



UNIVERSIDADE ESTADUAL DO NORTE DO PARANÁ
CAMPUS LUIZ MENEGHEL

RENAN ANGELINI FERRARI

**APLICAÇÃO DOS PROCESSOS DE MINERAÇÃO DE
DADOS EM UMA BASE DE DADOS DE AVALIAÇÃO
INSTITUCIONAL**

Bandeirantes

2017

RENAN ANGELINI FERRARI

**APLICAÇÃO DOS PROCESSOS DE MINERAÇÃO DE
DADOS EM UMA BASE DE DADOS DE AVALIAÇÃO
INSTITUCIONAL**

Trabalho de Conclusão de Curso apresentado ao Centro de Ciências Tecnológicas da Universidade Estadual do Norte do Paraná para obtenção do título de Bacharel em Sistemas de Informação e Licenciado em Computação.

Orientador: Prof. Me. Glauco Carlos Silva

Bandeirantes

2017

Dedico este trabalho a minha família que sempre esteve presente e incentivou meus estudos.

Agradecimentos

Agradeço primeiramente a UENP por disponibilizar e manter o curso de Sistemas de Informação e a todo o corpo docente que ministrou as aulas durante todo o período da minha graduação, sempre muito prestativos e direcionando o caminho dos discentes. Quero agradecer de maneira especial ao meu orientador Prof. Me. Glauco Carlos Silva por me auxiliar e direcionar neste trabalho que é muito importante na minha vida acadêmica e a todos que de certa forma me ajudaram neste trabalho e em minha graduação.

RESUMO

Este trabalho apresenta uma concepção da mineração de dados e casos de uso sobre a base de dados de avaliação institucional da UENP, aplicando processos da mineração de dados buscando regras de associação para revelar tendências. Foi encontrado correlações entre as características e demonstrado através de hipóteses o que estas correlações representam. Em seguida foram realizadas avaliações sobre estas hipóteses relatando a frequência com que estas regras de associação acontecem. Este conhecimento obtido possibilitou o encontro de soluções e respostas para questões que antes, sem o auxílio da mineração, seria difícil conseguir, no qual os gestores poderão agir com uma base sólida sabendo qual ponto precisa de alteração.

Palavras-Chave: Mineração de dados; Avaliação institucional.

ABSTRACT

This work presents a conception of data mining and use cases on the basis of data of institutional evaluation of UENP, processes of mining data seeking association rules to uncover trends. It was found correlations between the characteristics and demonstrated by means of assumptions that these correlations represent. Then evaluations were conducted on these hypotheses relating the frequency with which these association rules occur. This knowledge gained enabled the solutions and answers to questions that before, without the aid of mining, it would be difficult to achieve, in which managers will act with a solid base knowing which point needs to change.

Keywords: Data Mining; Institutional Evaluation.

Sumário

1. Introdução	1
1.1 Problema	2
1.2 Justificativa	3
1.3 Objetivo Geral	4
1.3.1 Objetivos Específicos	4
1.4 Metodologia	5
1.4.1 Seleção	6
1.4.2 Pré-Processamento	6
1.4.3 Transformação	7
1.4.4 Mineração de dados	7
1.4.4.1 Hipótese	7
1.4.4.2 A mineração	8
1.4.5 Avaliação	8
2. Fundamentação teórica	9
2.1 Introdução ao processo de Data Mining	9
2.2 Etapas da mineração de dados	10
2.3 Descrição de alguns métodos utilizados em Data Mining	12
2.3.1 Associação	12
2.3.2 Classificação	13
2.3.3 Padrões sequenciais	15
2.3.4 Agrupamento	16
2.4 Descrição de algumas técnicas utilizadas em Data Mining	17
2.4.1 Árvores de decisão	17
2.4.2 SVM – Support Vector Machines	18
2.4.3 Conjuntos aproximados	18
2.4.4 Classificação Bayesiana	18
2.4.5 Classificação Baseada em Regras (Rule-Based Classification)	19
2.5 Descrição de alguns algoritmos utilizados em Data Mining	19
2.5.1 C4.5	19
2.5.2 k – Means	20
2.5.3 Algoritmo Apriori	20
2.5.4 Algoritmo FP-Growth	21
2.5.5 Algoritmo MiRABIT	22
2.6 Ferramentas Data Mining	22
2.6.1 Oracle Data Mining	23

2.6.2 SQL Server Data Mining (SASS)	24
2.6.3 Apache Spark.....	24
2.6.4 RapidMiner.....	25
2.6.5 Weka.....	26
2.7 Avaliação Institucional	26
3. Desenvolvimento	28
3.1 A Mineração	28
3.1 Seleção	30
3.2 Pré-Processamento	30
3.2.1 Procedimentos realizados – Confiabilidade das Informações	31
3.3 Mineração de dados – Hipóteses.....	32
3.3.1 Hipótese 1	32
3.3.1.1 Observação.....	32
3.3.1.2 Características	32
3.3.2 Hipótese 2	33
3.3.2.1 Observação.....	33
3.3.2.2 Características	33
3.3.3 Hipótese 3	33
3.3.3.1 Observação.....	33
3.3.3.2 Características	34
3.3.4 Hipótese 4	34
3.3.4.1 Observação.....	34
3.3.4.2 Características	34
3.3.5 Hipótese 5	34
3.3.5.1 Observação.....	34
3.3.5.2 Características	34
3.3.6 Hipótese 6	34
3.3.6.1 Observação.....	35
3.3.6.2 Características	35
3.3.7 Hipótese 7	35
3.3.7.1 Observação.....	35
3.3.7.2 Características	36
3.3.8 Hipótese 8	36
3.3.8.1 Observação.....	36
3.3.8.2 Características	36
3.4 Avaliação dos dados	36

3.4.1 Hipótese 1	37
3.4.2 Hipótese 2	38
3.4.3 Hipótese 3	40
3.4.4 Hipótese 4	41
3.4.5 Hipótese 5	41
3.4.6 Hipótese 6	42
3.4.7 Hipótese 7	43
3.4.8 Hipótese 8	45
4. Conclusão	46
5. Trabalhos Futuros.....	47
6. Referências.....	50
ANEXO A – Quadros com as medidas referente aos fatos.....	55

Lista de Figuras

Figura 1 - Etapas do processo de mineração de dados.....	11
Figura 2 - Região do estado de São Paulo	17
Figura 3 - Etapas da mineração de dados	28

Lista de Quadros

Quadro 1 - Classificação climática de Köppen -Geiger	15
Quadro 2 - Avaliação Aluno	55
Quadro 3 - Avaliação Docente	56
Quadro 4 - Avaliação Aluno-Docente	58
Quadro 5 - Avaliação Estagiário.....	58
Quadro 6 - Avaliação Técnico	59

1. Introdução

A avaliação institucional é um mecanismo utilizado para avaliar e acompanhar os processos de uma instituição de ensino. Com isto pode-se reparar, incluir ou excluir itens que estão ligados aos processos de uma instituição de ensino com o objetivo de aperfeiçoar os serviços por ela prestados.

Com a mineração de dados, a avaliação e estudo destes processos se torna mais eficiente pois, conforme explica Cardoso (2008, p.497), a mineração de dados é “capaz de revelar, automaticamente, o conhecimento que está implícito em grandes quantidades de informações armazenadas nos bancos de dados de uma organização”.

A mineração de dados tem a proposta de encontrar nas informações conhecimentos necessários a organização, que auxiliem na tomada de decisão administrativa e forneça uma visão do caminho percorrido pela instituição ou conhecimentos relevantes ao objeto de estudo em questão.

Neste trabalho foi aplicado os processos da mineração de dados na base de dados de avaliação institucional da UENP com o objetivo de encontrar correlações entre as características para apoiar a tomada de decisão administrativa.

Segundo Schmitt (2005, p.16 apud FERNANDEZ 2003), a mineração de dados “automatiza o processo de descobrir relações e padrões nos dados e apresenta resultados que poderão ser utilizados em um sistema de suporte a decisão automatizado, ou acessado por tomadores de decisão”.

Conforme aponta Viana (2013, p.45), “devido ao grande montante de dados esses padrões dificilmente seriam descobertos com métodos mais tradicionais como consultas a base de dados ou relatórios.”

O modo como se trabalham com os dados vem se modificando com o tempo, no qual as entidades organizacionais em diferentes vertentes de negócio necessitam de novos processos de manipulação de dados.

De acordo com a organização DAMA International (2006, apud AMORIM, 2006, p.10) “a administração de dados consiste no desenvolvimento e execução de

estratégias, práticas e procedimentos para o processo de gerência do ciclo de vida completo dos dados em uma empresa”. Segundo Amorim (2006, p. 10), “é essencial o estudo da Administração de Dados para garantir a qualidade dos dados que são inseridos e recuperados em uma base de dados”.

1.1 Problema

Devido à grande quantidade de dados e informações geradas a cada momento pelas entidades organizacionais é necessário aos tomadores de decisão embutir inteligência ao negócio tendo em mãos informações úteis e corretas para trabalhar e agir. Grandes riquezas estão contidas nestas informações que quando bem administradas e trabalhadas, podem ser exploradas de maneira ágil e trazer soluções e respostas necessárias ao desenvolvimento da organização.

Neste trabalho foi buscado e encontrado correlações na base de dados de avaliação institucional da UENP com a aplicação dos processos da mineração de dados com o objetivo de melhorar a avaliação dos processos da instituição.

Realizar à avaliação dos processos da instituição somente com a consulta a base de dados ou relatórios pode ser ineficiente, no qual pode não ser encontrado relações ou respostas de interesse da instituição. Porém, através da mineração, à avaliação destes processos se torna muito mais eficiente.

Com os resultados obtidos através da mineração de dados pode-se maturar processos, práticas e metodologias utilizadas e resolver problemas existentes nos processos da instituição.

Deve-se fazer bom uso dos dados gerados pelas organizações, como a avaliação institucional, sendo assim, foram realizadas pesquisas que trouxeram resultados sólidos dos objetos de estudo em questão. Com isto, pode-se ajudar de alguma forma a organização a prestar um serviço (ensinar, transferir conhecimento, qualificação profissional) de melhor qualidade através das mudanças ocorridas com base nestes conhecimentos encontrados.

1.2 Justificativa

A mineração de dados é uma área de estudo em crescimento e de grande relevância que devido a sua utilidade prática obteve notoriedade no meio acadêmico e em produtos comerciais (CORTÊS et. al., 2006).

Com o patrimônio digital disponível de diversas naturezas, seja industrial, comercial, governamental, científico, educacional, entre outros, há muitas riquezas em potencial nestes ativos que devem ser exploradas e a mineração de dados surgiu justamente para dar este suporte.

A avaliação institucional é um destes ativos que devem ser explorados, no entanto, a busca de respostas e soluções utilizando somente relatórios ou pesquisas a base de dados pode se tornar uma trabalho árduo e deixar de encontrar respostas relevantes que muitas das vezes poderia ser a resposta para o que se procura.

Sendo assim, com o uso da mineração de dados este processo de busca por respostas e soluções se torna mais eficiente. Segundo Elder (2015), “a aprendizagem a partir dos dados é extremamente poderosa, e seu uso está transformando a tomada de decisões de empresas dos mais diversos setores em um ritmo acelerado, poupando dinheiro e até mesmo vidas”.

Na mineração de dados pode-se montar objetos de estudo, onde constrói-se uma base destinada a um fato ocorrido (vendas, variação de preço de ações, extinção ou proliferação de animais, entre outros) e através de metodologias e ferramentas especializadas pode-se encontrar informações ricas e de maneira rápida.

Este trabalho apresenta a aplicação dos processos da mineração de dados na base de dados de avaliação institucional da UENP no qual foi buscado e encontrado correlações entre as características. Foram identificadas características que incidem positivamente ou negativamente sobre os processos ou serviços da instituição.

Sendo assim, justifica-se esta produção de conhecimento, com o objetivo de se encontrar problemas, descobrir tendências e comportamentos, conhecer técnicas, práticas e metodologias que estão ligados aos processos da UENP e apresentar este conhecimento adquirido mostrando o resultado que a mineração de dados pode acrescentar as tarefas da organização.

1.3 Objetivo Geral

Como objetivo geral foi realizado a aplicação dos processos da mineração de dados na base de dados de avaliação institucional da UENP buscando correlações entre as características para que possam responder dúvidas ou solucionar problemas da instituição que sem o auxílio da mineração seria difícil conseguir.

1.3.1 Objetivos Específicos

O estudo apresentado tem como objetivo mostrar a importância da mineração de dados nas diferentes áreas do mercado, como a indústria, logística, agronegócio, pesquisas espaciais, entre outros, embora o estudo esteja direcionado ao campo educacional. Sua aplicação correta levanta valor para o negócio buscando informações, obtendo uma visão real do problema que pretende-se solucionar.

A mineração realizada sobre a base de dados da avaliação institucional produziu informações úteis aos tomadores de decisão e a todos os *stakeholders* do processo de ensino-aprendizagem para usufruir destes ativos gerados e aprimorar suas habilidades, práticas e metodologias no âmbito profissional, quanto alterar, caso necessário, no âmbito organizacional.

O estudo seguiu as etapas do processo de mineração conforme abaixo:

- Seleção: Tem como objetivo a identificação da(s) base(s) que serão utilizadas para coleta dos dados necessários a pesquisa.
- Pré-Processamento: Tem como finalidade formatar e corrigir os dados que não estão aptos a entrar no processo de análise.
- Transformação: A tarefa desta etapa é transformar os dados de acordo com o que se precisa para utilizá-los no processo de mineração.
- Mineração de Dados: O objetivo desta etapa é o de encontrar informações para o problema levantando no início do projeto.
- Avaliação: O ofício desta etapa é garantir a confiabilidade da informação levantada durante o processo de mineração.

