

Saúde do Trabalhador

Padrões em Afastamentos e Desligamentos

Ivair Nobrega Luques¹

¹Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET-RJ

`ivair.luques@eic.cefet-rj.br`

Resumo. *Planejar ações eficazes na garantia dos direitos da Saúde do Trabalhador é uma tarefa de grande complexidade. Extrair informações das bases de dados públicas, que apoiem a tomada de decisão dos gestores de instituições de defesa dos trabalhadores é essencial. Padrões Frequentes é uma metodologia de Mineração de Dados utilizada para verificar a existência de padrões em um conjunto de dados. O algoritmo a ser aplicado neste trabalho é o Apriori que tem por objetivo identificar regras de associação entre as diferentes classes de informação disponibilizadas.*

1. Introdução

“No ano de 1985, o número de acidentes de trabalho, segundo os dados oficiais da previdência social, foi de 1.077.861 casos, com 4.384 mortes, uma média de mais de 12 mortes por dia. E eu começava a entender que o acidente de trabalho era um gravíssimo problema de saúde pública, negligenciado, até hoje, como tal.”[de Vasconcellos and De 2007].

A questão da saúde do trabalhador não se restringe ao acompanhamento dos acidentes de trabalho:

“A área de Saúde do Trabalhador, no Brasil, tem uma conotação própria, reflexo da trajetória que lhe deu origem e vem constituindo seu marco referencial, seu corpo conceitual e metodológico. A princípio é uma meta, um horizonte, uma vontade que entrelaça trabalhadores, profissionais de serviços, técnicos e pesquisadores sob premissas nem sempre explicitadas.”[Minayo-Gomez and da Fonseca Thedim-Costa 1997]

Dentro deste cenário de complexidade, é fundamental obter uma visão mais clara das relações entre os diversos fatores que o compõem, através dos dados oficiais divulgados.

Em 1985, após a oitava Conferência Nacional de Saúde, foi realizada a primeira Conferência Nacional da Saúde do Trabalhador. Uma das ações realizadas posteriormente foi a criação do Centro de Estudos da Saúde do Trabalhador e Ecologia Humana (Cesteh), em 10 de dezembro de 1985, que integra a estrutura da Escola Nacional de Saúde Pública Sérgio Arouca (Ensp), da Fundação Oswaldo Cruz (Fiocruz).

O Cesteh atua nas áreas de Saúde, Trabalho e Ambiente, desenvolvendo atividades de Ensino, Pesquisa e Serviço. A área de Ensino é responsável pela formação de recursos humanos para o campo da Saúde do Trabalhador e Ecologia Humana. Por meio de cursos de pós-graduação e estágios, o Cesteh forma pesquisadores, professores e técnicos para atuação e fortalecimento do campo da Saúde do Trabalhador, no âmbito do SUS e de outras instituições.[da Silva and Da 2000]

O Cesteh desenvolve pesquisas de caráter integrador e interdisciplinar abordando temas tradicionais no campo, como estudos clássicos em setores industriais que geram exposição com risco à saúde dos trabalhadores e consequências socioambientais (benzeno, amianto, mercúrio, chumbo, agrotóxicos, entre outros), além de pesquisas sobre relações de produção e gênero, formação e comunicação em saúde do trabalhador, saúde de profissionais da saúde e da educação, até temas emergentes como nanotecnologia, cronobiologia, alterações genéticas e desastres ambientais.

No Serviço, o Cesteh conta com o Ambulatório de Saúde do Trabalhador e o Laboratório de Toxicologia como referências no campo da Saúde do Trabalhador. Esses estão articulados com outros setores das políticas públicas e da sociedade para enfrentar os problemas e doenças relacionadas ao trabalho. Seus profissionais, com formação em diferentes áreas do conhecimento, participam de diversas atividades de ensino e pesquisa.

O Ministério da Saúde(MS), o Ministério do Trabalho e Previdência Social(MTPS), o Instituto Brasileiro de Geografia e Estatística (IBGE), e alguns outros órgãos do governo estadual e municipal, disponibilizam dados públicos que podem ser acessados pelos cidadãos.

