que não houve tempo suficiente para que as bases reais chegassem a ser disponibilizadas.

Na Tabela 1 é apresentado o dicionário de dados do experimento após todo o pré-processamento.

|  |  |
| --- | --- |
| **Num** | **Nome da Variável** |
|  | Tipo\_aquisicao |
|  | Municipio |
|  | Origem |
|  | Comarca |
|  | Conta\_Contabil |
|  | Qtd\_Comprada |
|  | Valor\_a\_receber |
|  | Valor\_Comprado |
|  | Restituido (Target do modelo) |

Tabela 1 - Dicionário de dados  
Fonte: Autores

### Mineração e Experimentos

Para implementação dos métodos foram utilizados o scikit-learn, que é open soure e utilizada para exploração de dados com diversos algoritmos, possui praticamente tudo o que se precisa para aplicar as técnicas adequadas.

Na geração de modelos e testes do algoritmo foi utilizada a ferramenta Orange por sua facilidade com programação visual dos elementos do processo de mineração. Também foi utilizado o weka de forma a gerar estatísticas detalhadas do desempenho do modelo.

Inicialmente foram feita uma divisão na base com 70% dos dados para treinamento e 30% para teste. As configurações utilizadas para geração da Arvore de decisão são mostradas na figura 12.

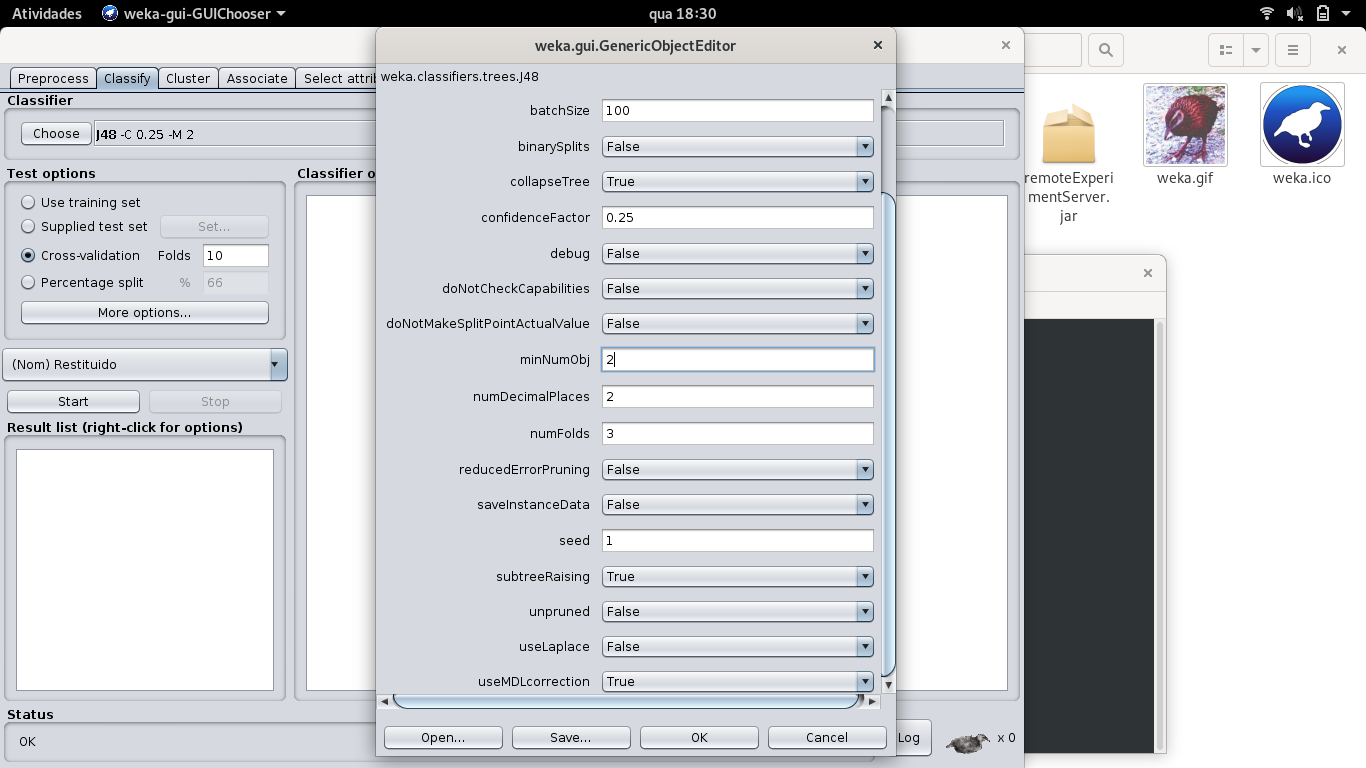


Figura 12 - Configurações da arvore de decisão

Também, para visualização do desempenho do modelo foi feita a análise utilizando os dados de treinamento para teste e utilizada a validação cruzada com 10 Folds (Padrão).

Inicialmente foi realizado um treinamento e teste com o mesmo conjunto de dados, e o desempenho do classificador foi superior a 90% conforme a figura 13.

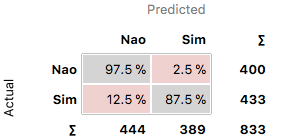


Figura 13 - Matriz de confusão para o conjunto de treinamento sendo utilizado como teste

Após a analise inicial utilizando o mesmo conjunto para treinamento e teste foi feita um treinamento com o mesmo conjunto só que desta vez utilizando o processo de K-Folds com valor 10.

A figura 13 mostra a matriz de confusão para o desempenho do algoritmo.