1.4 Metodologia

Metodologia é derivado de duas palavras, métodos e logos, que significam respectivamente, organização e estudo sistemático, ou seja, são os passos a seguir que irão levar a um resultado. (UFRGS, 2009 apud FONSECA, 2002).

Segundo Tartuce(2006) a metodologia científica trata de método e ciência, sendo o método o caminho para se chegar a um objetivo e a metodologia o estudo do método, ou seja, os procedimentos a seguir para realizar uma pesquisa.

Uma pesquisa é concebida segundo “à sua abordagem, sua natureza, seus objetivos e seus procedimentos” (UFRGS, 2009, p.31).

Quanto a abordagem este trabalho tem como alicerce uma pesquisa quantitativa em vista da necessidade de se mensurar as características do objeto de estudo com o intuito de classificá-lo e definir a que grupo o mesmo pertence perante suas semelhanças.

Quanto a sua natureza este trabalho resulta em uma pesquisa aplicada produzindo valores úteis ao campo da ciência dos dados e a administração de dados, demonstrando uma concepção, importância e aplicabilidade da mineração de dados nos diversos setores existentes e o quanto seu uso pode gerar riqueza em conhecimento nos mais diversos objetos de estudos.

Quanto ao objetivo trata-se de uma pesquisa exploratória visando encontrar informações úteis para o objeto de estudo em questão que irão auxiliar a organização a tomar um posicionamento e agir com uma base sólida de informações.

Quanto aos procedimentos leva-se a pesquisa bibliográfica pois “pesquisa bibliográfica é feita a partir do levantamento de referências teóricas já analisadas [...]” (UFRGS, 2009, p.12 apud FONSECA, 2002), podendo também ser considerada uma pesquisa de levantamento em vista dos resultados levantados através do estudo de uma população.

1.4.1 Seleção

A base de dados que foi utilizada neste projeto é a da avaliação institucional da UENP, na qual contém informações declaradas por grande parte dos envolvidos nos processos da instituição.

De acordo com Frizzo(2003), “a avaliação serve para pensar e planejar a prática docente e nessa reflexão torna-se um recurso para melhorar os processos pedagógicos” e afirma:

A avaliação como dimensão da prática docente e da vida institucional deve ser construída e aperfeiçoada permanentemente e ser colocada no âmbito dos processos que qualificam a função docente e o desempenho das instituições, num exercício de avaliar a avaliação para que assuma os contornos da coerência, da ética e da promoção do ser humano.

1.4.2 Pré-Processamento

Na base de dados da avaliação institucional há informações sobre a UENP em seus diversos contextos e atividades.

Parte destas informações não se encontravam preparadas para ser processadas e passaram sobre um processo de limpeza sendo corrigidas e formatadas.

Para garantir a confiabilidade das informações prestadas foram realizadas verificações de integridade que, conforme explica Batista (2003, p.42), “frequentemente envolve uma análise das relações permitidas entre os atributos”. Esta verificação é importante visto que existem atributos “dentro de faixas de valores aceitáveis, entretanto, a combinação dos valores dos atributos é muito improvável” (BATISTA,2003, p.43).

Sendo assim foi realizado várias regressões do conjunto de dados através de exclusões de respostas com punições (combinação de todos os registros com notas mínimas) ou que aparentemente não houve um raciocínio sobre as perguntas, ficando assim, prontas para entrarem para o processo de análise.

Camilo e Silva (2009, p.7) explicam o objetivo da limpeza dos dados conforme abaixo:

Frequentemente, os dados são encontrados com diversas inconsistências: registros incompletos, valores errados e dados inconsistentes. A etapa de limpeza dos dados visa eliminar estes problemas de modo que eles não influam no resultado dos algoritmos usados.

1.4.3 Transformação

Após o processo de limpeza e com os dados formatados de acordo com a necessidade foi realizado a transformação destes dados. Na base de avaliação institucional os dados sofreram uma transformação de acordo com o que se precisa para atingir o objetivo da pesquisa. Algumas transformações comum ao processo de mineração de acordo com Camilo e Silva (2009, p.7) são:

[...] suavização (remove valores errados dos dados), agrupamento (agrupa valores em faixas sumarizadas), generalização (converte valores muito específicos para valores mais genéricos), normalização (colocar as variáveis em uma mesma escala) e a criação de novos atributos (gerados a partir de outros já existentes).

1.4.4 Mineração de dados

Esta seção encontra-se dividida em 2 subseções: hipótese e a mineração. Na primeira está descrito o porquê o uso da mineração de dados é essencial em instituições de ensino e na segunda quais os procedimentos e recursos adotados para o desenvolvimento da mineração de dados.

1.4.4.1 Hipótese

Ao se utilizar da mineração de dados para analisar os dados fornecidos pelos participantes da organização sobre funções e atividades desempenhas pela UENP pode-se encontrar informações que irão auxiliar a administração em suas decisões

para aprimorar a execução de suas tarefas diárias, resultando em uma melhora em seu serviço prestado a sociedade. A mineração de dados é essencial para atingir o objetivo de uma instituição de ensino, pois fornece mecanismos para, conforme explica Jorge (2007, p.8), “procura de padrões úteis em grandes quantidades de dados”.

1.4.4.2 A mineração

A mineração de dados foi dividida em 8 hipóteses no qual está descrito a hipótese em si e quais as características envolvidas. Foi utilizada a ferramenta RapidMiner e o algoritmo Apriori para dar suporte na construção das regras de associação no qual de acordo com Amo(2005) “resolve o problema da mineração de itemsets frequentes”.

1.4.5 Avaliação

Na etapa de avaliação dos dados foi utilizado o algoritmo Apriori para validar o suporte e a confiança necessários no qual caso o itemset não for encontrado com o suporte e confiança mínima então a regra foi descartada.

Conforme explica Amo(2005), “para calcular a confiança de $A \rightarrow (L - A)$ basta dividir o suporte de L pelo suporte de A. Estes suportes já foram calculados antes, durante a execução do algoritmo Apriori”.

2. Fundamentação teórica

Esta seção foi dividida em 7 subseções. A primeira introduz ao processo da mineração de dados e a segunda demonstra as etapas da mineração de dados. Da seção 3 à 6 estão descritos respectivamente alguns métodos, técnicas, algoritmos e ferramentas utilizadas na mineração de dados. E por último, na seção 7, é abordado sobre a avaliação institucional.

2.1 Introdução ao processo de Data Mining

Salienta-se que com a grande quantidade de dados de diversas naturezas gerados a todo momento, a mineração de dados surgiu a partir de uma necessidade de uma busca mais eficiente nestes dados, tentando encontrar neles similaridades, padrões, tendências, classificações, entre outros. É fato que a organização que não valorizar esses ativos e deixá-los de lado ficará em desvantagem em relação a seus concorrentes, visto que, perdem a possibilidade de conhecer mais seu negócio e o caminho que está percorrendo. Entender o perfil de seus clientes, as características essenciais que levam alguns alunos a terem um maior desempenho, identificar similaridades de uma população com uma doença como Mal de Alzheimer ou câncer são conhecimentos que em diferentes áreas demonstram que tomar posse da informação pode levar a decisões mais corretas.

A mineração de dados é um campo de estudo que visa apresentar métodos de extração de conhecimento em uma base de dados. Segundo Fayyad et al. (1996):

A mineração de dados é um passo no processo de Descoberta de Conhecimento que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que, sob certas limitações computacionais, produzem um conjunto de padrões de certos dados.

A iniciativa neste processo requer o entendimento do negócio, ou seja, qual o objetivo do trabalho a ser executado e o problema ou questão que pretende-se encontrar respostas. De acordo com Abbott (2015), a parte mais importante de qualquer projeto de mineração de dados é definir claramente o problema a ser solucionado.

Conforme esclarece Chapman et. al. (1999) a tarefa do entendimento do negócio requer:

Identificar os objetivos dos usuários sob o ponto de vista do negócio sendo este conhecimento convertido dentro de uma definição de problemas da mineração de dados e um plano inicial projetado para alcançar estes objetivos.

É importante destacar a compreensão do problema e do negócio que pretende-se buscar, pois com isto, pode-se selecionar os dados corretos que irão para o processo de análise. Por conseguinte é necessário levantar quais dados são relevantes para a pesquisa mesmo que em bases e formatos diferentes.

No entanto, muitas das vezes esses registros se encontram inconsistentes ou com valores ausentes, sendo necessário aplicar sobre eles um processo de limpeza, retirando dados defeituosos ou duplicados e alocando dados faltantes. Como explica Neves (2003, p.45 apud GALHARDES et. al. 2000-a):

O problema de qualidade de dados podem ser encontrados em dados de uma única fonte, como arquivos e banco de dados, especialmente quando tratando de dados de múltiplas fontes como os que serão integrados em data warehouses.

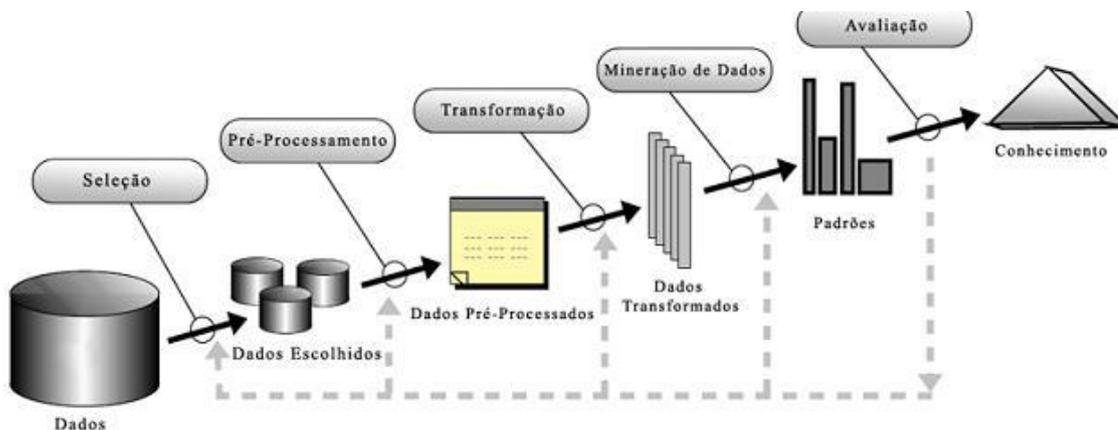
Em seguida realiza-se a integração destes dados depositados em bases e formatos distintos. Conforme aponta Schmitt (2005, p. 25 apud FERNANDEZ, 2003, p.6) “aproximadamente, 70% do tempo gasto na mineração é devido à preparação dos dados obtidos de diferentes bases”. No entanto, com metodologias ágeis construídas com o intuito de agilizar o processo de implementação de sistemas BI, este tempo acaba sendo menor, desse modo, reduzindo custos e trazendo retorno rápido.

Por conseguinte, deve-se utilizar de algoritmos junto de métodos e técnicas de extração para atingir o resultado planejado levantando informações necessários ao estudo do objeto em questão.

2.2 Etapas da mineração de dados

Nas etapas a serem seguidas num processo de mineração de dados, embora não exista uma definição global, pode ser exemplificada da seguinte maneira:

Figura 1- Etapas do processo de mineração de dados



Fonte: Fayyad et al., 1996

Primeiramente, sendo uma das etapas mais essenciais da mineração, é a definição do escopo do problema devendo ficar claro o objeto de estudo que pretende-se levantar e o objetivo da pesquisa. Em seguida, deve-se coletar as características necessárias ao estudo em questão e qual o impacto que as mesmas tem no objeto de estudo, logo após, localizar as bases em que as mesmas estão armazenadas. A próxima etapa é o pré-processamento e transformação, neste ponto os dados devem ser coletados e tratados levando em consideração que nem sempre os dados estão prontos para uso, esta fase abrange “funções relacionadas à captação, à organização, ao tratamento e à preparação dos dados para a etapa da Mineração de Dados” (BEZERRA; GOLDSCHMIDT; PASSOS, 2005, p. 26). Segue abaixo alguns tipos de transformação:

- i. **Categorização:** Os dados são categorizados de acordo com seu valor em grupos, no qual, para pertencer a determinado grupo, deve satisfazer a uma condição ou ter características semelhantes aquela população. Segue abaixo um exemplo:

A condição dos laboratórios de informática poderão ser categorizadas de acordo com suas médias:

Bom (7 – 10), Mediana (4 – 6.9) e Ruim (0 – 3.9)

- ii. Binarização: Os dados são categorizados em dois grupos conforme suas características, como por exemplo, em relação ao gênero (característica) pode ser masculino ou feminino (classificação). (PETERMAN, 2009)
- iii. Construção de atributos: Pode-se criar atributos novos visando simplificar o processo da mineração. (PETERMAN, 2009)

Posteriormente realiza-se a extração do conhecimento com o auxílio das técnicas de mineração de dados e uma avaliação para garantir a confiabilidade do conhecimento gerado.

2.3 Descrição de alguns métodos utilizados em Data Mining

Esta seção foi dividida em 4 subseções no qual foram descritos 4 métodos que são utilizados na mineração de dados, são eles: associação, classificação, padrões sequenciais e agrupamento.

2.3.1 Associação

Este método de pesquisa busca associações entre fatos. Como esclarece Tarapanoff (2001), Emassei e Navathe (2002) e Amo (2003) citado por Cardoso e Machado (2008, p.505), o método de associação:

É um padrão da forma $X \rightarrow Y$, em que X e Y são conjuntos de valores, ou seja, encontrar itens que determinem a presença de outros em uma mesma transação e estabelecer regras que correlacionam a presença de um conjunto de itens com outro intervalo de valores para outro conjunto de variáveis.