O Cesteh acessa estas bases de dados para, através do uso de técnicas de Estatística, obter resultados que possam apoiá-los em suas pesquisas e planejamento de ações em defesa à saúde do trabalhador. Uma das bases utilizadas contém as informações do Relatório Anual de Informações Sociais(RAIS), disponibilizada pelo MTPS. No entanto algumas das necessidades ainda não foram plenamente atendidas.

O objetivo deste trabalho é Identificar padrões de relação entre os diversos tipos de afastamento, e também dos desligamentos, e algumas informações sobre o trabalhador, informadas na RAIS.

Buscando uma complementação aos resultados obtidos pelos Métodos Estatísticos, utilizaremos a Mineração de Dados. Optamos pela metodologia de Padrões Frequentes com aplicação do algoritmo Apriori, para atingir o objetivo proposto.

Existem trabalhos anteriores que utilizam informações das bases de dados do INSS para estudar a viabilidade de seu uso na avaliações de desligamentos e afastamentos do trabalho. [Boff et al. 2002] Mas não com esse olhar específico demandado pelo Cesteh.[VERAS 2000]

Nas seções seguintes detalharemos o Relatório Anual de Informações Sociais(RAIS), Mineração de Dados e Padrões Frequentes. Podemos considerar estas 3 áreas do conhecimento como o eixo principal do Referencial Teórico que embasa este trabalho.

2. RAIS - Relatório Anual de Informações Sociais

A gestão governamental do setor do trabalho conta com o importante instrumento de coleta de dados denominado de Relação Anual de Informações Sociais - RAIS. Instituída pelo Decreto nº 76.900, de 23/12/75, a RAIS tem por objetivo:

- o suprimento às necessidades de controle da atividade trabalhista no País;
- o provimento de dados para a elaboração de estatísticas do trabalho;
- a disponibilização de informações do mercado de trabalho às entidades governamentais.

Os dados coletados pela RAIS constituem expressivos insumos para atendimento das necessidades:

- da legislação da nacionalização do trabalho;
- Realizar experimentos computacionais comparativos em relação a outras metodologias componentes do estado da arte.
- da legislação da nacionalização do trabalho;
- de controle dos registros do FGTS;
- dos Sistemas de Arrecadação e de Concessão e Benefícios Previdenciários;
- de estudos técnicos de natureza estatística e atuarial;
- de identificação do trabalhador com direito ao abono salarial PIS/PASEP.

2.1. Base de Dados da RAIS

A base de dados da RAIS é disponibilizada aos cidadãos pelo Ministério do Trabalho e da Previdência Social. Os dados de 1985 a 2015 podem ser obtidos. O volume de dados é grande, visto que apenas os dados de 2015 correspondem a 70 milhões de registros. Os arquivos estão no formato txt, com separação por ponto e vírgula, e agrupados por Unidade da Federação(DF) e ano. Os dados podem ser obtidos no link <ftp://ftp.mtps.gov.br/pdet/microdados/RAIS/>. Para obter melhores condições de ajuste ao escopo deste estudo deveremos realizar uma seleção de atributos e uma seleção de dados.

2.2. Seleção de Atributos

Cada arquivo de informações da RAIS possui uma estrutura abrangendo 45 atributos:

- Bairros do Município de São Paulo
- Bairros do município de Fortaleza
- Bairros do município do Rio de Janeiro
- Causa do primeiro afastamento do empregado/servidor no ano-base
- Causa do segundo afastamento do empregado/servidor no ano-base
- Causa do terceiro afastamento do empregado/servidor no ano-base
- Causa do desligamento
- Classificação Brasileira de Ocupações, criada em 2002 - atualizada em 23/08/2004
- Classe de Atividade Econômica, segundo classificação CNAE - versão 2.0
- Classe de Atividade Econômica segundo a classificação CNAE/95 (CNAE 1.0, revisada em 2002) (614 categorias)
- Indicador de vínculo ativo em 31/12
- Faixa Etária do trabalhador
- Faixa de horas contratuais
- Faixa de remuneração média de dezembro do trabalhador em salários mínimos
- Faixa de remuneração média do ano do trabalhador em salários mínimos
- Faixa de tempo de emprego
- Grau de instrução - a partir da RAIS2008
- Quantidade de horas contratuais por semana
- Idade do trabalhador (quando acumulada representa a soma das idades)
- Indicador de CEI vinculado
- Indicador de optante pelo SIMPLES
- mês da admissão do trabalhador
- mês da demissão do trabalhador

