



# M

## INERAÇÃO DE DADOS EM UM PROCESSO DE SELEÇÃO DE PESSOAS

Recebido: 05/02/2019

Aprovado: 12/05/2019

<sup>1</sup>Ítalo Alberto do Nascimento Sousa

<sup>2</sup>Edison Spina

<sup>3</sup>Antonio Geraldo da Rocha Vidal



### Resumo

Este relato tecnológico consiste em uma investigação do uso de técnicas de mineração de dados em uma das principais frentes da Gestão de Recursos Humanos, a seleção de pessoas. Ele é desenvolvido por meio de um estudo de caso em uma Organização da Sociedade Civil, onde métodos de exploração de dados são aplicados em um processo de seleção de pessoas. Avaliadores da organização de estudo indicam que o processo de avaliação dos candidatos possui muitos critérios de seleção, com diferentes graus de importância, e que no momento da decisão todos são considerados iguais, o que traz incertezas e tornam a seleção mais subjetiva. Nesse contexto, o objetivo desta pesquisa é tornar a seleção mais rápida e eficiente por meio da descoberta dos principais critérios de seleção com o auxílio do uso de mineração de dados. Com isso, almeja-se melhorar a aderência dos candidatos selecionados de acordo com os objetivos da seleção, de forma que ele atenda com mais assertividade aos requisitos estipulados pela organização. Tem-se também como objetivo, reduzir o esforço dos avaliadores na seleção, por meio da redução de tempo e etapas. Os resultados indicam que alguns critérios de avaliação influenciam mais na tomada de decisão que outros, como os critérios Comprometimento e Trabalho em equipe. Esses critérios podem ser utilizados para suporte na tomada de decisão dos candidatos incertos, pois indicam quais características pesam mais na seleção. O resultado de quais critérios são mais importantes na tomada de decisão são usados na organização, tornando seu processo de seleção mais rápido e assertivo.

**Palavras-chave:** Mineração de dados. Seleção de pessoas. Gestão de Recursos Humanos.

1Universidade de São Paulo – USP, São Paulo, (Brasil). E-mail: [italoengenhariaunb@gmail.com](mailto:italoengenhariaunb@gmail.com)  
Orcid id: <https://orcid.org/0000-0001-7370-6133>

2Universidade de São Paulo – USP, São Paulo, (Brasil). E-mail: [spina@usp.br](mailto:spina@usp.br)  
Orcid id: <https://orcid.org/0000-0002-9951-5738>

3Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo – FEA/USP, São Paulo, (Brasil). E-mail: [vidal@usp.br](mailto:vidal@usp.br) Orcid id: <https://orcid.org/0000-0002-8124-671X>

## **DATA MINING IN A PERSONNEL SELECTION PROCESS**

### **Abstract**

This technology report consists of an investigation of the use of data mining techniques on one of the main fronts of Human Resource Management, personnel selection. It is developed through a case study in a Civil Society Organization, where data mining methods are applied in a personnel selection process. Evaluators of the organization that is being studied indicate that the candidate evaluation process has many selection criteria, with different degrees of importance, and that at the moment of the decision all are considered equal, which brings uncertainties and make the selection more subjective. In this context, the objective of this research is to make selection faster and more efficient by discovering the main selection criteria with the use of data mining. This aims to improve the adherence of selected candidates in accordance with the objectives of the selection, so that it meets more assertively the requirements stipulated by the organization. It also aims to reduce the effort of evaluators in the selection, by reducing time and steps. The results indicate that some evaluation criteria influence decision making more than others, such as the Commitment and Teamwork criteria. These criteria can be used to support decision making of uncertain candidates, as they indicate which characteristics weigh more in the selection. The result of the importance of each criteria in decision making is used throughout the organization, making it faster and more assertive.

**Keywords:** Data Mining. Personnel Selection. Human Resource Management.

## INTRODUÇÃO

É notável um investimento crescente nos últimos anos das organizações na área da gestão de RH (Recursos Humanos), independente do porte ou da área de atuação. Essa área é reconhecida cada vez mais como um recurso estratégico na busca por vantagem competitiva (Baylão & Rocha, 2014). Segundo Caxito (2012), ela é responsável pela gestão das pessoas em uma organização e pode ser dividida em seis processos, como expresso pela figura 1.