Conforme aponta Bürkle (2006, p. 15) as regras de associação “permite encontrar padrões de relacionamentos entre itens de dados que ocorrem com determinada frequência em uma base de dados transacional (unidimensional), relacional (multidimensional), ou ainda em outros tipos de repositórios de dados”.

Com a mineração pode-se identificar regras de associação em grandes armazéns de dados. Segue um exemplo abaixo referente a mineração de dados aplicada na saúde (VIANNA et al., 2011, p. 539):

- a) Regra também relacionada à mãe adolescente (< 16 anos) – com escolaridade razoável (4-7 anos), sem nenhum filho morto, mas já com outro filho, apresenta risco para desconforto respiratório do recém-nascido;
- b) Regra relacionada à mãe adolescente (< 16 anos) – mesmo quando a criança apresenta bom peso ao nascer (2.500 a 3.500g), a mãe surge como alto critério de risco para o óbito perinatal;
- c) Regra que revela que o baixo peso ao nascer (< 2.500g), com pós-datismo (41 semanas de gestação ao nascer) e afetado por afecções maternas, está relacionado ao óbito em 30 dias de vida.

Bürkle (2006, p. 15) cita um exemplo de regra de associação atrelado ao ambiente acadêmico: “70% dos alunos inscritos em Sistemas Operacionais também estão inscritos em Estruturas de Dados” e explica que essa “afirmação pode ser utilizada pelo diretor do curso para alocar recursos como salas de aula e professores”.

Para se medir a importância que uma regra de associação se apresenta ao usuário pode se utilizar de medidas tais como o suporte e confiança, no qual, como esclarece Amo (2005, p. 4) o suporte é a “porcentagem de transações de um banco de dados de transações” e a confiança “mede o grau de certeza de uma associação”, ou seja, dado um item X e Y no qual $X \rightarrow Y$, o suporte é a relação entre a quantidade de vezes em que X e Y ocorrem juntos e a quantidade total de transações e a confiança é relação entre a quantidade de vezes em que X e Y ocorrem juntos e a quantidade de ocorrências de X.

Com isto o usuário pode buscar regras de associação utilizando de uma medida que irá dizer o quanto aquilo ocorre dentro das transações, ou seja, ter conhecimento do quanto uma regra de associação é frequente.

2.3.2 Classificação

A tarefa deste método é criar regras de classificação dos registros desejados, alocando o valor de uma característica a uma classe ou posição existente. Por exemplo:

a) Os alunos que obtiverem média final menor que 4 estão reprovados, entre 4 e 6,9 estão para exame e os que tiveram média maior ou igual a 7 estão aprovados.

b) Na classificação dos fundos pela ANBIMA (Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais), no nível 1 temos a classe de ativos, como:

i. Cambial: “Fundos que aplicam pelo menos 80% de sua carteira em ativos - de qualquer espectro de risco de crédito - relacionados diretamente ou sintetizados, via derivativos, à moeda estrangeira”. (ANBIMA, 2015, p.18).

ii. Ações:

Fundos que devem possuir, no mínimo, 67% (sessenta e sete por cento) da carteira em ações à vista, bônus ou recibos de subscrição, certificados de depósito de ações, cotas de fundos de ações, cotas dos fundos de índice de ações e Brazilian Depositary Receipts, classificados como nível I, II e III. (ANBIMA, 2015, p.18).

No nível de estratégia referente a classe de ativos de renda fixa, temos a classificação de fundos, conforme ANBIMA (2015, p. 11) em:

i. Soberano: “Fundos que investem 100% (cem por cento) em títulos públicos federais do Brasil”.

ii. Grau de Investimento:

Fundos que investem no mínimo 80% (oitenta por cento) da carteira em títulos públicos federais, ativos com baixo risco de crédito do mercado doméstico ou externo, ou sintetizados via derivativos, com registro das câmaras de compensação.

iii. Crédito Livre:

Fundos que objetivam buscar retorno por meio de investimentos em ativos de renda fixa, podendo manter mais de 20% (vinte por cento) da sua carteira em títulos de médio e alto risco de crédito do mercado doméstico ou externo.

Sendo assim, com uma análise de um fundo de investimento e de suas operações através da mineração de dados pode-se classificá-lo de acordo com as especificações da ANBIMA.

- c) A classificação climática de Köppen-Geiger para distinguir climas com diferentes variações de temperatura do ar, descreve subtipos para os climas árido (grupo B), temperado (grupo C) e continental (grupo D):

Quadro 1- Classificação climática de Köppen -Geiger

Código	Descrição	Grupos
Verão quente	Temperatura média do ar no mês mais quente > 22°C	C-D
Verão temperado	Temperatura média do ar no mês mais quente < 22°C Temperaturas médias do ar nos 4 meses mais quentes > 10°C	C-D
Verão curto e fresco	Temperatura média do ar no mês mais quente < 22°C Temperaturas médias do ar > 10°C durante menos de 4 meses Temperatura média do ar no mês mais frio > -38°C	C-D
Inverno muito	Temperatura média do ar no mês mais frio < -38°C	D
Seco e quente	Temperatura média anual do ar > 18°C Deserto ou semideserto quente (temperatura anual média do ar igual ou superior a 18°C)	B
Seco e frio	Temperatura média anual do ar < 18°C Deserto ou semideserto frio (temperatura anual média do ar inferior a 18°C)	B

Fonte: JORGE et al., 2014, p.3

Com a mineração pode-se analisar a variação do clima durante o ano e determinar em qual código o mesmo pertence.

2.3.3 Padrões sequenciais

Esta técnica tem como objetivo identificar padrões de sequência entre fatos, sendo assim, a ocorrência de um fato A leva a ocorrência de um fato B. De acordo Tarapanoff (2001), Emassei e Navathe (2002) e Amo (2003) citado por Cardoso e Machado (2008, p.506), o método padrão sequencial:

É uma expressão da forma $\langle I_1, \dots, I_n \rangle$, em que cada I_i é um conjunto de itens. A ordem em que estão alinhados os conjuntos reflete a cronologia com que aconteceram os fatos representados por eles. Encontrar padrões previsíveis em um período de tempo significa que um comportamento particular em um dado momento pode ter como consequência outro

comportamento ou sequência de comportamentos dentro de um mesmo período de tempo.

Os três exemplos a seguir não são verídicos e tem finalidade meramente explicativa:

- a) 70% dos consumidores que realizam a aquisição de computadores, em seguida, compram uma mesa de computador.
- b) 80% das pessoas que compram fraldas nas sextas-feiras compram cerveja logo em seguida.
- c) 95% dos clientes que adquirem iPhone não compram Samsung Galaxy.

Através desta técnica, os tomadores de decisão terão uma visão ampliada e uma base sólida para tomar iniciativas.

2.3.4 Agrupamento

Esta técnica visa agrupar um conjunto de objetos dispersos através de suas similaridades ou diferenças, levando em conta um ou mais critérios, em grupos. Como explica Tarapanoff (2001), Emassei e Navathe (2002) e Amo (2003) citado por Cardoso e Machado (2008, p.506), o método de agrupamento consiste em “identificar novos agrupamentos, que contenham características similares e agrupar os registros, ou seja, particionar (segmentar) uma dada população de eventos ou itens em conjuntos”.

Por exemplo:

- a) Em marketing, pode-se realizar o agrupamento dos clientes que possuem comportamentos de compras similares, sendo assim, pode-se enviar produtos em promoção para estes.
- b) Na estrutura do CEP, as regiões foram divididas tendo como critério fatores de desenvolvimento econômico e crescimento demográfico em comum. A região do estado de São Paulo estão agrupadas da seguinte maneira:

Figura 2 - Região do estado de São Paulo



Fonte: <https://www.correios.com.br/para-voce/precisa-de-ajuda/o-que-e-cep-e-por-que-usa-lo/estrutura-do-cep>

Sendo assim, as regiões de São José do Rio Preto, Ribeirão Preto, Araçatuba, Presidente Prudente, Bauru, Campinas, Sorocaba, São Paulo (RM), Taubaté e Santos pertencem ao grupo Região do estado de São Paulo.

2.4 Descrição de algumas técnicas utilizadas em Data Mining

Esta seção foi dividida em 5 subseções no qual foram descritas 5 técnicas que são utilizadas na mineração de dados, são elas: árvores de decisão, SVM (support vector machines), conjuntos aproximados, classificação bayesiana e classificação baseada em regras.

2.4.1 Árvores de decisão

Este método de classificação utiliza um fluxograma onde sobre cada nó é realizado um verificação sobre um valor e direciona-se o caminho conforme os testes

são executados. Tem-se o resultado da classificação percorrendo a árvore até se chegar a um valor final possível.

No entanto, caso o número de atributos vinculados a pesquisa sejam muitos (crescem fatorialmente) pode-se tornar inviável devido ao custo computacional envolvido, no qual para realizar todos os teste necessários levaria muito tempo e seria necessário grandes recursos computacionais envolvidos. (ZUBEN; ATTUX, 2010)

2.4.2 SVM – Support Vector Machines

É um técnica de aprendizado de máquina que vem sendo bastante utilizada na área de data mining, onde se é utilizado a inferência de indução para descrever grupos com base em dados de objetos particulares, sendo dividida em aprendizado supervisionado e não-supervisionado, no qual “têm como objetivo a determinação de limites de decisão que produzam uma separação ótima entre classes por meio da minimização dos erros”. (Vapnik, 1995)

2.4.3 Conjuntos aproximados

Esta técnica tem como objetivo a classificação através da equivalência alocando os itens em categorias por aproximação, sendo criado dois conjuntos, um de alta aproximação e outro de baixa aproximação. A técnica de conjuntos aproximados foi criada, “como um novo modelo matemático para representação do conhecimento e tratamento de incerteza, tendo sido usada, posteriormente, para o desenvolvimento de técnicas para classificação em aprendizado de máquina”. (SASSI, 2006, p. 3).

2.4.4 Classificação Bayesiana

É uma técnica de estatística baseada no teorema de Bayes onde são descritas possibilidades de um evento ocorrer com base na ocorrência de eventos anteriores, ou seja, pode-se prever a classe mais provável que um objeto pode pertencer com

base em características existentes. Tem como característica, conforme declara Maia (2005, p.4) “prover regras que expliquem como mudar o grau de crença atual em uma hipótese quando da constatação de novas evidências”.

2.4.5 Classificação Baseada em Regras (Rule-Based Classification)

Utiliza do uso de regras para tomar decisões sobre a alocação em categorias ou classes, utilizando de uma condição e uma ação com base neste resultado, sendo semelhante à técnica de árvores de decisão. Como salienta Patricio(2005), o sistema deve aprender estas regras que irão classificar os dados conforme seu valor.

2.5 Descrição de alguns algoritmos utilizados em Data Mining

Esta seção foi dividida em 5 subseções no qual foram descritos 5 algoritmos que são utilizados na mineração de dados, são eles: C4.5, k-Means, Apriori, FP-Growth e MiRABIT.

2.5.1 C4.5

O algoritmo C4.5 é um classificador que utiliza técnicas de árvores de decisão, onde para classificar um item é necessário percorrer toda árvore até chegar a uma folha. Os nós dessa árvore tem uma condição que quando satisfeita é levado a um nó mais abaixo, sendo assim, o algoritmo percorre todo um caminho até chegar a uma folha.

O C4.5 é capaz de aprender olhando para seus casos armazenados e tomar decisões com base neles, sendo um sistema em constante evolução aprendendo com casos novos e tomando decisões olhando seu histórico de casos.

O algoritmo C4.5 funciona da seguinte maneira, conforme relatam Zuben e Attux (2010, p. 28):

Inicia criando uma aresta para cada valor do atributo em questão. Após, são testadas todas as combinações possíveis de dois valores e, caso nenhuma dessas combinações produza um ganho maior que a divisão

anterior, o processo é interrompido e a divisão anterior é adotada como divisão final. Senão, é repetido o processo tendo como base a melhor das soluções.

A publicação deste algoritmo é datada de 1987 tendo como proprietário o desenvolvedor John Ross Quinlan.

2.5.2 k – Means

O Algoritmo K – Means é um classificador que utiliza de a uma tabela composta por linhas que são chamadas de pontos ou ocorrências e colunas chamadas dimensões para determinar e classifica-los em grupos (cluster).

Para fazer esta classificação cada linha deve possuir um valor e ser alocada no plano cartesiano, onde posteriormente, são determinados e criados K centroides neste plano que serão os grupos. Em seguida, o algoritmo irá realizar o cálculo da distância de cada ponto para cada centroide e gerar uma média e em seguida inserir o centroide neste coordenada e classificar os pontos ao grupo em que a distância for menor, sendo que este procedimento é realizado em loop até que nenhum ponto mude de grupo. No final tem-se cada linha classificada a um grupo.

Em suma, consiste em particionar observações em n partições, onde cada observação se assemelha mais a sua partição, ou seja, possui mais características em comum.

2.5.3 Algoritmo Apriori

De acordo com Amo (2005, p. 8), o algoritmo apriori “resolve o problema da mineração de itemsets frequentes, isto é, dados um banco de dados de transações D e um nível mínimo de suporte β , o algoritmo encontra todos os itemsets frequentes com relação a D e β ”.

O algoritmo se utiliza de regras de associação do tipo A leva a B, verificando quantas vezes a ocorrência de A leva a ocorrência de B e qual o suporte e confiança deste item, ou seja, se esta transação é repetida com frequência em relação ao total de transações (suporte) e em relação a premissa (confiança).