- Município onde o empregado esteja trabalhando ou prestando serviço
- Município de localização do estabelecimento
- Nacionalidade
- Natureza Jurídica
- Indicador se o empregado/servidor de portador de deficiência habilitado ou beneficiário reabilitado
- Quantidade total de dias de afastamento do empregado/servidor no ano-base
- Raça e Cor do Trabalhador
- Remuneração do trabalhador em dezembro (valor nominal)
- Remuneração de dezembro em salários mínimos (quando acumulada representa massa salarial)
- Remuneração média do trabalhador (valor nominal)
- Remuneração média do ano em salários mínimos (quando acumulada representa massa salarial)
- Subclasse de Atividade Econômica, segundo classificação CNAE - versão 2.0
- Sexo
- Tamanho do estabelecimento
- Tempo de emprego do trabalhador
- Tipo de admissão
- Tipo de estabelecimento
- Tipo de deficiência/Beneficiário habilitado
- Tipo de vínculo empregatício

O Cesteh indicou quais as informações dos trabalhadores seriam relevantes para este trabalho:

- Causa 1º Afastamento /ano
- Causa 2º Afastamento /ano
- Causa 3º Afastamento /ano
- Causa do Desligamento
- Quantidade de dias afastados no ano
- Classificação Brasileira de Ocupações
- Sexo
- Classe de Atividade Econômica
- Subclasse de Atividade Econômica
- Faixa Etária
- Raça / Cor
- Faixa tempo de emprego

2.3. Seleção dos Dados

Considerando o volume de dados de cada ano, e visando uma maior velocidade no processamento dos dados, diminuição da necessidades de configuração do computador a ser utilizado, optamos por utilizar o arquivo com as informações do ano de 2015 do estado do Amazonas.

3. Metodologia

A metodologia aplicada consiste na realização das etapas de Análise Exploratória de Dados, seguida de ações de Pré-Processamento dos Dados e posterior aplicação de uma ou mais técnicas de mineração de dados, com o apoio de Métodos Estatísticos, se necessário.

3.1. Análise Exploratória dos Dados

Conhecer os seus dados é a primeira grande tarefa em mineração de dados, sendo extremamente útil na etapa de processamento destes dados.[Han et al. 2012]. Dentre os atributos selecionados podemos identificar dois como sendo o alvo do comportamento que desejamos observar: Afastamentos e Desligamentos. Os demais serão considerados como variáveis que podem influenciar, mais ou menos, o comportamento dos atributos alvos.

Dentre os atributos de influência, se destacam os de Grau de Instrução e Raça/Cor. Eles apresentam uma distribuição de dados concentrada em um dos possíveis valores.

Esta circunstância poderá impactar em um peso maior na identificação das regras de associação. Avaliaremos se esta hipótese se confirma após realizarmos os experimentos desta pesquisa.

