

## Utilização de mineração de dados para obtenção de informações aparentemente ocultas sobre alunos

Use of data mining to obtain information apparently hidden from students

Rafael Vieira de Paula<sup>1\*</sup>; Kelly Pazolini<sup>2</sup>

<sup>1\*</sup> Bacharel em Engenharia Industrial de Controle e Automação – Rua Professora Maria Wanda Padilha, nº 120 – Jardim Belvedere - 27258-060 - Volta Redonda, RJ - Brasil  
<rafaelvieiradepaula@gmail.com>

<sup>2</sup> ESALQ/USP – Engenharia Agrônoma, M. Sc. em Produção Vegetal – Rua Guaporé 340 – Bairro Higienópolis – 13417-290 – Piracicaba, SP - Brasil

### Resumo

O processo de obtenção de informações em grandes bases de dados “Knowledge Discovery in Database” [KDD] foi utilizado por grandes empresas para sugerir livros, filmes ou músicas, respectivamente, de acordo com o gosto do usuário. Isso foi realizado por meio de técnicas que analisam padrões para facilitar a tomada de decisões. Este grupo de técnicas foi denominada como mineração de dados. O objetivo desse estudo foi utilizar técnicas de mineração de dados para descobrir padrões aparentemente ocultos em questionário referente à satisfação de alunos em doze áreas da vida. O conjunto de dados utilizado para esse estudo foi coletado por meio de um questionário respondido por um grupo de alunos dos cursos de “Master in Business Administration” [MBA] de uma instituição de ensino localizada na cidade de Piracicaba, no estado de São Paulo. Este conjunto de dados foi compilado e analisado com o auxílio do software “Waikato Environment for Knowledge Analysis” [WEKA], que através de técnicas de mineração de dados conseguiu estabelecer tendências de satisfação dos alunos em doze áreas da vida preestabelecidas. A ferramenta evidenciou a eficiência de técnicas de mineração de dados na descoberta de padrões para o cenário proposto.

**Palavras-chave:** árvores de decisão, associação, classificação, padrões de comportamento

### Abstract

The process of obtaining information in large KDD databases was used by large companies such as Amazon, Netflix and Spotify to suggest books, movies or songs, respectively, according to the user's profile. This was done through techniques that analyze patterns to facilitate decision making. This group of techniques was termed as data mining. The objective of this study was to use data mining techniques to discover patterns that are apparently hidden in a questionnaire regarding student satisfaction in twelve areas of life. The data set used for this study was collected through a questionnaire answered by a group of Master in Business Administration [MBA] students from a teaching institution located in the city of Piracicaba, in the state of São Paulo. This data set was compiled and analyzed with the help of the Waikato Environment for Knowledge Analysis [WEKA] software, which through data mining techniques managed to establish student satisfaction trends in twelve pre-established areas of life. The tool evidenced the efficiency of data mining techniques in the discovery of standards for the proposed scenario.

**Keywords:** decision trees, association, classification, behavior patterns

### Introdução

É frequente a busca das pessoas por mediadores até o caminho do sucesso em diversas áreas da vida, e, neste contexto, foi inserido o termo “coache”. Segundo Taie (2011), o processo de “coaching” foi definido como a arte e prática de inspirar, energizar e facilitar a performance dos seus clientes “coachees” para atingir o sucesso. Esta meta ou alvo a ser alcançado pelo cliente pode estar relacionado a qualquer área da vida.

Segundo Marques (2013), ao realizar uma análise das áreas de nossa vida, percebeu-se que há mais atenção a uma área em detrimento de outra. Foi notado que pessoas que se dedicam muito ao lado profissional, deixam a família de lado, alguns buscam realização financeira e deixam de cuidar da própria saúde, entre outros aspectos.

Racionalizando quais são as áreas que envolvem nossas vidas foi possível obter indicação e conduzir nossas ações para âmbitos mais desfavorecidos de atenção.

Ainda segundo o autor, a Roda da Vida foi uma das ferramentas mais utilizadas em processos de “coaching”, ajudou a mapear as áreas em desequilíbrio por meio de um questionário interrogando doze setores de nossas vidas em níveis de satisfação em relação aos recursos financeiros, relacionamentos amorosos, realização e propósito, criatividade hobbies e diversão, espiritualidade, plenitude e felicidade, saúde e disposição, contribuição social, vida social, desenvolvimento intelectual, equilíbrio emocional e relacionamentos familiares.

Foi observado por Witten et al. (2011) que a população estava cada dia mais sobrecarregada de dados e parece

que este caminho não tem reversão. Entretanto, a medida que este grupo de dados aumentava, o conhecimento das pessoas sobre os dados preocupantemente diminuía, sendo que o conhecimento contido nestes dados raramente foi apresentado de forma explícita e acabou não sendo aproveitado. Sendo assim, mesmo que seja gerada uma enorme quantidade de dados, encontra-se pobreza em informação (Haw e Kamber, 2006).

A dificuldade em estabelecer tendências ou identificar padrões de comportamento era oriunda da grande quantidade de dados encontrados, uma vez que nem sempre este conjunto de dados se apresentou de forma estruturada. Entretanto, existem algumas ferramentas que auxiliam a obtenção de informação em uma grande quantidade de dados, como a mineração. Segundo Larose (2005), a mineração de dados é o processo de descoberta significativa de novas correlações, padrões e tendências. Para isso, realiza-se uma determinada filtragem de informações chave em grandes quantidades de dados armazenados em repositórios, usando técnicas estatísticas, matemáticas e de aprendizado de máquina.

De acordo com Haw e Kamber (2006), a mineração de dados obteve grande importância na indústria da informação e na sociedade como um todo nos últimos anos em consequência da grande quantidade de dados, e a substancial necessidade de converter esses dados em informações e conhecimentos úteis. Estas informações e conhecimentos podem ser aplicados em análise de mercado, detecção de fraude, perfil de clientes, controle de produção, exploração científica, entre outros.

A Mineração de Dados foi usualmente apontada pela sua capacidade em realizar determinadas tarefas através de um “software”, e entre as tarefas mais comuns tem-se: associação, clusterização, regressão e classificação, e dentro de cada tarefa existem diversas possibilidades de técnicas a serem utilizadas nos algoritmos. O “Waikato Environment for Knowledge Analysis” [WEKA] é um software gratuito desenvolvido em Waikato na Nova Zelândia, e segundo Witten et al. (2011) ele foi projetado para que o usuário possa rapidamente experimentar os métodos existentes em uma plataforma amigável e flexível. Este oferece um amplo apoio a todo o processo de mineração de dados experimentais, incluindo a preparação dos dados de entrada, modelagem estatística, visualização e contém diversos algoritmos conhecidos entre as principais tarefas de mineração de dados existentes. Neste trabalho foi compilado um conjunto de algoritmos sobre os dados coletados na pesquisa.

O principal objetivo deste trabalho foi descobrir padrões de conhecimento aparentemente oculto no formulário respondido pelos alunos dos cursos de especialização com o uso de técnicas de mineração de dados, delineando a busca por padrões, correlações ou impactos de uma área da vida em relação à outra.

## Material e Métodos

Para a realização deste trabalho seguiu-se a metodologia descrita por Fayyad (1996), cujo sequenciamento de etapas foi: coleta de dados, seleção, ajustes e pré-processamento, transformação, mineração de dados, avaliação dos resultados e obtenção de conhecimento.

Para realizar a coleta de dados, o público alvo escolhido foi o de alunos de MBA dos cursos de Gestão de Negócios, Gestão de Projetos, Agronegócios, Marketing e Gestão Escolar em uma instituição de ensino localizado na cidade de Piracicaba, estado de São Paulo. Foi disponibilizado um questionário online, em que este grupo prestou as seguintes informações socioeconômicas na seção um do questionário: sexo, estado civil, curso em que está matriculado, estado que vive atualmente, idade, quantidade de filhos, situação ocupacional, renda bruta mensal, tipo de instituição que cursou ensino médio e superior, área de formação e área de atuação profissional.

Na segunda seção do questionário, o grupo preencheu o formulário com dados da ferramenta roda da vida que continham perguntas como está a situação do entrevistado quanto à satisfação: 1. pessoal com a saúde e disposição; 2. pessoal com o desenvolvimento intelectual; 3. pessoal com o equilíbrio emocional; 4. profissional com a realização e propósito; 5. profissional com os recursos financeiros; 6. profissional com a contribuição social; 7. com os relacionamentos familiares; 8. com os relacionamentos amorosos; 9. com os relacionamentos em sua vida social; 10. com a qualidade de vida em relação a criatividade, “hobbies” e diversão; 11. com a qualidade de vida em relação a plenitude e felicidade; 12. com a qualidade de vida em relação a espiritualidade.