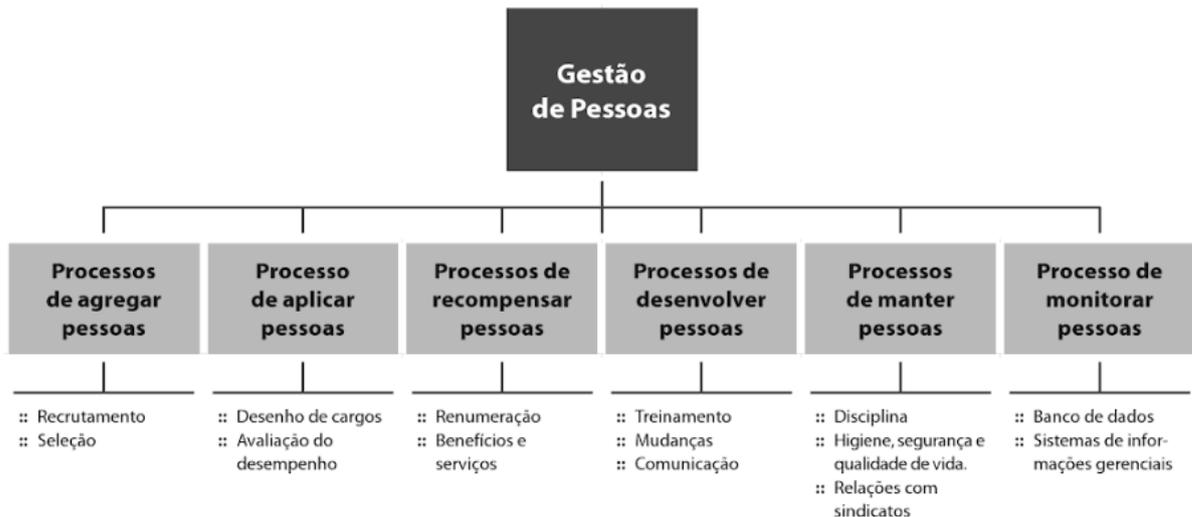


Figura 1: Divisões da área de gestão de pessoas (Caxito, 2012).

O processo de agregar pessoas é responsável pelo Recrutamento e Seleção de Pessoas; O processo de aplicar pessoas realiza a incorporação das pessoas à empresa, onde trabalha na socialização do funcionário e no desempenho de sua função; O processo de recompensar pessoas é o responsável pelas políticas de remuneração, de benefícios e incentivos; O processo de desenvolver pessoas é o responsável pelo treinamento e desenvolvimento das pessoas na organização; O processo de manter pessoas está relacionado com disciplina, higiene, saúde e qualidade de vida dos funcionários na organização; O processo de monitorar pessoas é responsável pela elaboração e manutenção de um banco de dados para auxiliar profissionais de RH no suporte das tomadas de decisão sobre as pessoas da organização. Nesse âmbito, cada um desses processos proporciona um diferente meio para se relacionar com pessoas na organização com objetivo de melhorar os objetivos do negócio (Caxito, 2012).

Estudos que envolvem a aplicação de técnicas de Mineração de dados para RH têm crescido de maneira expressiva nos últimos anos, como pode ser visto na figura 2 .