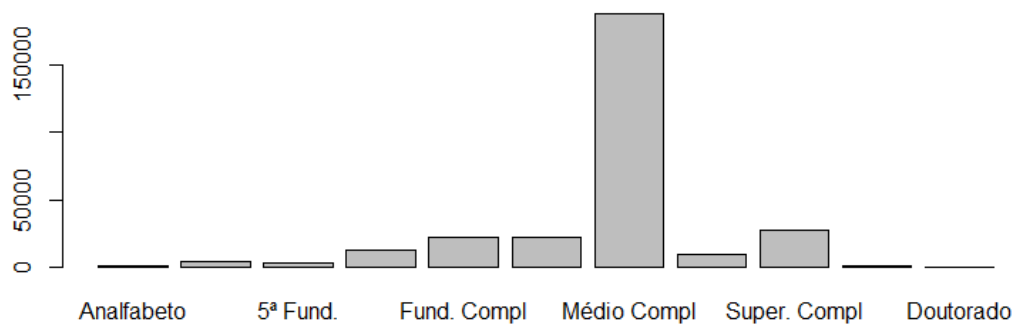


Figura 1. Grau de Instrução

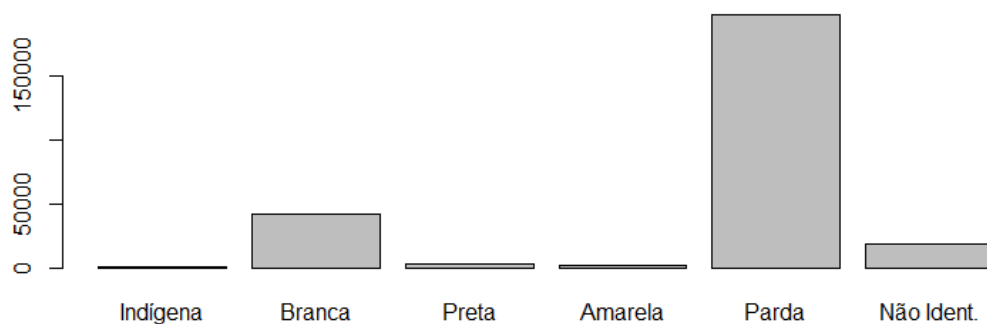


Figura 2. Raça / Cor

Esta circunstância poderá impactar em um peso maior na identificação das regras de associação. Avaliaremos se esta hipótese se confirma após realizarmos os experimentos desta pesquisa.

O comportamento da distribuição dos atributos alvo também aponta para uma concentração em alguma das possíveis categorias. Uma parcela considerável dos registros, cerca de XXX apresenta a categorização de Sem Afastamento. Se o objetivo é verificar as relações existentes nos casos de afastamento será necessário filtrar estes dados. No caso dos Desligamentos há uma maioria na categorização de Demissão sem justa causa, mas não com a mesma incidência que no caso dos afastamentos.

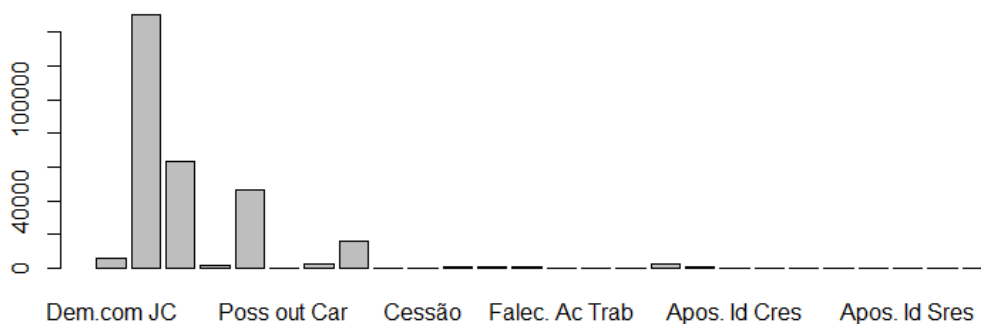


Figura 3. Desligamentos

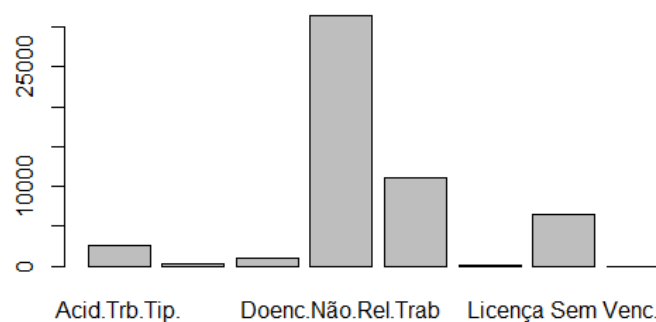


Figura 4. Afastamentos

3.2. Pré- Processamento

”Os Banco de Dados do mundo real de hoje são altamente suscetíveis a ruído, perdas e inconsistência de dados, além de serem tipicamente muito grandes e terem múltiplas, e

heterogêneas fontes de origem. Baixa qualidade dos dados leva a baixa qualidade de resultados de mineração de dados.”[Han et al. 2012] Desta forma faz-se necessário verificar a necessidade e alternativas de atuar sobre os dados.