Para as perguntas da segunda seção, o grupo atribuiu valores em uma escala crescente de felicidade de um a dez, de acordo com o nível de satisfação atual com as áreas da vida. A pesquisa contou com a participação de 885 alunos e o resultado foi transferido para uma planilha eletrônica.

A etapa de seleção de dados foi simplificada para este trabalho, pois o conjunto de informações de interesse foi previamente solicitado no questionário. Sendo assim, a primeira etapa realizada foi a remoção de dados duplicados e alunos com mais de três respostas em branco. Por fim, foi selecionado um grupo de 762 alunos como base de dados para a realização deste trabalho. A segunda etapa da metodologia consistiu na adequação para o formato apropriado ao tipo de tarefa a ser executada posteriormente.

Como o objetivo principal deste trabalho foi descobrir padrões de conhecimento aparentemente oculto em formulário, por motivos de preparação e padronização dos dados foi considerado que os alunos satisfeitos são

aqueles que atribuíram notas de oito a dez para cada área da vida. Sendo assim, foi realizada a substituição de valores pela palavra “SIM”. Por outro lado, o grupo de valores na escala de um a sete foi considerado para alunos não felizes e substituído pelo valor “NAO”. Em suma, SIM = “Satisfeito com esta área da vida” e, NAO = “Não satisfeito com esta área da vida”.

O software WEKA tem um formato de arquivo próprio chamado “Attribute-Relation File Format” [ARFF]. Este formato foi desenvolvido na Universidade de Waikato em um projeto do Departamento de Ciência da Computação. Sendo assim, foi necessária a conversão dos dados do formato “eXcel Spreadsheet” [XLS] para o formato ARFF.

### Aplicação de Técnicas com o software WEKA

Foi realizada a importação do arquivo contendo os dados do formulário ao software WEKA para realizar o pré-processamento. As tarefas mais utilizadas em mineração de dados foram descrição, estimativa, predição, classificação, agrupamento e associação (Larose, 2005). Cada tarefa apresentou uma lista de algoritmos disponíveis, sendo que para este, o enfoque foi dado para tarefas de classificação e associação.

Por simplificação de resultados, foi necessário realizar adaptações como: Engenharia Química = “Engenharia”; Engenharia Civil = “Engenharia”; Engenharia Elétrica = “Engenharia”; Engenharia de Produção = “Engenharia”; Administração com ênfase em comércio exterior = “Administracao”; Outras adaptações com dupla possibilidade de resposta, a área de atuação considerada foi a de maior abrangência no formulário. Exemplo: Projetos de engenharia = “Engenharia”.

O conjunto de dados adotado como modelo no WEKA foi composto por 762 instâncias (linhas) e 24 atributos (colunas). O conjunto de dados foi utilizado nas tarefas de classificação com o algoritmo J48 e associação com o algoritmo Apriori. A quantidade de regras gerada pelo algoritmo foi muito extensa, entretanto para fins de simplificação deste artigo, foi discutido apenas alguns insights obtidos em cada área da vida. Para cada simulação com o algoritmo J48 cada classe foi alternada de acordo com o alvo de pesquisa para determinado cenário. Para o algoritmo Apriori, foi exigido pelo menos 90% de confiança nas respostas.

### Associação

O algoritmo Apriori baseia-se em detectar quais atributos estão associados. A tarefa apresenta a configuração: Se atributo X, então atributo Y. A associação foi bastante conhecida em consequência de seus benefícios alcançados, principalmente nas análises de cestas de compras, do termo em inglês “Market Basket”, onde constatou quais produtos tendiam a ser

comprados juntos pelos consumidores em uma mesma transação (Berry e Linoff, 2004).

Segundo Aggarwall (2015), a definição de regra de associação é dada como: Dados dois itens A e B. Diz-se que a regra  $A \Rightarrow B$  é válida ao nível de suporte S e confiança C, se satisfazem duas condições seguintes:

O suporte [S] do item A é pelo menos S eq. (1)

$$S = \frac{NAB}{TR} \quad (1)$$

onde, NAB: é o número de registros que contém A e B; e, TR: é o número total de registros.

A confiança [C] de  $A \Rightarrow B$  é pelo menos C eq. (2)

$$C = \frac{NAB}{NA} \quad (2)$$

onde, NA: é o número de registros que contém A.

### Classificação

A área de aprendizado de máquina foi fundamentada em projetos de algoritmos que imitam o sistema de aprendizado humano, aprendendo padrões complexos automaticamente e tomando decisões. Os sistemas de aprendizado podem ser não supervisionados, semi-supervisionados ou supervisionados (Mitchell, 1997). Os métodos supervisionados são métodos que tentam descobrir a relação entre os atributos de entrada (às vezes chamados de variáveis independentes) e um atributo alvo (às vezes referido como uma variável dependente) (Rokach e Maimon, 2014).

Segundo Larose (2005), um modelo de classificação eficaz envolveu a construção de uma árvore de decisão, que se estendia através de ramificações e nós de decisão, tendo início na raiz e terminando nas folhas. O algoritmo J48 produziu um conjunto de regras e testam os atributos nos nós de decisão, gerando novos ramos. Cada ramo conduziu a decisão até o outro nó, repetindo o processo até atingir as folhas terminais (Figura 1).

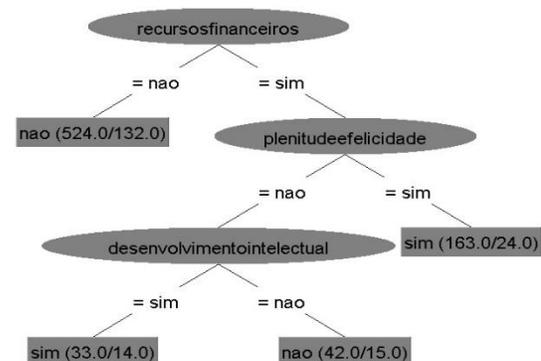


Figura 1. Exemplo de árvore de decisão

Utilizou o algoritmo J48, que é uma evolução e implementação em Java do algoritmo C.45 desenvolvida por Ross Quinlan (1993). Este algoritmo é um dos mais utilizados atualmente em tarefas de classificação de dados e foi projetado para um sistema de aprendizado supervisionado, ou seja, divide o conjunto de dados em amostras para treinamentos e testes.

Os três tipos mais comuns de aprendizado supervisionado na ferramenta WEKA foram: 1) utilizando-se a própria base de dados para realizar o treinamento, comparando ao final do treinamento a classe aprendida com a classe real, 2) separando cerca de um terço dos dados para treinamento e o restante para testes deste modelo, e ainda 3) validação cruzada, metodologia que foi utilizada neste trabalho. Esta metodologia consistiu em dividir os dados em vários conjuntos, sendo que um deles foi utilizado

para testes e os outros conjuntos para treinamento. Funcionando em sistema de alternância entre treino e teste, após o primeiro conjunto realizar o treino, o algoritmo verificou a performance no conjunto de testes, passando ao segundo conjunto (ou dobra) e repetindo os testes, e assim sucessivamente até o último conjunto de treino. O resultado final foi a média dos conjuntos de treinos aplicados ao conjunto de teste (Witten et al, 2011).

O algoritmo J48 apresentou diversas métricas e possibilidades para avaliar a performance do modelo gerado. Para este trabalho foi discutido sobre algumas métricas de avaliação do modelo, como acurácia, sensibilidade, especificidade, entre outras. Este conjunto de métricas apareceu destacada na Figura 2, baseando-se no sistema de treinamento de validação cruzada com dez dobras ou subconjuntos de dados.