Por exemplo, seja:

$$A = \{a, b, c\}$$
$$B = \{d\}$$

Define-se confiabilidade da regra:

$$\text{Número de ocorrência } (A \cup B) / \text{Número de ocorrência de } A$$

Define-se como suporte da regra:

$$\text{Número de ocorrência } (A \cup B) / \text{Total de transações.}$$

Sendo assim, o algoritmo vai eliminando itemset que não são frequentes e montando todas as possibilidades dos itemsets frequentes novamente. Para isto se utiliza do princípio de que qualquer subconjunto de um itemset frequente é frequente e que o contrário é também verdadeiro, ou seja, não é necessário fazer um subconjunto de candidatos com itemsets não frequentes.

Sendo assim, é definido pelo usuário um valor mínimo para o suporte e confiança e, em seguida, o algoritmo irá varrer todas as transações e determinar se os valores de suporte e confiança estão de acordo com o solicitado pelo usuário para cada itemset, retornando no final N regras que foram aprovadas.

2.5.4 Algoritmo FP-Growth

Júnior (2010, p. 76) descreve o algoritmo FP-Growth como “um método eficiente e escalável para a mineração de padrões frequentes sejam eles curtos ou longos”.

O algoritmo visa encontrar regras de associação no qual apresentou uma melhoria pelo fator de ter conseguido eliminar a geração de candidatos (HAN; PEI; YIN, 2000 apud WU et al., 2007, p.14), utilizando de outras maneiras para encontrar o conjunto de itemsets frequentes. Para isto ele codifica o conjunto de dados usando uma estrutura de dados compacta denominada FP-Tree (Frequent Pattern Tree) e extrai conjuntos de itens frequentes diretamente desta estrutura (TAN; STEINBACH; KUMAR, 2009, p.433).

Xavier (2010, p.51) explica que a construção da FP-Tree (ou árvore FP) utiliza “uma estratégia de forma que os itens de todas as transações percorridas fiquem organizados de forma mais compacta possível, criando menos ramos possíveis na árvore”, tendo como consequência um ganho de desempenho.

Conforme Xavier (2010, p.54) o algoritmo “encontra todos os conjuntos de itens frequentes, terminando com um prefixo específico e empregando uma estratégia dividir-para-conquistar fracionando o problema em subproblemas menores”.

2.5.5 Algoritmo MiRABIT

O algoritmo MiRABIT (Mineração de Regras de Associação em Banco de dados com Baixa média de Itens por Transação) tem como finalidade encontrar regras de associação tendo como principal característica, de acordo com Camargo e Engel (2002, p.6), “a geração dos conjuntos de itens candidatos a cada transação e não no início da passagem k tal qual o algoritmo Apriori e similares”.

Conforme explica Camargo (2002, p. 61), o principal objetivo desta característica é “melhorar o desempenho da tarefa de mineração de regras de associação em banco de dados com um baixo tamanho médio de transações”.

Camargo e Engel (2002, p. 10) salientam que “como o algoritmo Mirabit gera mais eficientemente os conjuntos de itens candidatos que o algoritmo Apriori, a quantidade destes conjuntos de itens tende a diminuir de forma quase não linear”.

2.6 Ferramentas Data Mining

Esta seção foi dividida em 5 subseções no qual foram descritas 5 ferramentas que são utilizadas na mineração de dados, são elas: Oracle Data Mining, SQL Server Data Mining (SASS), Apache Spark, RapidMiner e Weka.

2.6.1 Oracle Data Mining

O Oracle Data Mining é um dos produtos de mais excelência no mercado tendo sua implementação no Oracle Database Kernel visando alta escalabilidade e utilização eficiente dos recursos existentes. Dentro de suas características destacam-se:

1. **Não há movimentação dos dados:** Os dados são processados diretamente da base de dados do Oracle.
2. **Segurança:** O Oracle garante a segurança das informações e dos dados processados. Além do mais, é necessário ter privilégios para a utilização de determinados modelos de mineração.
3. **Facilidade de atualização dos dados:** Como a base de dados acessada é diretamente a de produção os dados estão sempre atualizados, dispensando o uso da extração em base de terceiros e ganhando tempo.
4. **Ambiente de domínio:** As etapas de pré-processamento e pós-processamento são realizadas no mesmo ambiente.

Oracle Data Mining suporta os seguintes algoritmos supervisionados, no formato algoritmo – função.

1. Árvores de decisão – Classificação
2. Modelos lineares generalizados – Classificação e Regressão
3. Comprimento mínimo da descrição (MDL) – Importância do atributo
4. Naive Bayes (NB) – Classificação
5. Máquina do Vetor do apoio – Classificação e Regressão

Oracle Data Mining suporta os seguintes algoritmos supervisionados, no formato algoritmo – função.

1. Apriori – Associação
2. k Means – Agrupamento
3. Factorização de Matriz Não-Negativa(NMF) – Extração de recursos
4. Máquina de vetor de suporte de classe – Detecção de anomalias
5. Clonagem de particionamento Ortogonal – Agrupamento

2.6.2 SQL Server Data Mining (SASS)

Um das tecnologias mais conceituadas do mercado, fornecendo o processo de mineração através do Analysis Services, sendo capaz de realizar a limpeza dos dados, gerir o aprendizado de máquina e produzir relatórios excelentes. Dentro de suas características destacam-se:

1. **Várias fontes de dados:** Os dados pode ser extraídos de diversas bases, como planilhas e arquivos de textos.
2. **Limpeza dos dados integrada:** Esta etapa é uma das mais complexas de se realizar, porém esta ferramenta disponibiliza um processo de limpeza eficiente diminuindo dados inconsistentes ou faltantes na base de dados.
3. **Segurança:** Garante a segurança dos dados durante o processo de Data Mining.
4. **Drillthrough:** Uma função que permite que através de uma célula presente num cubo o usuário consiga acessar as informações de origem daquela célula, abrangendo um universo maior de dados para se trabalhar.

O SASS disponibiliza o uso de diversos algoritmos de classificação, regressão, segmentação, associação e análise de sequência, segue abaixo alguns deles no formato algoritmo – método:

1. Algoritmo de associação Microsoft - Associação
2. Algoritmo de Naive Bayes – Classificação
3. Algoritmo MSC – Sequência
4. Algoritmo de regressão linear Microsoft – Regressão

2.6.3 Apache Spark

É uma aplicação de código aberto desenvolvida pela AMPLab da Universidade da Califórnia para processar grande quantidade de dados de maneira rápida, focando na facilidade de uso e obtenção de resultados consistentes. É capaz de realizar análises sofisticadas buscando dados de quaisquer natureza (textos, documentos, grafos, entre outros) em diferentes bases de dados (HBase, Cassandra, HDFS, entre

outros). Adere suporte a APIs das tecnologias Java, Python e Scala garantindo uma escalabilidade maior de utilização.

Suas principais características são:

1. Armazena dados na memória, garantindo um processamento muito mais rápido.
2. Fornece APIs para auxílio no desenvolvimento nas linguagens Java, Scala e Python.
3. Disponibiliza bibliotecas como Spark Streaming , Spark SQL, SparkMLlib e Spark GraphX.
4. Tem seu conceito central em conjunto de dados resilientes e distribuídos (RDD) armazenando os dados em diferentes partições, otimizando seu processamento.
5. Oferece shell interativo para Scala e Python.

2.6.4 RapidMiner

Como salienta Bittencourt(2014) o RapidMiner é “uma plataforma de software que fornece um ambiente integrado para a aprendizagem de máquina, mineração de dados, mineração de texto, análise preditiva e análise de negócios”.

Esta ferramenta otimiza os processos da mineração aderindo suporte a visualização, preparação, validação e otimização dos dados. Possui uma excelente interface gráfica para o usuário e oferece feedback constante alertando as falhas que irão ou podem acontecer durante a construção do projeto. Além do mais, oferece diversos recursos da mineração prontos para uso, fazendo seu usuário focar no desenho do processo ao invés de gastar tempo resolvendo um problema específico de uma parte da mineração.

O RapidMiner visa atender a necessidade de um processo ágil, ou seja, possui recursos que auxiliam na tarefa de tornar a mineração simples, rápida e com redução de custos.

2.6.5 Weka

A suíte Weka (Waikato Environment for Knowledge Analysis) é formado por um conjunto de implementações de algoritmos de diversas técnicas de Mineração de Dados (UNIVERSITY OF WAIKATO, 2010).

Disponível sob a licença GNU General Public License (GPL), o Weka está implementado na linguagem Java podendo assim ser utilizado em diversas plataformas.

Com uma GUI amigável, no Weka o usuário pode gerar relatórios referente ao que foi minerado e conforme explica Silva (2004, p.15) “grande parte de seus recursos é acessível via sua GUI, sendo que os demais podem ser utilizados programaticamente através de API’s”.

Damasceno (2010, p.5) descreve alguns métodos e algoritmos suportados pelo Weka como: métodos de associação, agrupamento, predição numérica e classificação, sendo que, no método de associação, alguns algoritmos suportados são o Apriori, FPGrowth, PredictiveApriori e Tertius.

2.7 Avaliação Institucional

A avaliação institucional é um mecanismo utilizado por autoridades competentes para avaliar e acompanhar o processo de ensino de uma instituição pública ou privada. Com os dados coletados pelos envolvidos na pesquisa pode-se corrigir, modificar ou excluir regras de negócio, processos, planejamentos, metodologias, entre outros, visando manter as atividades da organização alinhada com sua missão.

O sistema nacional de avaliação da Educação Superior (SINAES) foi instituído pela lei nº 10861 de 14 de Abril de 2004 tendo como finalidade o descrito no primeiro parágrafo do artigo nº 1:

§ 1o O SINAES tem por finalidades a melhoria da qualidade da educação superior, a orientação da expansão da sua oferta, o aumento permanente da sua eficácia institucional e efetividade acadêmica e social e, especialmente, a promoção do aprofundamento dos compromissos e responsabilidades sociais das instituições de educação superior, por meio da valorização de sua missão pública, da promoção dos valores democráticos,

do respeito à diferença e à diversidade, da afirmação da autonomia e da identidade institucional.

O uso deste instrumento pelo governo é de grande valia, onde pode-se constatar quais são as regiões que tem maiores dificuldades no ensino superior, quais são os principais problemas apresentados, quais os itens que mais agregaram desenvolvimento na educação, levantar as principais metodologias de ensino, entre outros. Com isto, as entidades competentes tomam conhecimento de quais pontos necessitam de alteração, podendo agir com uma visão mais consolidada dos problemas que se pretende resolver.

Os participantes dessa pesquisa são discentes, docentes e funcionários técnicos-administrativos como descrito no inciso quarto do artigo nº 2:

Art. 2º O SINAES, ao promover a avaliação de instituições, de cursos e de desempenho dos estudantes, deverá assegurar:

IV – a participação do corpo discente, docente e técnico-administrativo das instituições de educação superior, e da sociedade civil, por meio de suas representações.

O objetivo da pesquisa é conhecer as características do ensino oferecido aos discentes para assim poder tomar decisões e agir, conforme informado no artigo nº 4:

Art. 4º A avaliação dos cursos de graduação tem por objetivo identificar as condições de ensino oferecidas aos estudantes, em especial as relativas ao perfil do corpo docente, às instalações físicas e à organização didático-pedagógica.

Com o auxílio da mineração de dados pode-se varrer este agregado de informação e tentar encontrar padrões de comportamento, fazer classificações dos itens encontrados, montar objetos de estudo para solucionar problemas, associar mecanismos de ensino ao rendimento dos discentes, encontrar sequências de fatos que levam a um elevado ou baixo desempenho dos discentes, entre outros.

3. Desenvolvimento

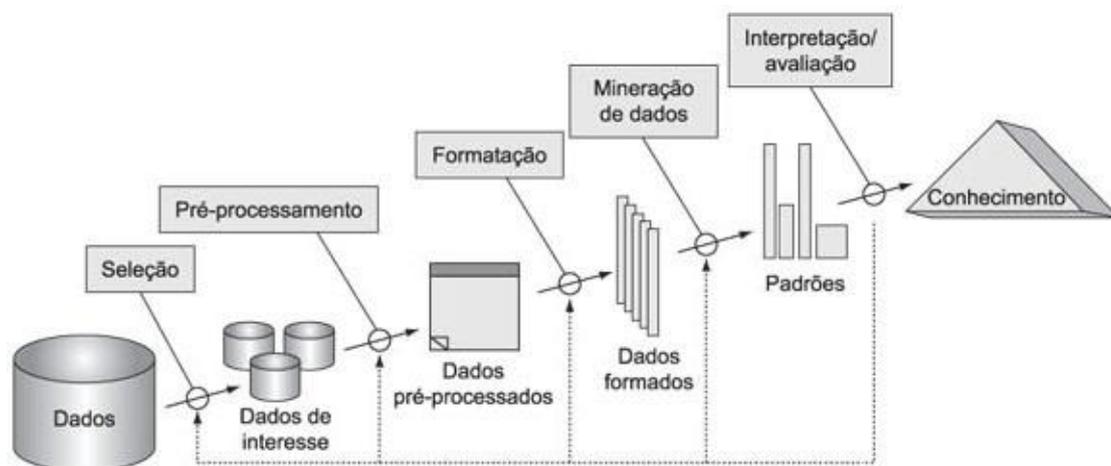
Esta seção apresenta na primeira subseção uma abordagem de como foi realizado os etapas da mineração de dados para se alcançar os objetivos. Nas outras subseções estão descritos como foi realizado o desenvolvimento de cada etapa da mineração de dados.