As ações de Pré-Processamento de Dados realizadas neste caso foram relativas a Limpeza de Dados, Redução dos Dados e Adaptações nas Observações. [Elmasri and Navathe 2015]

3.2.1. Limpeza de dados

Como nossos atributos alvo são afastamentos e desligamentos, os registros de nossa Base de Dados que não possuam preenchimento de informações em nenhum destes atributos devem ser eliminados. Não fazem parte do escopo de nossa pesquisa, neste momento. Esta ação impactou na redução de 901.720 registros para 332.773 em nossa base.

3.2.2. Redução de dados

Como o arquivo original da RAIS continha 45 atributos e a nossa pesquisa abrangeia a ação de apenas 10 destes foi realizada uma Redução de dimensionalidade para que os atributos sem relevância para o Cesteh fossem desconsiderados em nossos experimentos. Restaram apenas os atributos de desligamento, afastamento, sexo, faixa etária, número de funcionários na empresa, tempo de serviço, CBO e Raça/Cor.

3.2.3. Adaptações nas Observações

A estrutura original dos registros da RAIS prevê em um único registro até três afastamentos e um desligamento. Com o objetivo de verificar o impacto de cada ocorrência em separado, realizamos a operação de gerar, a partir de um único registro, um registro específico para cada ocorrência.

Por exemplo, se um registro A possui afastamento1 preenchido, afastamento2 preenchido, iremos gerar então dois registro destintos com apenas um atributo de afastamento.

3.2.4. Classificação Brasileira de Ocupação (CBO)

A Composição do código da CBO é de cinco dígitos, representando a seguinte estrutura e quantidades :

1 - 10 Grandes Grupos 2 - 47 Subgrupos Principais 3 - 192 Subgrupos 4 - 596 Grupos de Base 5 - 2422 Ocupações

Consideraremos uma visão por Subgrupo para a CBO para reduzir a dimensionalidade dos dados a serem avaliados.

3.3. Dados da RAIS de 2015 no estado do Amazonas

Após as operações de pré-processamento chegou-se uma distribuição de atributos por categoria modificada para as informações contidas no arquivo da RAIS de 2015 para o Amazonas.

Podemos destacar um equilíbrio na distribuição pelas categorias de número de funcionários da empresa e em pelo menos quatro opções da Classificação Brasileira de Ocupação (CBO). Há uma categoria que se destaca em Grau de Instrução, com predominância significativa dos que possuem Ensino Médio Completo e da Raça Parda. Sesenta por cento dos trabalhadores são do sexo Masculino e um terço deles possuem entre 30 e 39 anos.

MotivDesliga	sexo	raca	Idade	Escolar
Dem.sem JC :150.203	masc.:197.565	Indígena : 1.024	30 a 39:110.099	Médio Compl :212.407
Term. Contr. : 63.051	fem. :135.208	Branca : 47.682	18 a 24: 74.241	Super. Compl : 35.608
Desl.sem JC : 46.232		Preta : 3.501	25 a 29: 66.240	Fund. Compl : 25.245
Trans s/ Onus: 16.003		Amarela : 2.492	40 a 49: 50.346	Médio Incompl.: 24.658
Dem.com JC : 5.907		Parda :224.446	50 a 64: 27.238	6ª a 9ª Fund. : 14.648
(Other) : 9.735		Não Ident.: 20.083	65 ou +: 2.703	Super.Incompl.: 10.759
NA's : 41.642		NA's : 33.545	(Other): 1.906	(Other) : 9.448

NumFunc	Afastamento	CBO
250 a 499:51.034	Sem Afasta. :279.621	521 : 30.214
50 a 99 :44.965	Doenc.Não.Rel.Trab: 31.367	411 : 25.627
500 a 999:39.804	Licença Mater. : 11.039	731 : 25.193
1000 ou +:38.644	Licença Sem Venc. : 6.524	514 : 21.358
100 a 249:31.045	Acid.Trb.Tip. : 2.655	513 : 14.278
(Other) :78.852	Doenc.Relac.Trab : 1.097	(Other):216.069
NA's :48.429	(Other) : 470	NA's : 34