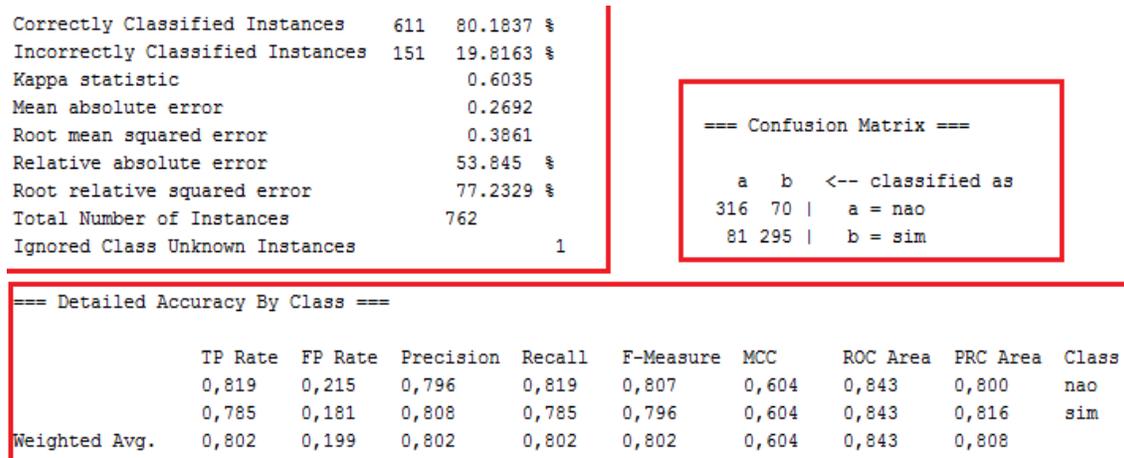


Figura 2. Representação dos resultados gerados através do algoritmo J48 obtido a partir do software Waikato Environment for Knowledge Analysis [WEKA]

Para avaliar o desempenho do classificador J48 foi citado um conjunto contendo algumas das principais métricas de desempenho segundo Rokach e Maimon (2014):

Matriz de confusão: A matriz de confusão foi considerada um dos melhores aspectos para verificar rapidamente a resposta do classificador. Indicou as classificações corretas do classificador e as classificações preditas, onde o número de acertos de cada classe se localiza na diagonal principal e os demais elementos apontam os erros na classificação.

Acurácia [A]: Foi definida como a porcentagem de amostras positivas classificadas corretamente [VP] e negativas classificadas corretamente [VN] sobre o total de amostras classificadas corretamente somadas com amostras com erros de predição denominadas Falsas Positivas [FP] ou Falsas Negativas [FN] de acordo com a eq. (3)

$$A = \frac{VP+VN}{VP+VN+FP+FN} \quad (3)$$

onde, VP: é o número de amostras positivas classificadas corretamente; VN: é o número de amostras negativas classificadas corretamente; FP: é o número de amostras positivas classificadas incorretamente; e, FN: é o número de amostras negativas classificadas incorretamente.

Sensibilidade [S] ou (“Recall”): foi a porcentagem de amostras positivas classificadas corretamente sobre o total de amostras positivas de acordo com a eq. (4)

$$S = \frac{VP}{VP+FN} \quad (4)$$

Especificidade [E]: foi a porcentagem de amostras negativas classificadas corretamente sobre o total de amostras negativas de acordo com a eq. (5)

$$E = \frac{VN}{VN+FP} \quad (5)$$

Precisão [P]: foi a quantidade de amostras classificadas corretamente sobre o total de amostras classificadas como positivas de acordo com a eq. (6)

$$P = \frac{VP}{VP+FP} \quad (6)$$

Área da “Receiver Operating Characteristic Curve” [Curva ROC]: área da curva ROC foi dada por um gráfico com os eixos TPR e FPR, onde em um modelo ideal o TPR foi aproximadamente igual a 1, o FPR foi aproximadamente igual a zero.

TPR foi o mesmo que recall ou a porcentagem de amostras corretamente classificadas como positivas dentre todas as positivas reais e FPR foi a porcentagem de amostras erroneamente classificadas como positivas dentre todas as negativas reais de acordo com a eq. (7)

$$FPR = \frac{FP}{VN + FP} \quad (7)$$

### Resultados e Discussão

Em relação às informações socioeconômicas dos entrevistados, percebeu-se um equilíbrio no atributo sexo, mesmo com ligeira superioridade numérica de alunos do sexo masculino. No atributo estado civil, a maioria deles também foi dividida entre solteiros e casados. Entretanto, nas variáveis curso, filhos, idade, situação ocupacional e estado que habitam fica evidente maior participação na pesquisa de alunos que cursam Especialização em Gestão de Projetos, solteiros, que não tem filhos e que trabalham em regime de Consolidação das Leis do Trabalho [CLT] (Tabela 1).

Tabela 1. Descrição dos resultados de parâmetros como: sexo, nome do curso, estado civil obtidos através do questionário socioeconômico aplicado aos alunos de uma instituição de ensino localizada na cidade de Piracicaba, SP

Parâmetros	Detalhamento	Participação
		----- % -----
Sexo	Masculino	57
	Feminino	43
Cursos de Especialização	Gestão de Negócios	23
	Gestão de Projetos	34
	Marketing	14
	Gestão Escolar	14
	Agronegócios	15
Estado Civil	Solteiro	48
	Casado	44
	Companheiro	3
	Viúvo	1
Filhos	Divorciado	4
	Não tem filhos	71
	01 filho	16
	02 filhos	10
Situação Ocupacional	03 ou mais filhos	3
	Autônomo	8
	Consolidação das Leis do Trabalho [CLT]	66
	Desempregado	7
	Funcionário Público	8
	Pessoa Jurídica	6
	Outros	5

Quanto à formação dos alunos, existiu grande variabilidade de cursos superiores entre os entrevistados, entretanto cerca de um terço dos alunos são formados em engenharia ou administração de empresas (Figura 3).

Observou-se uma grande quantidade de classes para

as categorias Área de formação e Área de atuação, sendo que um terço do grupo de entrevistados foi formado por engenheiros e grande parte trabalhava na área de educação (Figura 3 e 4).