3.1 A Mineração

Este trabalho levantou um estudo sobre a base de dados da avaliação institucional da UENP através de técnicas e processos da mineração de dados e produziu resultados sólidos para tomadores de decisão avaliarem as informações e com isto maturar processos de ensino-aprendizagem institucionais.

A mineração de dados seguiu as etapas conforme abaixo:

Figura 3 - Etapas da mineração de dados



Fonte: Fayyad et al., 1996

A etapa de seleção dos dados foi sobre um arquivo Microsoft Office Access que posteriormente foi transferido para uma base MySQL sem alteração de estrutura e valor. Em seguida, foi criado o modelo dimensional da base de dados com cinco fatos: avaliação aluno, avaliação docente, avaliação aluno-docente, avaliação estagiário e avaliação técnico e para cada fato são encontradas diversas medidas, que estão descritas no “ANEXO A” deste trabalho (CRISTOVO, 2016, p.32-44).

Estas medidas auxiliaram na busca das informações durante o processo de mineração de dados, no qual, após a definição do escopo do problema, foram definidas as medidas que foram analisadas para se alcançar o objetivo do projeto que é fornecer informações úteis para prover soluções aos tomadores de decisão.

Foi criado somente uma dimensão de tempo devido ao dado ser somente o ano em que a avaliação institucional foi aplicada e também as seguintes dimensões convencionais: professor, disciplina, campus e curso (CRISTOVO, 2016).

Na etapa de pré-processamento foram extraídos características referente a objetos de estudo aplicando sobre estes dados processos de limpeza, ajustando valores com desconformidades, valorando características nulas, descartando itens desnecessários, entre outros. Também foram realizadas regressões dos conjuntos de dados para conjuntos de dados com maior confiabilidade, ou seja, através de verificações de integridade foram retiradas respostas sem raciocínio ou com punições e formulários que não foram finalizados.

Na etapa de transformação foi convertido os dados em numéricos ou categóricos, de acordo com a necessidade, realizando processos de agrupamento nos itens (sumarização, contagem, média, desvio padrão), generalização (transformar valores muito detalhados em valores mais genéricos), distribuir os valores em escalas e criar novos atributos, sendo que, cada processo também varia com a necessidade e o objetivo que se pretendia alcançar.

Na etapa de mineração foi escolhido o método de associação junto do algoritmo Apriori e a ferramenta RapidMiner para encontrar as regras de associação.

O método de associação foi escolhido devido a possibilidade de relacionar diversos fatos presentes na base, podendo correlacionar respostas de perguntas e analisar o quanto é frequente esta ocorrência, validando através de um suporte e confiança o quão confiável é aquela relação. Em vista disto, foi determinado e utilizado o algoritmo Apriori para suprir esta necessidade de encontrar e validar regras de associação com um suporte e confiança mínimos, pois como é uma base relativamente pequena, não há necessidade de se utilizar algoritmos que aderem velocidade ao processo e sim algoritmos simples que possuem funcionalidades suficientes para suprir a necessidade. E por último, foi utilizado a ferramenta RapidMiner pois é uma excelente ferramenta que tem como foco a mineração, possui

uma interface amigável e os algoritmos de mineração estão prontos para serem utilizados, ou seja, o foco é voltado na mineração de dados.

O resultado da mineração passou sobre um processo de avaliação com o intuito de garantir a confiabilidade das informações coletadas no qual foram informados o suporte e confiança mínimos para cada regra de associação levantada.

3.1 Seleção

A base de dados se encontra no formato do Microsoft Access, porém foi transferida para uma base MySQL para poder ser analisada. Para isto foi realizado o download e instalação do arquivo “mysql-connector-odbc-5.3.8-winx64.msi” e feita a conexão com a base MySQL com o auxílio do programa “Administrador de fonte de dados ODBC” do Windows. Em seguida, foi realizado a exportação no programa Microsoft Access através da opção “Dados Externos” -> “Mais” -> “Banco de dados ODBC”.

3.2 Pré-Processamento

Na base de dados da avaliação institucional encontrou-se respostas que estavam nulas, então foram formatadas para o valor zero que diz respeito a não ter condições de avaliar a pergunta.

Para garantir a confiabilidade das informações prestadas pelos participantes e a qualidade dos dados foram realizadas análises nas tabelas com o objetivo de excluir dados com desconformidades.

Sendo assim foram realizadas verificações de integridade nestes dados com o objetivo de excluir da análise dados duvidosos ou que possuíam uma combinação de valores improváveis, regredindo o conjunto de dados para um conjunto com maior confiabilidade, tornando assim o resultado da mineração o mais próximo da realidade. Para isto, segue abaixo os procedimentos que foram realizados e a quantidade de valores excluídos de cada quadro em relação a cada procedimento.

3.2.1 Procedimentos realizados – Confiabilidade das Informações

Foram realizadas três validações nos dados para garantir a qualidade das informações prestadas, sendo validados todos os quadros exceto o quadro “Avaliação Estagiário” pois não há nenhuma hipótese levantada neste trabalho com base nas características deste quadro.

A primeira validação nas tabelas foi se o participante chegou a responder ao menos uma pergunta. Sendo assim, caso não haver contabilizado nenhuma resposta então os dados foram excluídos.

A segunda validação foi a respeito das perguntas que tinham um padrão de repetição nas respostas no qual era comparado um “pedaço” das respostas de tamanho $x \leq 3$ com os pedaços seguintes. Para isto foi realizado uma concatenação nas respostas e criados 2 loops, o primeiro para validar padrões de tamanho $x \leq 3$ e o outro para testar estes pedaços com os pedaços seguintes. Em seguida, caso constatado a repetição, era verificado se o tempo de conclusão do questionário do participante era em média menor ou igual a 6 segundos por pergunta, no qual, dado por verdadeiro, pode-se concluir que o participante marcou o mesmo tipo de respostas sem ler, sem prestar atenção ou sem raciocinar o suficiente sobre as perguntas. Sendo assim, estas respostas foram excluídas.

A terceira validação verifica se, após serem aplicadas a primeira e segunda validação, existir algum questionário respondido com o valor mínimo de nota, no qual se existir então foi retirado.

Na primeira validação referente ao quadro “Avaliação Aluno” foram retirados 141 registros sem nenhuma resposta. Como o quadro “Avaliação Aluno” é pai do quadro “Avaliação Aluno-Docente” então foram excluídos os seus registros que estavam vinculados nesta tabela, retirando assim 1305 registros da tabela referente ao quadro “Avaliação Aluno-Docente”. Em seguida, para o quadro “Avaliação Aluno-Docente” foram encontrados e excluídos 731 registros. Para os quadros “Avaliação Técnico” e “Avaliação Docente” não foram encontrados registros inconsistentes referente a este tipo de validação.

Na segunda validação referente ao quadro “Avaliação Aluno” foram retirados 4 registros que estavam com um padrão de repetição nas respostas e que as questões

foram respondidas com um tempo médio menor ou igual a 6 segundos por pergunta. Com isto também foram excluídos mais 36 registros referente ao quadro “Avaliação Aluno-Docente”. Em seguinte para os quadros “Avaliação Aluno-Docente”, “Avaliação Docente” e “Avaliação Técnico” foram excluídos respectivamente, 1699, 4 e 2 registros.

Na terceira validação não foram encontrados para nenhum dos quadros valores a serem retirados, no entanto, a título de observação, foi criado uma base em separado para alocar a terceira validação no lugar da segunda e verificar a existência de valores mínimos que foram retirados pela segunda validação e foram encontrados para os quadros “Avaliação Aluno”, “Avaliação Aluno-Docente” e “Avaliação Docente” respectivamente as quantidades 1, 29 e 1 de registros e para o quadro “Avaliação Técnico” não foram encontrados valores mínimos.

3.3 Mineração de dados – Hipóteses

Para melhor entendimento das pesquisas o desenvolvimento foi dividido em várias hipóteses que contém o objeto de estudo (o que foi observado) e as características envolvidas. Na próxima seção que foi avaliado se a hipótese levantada tem um suporte e confiança válido para não ser descartada.

3.3.1 Hipótese 1

3.3.1.1 Observação

Disciplinas que possuem de 2 a 3 professores para ministra-las, os alunos avaliaram às perguntas solucionar de dúvidas em sala de aula, cumprimento de conteúdo programático e professor mantém uma boa relação com os alunos com nota maior que ou igual a “bom”.

3.3.1.2 Características

As características utilizadas são os números itens 2, 10 e 12 dispostos no quadro de medidas “Avaliacao Aluno-Professor”.

3.3.2 Hipótese 2

3.3.2.1 Observação

A hipótese 2 foi dividida em 4 hipóteses referente a questão do ambiente de sala de aula. Segue abaixo os itens:

- a) Os alunos que assinalaram “ruim” ou “péssimo” referente a iluminação das salas de aulas também informaram ter um “bom” ou “ótimo” interesse e motivação pelas disciplinas.
- b) Os alunos que assinalaram “ruim” ou “péssimo” referente a salas arejadas também informaram ter um “bom” ou “ótimo” interesse e motivação pelas disciplinas.
- c) Os alunos que assinalaram “ruim” ou “péssimo” referente a sala ter boa acústica também informaram ter um “bom” ou “ótimo” interesse e motivação pelas disciplinas.
- d) Os alunos que assinalaram “ruim” ou “péssimo” referente a salas arejadas também informaram ter um “bom” ou “ótimo” interesse e motivação pelas disciplinas.

Pode-se observar que as características referente ao ambiente de sala de aula não impactam a motivação e interesse dos alunos pelas disciplinas.

3.3.2.2 Características

As características envolvidas nesta análise são os números itens 7, 8, 9 e 10 dispostas no quadro de medidas “Avaliação Aluno” e o número item 20 disposto no quadro de medidas “Avaliação Aluno-Docente”.

3.3.3 Hipótese 3

3.3.3.1 Observação

Quando existem duas ou mais turmas por disciplina então a média das respostas das perguntas solucionar de dúvidas em sala de aula, cumprimento de conteúdo programático e professor mantém uma boa relação com os alunos é igual ou maior que “regular”.

3.3.3.2 Características

As características envolvidas nesta análise são os números itens 2, 10 e 12 dispostas no quadro de medidas “Avaliação Aluno-Docente” e uma característica criada que é a número de turmas por disciplina.

3.3.4 Hipótese 4

3.3.4.1 Observação

Quando um professor ministra 3 ou mais disciplinas então os alunos respondem “bom” ou “ótimo” referente aos questionamentos disponibilidade de atendimento extraclasse e pontualidade do professor.

3.3.4.2 Características

As características envolvidas nesta análise são os números itens 10 e 13 dispostas no quadro “Avaliação Aluno-Docente” e uma característica criada que é a quantidade de disciplinas ministradas por professor.

3.3.5 Hipótese 5

3.3.5.1 Observação

Quando um curso possui a razão de 2 disciplinas ou menos por professor então a nota destes professores destes cursos em relação a participação nas decisões da UENP é maior ou igual a “regular”.

3.3.5.2 Características

Foi criada uma característica que é a relação de disciplinas por professor em um curso e utilizada a característica 70 do quadro “Avaliação Docente”.

3.3.6 Hipótese 6

3.3.6.1 Observação

A hipótese 6 foi dividida em 4 hipóteses referente a questão de limpeza e conservação do local de trabalho de um técnico. Segue abaixo os itens:

- a) O técnico que respondeu “bom” ou “ótimo” referente a limpeza do setor onde trabalha também respondeu “bom” ou “ótimo” referente a satisfação com as atividades que desenvolve.
- b) O técnico que respondeu “bom” ou “ótimo” referente ao estado de conservação do setor onde trabalha também respondeu “bom” ou “ótimo” referente a satisfação com as atividades que desenvolve.
- c) O técnico que respondeu “bom” ou “ótimo” referente a limpeza do campus também respondeu “bom” ou “ótimo” referente a satisfação com as atividades que desenvolve.
- d) O técnico que respondeu “bom” ou “ótimo” referente ao estado de conservação do campus também respondeu “bom” ou “ótimo” referente a satisfação com as atividades que desenvolve.

3.3.6.2 Características

As características envolvidas nesta análise são os números itens 19, 21, 22, 23 e 24 dispostas no quadro “Avaliação Técnico”.

3.3.7 Hipótese 7

3.3.7.1 Observação

A hipótese 7 foi dividida em 4 hipóteses referente a questão do acesso ao portal da UENP pelos discentes. Segue abaixo os itens:

- a) O discente que assinalou uma nota maior que “regular” referente ao acesso ao site da UENP também assinalou uma nota maior que regular para o questionamento se os meios de comunicação e informação da UENP atendem a divulgação de normas, eventos e projetos.
- b) O discente que assinalou uma nota maior que “regular” referente ao acesso ao site da UENP também assinalou uma nota maior que regular para o questionamento a UENP incentiva a iniciação científica.

- c) O discente que assinalou uma nota maior que “regular” referente ao acesso ao site da UENP também assinalou uma nota maior que regular para o questionamento boa imagem pública da UENP.
- d) O discente que assinalou uma nota maior que “regular” referente ao acesso ao site da UENP também assinalou uma nota maior que regular para o questionamento a UENP contribui para o desenvolvimento local, regional e do Estado do Paraná.

3.3.7.2 Características

As características envolvidas nesta análise são os números itens 1,3,4,5 e 6 encontradas no Quadro “Avaliação Aluno”.

3.3.8 Hipótese 8

3.3.8.1 Observação

Os docentes da UENP – Campus Jacarezinho que deram uma nota total maior que a nota total média referente as perguntas sobre atividades de extensão e cultura também assinalaram “bom” ou “ótimo” referente as perguntas sobre a reitoria de extensão e cultura (PROEC).