Figura 5. Dados de 2015 no Amazonas

4. Metodologias de Mineração de Dados

Mineração de Dados deveria ser melhor nomeado como Mineração de Conhecimento a partir de dados. Várias são as abordagens de extração de conhecimento. Devido ao objetivo desta pesquisa precisamos identificar relações entre os atributos de Afastamentos e os demais atributos, e o mesmo em relação aos Desligamentos. Optamos por identificação de RAs (Regras de Associação) utilizando o conceito de identificação de Padrões Frequentes.

4.1. Padrões Frequentes

Padrões Frequentes é um método de se minerar conhecimento procurando por padrões recorrentes. As associações e correlações entre os atributos envolvidos irão identificar padrões escondidos.[James et al. 2014] Existem várias técnicas diferentes de se minerar padrões frequentes. Utilizaremos o algoritmo Apriori.[Agrawal et al. 1994]

4.2. Algoritmo Apriori

O Algoritmo Apriori visa identificar Regras de Associação, ou seja, encontrar elementos que implicam na presença de outros elementos numa mesma transação. Há um princípio

básico deste algoritmo que apoia este objetivo: Qualquer subconjunto de um conjunto de itens frequentes devem ser frequentes.[Han et al. 2012]

Suporte e Confiança são medidas parametrizáveis de condições para execução do Apriori. Suporte mede a probabilidade da transação envolver a união entre os dois atributos, enquanto a confiança mede a probabilidade de caso um ocorra, o outro ocorra também. Os ajustes nos valores de suporte e confiança acarretam resultados diferentes na identificação das regras de associação.

Uma medida resultante, que relaciona confiança e suporte é a lift.[Han et al. 2007] A importância da lift é que se o valor da mesma é menor que 1 isto sinaliza que não a associação entre as mesmas é fraca. Quanto maior que 1 ela for mais representativa é a associação entre os atributos. Por exemplo, um valor de lift de 1,83 entre os atributos A e B indica uma forte correlação entre eles, portanto se A ocorre existe uma chance 83 por cento maior de que B ocorra.

A lógica do algoritmo pode ser representada por um pseudo-código.

C_k : Candidatos a Itens-frequentes de tamanho k

L_k : Itens-frequentes de tamanho k

$L_1 = \{\text{itens frequentes}\};$

para ($k = 1; L_k \neq \emptyset; k++$)

C_{k+1} = candidatos gerados de L_k para cada transação t na base

Incrementar o contador de todos os candidatos em C_{k+1} que estejam contidos em t

L_{k+1} = candidatos em C_{k+1} com suporte mínimo

Retorne k, L_k

Figura 6. Algoritmo Apriori

5. Experimentos Realizados

Algumas opções de parametrização dos valores de suporte e confiança foram testados para identificar quais valores nos gerariam uma saída contendo um número de regras de associação que nos permitisse avaliar o comportamento desejado.

Por exemplo, quando executamos com um parâmetro de confiança igual a 0.9 para os desligamentos, foram geradas apenas 3 regras e nenhuma delas envolvendo o atributo alvo. Por outro lado, quando aumentamos a condição de suporte para 0.5, foram geradas 212 regras, o que prejudicou consideravelmente a avaliação.

A parametrização que gerou os resultados que apresentaremos nesta pesquisa foi de Suporte = 0,1 e Confiança = 0,5.

5.1. Resultados Afastamentos

O total de registros processados na execução relativa aos afastamentos foi de 332.773 observações. O valores mínimos de suporte foi de 0.10001, bem próximo ao valor do parâmetro, e variou até o valor máximo de 0.6629, evidenciando uma boa variação.

O valor mínimo de confiança foi de 0.5158, também próximo ao parâmetro, no entanto o valor máximo foi de 1.0000. Este valor deve ser avaliado com bastante atenção pois indica uma relação de 100 por cento dos casos para os atributos relacionados. A relação que gerou a confiança de 100 por cento é entre os atributos sexo = feminino e afastamento por licença maternidade. Desta forma está justificado o resultado que o modelo gerou.