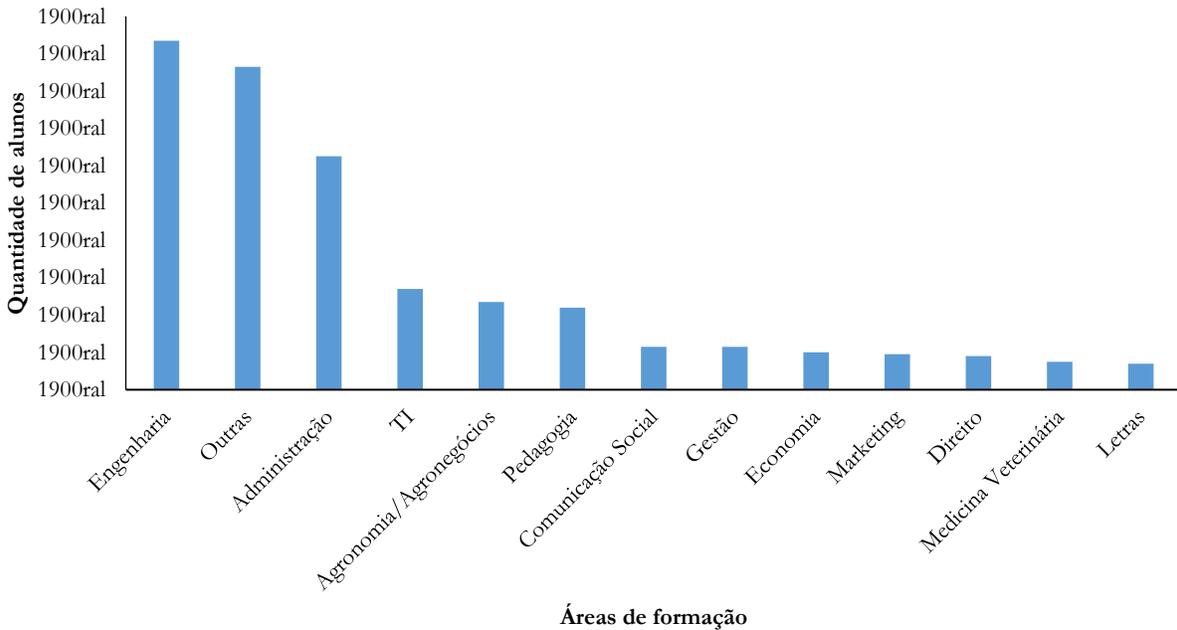


Figura 3. Áreas de formação dos alunos entrevistados

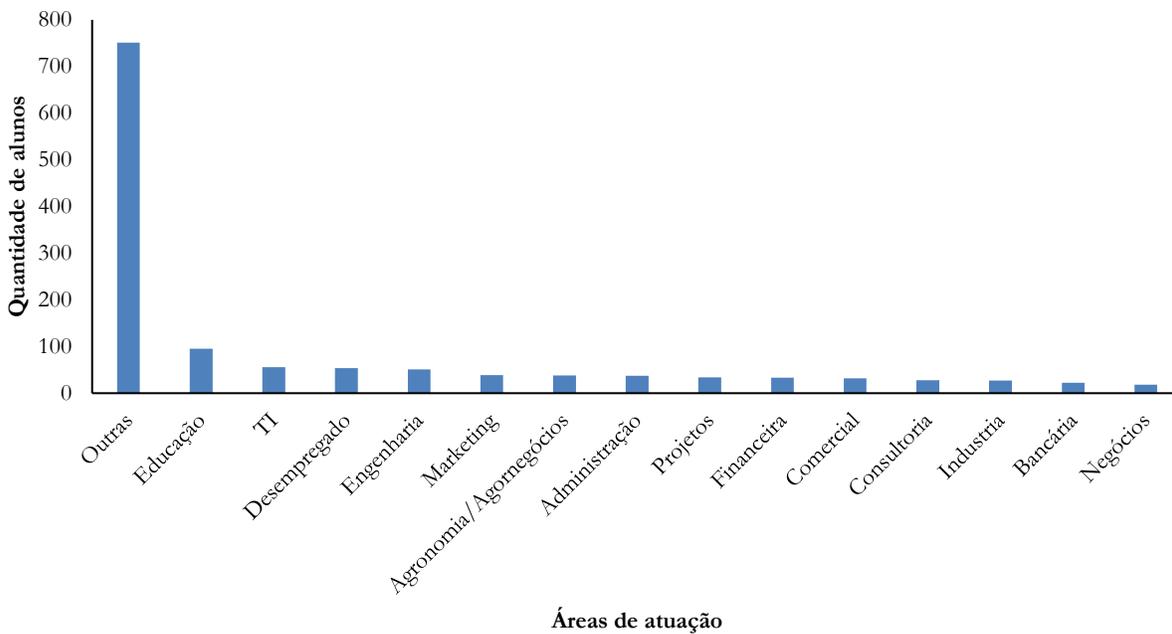


Figura 4. Áreas de atuação dos alunos entrevistados

Observou-se haver equilíbrio entre alunos de Especialização que cursaram o ensino médio e superior em instituições públicas e privadas. A renda bruta dos alunos apresentou certa proporcionalidade com pequena margem percentual a mais de alunos com renda superior a R\$8000,00, e a faixa etária mais frequente foi de alunos entre 26 e 30 anos (Tabela 2).

Tabela 2. Descrição dos resultados de parâmetros como: idade, escolaridade e renda bruta obtidos através do questionário socioeconômico aplicado aos alunos de uma instituição de ensino localizada na cidade de Piracicaba, SP

Parâmetros	Detalhamento	Participação
		---- % ----
Idades dos Alunos	Até 25 anos	14
	De 26 a 30 anos	35
	De 31 a 35 anos	25
	De 36 a 40 anos	12
	De 41 a 45 anos	6
	de 46 a 50 anos	4
	Maior que 51 anos	4
Cursou o ensino médio	Todo em instituição privada	42
	Todo em instituição pública	42
	Maior parte em instituição privada	6
	Maior parte em instituição pública	6
	Outra Situação	3
Cursou o ensino superior	No exterior	1
	Instituição Privada	54
	Instituição Pública	42
	No exterior	1
	Outra Situação	3
Renda Bruta Mensal	Até R\$2000,00	11
	Entre R\$2000,01 e R\$4000,00	26
	Entre R\$4000,01 e R\$6000,00	22
	Entre R\$6000,01 e R\$8000,00	14
	Maior que R\$8000,00	27

O outro enfoque abordado no questionário foi em relação a questões da ferramenta Roda da Vida. Através da análise do gráfico de radar (Figura 5) foi obtido um conjunto de indicadores de dados mais frequentes (moda), as médias das respostas e o desvio padrão. Uma conclusão foi que, de maneira global, a maior parte dos alunos apresentou-se extremamente feliz com seus Relacionamentos Amorosos e apresentaram maior insatisfação com a sua Contribuição Social.

Notou-se que as classes Relacionamentos Amorosos e Relacionamentos Familiares em conjunto, obteve pontuação mais alta de concentração de dados se comparado com

outras variáveis (Figura 6). A pontuação com concentração mais baixas encontrou-se na classe Contribuição Social.

### Recursos Financeiros

De acordo com os critérios impostos, o grupo que atribuiu valor menor ou igual a sete para cada área foi considerado como não satisfeito ou não feliz com a respectiva área (=não). Por outro lado, valores maiores ou iguais a oito indicou o grupo como satisfeito ou feliz com a respectiva área da vida (=sim). O resultado dos caminhos tomados pela árvore de decisão da classe Recursos Financeiros foi indicado na Figura 7.

O modelo gerado pelo classificador após treinamento do tipo validação cruzada com 10 dobras gerou o seguinte caminho: para o grupo alunos que respondeu não para Realização e Propósito, o algoritmo classificou automaticamente que o grupo de alunos também não estava feliz com os Recursos Financeiros. Para o grupo de alunos felizes com sua Realização e Propósito, o algoritmo investigou a satisfação com a Contribuição Social, e para o grupo que respondeu sim para a Contribuição Social, o algoritmo também os classificou como felizes com os Recursos Financeiros (Figura 7).

O grupo que respondeu não para Contribuição Social foi direcionado para outra ramificação para investigar os Relacionamentos Amorosos, e, para aquele que respondeu não, o algoritmo classificou como não feliz com os Recursos Financeiros. Para o grupo de alunos que respondeu sim para os Relacionamentos Amorosos, o algoritmo ramificou-se para investigar os Relacionamentos Familiares. O grupo de alunos não satisfeitos com os Relacionamentos Familiares apresentou-se satisfeito com os Recursos Financeiros, o grupo satisfeito e feliz com seus Relacionamentos Familiares, não obteve a mesma satisfação com seus Recursos Financeiros.

Uma das possibilidades de conclusões foi que este grupo não buscou mais o emprego somente como fonte de renda, e sim que aquilo fizesse sentido em termos de Realização e Propósito. Pode-se concluir também que o grupo buscou algum trabalho que possibilite ter realização financeira, mas não abandonasse os laços amorosos e familiares. O resultado apontou que as pessoas se sentem melhores quando conseguem obter satisfação financeira, juntamente com benefícios de Contribuição Social.

Existem diversas métricas para performance dos classificadores. Os resultados de algumas métricas foram ser observados na Figura 8. Para a classe Atributo e Felicidade o algoritmo obteve um percentual de 75,32% classificações corretas (574 instâncias) e 24,67% (188 instâncias) classificadas incorretamente. O quadro de métricas em destaque foi oriundo da matriz de confusão. A área da Curva ROC ideal tem valor próximo de um (Rokach e Maimon, 2014).

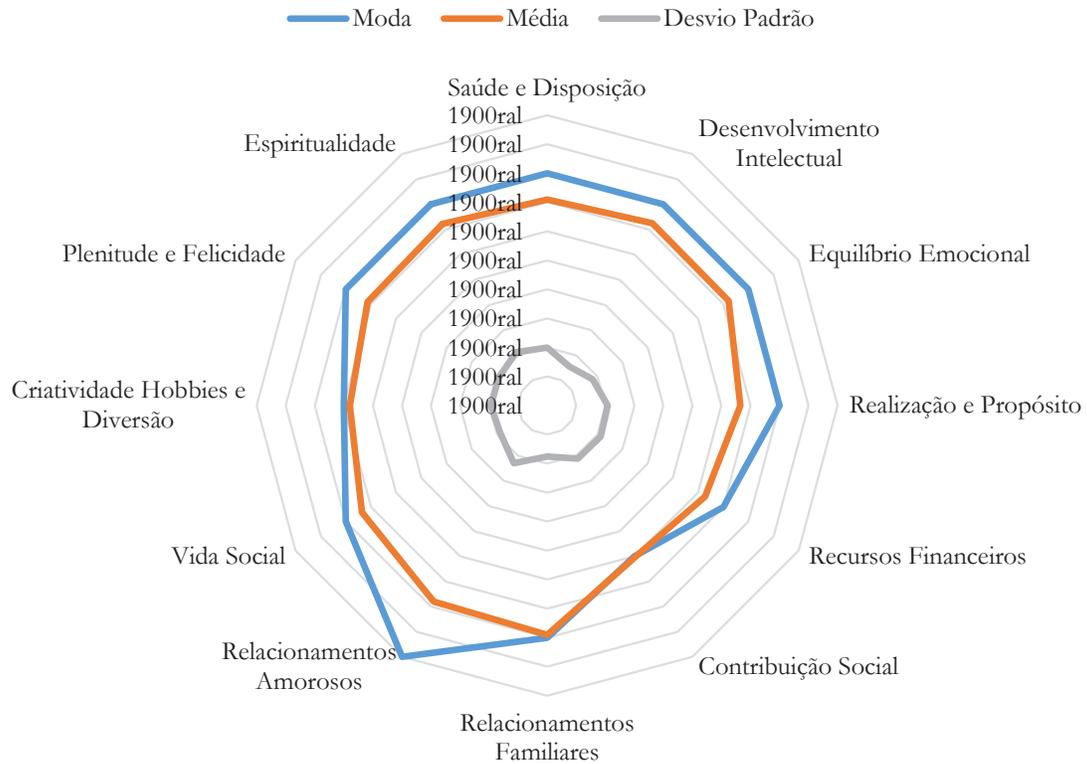


Figura 5. Roda da Vida obtida através do questionário respondido pelos alunos dos cursos de especialização