3.3.8.2 Características

As características envolvidas nesta análise são os números itens compreendidos entre 16 e 24 e o número item 41 do Quadro “Avaliação Professor” e através das características 16 a 24 foram criados um novo atributo que é nota total e a nota total média.

3.4 Avaliação dos dados

Na etapa de avaliação dos dados foi informado para cada hipótese o suporte e a confiança. Para a hipótese ser válida deve ter um mínimo de 20% de suporte e 50% de confiança, sendo que se não atingir o mínimo para ambos a hipótese foi descartada.

Conforme explica Lima e Peres (2013, p.16) estes valores variam de projeto para projeto. Neste projeto foi utilizado os mínimos de 20% de suporte e 50% de confiança para acrescentar maior credibilidade e consistência as regras de associação encontradas.

Seja os itens X e Y no qual $X \rightarrow Y$, o mínimo de 20% em relação ao suporte significa que dentro da ocorrência de, por exemplo, 10 transações então os itens X e Y devem aparecer juntos no mínimo em 2 transações. E para acrescentar maior credibilidade as hipóteses aprovadas foi considerado 50% de confiança que diz respeito que dentro das ocorrências de X o valor Y deve estar presente em no mínimo metade das transações.

Mesmo que uma hipótese tenha sido reprovada foi descrito sua representativa da mesma maneira que as que foram aprovadas podendo ficar a critério do leitor, através do suporte e confiança exposto na hipótese, dar um grau de credibilidade de acordo com sua linha de raciocínio.

3.4.1 Hipótese 1

Suporte: 5,48%

Confiança: 71,70%

Suporte Aprovado? () Sim (X) Não

Confiança Aprovada? (X) Sim () Não

Hipótese Descartada.

Pode-se observar que o suporte a esta regra é pequeno, ou seja, é baixa a frequência em que ocorre 2 a 3 professores por disciplina e nota maior ou igual a “bom” em relação as respostas das perguntas solucionar de dúvidas em sala de aula, cumprimento de conteúdo programático e professor mantém uma boa relação com os alunos.

Porém notou-se que dentro deste conjunto (disciplinas com a 2 a 3 professores para lecioná-las) há uma confiança de 71,70% no qual se pode constatar que quando existem de 2 a 3 professores por disciplina os alunos avaliaram às perguntas com nota maior que ou igual a “bom”.

Esta hipótese não atingiu o necessário em relação ao suporte, isto quer dizer que, não há ocorrência suficiente desta associação para garantir a confiabilidade da regra, no entanto, a confiança mostra que, dentro do conjunto de ocorrência da premissa apareceram em 71,70% dos casos notas maior que ou igual a “bom” referente as características já citadas.

Sendo assim esta regra representa que quando há mais professores para lecionar a disciplina (isto não quer dizer que há 2 professores dentro da sala de aula ou que há divisão de turmas) os alunos tendem a informar estas características com uma nota maior que regular.

3.4.2 Hipótese 2

a)

Suporte: 15,75 %

Confiança: 71,94 %

Suporte Aprovado? () Sim (X) Não

Confiança Aprovada? (X) Sim () Não

Hipótese Descartada.

b)

Suporte: 26,82 %

Confiança: 72,04 %

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

c)

Suporte: 14,57 %

Confiança: 70,61 %

Suporte Aprovado? () Sim (X) Não

Confiança Aprovada? (X) Sim () Não

Hipótese Descartada.

d)

Suporte: 10,05%

Confiança: 68,96 %

Suporte Aprovado? () Sim (X) Não

Confiança Aprovada? (X) Sim () Não

Hipótese Descartada.

Estas regras representam o fato de que respostas negativas sobre o ambiente da sala de aula não levam o aluno a se desmotivar em relação as disciplinas. No entanto, apresentou um suporte inferior a 20% (com exceção da característica salas arejadas), ou seja, não há uma frequência considerável em que aparecem respostas negativas sobre ambiente de sala de aula e respostas positivas sobre a motivação e interesse pelas disciplinas, porém, dentro do conjunto de ocorrências em que

aparecem respostas negativas sobre o ambiente de sala de aula também apareceram em 68% ou mais dos casos respostas positivas sobre a motivação e interesse pela disciplina.

Vale ressaltar que isto não quer dizer que um ambiente de sala ruim não interfere na sua motivação e interesse pela disciplina mas que não existe um impacto direto ou uma correlação forte entre as duas características.

3.4.3 Hipótese 3

Suporte: 15,28%

Confiança: 79,13 %

Suporte Aprovado? () Sim (X) Não

Confiança Aprovada? (X) Sim () Não

Hipótese Descartada.

Esta regra busca uma correlação entre a quantidade de turmas por disciplina e respostas positivas sobre o solucionar de dúvidas, cumprimento de conteúdo programático e boa relação professor-aluno. Possui um suporte abaixo do requerido, ou seja, é baixa a frequência em que acontece duas ou mais turmas por disciplina e a média do aluno com nota maior ou igual a “regular”.

No entanto, nas transações em que acontecem duas ou mais turmas por disciplina os alunos respondem em média positivamente em relação a estas características.

Convém relatar que, através do resultado encontrado, a divisão por turmas não vai resolver os problemas referente a estas características por completo, porém foi encontrado uma certa similaridade envolvendo a premissa e as outras características.

3.4.4 Hipótese 4

Suporte: 56,67%

Confiança: 85,40 %

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

Esta regra representa que quando a quantidade de disciplinas que um professor ministra é maior ou igual a 3 então o professor ainda possui disponibilidade de atendimento extraclasse e pontualidade alta.

O suporte demonstra que há uma frequência acima da mínima (20%) então pode-se concluir que existem bastante professores que lecionam 3 ou mais disciplinas e recebem nota acima de regular referente as duas características.

A confiança confirma mais ainda a regra, visto que, dentro do conjunto de professores que lecionam 3 ou mais disciplinas 85,40% deles receberam nota maior que regular referente a disponibilidade de atendimento e a pontualidade.

3.4.5 Hipótese 5

Suporte: 73,00%

Confiança: 73,83 %

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

Está hipótese possui um suporte e confiança de respectivamente 73% e 73,83% demonstrando que um curso que possui a razão de 2 disciplinas ou menos por professor leva o docente a participar mais das decisões na UENP, pois leva a acreditar que como a relação de disciplinas por professores é baixa então irá sobrar mais tempo para o professor participar das reuniões e como consequência das decisões a se tomar na universidade.

3.4.6 Hipótese 6

a)

Suporte: 35,19%

Confiança: 70,37 %

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

b)

Suporte: 38,89%

Confiança: 77,78 %

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

c)

Suporte: 37,04%

Confiança: 74,07%

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

d)

Suporte: 37,04%

Confiança: 74,07%

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

Estas regras de associação representam uma correlação entre as características limpeza e conservação do setor e campus onde um técnico trabalha com a satisfação com as atividades que ele desenvolve, no qual pode-se constatar que com um suporte e confiança acima do mínimo que, o técnico que informa positivamente sobre a questão de limpeza e conservação também assinala positivamente sobre a satisfação com as atividades que desenvolve, demonstrando que o local de trabalho é um critério importante a se considerar em relação a satisfação dos funcionários.

3.4.7 Hipótese 7

a)

Suporte: 50,06%

Confiança: 62,37 %

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

b)

Suporte: 58,69%

Confiança: 73,12%

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

c)

Suporte: 69,17%

Confiança: 86,18%

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

d)

Suporte: 69,67%

Confiança: 86,79%

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

Estas associações relacionam as notas do acesso ao portal da UENP com as notas referente a UENP, no qual se constatou que os alunos que acessam frequentemente o portal tomam um conhecimento maior dos serviços prestados a sociedade e com isto avaliam positivamente as características sobre a UENP.

Pode-se notar que todas as associações levantadas possuem mais de 50% de suporte e confiança garantindo confiabilidade a hipótese levantada.

3.4.8 Hipótese 8

Suporte: 42,50%

Confiança: 73,91%

Suporte Aprovado? (X) Sim () Não

Confiança Aprovada? (X) Sim () Não

Hipótese Aprovada.

Esta hipótese relaciona quando a soma das respostas das avaliações dos docentes sobre as atividades de extensão da UENP – Campus Jacarezinho for maior ou igual a média então o docente também afirma uma nota maior que regular sobre a reitoria de extensão e cultura, ou seja, pode-se concluir que há uma associação com suporte e confiança acima dos mínimos no qual os docentes que avaliaram positivamente as condições de atividades de extensão atribuíram sobre a reitoria um papel fundamental nesta conquista.

4. Conclusão

A avaliação institucional é um excelente instrumento para acompanhar o funcionamento, planejamento e desenvolvimento de uma universidade e com isto garantir que sua missão esteja sendo executada de maneira correta trazendo os valores necessários a sociedade.

Com a mineração de dados foi feito buscas neste grande armazém de dados e foi encontrado informações úteis aos tomadores de decisão, para poderem agir e ter uma visão mais robusta destas informações coletadas.

A mineração de dados é uma área que auxilia na geração de valores com os dados armazenados, encontrando conhecimentos que sem sua ajuda seria difícil conseguir. Nesta base de avaliação institucional pode-se concluir a partir das hipóteses levantadas que existem muitas possibilidades de objetos de estudo e que através de uma análise pode-se encontrar respostas para diversas perguntas.

Com a mineração de dados concluída pode-se agir com mais segurança e tomar um posicionamento mais sólido referente a diversas situações. Por exemplo:

- a) Sobre a hipótese 2 notou-se que as características referente ao ambiente de sala de aula não há um impacto direto sobre a motivação e interesse dos alunos pelas disciplinas, ou seja, os alunos que tem interesse pela disciplina mantêm ou adquirem este interesse independente destas características, no entanto, não quer dizer que não existe impacto algum e sim que não existe um impacto direto.
- b) Conforme hipótese 8 observa-se que para as atividades de extensão e cultura os professores informaram que a participação da reitoria é um dos papéis fundamentais para o funcionamento de um setor da universidade.
- c) Referente a hipótese 6 nota-se em relação aos técnicos que a satisfação das atividades que desenvolve está correlacionado com a limpeza e o estado de conversão do campus e de seu local de trabalho, porém não quer dizer que tais características motivam ou desmotivam o funcionário, mas que existe sim um determinado impacto.

- d) Através da hipótese 7 nota-se uma correlação entre o acesso ao site da UENP com as respostas referente a universidade, logo poderá se divulgar mais o acesso ao site para os envolvidos tomarem conhecimento das atividades e serviços prestados da UENP.

Conclui-se que o estudo dos dados em uma organização é de grande importância e que pode auxiliar em suas tarefas diárias e no aperfeiçoamento de seus serviços.

5. Trabalhos Futuros

Uma questão a ser levantada em uma base de avaliação institucional é sobre a confiabilidade das respostas pois é uma base que não trabalha com fatos e sim com a visão de diferentes papéis envolvidos sobre as perguntas expostas no questionário.

Neste trabalho para dar mais confiabilidade e veracidade aos dados do questionário foi aplicado no pré-processamento validações de notas mínimas e respostas padronizadas em pouco tempo, ou seja, foram excluídas respostas com punições (responder “péssimo” em todas as características de determinado professor, por exemplo) e respostas com um padrão de repetição (responder tudo “ótimo” ou “regular” em pouco tempo leva a acreditar que não houve um raciocínio na pergunta).

No entanto quanto tempo poderia se considerar após ser identificado um padrão de repetição que não houve raciocínio sobre as perguntas? Neste trabalho foi considerado 6 segundos ou menos. Porém para trabalhos futuros poderia ser criada uma função que identifica quantas palavras existem numa questão e levantar um tempo médio de leitura do ser humano para cada palavra ou um conjunto de palavras, sendo assim, poderia se determinar com mais precisão se houve um raciocínio sobre as perguntas ou não.

Uma outra questão é a verificação de contradições nas respostas onde poderia ser levantada em uma outra tabela correlações entre as próprias perguntas no qual quando respondido positivamente sobre um questionamento este outro também deve ser respondido positivamente. Por exemplo:

O aluno que responder cumprimento de conteúdo programático, o professor se faz compreender em sala de aula e o professor soluciona a dúvida dos alunos durante

a aula com nota maior ou igual a regular então deveria responder também que o professor demonstra domínio da disciplina com nota maior ou igual a regular. Motivo:

a) Se o professor está cumprindo o conteúdo programático quer dizer que está ensinando o que deveria em relação a disciplina.

b) Se o professor se faz compreender em sala de aula então se faz compreender em relação ao que está ensinando, ou seja, ao conteúdo da disciplina, pois já foi afirmado que o professor está transmitindo o seu conhecimento do conteúdo programático.

c) Se o professor soluciona a dúvida em sala de aula então soluciona a dúvida em relação ao que está ensinando, ou seja, retira dúvidas dos alunos em relação a disciplina, o que requer conhecimento em relação a disciplina.

Sendo assim leva-se a acreditar que o aluno deveria assinalar nota maior ou igual a regular referente ao professor demonstrar domínio da disciplina. Em relação a esta contradição foram encontrados 19 registros.

No entanto até quanto isto é válido ou o quanto de confiança é necessário para anular uma resposta com base na contradição de outra resposta?

Sobre o questionário pode-se levantar a seguinte questão: Esta maneira é a correta de se avaliar? Com esta dúvida, pode-se ser produzido futuros trabalhos com o intuito de melhorar a maneira como é realizado o questionário, visto que também há perguntas sobre isto dentro do próprio questionário.