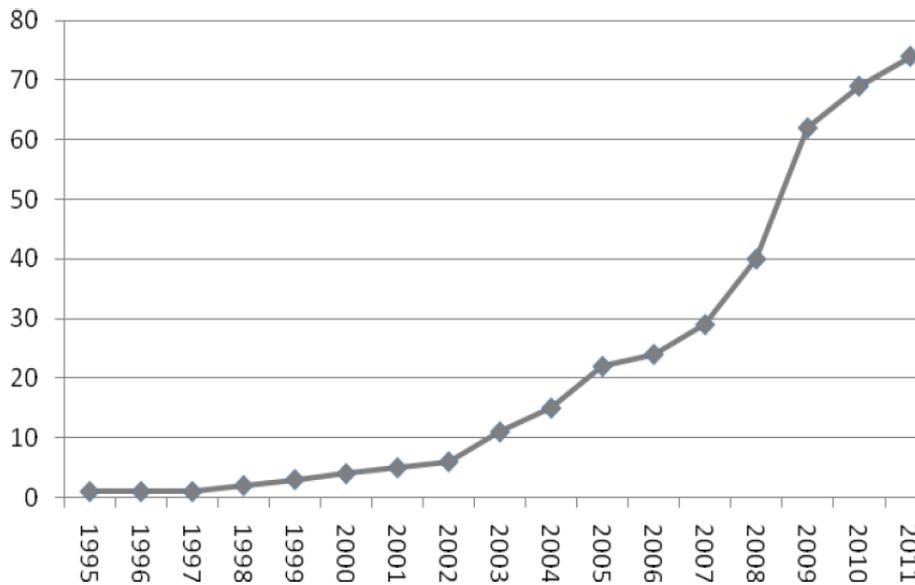


Figura 2:

Figura 2: Publicação de artigos em *Mineração de Dados para Recursos Humanos* (Strohmeier e Piazza, 2012).

O processo de agregar pessoas, área que atua com Recrutamento e Seleção destaca-se entre os processos de RH citados, pois é responsável por agregar os talentos necessários na busca por vantagem competitiva diante de um contexto altamente dinâmico e complexo. Apesar desses dois termos serem comumente abordados como uma só função, as atividades de cada uma variam. O Recrutamento consiste na procura e atração de candidatos para uma determinada função na organização. Já a Seleção, que será trabalhada nesta pesquisa, consiste na tomada de decisão do candidato a ser escolhido para atuar em determinada função da organização, ou seja, no tipo de talento a ser adquirido (Baylão & Rocha, 2014; Caxito, 2012).

A partir disso, esta pesquisa tem como objetivo investigar sobre o uso da mineração de dados com métodos exploratórios em um processo real de seleção de pessoas de uma organização para avaliar se pode ser possível tornar a seleção mais rápida e eficiente. Portanto, almeja-se melhorar a aderência dos candidatos selecionados no processo, de forma que possam atender com mais assertividade aos requisitos estipulados pela organização.

#### CONTEXTO DA REALIDADE INVESTIGADA

Este relato tecnológico é baseado em dados reais de uma Organização da Sociedade Civil que atua com assistência social. Em sua missão, ela parte da ideia do ditado popular “Não dê o peixe, mas ensine a pescar” como forma de atuação. Ensinar a pescar, do ponto de vista desta organização, é trabalhar na melhoria das chamadas habilidades socioemocionais, que são compostos por habilidades intrapessoais e interpessoais, como autoconhecimento, facilidade de comunicação, inteligência emocional, planejamento pessoal e de carreira, prioritariamente focando em estudantes de baixa renda. A motivação da organização é que o pleno desenvolvimento dessas habilidades possa tornar os profissionais mais

competitivos no mercado de trabalho, e que assim, possam ter mais condições de crescer profissionalmente.

Para viabilizar sua missão, a organização conta com profissionais líderes do mercado nas diferentes linhas de habilidades socioemocionais que contribuem voluntariamente na realização de uma formação gratuita para 55 estudantes por um período de quatro meses. Como forma de retorno, e também como parte do programa de formação, esses estudantes se reúnem em cinco grupos de onze pessoas e trabalham colaborativamente em projetos que impactam positivamente a sociedade nesse período de quatro meses. Esses projetos devem priorizar as áreas mais carentes da sociedade, tais como órfãos, idosos abandonados, moradores de rua, viciados em drogas, deficientes físicos, doentes mentais ou refugiados.

O público-alvo da organização são estudantes que possuem bons níveis de habilidades comportamentais básicas, como conseguir trabalhar em equipe, interagir sem ofender, organizados, flexíveis e comprometidos com sua formação tanto pessoal quanto profissional. Tais características são almejadas pois durante o programa eles conduzem suas atividades atuando com questões sociais críticas, lidando diretamente com pessoas em vulnerabilidade, atividades que exigem bom rendimento para que possa ser gerado um efetivo retorno social. Baseado nisso, a organização estabelece os objetivos principais do processo de seleção, pelos quais busca-se descobrir sobre os candidatos: o nível de interesse para aprimorar suas habilidades socioemocionais; o nível de suas habilidades comportamentais básicas; o nível de engajamento para participar do programa de formação nos quatro meses de duração.