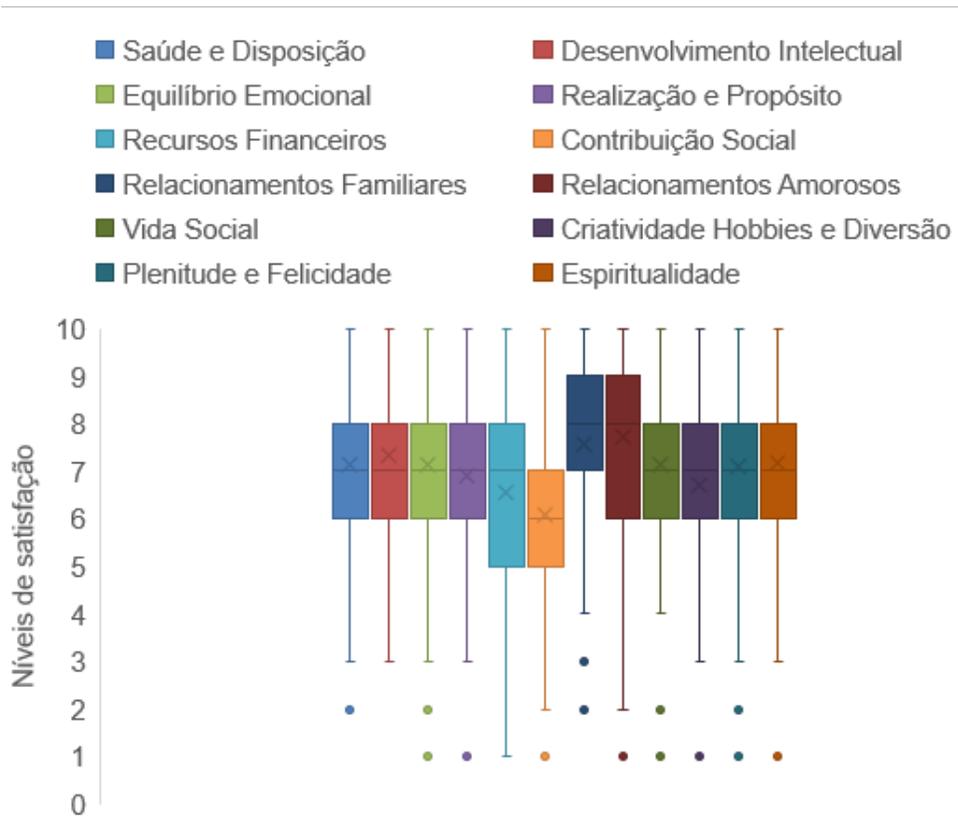


Figura 6. Gráficos de caixa aos atributos da Roda da Vida

## J48 pruned tree

```

-----

realizacaoeproposito = sim
|   contribuicaosocial = nao
|   |   relacionamentosamorosos = nao: nao (43.0/16.0)
|   |   relacionamentosamorosos = sim
|   |   |   relacionamentosfamiliares = nao: sim (23.0/6.0)
|   |   |   relacionamentosfamiliares = sim: nao (103.0/44.0)
|   |   contribuicaosocial = sim: sim (136.0/40.0)
realizacaoeproposito = nao: nao (457.0/65.0)

Number of Leaves   :    5

Size of the tree   :    9

```

Figura 7. Alternativa textual da árvore de decisão em relação a classe Recursos Financeiros

Correctly Classified Instances	574	75.3281 %
Incorrectly Classified Instances	188	24.6719 %
Kappa statistic	0.3771	
Mean absolute error	0.3275	
Root mean squared error	0.4169	
Relative absolute error	76.1927 %	
Root relative squared error	89.9459 %	
Total Number of Instances	762	
Ignored Class Unknown Instances	1	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,885	0,538	0,784	0,885	0,832	0,387	0,723	0,801	nao
	0,462	0,115	0,647	0,462	0,539	0,387	0,722	0,549	sim
Weighted Avg.	0,753	0,406	0,741	0,753	0,740	0,387	0,722	0,722	

=== Confusion Matrix ===

a	b	<-- classified as
464	60	a = nao
128	110	b = sim

Matriz de confusão

		Predição	
		nao	sim
Respostas	nao	VN	FP
	sim	FN	VP

Figura 8. Resultados de performance do algoritmo gerado para a classe Recursos Financeiros

### Insights obtidos por intermédio do algoritmo Apriori para Recursos Financeiros

O algoritmo Apriori gerou 41522 regras com, pelo menos, 80% de confiança. Exemplos de alguns “insights” relacionados à satisfação com Recursos Financeiros: a) O grupo de alunos com renda bruta entre R\$2.000,00 e R\$4.000,00 que não estava satisfeito com Equilíbrio Emocional e não estava satisfeito com

Realização e Propósito, também não estava feliz com seus Recursos Financeiros. Isto aconteceu com 85 a cada 87 alunos neste cenário, ou seja, com confiança de 98%; b) O grupo de alunos com renda bruta entre R\$2.000,00 e R\$4.000,00 que não estava satisfeito com o Desenvolvimento Intelectual e não estava satisfeito com Realização e Propósito, também não estava feliz com seus Recursos Financeiros. Este cenário aconteceu



alunos. De modo geral, pode-se concluir pelo modelo que o grupo de alunos que tinha boa relação entre seus parceiros e familiares. Vale destacar que para o grupo de alunos não satisfeitos com Relacionamentos Amorosos, o algoritmo necessitou de menor número de análises (ramos) em relação ao sexo feminino para tomar decisão. Ou seja, o grupo de mulheres com Relacionamentos Amorosos não satisfatórios, porém com elevada espiritualidade, estava feliz em seus Relacionamentos Familiares.

### Relacionamentos Amorosos

A árvore da classe Relacionamentos Amorosos (Figura 11) teve sua raiz decisória na classe Relacionamentos

Familiares, conforme dito anteriormente. A classe Relacionamentos Amorosos foi a que apresentou melhor resultado geral de satisfação entre alunos (Figura 5). De acordo com o modelo, outro pilar de relacionamentos amorosos foi na classe no Equilíbrio Emocional, o que levou a concluir que o grupo de casais sem satisfação familiar e que apresentou qualquer instabilidade emocional por algum motivo, tendeu a impactar negativamente os Relacionamentos Amorosos. Vale destacar que mesmo para alunos equilibrados emocionalmente, se o estado civil foi igual a solteiro, este grupo também não estava feliz em seus Relacionamentos Amorosos. O algoritmo classificou corretamente aproximadamente 75,3% dos alunos.

```

relacionamentosamorosos = nao
  espiritualidade = sim
    Sexo = Feminino: sim (36.0/8.0)
    Sexo = Masculino
      estadocivil = solteiro: sim (31.0/12.0)
      estadocivil = casado: nao (11.0/4.0)
      estadocivil = separado: sim (1.0)
      estadocivil = companheiro: nao (1.0)
      estadocivil = divorciado: nao (2.0)
      estadocivil = uniao: sim (0.0)
      estadocivil = clerigo: sim (0.0)
      estadocivil = naoresp: sim (0.0)
    Sexo = naoresp: sim (0.0)
  espiritualidade = nao
    equilibrioemocional = sim: sim (36.0/15.0)
    equilibrioemocional = nao: nao (146.0/28.0)
relacionamentosamorosos = sim: sim (498.0/96.0)

Correctly Classified Instances   578   75.853 %
Incorrectly Classified Instances 184   24.147 %

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,456    0,085    0,733     0,456   0,562     0,429    0,691    0,549    nao
0,915    0,544    0,765     0,915   0,833     0,429    0,692    0,762    sim
0,759    0,388    0,754     0,759   0,741     0,429    0,692    0,689