Para futuros trabalhos também poderia ser realizada a integração da base de dados de avaliação institucional com outras bases.

Poderia se cruzar as bases que contém informações referente aos envolvidos no questionário ou com as bases de avaliação institucional de outras universidades.

Com as bases que contém informações referente aos envolvidos poderia ser obtido informações referente ao docente, discente, entre outros e tentar correlacionar características específicas destes envolvidos com as características da base de dados de avaliação institucional. Por exemplo:

a) Poderia ser obtido dados referente ao vínculo (CRES, Efetivo, entre outros) do docente na universidade e realizar novamente a correlação feita na hipótese 4

que trata das características referente disponibilidade de atendimento extraclasse e pontualidade do professor com a quantidade de disciplinas que o professor com determinado vínculo ministra.

- b) Também se poderia extrair informações referente as certificações, titulação ou tempo de serviço de um docente e relacionar com a hipótese 1 verificando as notas dos discentes referente as disciplinas a partir de disciplinas que contem docentes com determinado tipo de característica.
- c) Um outra possibilidade seria de relacionar as notas dos discentes que já possuem uma graduação com as notas dos discentes que estão cursando a primeira graduação.
- d) Outra opção seria relacionar uma base que possui as notas médias do discente no ensino fundamental e médio classificando as correlações levantadas através de um intervalo das notas do discente no ensino fundamental e médio.

Referente ao cruzamento de bases entre as universidade poderia ser feita uma seleção dos discente que iniciam e terminam com um bom desempenho na universidade e os que iniciam mas não terminam com bom desempenho. Através disto levantar um questionário para eles com o objetivo de identificar o que ajudou ou não ajudou a manter o desempenho.

Deste modo para trabalhos futuros fica a questão da confiabilidade e veracidade dos dados prestados, de como tratar este dilema com mais precisão e a integração da base de dados de avaliação institucional da UENP com outras bases.

6. Referências

- ABBOTT, Dean. **Definição de Mineração de Dados:** Mineração de dados: tecnologias, ferramentas e técnicas. 2015. Disponível em: <http://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html>. Acesso em: 11 out. 2016.
- AMO, Sandra de. **Técnicas de Mineração de Dados.** 2005. Disponível em: <<http://www.deamo.prof.ufu.br/arquivos/JAI-cap5.pdf>>. Acesso em: 17 dez. 2016.
- AMORIM, Thiago. **Conceitos, técnicas, ferramentas e aplicações de Mineração de dados para gerar conhecimento a partir de bases de dados.** 2006. 49 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade Federal de Pernambuco, Pernambuco, 2006. Cap. 1. Disponível em: <<http://docplayer.com.br/1201836-Conceitos-tecnicas-ferramentas-e-aplicacoes-de-mineracao-de-dados-para-gerar-conhecimento-a-partir-de-bases-de-dados.html>>. Acesso em: 11 out. 2016.
- ANBIMA (Brasil). **Classificação de Fundos.** [s. L.]: Anbima, 2015. 19 p. Disponível em: <http://www.anbima.com.br/data/files/E3/62/8C/0B/242085106351AF7569A80AC2/NovaClassificacaodeFundos_PaperTecnico_1_.pdf>. Acesso em: 17 dez. 2016.
- BITTENCOURT, Raphael. **RapidMiner: Aprenda a Usar.** 2014. Disponível em: <<https://prezi.com/-yo8qjamdbbq/rapidminer-aprenda-a-usar/>>. Acesso em: 08 abr. 2017.
- BRASIL. Constituição (2004). Lei nº 10861, de 14 de abril de 2004. **Institui O Sistema Nacional de Avaliação da Educação Superior – Sinaes e Dá Outras Providências.** Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/10.861.htm>. Acesso em: 05 dez. 2016.
- BÜRKLE, Paula Yamada. **Um Método de Pós-processamento de Regras de Associação com Base nas Relações de Dependência entre os Atributos.** 2006. 91 f. Dissertação (Mestrado) - Curso de Computação, Universidade Federal Fluminense, Niterói, 2006. Disponível em: <http://www.addlabs.uff.br/Novo_Site_ADDLabs/images/documentos/publicacoes/teses_dissertacoes/Dissertacao_Paula.pdf>. Acesso em: 26 jun. 2017.
- CAMARGO, Sandro da Silva; ENGEL, Paulo Martins. **MIRABIT: Um Novo Algoritmo para Mineração de Regras de Associação.** 2002. Disponível em: <http://sedici.unlp.edu.ar/bitstream/handle/10915/22997/Documento_completo.pdf?sequence=1>. Acesso em: 26 jun. 2017.
- CARMAGO, Sandro da Silva. **Mineração de Regras de Associação no Problema da Cesa de Compras Aplicada ao Comércio Varejista de Confecção.** 2002. 101 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002. Disponível em:

<<https://www.lume.ufrgs.br/bitstream/handle/10183/3110/000332309.pdf?sequence=1>>. Acesso em: 26 jun. 2017.

CAMILO, Cássio Oliveira; SILVA, João Carlos. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. 2009. Disponível em:

<http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf>. Acesso em: 11 out. 2016.

CHAPMAN, P. et al. **The CRISP-DM. Process Model**. 1999. Disponível em:

<<http://www.crisp-dm.org>>

CÔRTEZ, Sérgio da Costa; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de dados - Funcionalidades, Técnicas e Abordagens**. 2002. 34 f. Monografia

(Especialização) - Curso de Ciência da Computação, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2002. Disponível em:

<ftp://obaluae.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf>. Acesso em: 17 dez. 2016.

CRISTOVO, Eduardo. **IMPLANTANDO A METODOLOGIA AGILE ROLAP - UMA METODOLOGIA ÁGIL DE DESENVOLVIMENTO PARA SISTEMAS DE INTELIGÊNCIA DO NEGÓCIO**. 2016. 94 f. TCC (Graduação) - Curso de Sistemas de Informação, Centro de Ciências Tecnológicas, Universidade Estadual do Norte do Paraná, Bandeirantes, 2016.

DAMA INTERNATIONAL - Data Management Association International. 2006. In: AMORIM, Thiago. **Administração de dados - Conceitos, técnicas, ferramentas e aplicações de Data Mining para gerar conhecimento a partir de bases de dados**. 2006. 50 f. TCC (Graduação) - Curso de Ciência da Computação, Centro de

Informática, Universidade Federal de Pernambuco, Recife, 2006. Disponível em:

<<http://docplayer.com.br/1201836-Conceitos-tecnicas-ferramentas-e-aplicacoes-de-mineracao-de-dados-para-gerar-conhecimento-a-partir-de-bases-de-dados.html>>.

Acesso em: 16 dez. 2016.

DAMASCENO, Marcelo. INTRODUÇÃO A MINERAÇÃO DE DADOS UTILIZANDO O WEKA. In: CONGRESSO DE PESQUISA E INOVAÇÃO DA REDE NORTE NORDESTE DE EDUCAÇÃO TECNOLÓGICA, 5., 2010, Maceió. **Anais...** . Maceió:

Ifal, 2010. p. 1 - 14. Disponível em:

<<http://connepi.ifal.edu.br/ocs/index.php/connepi/CONNepi2010/paper/viewFile/258/207>>. Acesso em: 08 jul. 2017.

ELDER, John. **Mineração de Dados. O que é e por que é importante?:** História da mineração de dados e avanços recentes. 2015. Disponível em:

<http://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html>. Acesso em: 17 dez. 2016.

FAYYAD, U; PIATETSKY-SHAPIRO, G; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**. American Association for Artificial Intelligence, 1996.

FAYYAD, Usama M. et al. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: KDD. 1996. p. 82-88.

FERNANDEZ, G. Data Mining using SAS applications. USA: Chapman & Hall, 2003.

FONSECA, João José Saraiva de. **Metodologia da Pesquisa Científica**. Fortaleza: Uece, 2002. 127 p. Disponível em: <<http://www.ia.ufrj.br/ppgea/conteudo/conteudo-2012-1/1SF/Sandra/apostilaMetodologia.pdf>>. Acesso em: 17 dez. 2016.

FRIZZO, Micheline. A importância da avaliação institucional para a melhoria da qualidade de ensino - a experiência de uma instituição de ensino superior. In: XXIII ENCONTRO NAC. DE ENG. DE PRODUÇÃO, 23., 2003, Ouro Preto. **Anais...**. Ouro Preto: Enegep, 2003. p. 3 - 4. Disponível em: <http://www.abepro.org.br/biblioteca/ENEGETP2003_TR0208_1317.pdf>. Acesso em: 27 jun. 2017.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel; BEZERRA, Eduardo. **Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações**. 2. ed. Rio de Janeiro: Elsevier Editora Ltda, 2015. 269 p. Disponível em: <<https://books.google.com.br/books?id=HN6sCQAAQBAJ&pg=PT42&lpg=PT42&dq=funções+relacionadas+à+captação,+à+organização,+ao+tratamento+e+à+preparação+o+dos+dados+para+a+etapa+da+Mineração+de+Dados&source=bl&ots=3ryG6RgOK9&sig=XmJ5nyQXlxCvk42UQRb2Kr695To&hl=pt-BR&sa=X&ved=0ahUKEwiZ2v6T7PvQAhXGx5AKHfN4ByQQ6AEIGjAA#v=onepage&q=funções+relacionadas+à+captação,+à+organização,+ao+tratamento+e+à+preparação+dos+dados+para+a+etapa+da+Mineração+de+Dados&f=false>>. Acesso em: 17 dez. 2016.

GRILO JÚNIOR, Tarcísio Ferreira. **Aplicação de técnicas de Data Mining para auxiliar no processo de fiscalização no âmbito do Tribunal de Contas do Estado da Paraíba**. 2010. 102 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Centro de Tecnologia, Universidade Federal da Paraíba, João Pessoa, 2010. Disponível em: <<http://tede.biblioteca.ufpb.br/bitstream/tede/5238/1/arquivototal.pdf>>. Acesso em: 26 jun. 2017.

BATISTA, Gustavo Enrique de Almeida Prado Alves. **Pré-Processamento de Dados em Aprendizado de Máquina Supervisionado**. 2003. 204 f. Tese (Doutorado) - Curso de Ciências de Computação e Matemática Computacional, Universidade de São Paulo, São Carlos, 2003. Disponível em: <<http://repositorio.ufsc.br/handle/123456789/101803>>. Acesso em: 09 jul. 2017.

HAN, J.; PEI, J.; YIN, Y. Mining frequent patterns without candidate generation. In: Proceedings of ACM SIGMOD international conference on management of data, p. 1-12, 2000.

JORGE, Alípio. Introdução ao Data Mining: Introdução e conceitos exemplos, relação com outras áreas. Ceará: Doutorado em Informática – MAP I, 2007.

JORGE, Heitor C. et al. **Classificação climática de Köppen-Geiger**. 2014. Disponível em: <https://portais.ufg.br/up/68/o/Classifica____o_Clim__tica_Koppen.pdf>. Acesso em: 18 dez. 2016.

LIMA, Clodoaldo Aparecido de Moraes; PERES, Sarajane Marques. **Mineração de Dados**. 2013. Disponível em: <http://www.each.usp.br/digiampietri/SIN5001/SIN5001_Parte3_aula3.pdf>. Acesso em: 08 jul. 2017.

MAIA, Roberto Bomeny. **DETECÇÃO DA INTRUSÃO UTILIZANDO CLASSIFICAÇÃO BAYESIANA**. 2005. 137 f. Dissertação (Mestrado) - Curso de Ciências em Engenharia Elétrica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005. Disponível em: <<http://www.eletrica.ufrj.br/teses/textocompleto/2005042802.pdf>>. Acesso em: 18 dez. 2016.

MICROSOFT CORPORATION (Washington). **Mineração de dados (SSAS)**. 2016. Disponível em: <<https://msdn.microsoft.com/pt-br/library/bb510516.aspx>>. Acesso em: 21 nov. 2016.

NEVES, Rita de Cássia David das. **Pré-Processamento no Processo de Descoberta de Conhecimento em Banco de Dados**. 2003. 131 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003. Cap. 2. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/2701/000375412.pdf?...1>>. Acesso em: 11 out. 2016.

ORACLE CORPORATION (Califórnia). **Introducing Oracle Data Mining**. 2016. Disponível em: <https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/intro_concepts.htm#BHCJIDED>. Acesso em: 21 nov. 2016.

PATRÍCIO, Cristian Mara Mazzini Medeiros. **DETECÇÃO DE FRAUDE OU ERRO DE MEDIÇÃO EM GRANDES CONSUMIDORES DE ENERGIA ELÉTRICA UTILIZANDO ROUGH SETS BASEADO EM DADOS HISTÓRICOS E EM DADOS EM TEMPO REAL**. 2005. 120 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Universidade Federal do Mato Grosso do Sul, Campo Grande, 2005. Disponível em: <[http://200.129.202.51:8080/jspui/bitstream/123456789/628/1/Cristian Mara Mazzini Medeiros Patricio.pdf](http://200.129.202.51:8080/jspui/bitstream/123456789/628/1/Cristian%20Mara%20Mazzini%20Medeiros%20Patricio.pdf)>. Acesso em: 18 dez. 2016.

PENCHIKALA, Srini. **Big Data com Apache Spark - Parte 1: Introdução**. 2015. Traduzido por Luiz Santana em 14 Nov 2015. Disponível em: <<https://www.infoq.com/br/articles/apache-spark-introduction>>. Acesso em: 27 nov. 2016.