Para representar esses objetivos, são elaboradas pelos membros da organização alguns critérios de avaliação que buscam aumentar o conhecimento sobre os candidatos e sintetizar o que é esperado de um candidato para exercer com bom desempenho as atividades propostas. Ou seja, são identificados elementos que traduzem o que a organização gostaria de ter como valores fundamentais na execução de suas atividades por meio dos candidatos selecionados.

Nesse contexto, a organização elaborou critérios de avaliação psicométricos com intuito de mensurar os níveis comportamentais de cada candidato no processo de seleção. Resultados da avaliação no processo de seleção da organização trabalhada fornecem os dados utilizados na análise desta pesquisa. Em seu processo, são realizadas duas etapas de avaliação quem geram pontuações finais no processo, a dinâmica em grupo e a entrevista individual. Essas pontuações são coletadas por meio da avaliação dos critérios e inseridas em formulários preenchidos pelos avaliadores das duas etapas, que ocorrem de forma independente uma da outra na avaliação. Na dinâmica em grupo são utilizadas os seguintes critérios e pontuações:

Critérios da Dinâmica	Pontuação
Comprometimento	0 - 5
Comunicação	0 - 5
Motivação	0 - 5
Trabalho em Equipe	0 - 5
Organização	0 - 5
Iniciativa	0 - 5

Figura 3: Critérios e pontuações da dinâmica em grupo.

Na dinâmica em grupo, é colocado um critério a mais para geração da pontuação, a Iniciativa, pois os membros da organização indicam que pode compensar o contato pessoal da entrevista. Na entrevista individual são utilizadas os seguintes critérios e pontuações:

Critérios da Entrevista	Pontuação
Comprometimento	0 - 5
Comunicação	0 - 5
Motivação	0 - 5
Trabalho em Equipe	0 - 5
Organização	0 - 5

Figura 4: Critérios e pontuações da entrevista individual.

Os dois passos ocorrem de forma independente, pelos quais os avaliadores não interagem entre si e não influenciam na tomada de decisões um do outro. Cada etapa conta com status (resultado final) de aprovação e porcentagem de pontuação como resultado. A porcentagem de pontuação é calculada pela relação entre a soma total dos critérios do candidato pela soma total máxima. A entrevista possui o status de Verificação para deixar a decisão final do status de aprovação do candidato para o avaliador no fim do processo.

Etapa de seleção	Status	Pontuação
Dinâmica em grupo	Aprovado (A) Reprovado (R)	0 a 100%
Entrevista individual	Aprovado (A) Reprovado (R) Verificação (V)	0 a 100%

Figura 5: Denominação do estado de aprovação.

Na entrevista, tem-se a avaliação da Disponibilidade do candidato, que é medida de forma isolada e não contribui na geração da pontuação para seleção dos candidatos, influencia apenas como suporte na decisão nos desempates ou da verificação. Esse critério é gerado por meio de uma pergunta simples

e direta sobre a participação de cada candidato na formação. Ela é definida de forma qualitativa pelo avaliador com as opções de resposta:

Nível	Descrição
A	Candidatos que têm certeza de que irão até o final e absolutamente não faltarão
B	Candidatos que pensam que irão até o fim, mas talvez venham a faltar algum dia
C	Candidatos que não têm muita certeza se irão até o final

Figura 6: Índices de disponibilidade e sua descrição.

### DIAGNÓSTICO DE OPORTUNIDADE

Segundo os membros da organização, a cada processo há mais inscritos e fica mais difícil a seleção dos melhores candidatos. Também, como os recursos são limitados, eles só podem trabalhar com esse número máximo de 55 estudantes. Ou seja, a seleção dos candidatos a serem escolhidos no processo deve ser bem elaborada. Se ela não for elaborado de uma forma bem articulada, a tendência é tanto de aumentar o número de evasão no decorrer dos meses do programa, como também de uma má realização dos projetos sociais.

A partir disso, tem-se como objetivo, tornar o processo de seleção mais rápido e eficiente. Isso é feito por meio da determinação dos principais critérios de avaliação e da diminuição de subjetividade de escolha. Com isso, almeja-se melhorar a aderência dos candidatos selecionados de acordo com os objetivos da seleção, de forma que ele atenda com mais assertividade aos requisitos estipulados pela organização. Tem-se também como objetivo, reduzir o esforço dos avaliadores na seleção, por meio da redução de tempo e etapas.