```

Figura 10. Resultados encontrados para a classe Relacionamentos Familiares

```

=== Classifier model (full training set) ===

J48 pruned tree
-----

relacionamentosfamiliares = nao
  equilibrioemocional = sim
    estadocivil = solteiro: nao (27.0/12.0)
    estadocivil = casado: sim (34.0/8.0)
    estadocivil = separado: sim (0.0)
    estadocivil = companheiro: sim (1.0)
    estadocivil = divorciado: nao (1.0)
    estadocivil = uniao: sim (0.0)
    estadocivil = clerigo: sim (0.0)
    estadocivil = naoresp: sim (0.0)
  equilibrioemocional = nao: nao (196.0/57.0)
relacionamentosfamiliares = sim: sim (503.0/101.0)

Correctly Classified Instances   574   75.3281 %
Incorrectly Classified Instances 188   24.6719 %

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,542    0,135    0,681     0,542   0,603     0,434    0,690    0,544    nao
0,865    0,458    0,781     0,865   0,821     0,434    0,691    0,753    sim
0,753    0,346    0,746     0,753   0,746     0,434    0,690    0,681

```

Figura 11. Resultados encontrados para a classe Relacionamentos Amorosos

### Equilíbrio Emocional

A árvore da classe Equilíbrio Emocional (Figura 12) estabeleceu forte correlação principalmente com a variável Plenitude e Felicidade, tornando essa classe atrativa para pesquisas e trabalhos futuros em relação aos alunos entrevistados. De acordo com o modelo gerado, nenhum aluno desequilibrado emocionalmente apresentou satisfação com sua plenitude e felicidade. Esta árvore gerou uma classificação interessante e inesperada, uma vez que o grupo de alunos não satisfeitos com seus Relacionamentos Amorosos e não satisfeitos com a sua Vida Social, apresentou elevado Equilíbrio Emocional. Neste mesmo grupo, o que apresentou satisfação com a Vida Social, não apresentou equilíbrio emocional. Pode-se concluir que este grupo de alunos podia estar buscando em sua Vida Social uma espécie de refúgio para manter o Equilíbrio Emocional, porém sem resultados satisfatórios.

Este modelo também buscou, de forma inédita neste trabalho, relações de onde o aluno cursou o ensino médio com a sua Plenitude e Felicidade. Porém, apesar da taxa de acerto de 100% neste ramo da árvore, o modelo foi somente testado perante 6 alunos. Esta árvore classificou corretamente 72,7% dos alunos.

### Realização e Propósito

A classe da árvore Realização e Propósito (Figura 13) estabeleceu uma forte correlação com a questão de satisfação com Recursos Financeiros. A partir desta análise, um conjunto de conclusões pode ser discutida, como o fato de que um grupo de alunos que trabalhou com o que gosta, porém não se sentia realizado por não se sentir satisfeito com os rendimentos financeiros. Esta árvore classificou corretamente 73% dos alunos.

```
plenitudeefelicidade = nao: nao (386.0/98.0)
plenitudeefelicidade = sim
  saudeedisposicao = sim
    relacionamentosamorosos = nao
      vidasocial = sim: nao (21.0/8.0)
      vidasocial = nao: sim (14.0/2.0)
    relacionamentosamorosos = sim: sim (219.0/38.0)
  saudeedisposicao = nao
    contribuicaosocial = nao: nao (88.0/39.0)
    contribuicaosocial = sim
      relacionamentosfamiliares = nao
        cursouensinomedio = todopublica: nao (1.0)
        cursouensinomedio = todoprivada: sim (2.0)
        cursouensinomedio = outrasiituacao: nao (3.0)
        cursouensinomedio = maiorpartepublica: nao (0.0)
        cursouensinomedio = maiorparteprivada: nao (0.0)
        cursouensinomedio = noexterior: nao (0.0)
        cursouensinomedio = naoresp: nao (0.0)
      relacionamentosfamiliares = sim: sim (28.0/2.0)

Correctly Classified Instances   554   72.7034 %
Incorrectly Classified Instances  208   27.2966 %

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,615    0,169    0,771     0,615   0,684     0,458    0,739    0,699    sim
0,831    0,385    0,700     0,831   0,760     0,458    0,740    0,695    nao
0,727    0,281    0,734     0,727   0,723     0,458    0,740    0,697
```

=== Confusion Matrix ===

a	b	<-- classified as
225	141	a = sim
67	329	b = nao

Figura 12. Resultados encontrados para a classe Equilíbrio Emocional

```
recursosfinanceiros = nao: nao (524.0/132.0)
recursosfinanceiros = sim
  plenitudeefelicidade = nao
    desenvolvimentointelectual = sim: sim (33.0/14.0)
    desenvolvimentointelectual = nao: nao (42.0/15.0)
  plenitudeefelicidade = sim: sim (163.0/24.0)

Correctly Classified Instances   557   73.0971 %
Incorrectly Classified Instances  205   26.9029 %

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,538    0,140    0,719     0,538   0,615     0,425    0,768    0,661    sim
0,860    0,462    0,736     0,860   0,793     0,425    0,768    0,794    nao
0,731    0,333    0,729     0,731   0,722     0,425    0,768    0,741
```

=== Confusion Matrix ===

a	b	<-- classified as
164	141	a = sim
64	393	b = nao

Figura 13. Resultados encontrados para a classe Realização e Propósito

### Contribuição Social

A árvore da classe Contribuição Social (Figura 14) estabeleceu forte relação com a questão de Realização e Propósito. Vale destacar que, de acordo com os dados coletados na pesquisa (Figura 5 e 6), foi a classe com maior frequência de valores mais baixos atribuídos pelos alunos na Roda da Vida em comparação com as outras classes. O modelo detectou que um grupo de alunos somente se sentiu satisfeito com sua Contribuição Social se a área de Realização e Propósito da vida também fosse satisfeita. Outro aspecto interessante foi que o grupo de alunos não satisfeito financeiramente, estava feliz com a área Contribuição Social se a situação ocupacional fosse funcionário público ou pessoa jurídica. Possível conclusão foi que o grupo de empresários se sentiu mais feliz com sua Contribuição Social ao empregar pessoas e que o grupo de funcionários públicos percebeu, de forma mais direta, sua contribuição para a sociedade. Esta Contribuição Social em níveis os e predominantes

na pesquisa se devem ao fato de 66% do grupo de aluno que respondeu esta pesquisa trabalhava em regime CLT. Esta árvore classificou corretamente 76,11% dos alunos.

### Vida Social

A classe da árvore Vida Social (Figura 15) foi construída com base em três aspectos para tomada de decisão: Plenitude e Felicidade, Relacionamentos Familiares e Relacionamentos Amorosos. O modelo entendeu que satisfação em Plenitude e Felicidade implicou em satisfação na Vida Social, a insatisfação quanto a Plenitude e felicidade implicou em satisfação na Vida Social desde que o par Relacionamentos Familiares e Amorosos também estivesse satisfeito, funcionando com pilar do convívio social destes alunos. O algoritmo classificou corretamente aproximadamente 74,7% dos alunos, mesmo com uma árvore razoavelmente simples e pequena em comparação às outras.

```

realizacaoeproposito = sim
|   recursosfinanceiros = nao
|   |   plenitudeefelicidade = nao: nao (40.0/5.0)
|   |   plenitudeefelicidade = sim
|   |   |   situacaoocupacional = Outros: nao (8.0/2.0)
|   |   |   situacaoocupacional = funcpublico: sim (16.0/5.0)
|   |   |   situacaoocupacional = clt: nao (50.0/14.0)
|   |   |   situacaoocupacional = autonomo: nao (11.0/4.0)
|   |   |   situacaoocupacional = naoresp: nao (0.0)
|   |   |   situacaoocupacional = Desempregado: nao (1.0)
|   |   |   situacaoocupacional = pj: sim (6.0/2.0)
|   |   recursosfinanceiros = sim
|   |   |   saudeedisposicao = sim: sim (110.0/38.0)
|   |   |   saudeedisposicao = nao: nao (63.0/24.0)
|   realizacaoeproposito = nao: nao (457.0/50.0)
Correctly Classified Instances   580 76.1155 %
Incorrectly Classified Instances 182 23.8845 %

==== Confusion Matrix ====
      a  b  <-- classified as
505  71  |  a = nao
111  75  |  b = sim

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0,877    0,597    0,820    0,877    0,847    0,306  0,685    0,827    nao
0,403    0,123    0,514    0,403    0,452    0,306  0,684    0,409    sim
0,761    0,481    0,745    0,761    0,751    0,306  0,684    0,725