PETERMAN, Renato. **Pré-processamento em Data-Mining**. 2009. Disponível em: <<http://www.din.uem.br/~gpea/linhas-de-pesquisa/mineracao-de-dados/pre-processamento/pre-processamento-em-data-mining/>>. Acesso em: 18 dez. 2016.

SASSI, Renato José. **Uma Arquitetura Híbrida para Descoberta de Conhecimento em Bases de Dados: Teoria dos Rough Sets e Redes Neurais Artificiais Mapas Auto-Organizáveis**. 2006. 169 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Departamento de Engenharia de Sistemas Eletrônicos,

Universidade de São Paulo, São Paulo, 2006. Disponível em:
<<http://www.teses.usp.br/teses/disponiveis/3/3142/tde-16032007-163930/publico/teseversaorevisada.pdf>>. Acesso em: 18 dez. 2016.

SCHMITT, Jeovani. **Pré-Processamento para a Mineração de Dados: Uso da Análise de Componentes Principais com Escalonamento Ótimo**. 2005. 146 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2005. Cap. 2. Disponível em:
<<https://repositorio.ufsc.br/bitstream/handle/123456789/101803/223783.pdf?sequencia=1>>. Acesso em: 11 out. 2016.

SILVA, Marcelino Pereira dos Santos. **Mineração de Dados - Conceitos, Aplicações e Experimentos com Weka**. 2004. Disponível em:
<<http://www.lbd.dcc.ufmg.br/colecoes/erirjes/2004/004.pdf>>. Acesso em: 08 jul. 2017.

TAN, P-N; STEINBACH, M.; KUMAR, V. **Introdução ao Data Mining Mineração de Dados**. Rio de Janeiro: Ciência Moderna, 2009. ISBN: 978-85-7393-761-9.

TARTUCE, T. J. A. **Métodos de pesquisa**. Fortaleza: UNICE – Ensino Superior, 2006. Apostila.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (Rio Grande do Sul). **Métodos de pesquisa**. Porto Alegre: Ufrgs, 2009. 114 p. Disponível em:
<<http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>

>. Acesso em: 17 dez. 2016.

UNIVERSITY OF WAIKATO. Weka 3 – Machine Learning Software in Java. Disponível no site da University of Waikato (2010). URL: <http://www.cs.waikato.ac.nz/ml/weka>

Vapnik, V. **The Nature of Statistical Learning Theory**. New York: Springer-Verlag, 1995.

XAVIER, Sandro Luiz da Paixão. **SAMIRA – UMA PROPOSTA DE SISTEMA DE APOIO À MINERAÇÃO DE REGRAS DE ASSOCIAÇÃO**. 2010. 112 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Centro de Ciências Tecnológicas - CCT, Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Fortaleza, 2010. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-29032012-090714/publico/disalinson.pdf>>. Acesso em: 26 jun. 2017.

VIANNA, Rossana Cristina Xavier Ferreira et al. **Mineração de dados e características da mortalidade infantil**. 2011. Disponível em:
<http://www.universidadesaudavel.com.br/wp-content/uploads/downloads/2011/07/Rossana-Vianna-et-al_publicado.pdf>. Acesso em: 18 dez. 2016.

ZUBEN, Fernando Von; ATTUX, Romis. **Árvores de Decisão**. 2010. Disponível em:
<ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia004_1s10/notas_de_aula/topico7_IA004_1s10.pdf>. Acesso em: 17 dez. 2016

ANEXO A – Quadros com as medidas referente aos fatos

Quadro 2 - Avaliação Aluno

Nr item	Descrição Medida
1	Media estagios e projetos de extensao atendem e beneficiam a comunidade externa
2	Media meios de comunicacao da UENP atendem as necessidades de divulgacao
3	Media UENP incentiva a iniciacao cientifica
4	Media imagem publica da UENP na sociedade
5	Media contribuicao UENP para a sociedade
6	Media acessa portal da UENP frequentemente
7	Media salas sao arejadas
8	Media salas sao iluminadas
9	Media salas tem boa acustica
10	Media sao amplas para comportar o numero de alunos
11	Media salas apresentam mobiliario adequado
12	Media salas dispoem de recursos audiovisuais suficientes
13	Media biblioteca possui acervo com qualidade
14	Media biblioteca possui numero suficiente de exemplares
15	Media biblioteca possui instalacoes satisfatorias
16	Media eficiencia sistema de busca de livros da biblioteca
17	Media numero de equipamentos atendem aos de alunos
18	Media condicoes dos equipamentos atendem aos de alunos
19	Media laboratorios de informatica atendem as necessidades
20	Media limpeza salas de aula
21	Media conservacao salas de aula
22	Media limpeza do campus
23	Media conservacao do campus
24	Media limpeza dos banheiros do campus
25	Media conservacao dos banheiros do campus
26	Media iluminacao do campus
27	Media seguranca do campus
28	Media atuacao do coordenador de colegiado de curso
29	Media atuacao do diretor de centro de estudos
30	Media atuacao da diretoria do campus / unidade
31	Media atuacao do Vice-reitor em exercicio da reitoria
32	Media atendimento das secretarias academicas
33	Media atendimento da telefonista
34	Media atendimento do protocolo
35	Media atendimento do servicos de fotocopia
36	Media atendimento da biblioteca
37	Media avaliacao da avaliacao

38	Quantidade alunos
39	Quantidade alunos finalizaram algum questionario
40	Quantidade alunos nao finalizaram algum questionario

Fonte: CRISTOVO, 2016, p. 32-34

Quadro 3 - Avaliação Docente

Nr item	Descrição Medida
1	Media divulgacao das normas
2	Media mudancas a partir das decisoes do colegiado de curso
3	Media praticas institucionais que estimulam a melhoria do ensino
4	Media relacao grade, demandas sociais, necessidades alunos
5	Media integracao pesquisa, ensino e extensao
6	Media condicoes e incentivo para pesquisa pelo professor
7	Media apoio participacao de eventos cientificos de apresentacao
8	Media relacao entre pesquisa e desenvolvimento da regio
9	Media incentivo a realizacao de eventos cientificos
10	Media tramite dos projetos de pesquisa na UENP
11	Media execucao dos projetos de pesquisa na UENP
12	Media integracao pesquisa com a atividade de ensino da UENP
13	Media politicas criacao e expansao pos-graduacao lato sensu
14	Media politicas criacao e expansao pos-graduacao stricto sensu
15	Media politicas manutencao pos-graduacao stricto sensu
16	Media condicoes e incentivo para desenvolvimento de projetos de extensao
17	Media tramite projetos de extensao na UENP
18	Media dos projetos de extensao na UENP
19	Media relevancia dos projetos de extensao UENP para desenvolvimento local
20	Media integracao extensao e atividades de ensino da UENP
21	Media incentivo a participacao de alunos em projetos de extensao
22	Media incentivo realizacao eventos extensao
23	Media apoio UENP na participacao em eventos de extensao
24	Media incentivo a realizacao de eventos culturais
25	Media politicas de qualificacao e titulacao docente
26	Media politicas atendimento portadores necessidades especiais
27	Media politicas atendimento discentes
28	Media acesso ao telefone
29	Media acesso a microcomputadores
30	Media atendimento das secretarias
31	Media atendimento da telefonista
32	Media atendimento do protocolo
33	Media atendimento do servico de fotocopia
34	Media atendimento da biblioteca
35	Media atuacao do coordenador de colegiado de curso

36	Media atuacao do diretor de centro de estudos
37	Media atuacao da diretoria do campus
38	Media atuacao do Vice-reitor em exercicio da reitoria
39	Media Pro-Reitoria (PROAF)
40	Media Pro-Reitoria (PROPG)
41	Media Pro-Reitoria (PROEC)
42	Media Pro-Reitoria (PROGRAD)
43	Media Pro-Reitoria (PRORH)
44	Media Pro-Reitoria (PROPAV)
45	Media satisfacao com as atividades que desenvolve
46	Media condicoes espaco fisico onde desenvolve atividades de permanência
47	Media condicoes espaco fisico onde desenvolve atividades de ensino
48	Media condicoes espaço fisico onde desenvolve atividades de pesquisa e extensao
49	Media salas sao arejadas
50	Media salas sao iluminadas
51	Media salas tem boa acustica
52	Media salas sao amplas para comportar o numero de alunos
53	Media salas apresentam mobiliario adequado
54	Media salas dispoem de recursos audiovisuais
55	Media sala permanencia professores para atendimento alunos está adequada
56	Media biblioteca possui acervo com qualidade
57	Media biblioteca possui numero suficiente de exemplares
58	Media possui instalacoes para a leitura e estudo satisfatorias
59	Media numero dos equipamentos disponiveis atendem aos de alunos
60	Media condicoes dos equipamentos disponiveis atendem aos de alunos
61	Media situacao dos laboratorios de informatica
62	Media limpeza das salas de aula
63	Media conservacao das salas de aula
64	Media limpeza do campus
65	Media conservacao do campus
66	Media limpeza dos banheiros do campus
67	Media conservacao dos banheiros do campus
68	Media Iluminacao do campus
69	Media seguranca do campus
70	Media participaco nas decisao da UENP
71	Media clareza funcionamento administrativo UENP
72	Media cumprimento (PDI) da UENP
73	Media imagem e contribuicao UENP na sociedade
74	Media avaliacao da avaliacao
75	Quantidade docentes
76	Quantidade docentes finalizaram algum questionario
77	Quantidade docentes nao finalizaram algum questionario

Fonte: CRISTOVO, 2016, p. 37-41

Quadro 4 - Avaliação Aluno-Docente

Nr item	Descrição Medida
1	Media professor apresenta plano de ensino
2	Media professor cumpre conteudo programatico
3	Media professor varia tecnicas de ensino
4	Media professor utiliza diversos recursos
5	Media aplicacao pratica do conteudo
6	Media dominio teorico professor
7	Media dominio pratico professor
8	Media dominio pratica clinica professor
9	Media professor domina a disciplina
10	Media relacao professor alunos
11	Media aulas compreensíveis
12	Media professor soluciona duvidas durante a aula
13	Media professor atende extra-classe
14	Media avaliaco es coerentes com conteudo
15	Media extensao da prova adequada
16	Media professor pontual
17	Media professor nao falta
18	Media professor incentiva os alunos a participar de eventos
19	Media conhecimento aluno para acompanhar a disciplina
20	Media motivacao aluno para acompanhar a disciplina
21	Media cumprimento horario aluno
22	Media participacao aluno na classe e extra-classe
23	Quantidade alunos
24	Quantidade alunos finalizaram algum questionario
25	Quantidade alunos nao finalizaram algum questionario
26	Media geral professor
27	Media geral aluno

Fonte: CRISTOVO, 2016, p. 35-36

Quadro 5 - Avaliação Estagiário

Nr item	Descrição Medida
1	Media oferta de cursos e palestras pela instituicao
2	Media divulgacao das normas sobre estagio
3	Media pagamento em dia da bolsa auxilio
4	Media agamento em dia do vale transporte
5	Media reducao da carga horaria em epoca de prova
6	Media cumprimento do recesso remunerado
7	Media atuacao do Reitor em exercicio na reitoria
8	Media atuacao da Diretoria do Campus

9	Media atuacao do Supervisor/Chefe Imediato/Diretor de Centro
10	Media atuacao do Gestor de Estagio
11	Media satisfacao com o estagio
12	Media condicoes espaco onde desenvolve atividades
13	Media limpeza do setor onde trabalha
14	Media conservacao do setor onde trabalha
15	Media clareza funcionamento administrativo da UENP
16	Media equipamentos e materiais disponiveis atividades
17	Media imagem publica e contribuicoes da UENP na sociedade
18	Media avaliacao da avaliacao
19	Quantidade estagiarios
20	Quantidade estagiarios finalizaram algum questionario
21	Quantidade estagiarios nao finalizaram algum questionario

Fonte: CRISTOVO, 2016, p. 41-42

Quadro 6 - Avaliação Técnico

Nr item	Descrição Medida
1	Media programa qualificacao profissional
2	Media melhoria qualidade de vida na periodo de trabalho
3	Media valorizacao do servidor
4	Media programa de qualificacao profissional e melhoria da qualidade de vida
5	Media nota quantidade de pessoal que trabalha no seu setor
6	Media divulgacao das normas
7	Media circulacao de informacao na UENP
8	Media acesso a equipamentos de comunicacao
9	Media acesso a equipamentos de informacao
10	Media protocolo fluxo e distribuicao de documentos
11	Media atuacao da diretoria do campus
12	Media atua do Vice-reitor em exercicio da reitoria
13	Media pro-Reitoria (PROAF)
14	Media pro-Reitoria (PROPG)
15	Media pro-Reitoria (PROEC)
16	Media pro-Reitoria (PROGRAD)
17	Media pro-Reitoria (PRORH)
18	Media pro-Reitoria (PROPAV)
19	Media satisfacao com as atividades que desenvolve
20	Media condicoes espaço fisico onde desenvolve atividades
21	Media limpeza setor onde trabalha
22	Media conservacao setor onde trabalha
23	Media limpeza do campus
24	Media conservacao do campus
25	Media limpeza dos banheiros do campus

26	Media conservacao dos banheiros do campus
27	Media iluminacao do campus
28	Media seguranca do campus
29	Media participacao nas decisos da UENP
30	Media clareza funcionamento administrativo da UENP
31	Media equipamentos e materiais disponiveis para atividades
32	Media imagem e contribuição UENP na sociedade
33	Media cumprimento (PDI) da UENP
34	Media avaliacao da avaliacao
35	Quantidade tecnicos
36	Quantidade tecnicos finalizaram algum questionario
37	Quantidade tecnicos nao finalizaram algum questionario

Fonte: CRISTOVO, 2016, p. 43-44