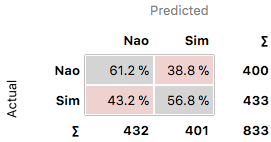


Figura 14 - Matriz de confusão com a validação cruzada

### Análise e discussão

Nesta etapa serão apresentadas as estatísticas dos experimentos realizados com o Algoritmo de Arvore de Decisão.

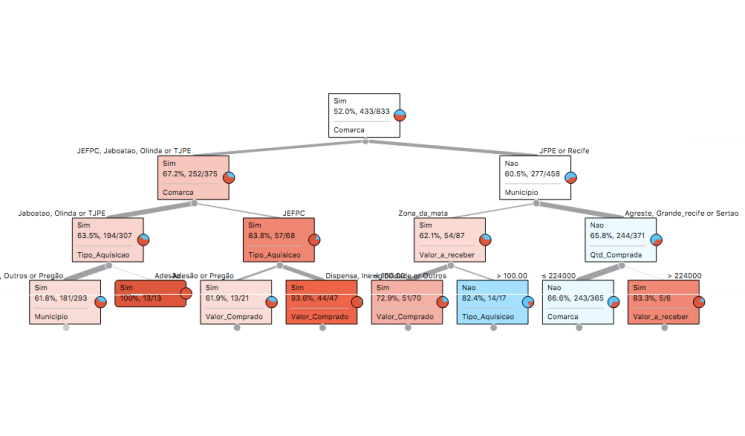
Na figura 15 tem-se a arvore de decisão gerada e com ela, todos os testes dos experimentos foram efetuados.

Figura 15 - Arvore de decisão gerada

Como os dados utilizados foram gerados aleatoriamente, não se consegue de fato treinar um modelo de forma consistente já que as relações criadas dessa forma não são reais.

Através da ferramenta Weka pode-se gerar as estatísticas de precisão do modelo. A figura 15 mostra a baixa taxa de acerto do algoritmo por motivos de geração de muitos dados aleatórios.

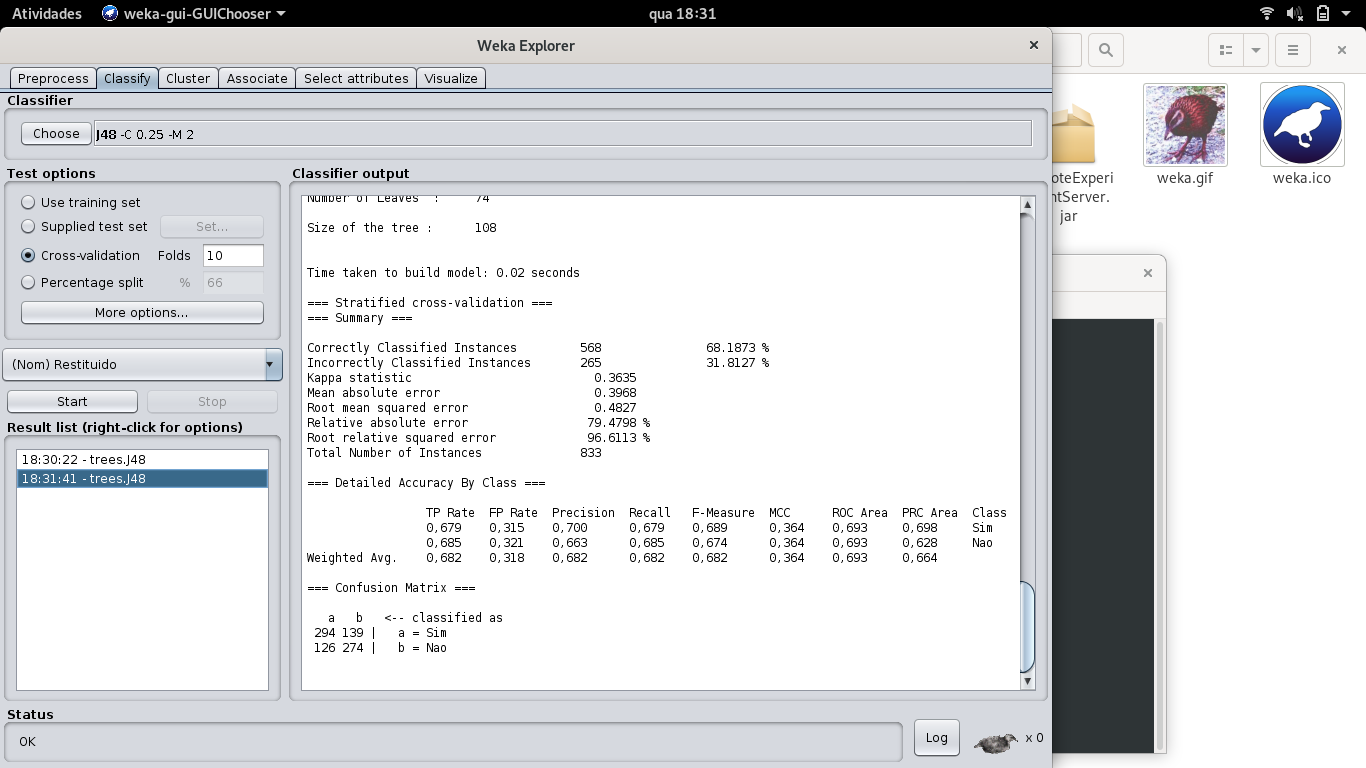


Figura 16 - Estatisticas do algoritmo

### Conclusão

Este trabalho verificou a possibilidade de aplicação do algoritmo