```

Figura 14. Resultados encontrados para a classe Contribuição Social

```

plenitudeefelicidade = nao
|   relacionamentosfamiliares = nao: nao (204.0/29.0)
|   relacionamentosfamiliares = sim
|   |   relacionamentosamorosos = nao: nao (61.0/17.0)
|   |   relacionamentosamorosos = sim: sim (121.0/57.0)
|   plenitudeefelicidade = sim: sim (376.0/82.0)
Correctly Classified Instances   569 74.6719 %
Incorrectly Classified Instances 193 25.3281 %

==== Confusion Matrix ====
      a  b  <-- classified as
328  76  |  a = sim
117 241  |  b = nao

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0,812    0,327    0,737    0,812    0,773    0,491  0,763    0,724    sim
0,673    0,188    0,760    0,673    0,714    0,491  0,762    0,727    nao
0,747    0,262    0,748    0,747    0,745    0,491  0,763    0,725

```

Figura 15. Resultados encontrados para a classe Vida Social

### Criatividade “Hobbies” e Diversão

A árvore da classe Criatividade, “Hobbies” e Diversão indicou que esta foi a classe mais complexa. O algoritmo gerou 114 folhas (conclusões), sendo a grande maioria a partir da análise das áreas de atuação. Isso motivou uma análise mais profunda e investigativa para trabalhos futuros, em artigos e teses que correlacionam variáveis relacionadas às áreas de atuação com a classe criatividade, hobbies e diversão, uma vez que foi classificado corretamente aproximadamente 78,9% do grupo de alunos (Figura 16).

A complexidade da árvore surgiu do fato que a análise foi fundamentada na área de atuação de cada

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,848	0,304	0,814	0,848	0,831	0,551	0,800	0,832	nao
0,696	0,152	0,744	0,696	0,719	0,551	0,800	0,651	sim
0,789	0,245	0,787	0,789	0,787	0,551	0,800	0,762	

```

=== Confusion Matrix ===
      a  b  <-- classified as
    395  71 | a = nao
     90 206 | b = sim
  
```

Figura 16. Resultados da classe Criatividade, “Hobbies” e Diversão

```

plenitudeefelicidade = nao: nao (386.0/43.0)
plenitudeefelicidade = sim
saudeedisposicao = sim: sim (254.0/57.0)
saudeedisposicao = nao
areadeatuacao = educacao
| contribuicaosocial = nao
| | equilibrioemocional = sim: sim (3.0)
| | equilibrioemocional = nao
| | espitirualidade = sim: nao (3.0)
| | espitirualidade = nao: sim (2.0)
| contribuicaosocial = sim: nao (12.0)
areadeatuacao = militar: nao (0.0)
areadeatuacao = industria: nao (6.0/1.0)
areadeatuacao = administrativa: nao (8.0/1.0)
areadeatuacao = projetos
| rendabruta = entre4mile6mil: nao (1.0)
| rendabruta = maiorque8mil: nao (2.0)
| rendabruta = ate2mil: nao (0.0)
| rendabruta = entre2mile4mil: sim (2.0)
| rendabruta = naoresp: nao (0.0)
| rendabruta = entre6mile8mil: nao (0.0)
areadeatuacao = TI
| relacionamentosfamiliares = nao: nao (2.0)
| relacionamentosfamiliares = sim: sim (5.0)
areadeatuacao = desempregado: sim (7.0)
areadeatuacao = engenharia: nao (5.0/1.0)
areadeatuacao = naoresp: nao (1.0)
areadeatuacao = negocios: sim (2.0)
areadeatuacao = varejo: sim (1.0)
areadeatuacao = ti: nao (0.0)
areadeatuacao = Militar: nao (0.0)
areadeatuacao = Financeira
| estadocivil = solteiro: nao (2.0)
| estadocivil = casado: sim (3.0)
| estadocivil = separado: sim (0.0)
| estadocivil = companheiro: sim (1.0)
| estadocivil = divorciado: sim (0.0)
| estadocivil = uniao: sim (0.0)
| estadocivil = clerigo: sim (0.0)
| estadocivil = naoresp: sim (0.0)
areadeatuacao = saneamento: nao (1.0)
areadeatuacao = comunicacao: nao (2.0/1.0)
areadeatuacao = qualidade: nao (0.0)
areadeatuacao = Arquitetura: nao (0.0)
  
```

Figura 17. Parte da árvore com os resultados encontrados para a classe Criatividade, “Hobbies” e Diversão

aluno, o que levou a concluir que cada área de atuação teve impacto direto com a classe Criatividade “Hobbies” e Diversão. Deve-se considerar que um grupo de profissões possui poucos exemplos para o modelo aprender. Outro aspecto a destacar foi que o grupo de alunos que estava satisfeito com a área de Saúde e Disposição também estava satisfeito com Criatividade, “Hobbies” e Diversão. Uma das conclusões possíveis foi o prazer de se sentir bem com o corpo em forma e, possivelmente, o grupo de alunos pessoas que praticava atividade física considerou esta como um passatempo ou diversão (Figura 17).

### Espiritualidade

A árvore gerada pela classe Espiritualidade estabeleceu forte relação com a Plenitude e Felicidade. Neste modelo, o algoritmo encontrou dificuldades para classificar corretamente as alunas do sexo feminino. O algoritmo necessitou estender a investigação em outras variáveis para aprender e tomar decisões (Figura 18). Uma possível conclusão foi de que alunos do sexo feminino apresentou maior complexidade de entendimento considerando a classe Espiritualidade. O algoritmo classificou corretamente aproximadamente 73% desta classe.

### Desenvolvimento Intelectual

A árvore da classe Desenvolvimento Intelectual (Figura 19) foi a que obteve menor percentual de acertos entre as classes, com aproximadamente 69,4% de acertos. Entretanto, observando a matriz de confusão, verificou-se a elevada porcentagem de amostras positivas classificadas corretamente (78,2%). Uma conclusão para este modelo foi que os alunos com Saúde e Disposição também apresentou satisfação intelectual, enfatizando a importância de o corpo estar bem para a mente também apresentar bom comportamento e desenvoltura.

Resultados significativos utilizando a mesma técnica de classificação por árvores de decisão foi realizada por (Sato et al, 2013), que também utilizou o software WEKA para classificação de imagens.

```

plenitudeefelicidade = nao: nao (386.0/103.0)
plenitudeefelicidade = sim
  relacionamentosfamiliares = nao
    Sexo = Feminino
      Filhos = naotemfilhos
        saudeedisposicao = sim: nao (7.0)
        saudeedisposicao = nao
          realizacaoeproposito = sim
            desenvolvimento intelectual = sim
              vidasocial = sim: nao (4.0)
              vidasocial = nao: sim (2.0)
            desenvolvimento intelectual = nao: sim (3.0)
          realizacaoeproposito = nao: nao (3.0)
        Filhos = 01filho: nao (3.0/1.0)
        Filhos = 02filhos: sim (2.0)
        Filhos = naoresp: nao (0.0)
        Filhos = 03oumaisfilhos: nao (0.0)
      Sexo = Masculino: sim (31.0/10.0)
      Sexo = naoresp: sim (0.0)
    relacionamentosfamiliares = sim: sim (321.0/69.0)

```

```

=== Confusion Matrix ===
      a  b  <-- classified as
    271 113 |   a = sim
      93 285 |   b = nao

```

Correctly Classified Instances	556	72.9659 %
Incorrectly Classified Instances	206	27.0341 %

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,706	0,246	0,745	0,706	0,725	0,460	0,721	0,699	sim
0,754	0,294	0,716	0,754	0,735	0,460	0,721	0,659	nao
0,730	0,270	0,730	0,730	0,730	0,460	0,721	0,679	

Figura 18. Resultados encontrados para a classe Espiritualidade

```

saudeedisposicao = sim: sim (362.0/108.0)
saudeedisposicao = nao
  contribuicaosocial = nao
    realizacaoeproposito = sim
      equilibrioemocional = sim: sim (35.0/14.0)
      equilibrioemocional = nao: nao (43.0/15.0)
    realizacaoeproposito = nao: nao (267.0/52.0)
  contribuicaosocial = sim: sim (55.0/21.0)