Os resultados de lift mostram que houve regras de saída que não indicavam uma forte relação de associação, visto que o valor mínimo apontado foi de 0.874, ou seja, menor que 1. Todavia foi registrado valor de lift máximo de 1.813, casos com aumento de 81 por cento de casos em uma relação.

Sobre as regras apontadas com valores mais altos de confiança, considerando as 5 mais bem ranqueadas, vimos que a de mais alto valor foi a já avaliada anteriormente. Dentre as demais destacam-se as que relacionam com os afastamentos e os demais atributos com considerável incidência na base: raça parda, nível de instrução ensino médio completo, sexo masculino e idade entre 30 e 39 anos. Vale ressaltar no entanto que houve diferenciação entre o afastamentos relacionados, o que deve ser considerado na avaliação de possíveis regras de associação.

suporte	confiança	lift
Min. :0.1001	Min. :0.5158	Min. :0.874
1st Qu.:0.1164	1st Qu.:0.6012	1st Qu.:1.006
Median :0.1519	Median :0.6517	Median :1.079
Mean :0.2042	Mean :0.6660	Mean :1.109
3rd Qu.:0.2514	3rd Qu.:0.7240	3rd Qu.:1.141
Max. :0.6629	Max. :1.0000	Max. :1.813

Figura 7. Sumário dos Afastamentos

Regra de Associação	Suporte	confiança	Lift
sexo=Fem. => Afastamento=Licença Mater	0,1315096	1	1,11812948
sexo=masc. raca=Parda, Escolar=Médio Compl => Afastamento=Doenc.Não.Rel.Trab	0,1550459	0,7966167	1,13498827
Idade=50 a 64 => Afastamento=Doenc.Não.Rel.Trab	0,1039095	0,7796443	1,13211226
sexo=masc., Escolar=Médio Compl, Afastamento=Doenc.Não.Rel.Trab => raca=Parda	0,1550459	0,7729319	1,11659017
Idade=30 a 39, Afastamento=Doenc.Não.Rel.Trab => raca=Parda	0,1400888	0,7408218	1,1174663

Figura 8. Exemplo dos Resultados dos Afastamentos

5.2. Resultados Desligamentos

O total de registros processados na execução relativa aos afastamentos foi de 291.131 observações. O valores mínimos de suporte foi de 0.1022, bem próximo ao valor do parâmetro, e variou até o valor máximo de 0.6799, evidenciando uma boa variação e quase idêntico ao encontrado no processamento dos afastamentos.

O valor mínimo de confiança foi de 0.5104, mais uma vez próximo ao encontrado nos afastamentos, no entanto o valor máximo desta vez foi de 0.7373, que não indica nenhuma questão bem específica e rara.

Os resultados de lift mostram que com os desligamentos também ocorreram regras com valores menores que 1, mas o que chama mais atenção neste caso é a ocorrência de um valor máximo de 2.599, indicando uma relação de alta associação entre os atributos. A situação em questão envolve os atributos de tempo de empresa igual a 1 ano e o desligamento por término de contrato. Podemos avaliar que há muitos contratos cujo prazo de validade seja de 1 ano ou menos que 2 anos, e que ao término destes ocorra um desligamento, já previsto pelo término do contrato.

Em relação aos desligamentos relacionamos em destaque as 5 mais bem ranqueadas, mas que envolvam necessariamente um atributo alvo. O resultado demonstra claramente o destaque para o desligamento por demissão sem justa causa. Isto se deve ao fato de que um percentual de 47 por cento das ocorrências de desligamento sejam por este motivo. A única que aparece com um atributo alvo diferente é justamente a que possui alto valor de lift.

support		confidence		lift	
Min.	:0.1022	Min.	:0.5104	Min.	:0.8909
1st Qu.	:0.1221	1st Qu.	:0.6104	1st Qu.	:1.0037
Median	:0.1472	Median	:0.6518	Median	:1.0354
Mean	:0.2043	Mean	:0.6481	Mean	:1.0725
3rd Qu.	:0.2489	3rd Qu.	:0.6996	3rd Qu.	:1.0720
Max.	:0.6799	Max.	:0.7373	Max.	:2.5995