No processo, os dados fornecidos da avaliação não estavam em uma formatação adequada. Portanto, foi realizada uma estruturação desses dados para transformação em dados prontos para análise. Dentre os problemas analisados, o primeiro foi de candidatos que se cadastraram, fizeram o teste online, mas não participaram do estágio presencial do processo. O segundo problema foi em relação à formatação da tabela, que não seguiu linhas ordenadas de tópicos e dados. Algumas células tinham erros devido a resultados desconfigurados de cálculo. Assim também como dados perdidos e observações dispersas na planilha. Para resolver esse problema e aumentar a qualidade dos dados, foram realizados alguns tratamento, como limpeza de dados faltantes, estimação de dados baseado em correspondentes anteriores ou posteriores, exclusão de linhas quando se encontram significativos dados faltantes. Foi realizado um tratamento e preparação desses dados, trazendo uma nova planilha de dados analíticos que pode ser utilizada agora em um processo de mineração de dados.

As abordagens, ou metodologias, de mineração de dados descrevem como o processo de mineração de dados deve ser conduzido. Essas abordagens podem ser do tipo *top-down* ou *bottom-up*, como pode ser visto na figura 7 (Côrtes, Porcaro, & Lifschitz, 2002).

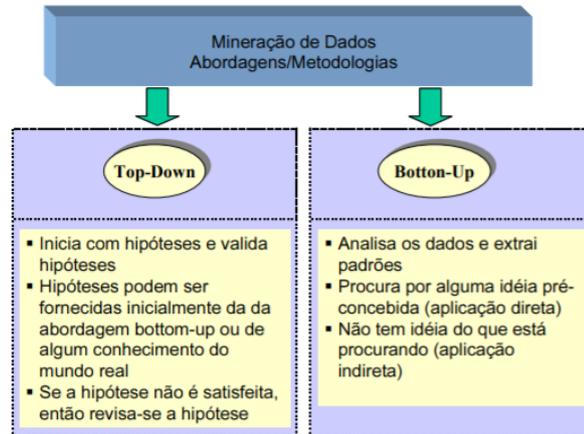


Figura 7: Tipos de abordagens de Mineração de Dados (Côrtes, Porcaro e Lifschitz, 2002).

A abordagem *top-down* funciona como uma forma avaliativa de teste de hipóteses, pois ela é utilizada quando há um cenário pré-concebido que pode ser confirmado ou refutado. Já a abordagem *bottom-up* funciona como uma busca de conhecimento, pois se inicia o processo com uma exploração de dados com o intuito de descobrir padrões que sejam úteis ao conhecimento. Há também uma terceira abordagem chamada de metodologia híbrida que incorpora uma combinação das duas abordagens (Côrtes, Porcaro, & Lifschitz, 2002).

Neste projeto de análise, utiliza-se a condução segundo uma abordagem *bottom-up*. Nela, busca-se analisar os dados sobre a seleção de forma exploratória, que permita gerar alguma ideia pré-concebida sobre seus elementos, para extrair padrões e melhorar o processo. Para isso, primeiramente, busca-se compreender o comportamento dos dados em relação a uma variável-chave em um processo de seleção, o status de aprovação. Pode-se observar que em ambas as distribuições, nas figuras 8 e 9, o status de aprovação tende a ser maior em valores de pontuação mais altos e os pontos de probabilidade vão nessa direção. Mesmo assim, a diferença da média entre aprovação e reprovação na primeira é maior que na segunda. Percebe-se também, na entrevista individual, que a distribuição dos candidatos aprovados e sujeitos à Verificação é semelhante, podendo até tornar-se o mesmo parâmetro.

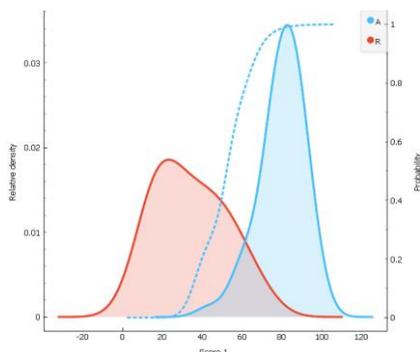


Figura 8: Distribuição da pontuação da Dinâmica em grupo em relação ao status de aprovação.