```

Correctly Classified Instances	529	69.4226 %
Incorrectly Classified Instances	233	30.5774 %

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,782	0,391	0,661	0,782	0,716	0,396	0,710	0,640	sim
0,609	0,218	0,741	0,609	0,669	0,396	0,709	0,700	nao
0,694	0,304	0,702	0,694	0,692	0,396	0,709	0,671	

```

=== Confusion Matrix ===
      a  b  <-- classified as
    294  82 |   a = sim
    151 235 |   b = nao

```

Figura 19. Resultados encontrados para a classe Desenvolvimento Intelectual

### Insights gerados pelo algoritmo Apriori

Foi realizada uma simulação com o algoritmo Apriori com as respostas originais da Roda da Vida, com valores pontuados de 1 a 10. O algoritmo gerou 102 regras com confiança mínima de 90%. Os resultados foram: a) Se estado civil fosse igual a solteiro, e atribuído valor oito para Plenitude e Felicidade, então o grupo de alunos não tinha filhos. Isto ocorreu em 83 vezes dos 85 possíveis, com confiança de 98%; b) Se estado civil fosse igual a solteiro, foi atribuído valor nove para satisfação com Relacionamentos Familiares, então o grupo de alunos não tinha filhos. Isto ocorreu em 78 de 80 casos possíveis, com confiança de 97%; c) Se estado civil fosse igual a solteiro, foi atribuído valor oito para Desenvolvimento Intelectual, então o grupo de alunos não tinha filhos. Isto aconteceu em 109 dos 112 casos

possíveis, com confiança de 97%; d) Se o estado civil fosse igual a solteiro, foi atribuído valor oito para Saúde e Disposição, então o grupo de alunos não tinha filhos, isto aconteceu em 86 de 90 casos possíveis, com confiança de 96%; e) Se estado civil fosse igual a solteiro, foi atribuído valor oito para Equilíbrio Emocional, então o grupo de alunos não tinha filhos, isto aconteceu em 82 de 87 casos possíveis, com confiança de 94%; f) Se estado civil fosse igual a solteiro, foi atribuído valor oito para Vida Social, então o grupo de alunos não tinha filhos. Isto aconteceu em 78 dos 83 casos com confiança de 94%; g) Se idade era entre 26 e 30 anos, e o grupo atribuiu satisfação dez aos Relacionamentos Amorosos, então o grupo de alunos não tinha filhos. Isto ocorreu em 78 de 86 casos possíveis, com confiança de 91%.

Apesar de apenas 29% dos alunos terem filhos, o algoritmo encontrou regras com valores mais elevados de satisfação para essas pessoas. Com o objetivo de encontrar outro cenário, foi gerada outra simulação reduzindo a confiança mínima para 70% a fim de encontrar padrões de comportamento de alunos que possuem filhos. Porém a única regra que o algoritmo gerou foi: Alunos que tem um filho são casados. Isto ocorre em 94 dos 121 possíveis, com confiança de 78%.

Vale destacar que existem diversos estudos sobre o tema como, por exemplo, um estudo conduzido por Gabb et al. (2013), no Reino Unido, que recomenda não ter filhos para uma vida mais próspera no casamento. Por outro lado, há diversos estudos conflitantes, como o site de um conselho de famílias contemporâneas (Glass, 2016), que revelou que estas relações de felicidade relacionados a paternidade é dada em função do país que vivem.

Os algoritmos de mineração de dados mostram-se muito úteis em sistemas de apoio à tomada de decisão. Os algoritmos de classificação em árvore utilizados neste trabalho apresentaram excelentes resultados de acordo com os modelos validados em outros conjuntos de dados utilizando-se das mesmas técnicas mencionadas pelos autores (Agarwal, 2015; Witten et al, 2011 e Dean, 2014). De acordo com os autores, os algoritmos se mostraram muito úteis para classificar e desenvolver modelos preditivos.

Baker (2015) comparou os desempenhos de modelos tradicionais de avaliação de crédito com modelos de avaliação de crédito utilizando-se de modelos de mineração de dados. Os modelos similares aos utilizados neste trabalho com conjuntos de dados com estrutura similar, apresentaram resultados melhores se comparados com modelos tradicionais como regressão logística. Uma das bases de dados mais conhecidas e testadas encontram-se disponíveis e são disponibilizadas no repositório da UCI para validação de modelos de mineração de dados. O modelo de dados utilizado neste trabalho apresenta características semelhantes a modelos clássicos como Iris, “Wine Quality”, “Breast Cancer Wisconsin (Diagnostic)”.

### Conclusão

De acordo com as análises realizadas, o uso de mineração de dados através da ferramenta WEKA mostra-se eficiente na descoberta de padrões e relações entre as diversas áreas da Roda da Vida, sendo possível a aplicação em outros cenários com o objetivo de extrair conhecimentos e informações ocultas conforme abordado na literatura sobre o assunto referenciadas neste trabalho. A capacidade de descobrir tendências com base em informações faz com que a ferramenta esteja cada vez mais utilizada para resolver problemas de análise

preditiva e prescritiva. Outras tarefas de mineração de dados podem ser empregadas em trabalhos futuros, como por exemplo a clusterização, e com isso, segmentar perfis de comportamento de alunos e descobrir similaridades e diferenças aparentemente ocultas entre estes também, e com isso, adotar estratégias mais personalizadas no relacionamento com estes alunos.

### Referências

- Aggarwal, C.C. 2015. *Data mining: the textbook*. Springer, New York, NY, USA. Disponível em: <<https://rd.springer.com/book/10.1007/978-3-319-14142-8>>. Acesso em: 11 maio 2017.
- Berry, M.J.A.; Linoff, G.S. 2004. *Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management*. 2ed. Wiley Publishing, Indianapolis, Indiana, USA.
- Baker, B. 2015. *Consumer Credit Risk Modeling*. Disponível em: <<https://bowenbaker.github.io/assets/credit-score.pdf>>. Acesso em: 02 abr. 2018.
- Brookshear, J.G. 2013. *Ciência da Computação: Uma Visão Abrangente*. 11ed. Bookman, Porto Alegre, RS, Brasil.
- Dean, J. 2014. *Machine Learning: Value Creation for Business Leaders and Practitioners*. John Wiley & Sons. New Jersey, USA.
- Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. 1996. From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence. AI Magazine*. 17(3): 37-54.
- Gabb, J.; Klett-Davies, M.; Fink, J.; Thomae, M. 2013. *Enduring Love? Couple Relationships in the 21st Century*. Disponível em: <<http://www.open.ac.uk/researchprojects/enduringlove/sites/www.open.ac.uk/researchprojects/enduringlove/files/files/ecms/web-content/Final-Enduring-Love-Survey-Report.pdf>>/. Acesso em: 07 jun. 2017.
- Glass, J. 2016. *CCF BRIEF: Parenting and Happiness in 22 Countries*. Disponível em: <<https://contemporaryfamilies.org/brief-parenting-happiness/>>. Acesso em: 01 jun. 2017.
- Ham, J.; Kamber, M. 2006. *Data Mining: Concepts and Techniques*. 2ed. Elsevier. San Francisco, CA, USA.
- Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons. New Jersey, USA.
- Marques, J.R. 2013. *Roda da Vida: O que é e como funciona?* Disponível em: <<http://www.ibccoaching.com.br/portal/coaching/conheca-ferramenta-roda-vida-coaching/>>. Acesso em: 05 fev. 2017.
- Mitchel, T. 1997. *Machine Learning*. Mc Graw Hill, Singapore.
- Rokach, L.; Maimon, O. 2015. *Data Mining With Decision Trees: Theory and Applications*. 2ed. World Scientific. Singapore.
- Sato, L.Y.; Shimabukuro, Y.E.; Kuplich, T.M.; Gomes, T.C.F. 2013. Análise comparativa de algoritmos de árvore de decisão do sistema WEKA para classificação do uso e cobertura da terra. *Simpósio Brasileiro de Sensoriamento Remoto* 16: 2353-2359.

Taie, E.S. 2011. Coaching as an Approach to Enhance Performance. *The Journal for Quality and Participation*: 34(1): 34-38.

Witten, I.H.; Frank, E.; Hall, M.A. 2011. *Mark A. Data Mining: Practical Learning Machine Tools and Techniques*. 3ed. Elsevier. New York, NY, USA.