Figura 9. Sumário dos Desligamentos

Regra de Associação	Suporte	confiança	Lift
Idade=30 a 39,Escolar=Médio Compl => MotivDesliga=Dem.sem JC	0,1226733	0,5677992	1,1005369
sexo=fem., Idade=18 a 25 => MotivDesliga=Dem.sem JC	0,1166279	0,5819721	1,1280076
sexo=masc.,Escolar=Sup Compl => MotivDesliga=Dem.sem JC	0,2154975	0,5511068	1,0681829
sexo=masc.,raca=Parda,Escolar=Médio Compl => MotivDesliga=Dem.sem JC	0,1580113	0,5709144	1,106575
TempoEmpr=01 => MotivDesliga=Term.	0,1335756	0,5629904	2,5995458

Figura 10. Exemplo dos Resultados dos Desligamentos

6. Considerações Finais

Após a realização dos experimentos relacionados e dos resultados obtidos podemos chegar a algumas conclusões e também indicações relevantes de continuidade deste trabalho.

6.1. Conclusões

O uso de padrões frequentes mostrou ser uma ferramenta útil na identificação de associações entre os dados e o grau de dependência entre os mesmos ao ser aplicada na base de dados da RAIS.

Podemos confirmar a hipótese inicial de que algumas das associações obtiveram destaque em razão do significativo volume de dados de determinado atributo.

Existem inúmeras demandas do CESTEHE que seriam atendidas com a aplicação de padrões frequentes, e outras técnicas de Mineração de Dados.

6.2. Trabalhos Futuros

O aprendizado com a experiência desta pesquisa nos permitiu perceber uma gama considerável de alternativas de continuidade para este trabalho. Relacionamos algumas delas, sinalizando a real intenção de dar sequência a esta pesquisa:

- Realizar os experimentos deste trabalho para o ano de 2015 abrangendo todas as Unidades da Federação.
- Realizar os experimentos deste trabalho com os algoritmos FP-Tree e ECLAT, de padrões frequentes.
- Realizar agrupamentos dos afastamentos e dos desligamentos por CBO, utilizando k-means.
- Identificar regras de associações apenas entre afastamentos e desligamentos, sem incluir outros atributos.

7. References

Referências

- Agrawal, R., Srikant, R., et al. (1994). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, volume 1215, pages 487–499.
- Boff, B. M., Leite, D. F., and Azambuja, M. I. R. (2002). Morbidade subjacente à concessão de benefício por incapacidade temporária para o trabalho. *Revista de Saúde Pública*, 36(3):337–342.
- da Silva, C. T. and Da, C. T. (2000). Saúde do trabalhador: um desafio para qualidade total no Hemorio.
- de Vasconcellos, L. C. F. and De, L. C. F. (2007). Saúde, trabalho e desenvolvimento sustentável: apontamentos para uma política de Estado.
- Elmasri, R. and Navathe, S. B. (2015). *Fundamentals of database systems*. Pearson.
- Han, J., Cheng, H., Xin, D., and Yan, X. (2007). Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery*, 15(1):55–86.
- Han, J., Kamber, M., and scientist) Pei, J. C. (2012). *Data mining : concepts and techniques*. Elsevier/Morgan Kaufmann.
- James, G., Witten, D., and Hastie, T. (2014). An introduction to statistical learning: With applications in r.
- Minayo-Gomez, C. and da Fonseca Thedim-Costa, S. M. (1997). *A construção do campo da saúde do trabalhador: percurso e dilemas*. Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz.
- VERAS, M. (2000). Afastamentos e licenciamento segundo o sistema rais-caged. *SEMINÁRIO NACIONAL “ESTATÍSTICAS SOBRE DOENÇAS E ACIDENTES DO TRABALHO NO BRASIL: SITUAÇÃO E PERSPECTIVAS”*. Anais. São Paulo: Fundacentro, pages 47–51.