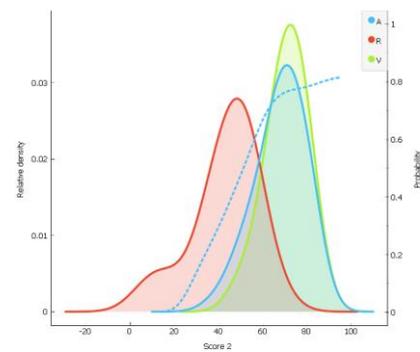


Figura 9: Distribuição da pontuação da Entrevista Individual em relação ao status de aprovação.

## PROPOSTA DE INTERVENÇÃO

Para explorar os dados do processo de seleção e conseguir melhorá-lo, utilizam-se técnicas de mineração de dados que melhor se encaixam ao problema. Inicia-se com uma técnica chamada de

Agrupamento, nela candidatos com pontuação geral de critérios semelhantes são agrupados e grupos com pontuação geral diferentes são separados. De acordo com os dados da seleção, tem-se como resultado o agrupamento representado pela figura 10. Nela, grupos não-destacados são compostos totalmente pelos candidatos aprovados. Já grupos destacados em azul e vermelho representam grupos de candidatos amplamente rejeitados.

Com esses resultados, verifica-se que o grupo azul representa todos os candidatos reprovados e o grupo vermelho representa os candidatos que possuem características semelhantes de reprovação mas que possuem alguns candidatos bem sucedidos. Essa composição demonstra que o grupo vermelho possui no geral candidatos com pontuações gerais semelhantes nos critérios de avaliação mas que por algum motivo são aprovados.

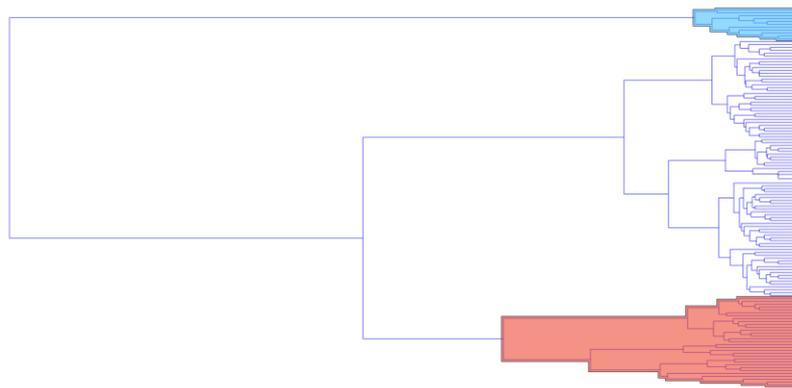


Figura 10: Agrupamento pelas pontuações dos critérios.

Para explorar esse acontecimento, busca-se saber no agrupamento vermelho o que difere os candidatos aprovados dos demais. Para isso, realiza-se um ranking dos critérios que trazem o maior ganho de informação (Info. Gain) para a análise de aprovação. Os resultados obtidos trazem a resposta de quais critérios possuem maior importância na avaliação.



Figura 11: Ranking do ganho de informação dos critérios.

Na entrevista individual, o melhor ganho de informação vem do Comprometimento (Commitment), seguido pela Motivação (Motivation) e Trabalho em Equipe (Team work), figura 11. Na dinâmica em grupo, ele vem do Trabalho em equipe, depois da Comunicação e Iniciativa. Esse ranking demonstra a importância de cada critério no momento da decisão.

Na análise do quão diferentes esses candidatos bem sucedidos diferem dos demais, o gráfico de dispersão é usado para verificar qual é a relação entre esses candidatos e as pontuações dos critérios. Ao verificar cada item, nota-se que aqueles que oferecem maiores valores de Comprometimento ou Trabalho em Equipe no

eixo horizontal, apesar de terem características predominantemente semelhantes aos reprovados, possuem maiores tendências de serem aprovados na seleção por possuírem maiores incidências de aprovação.

Como o critério Comprometimento é o que traz o maior ganho de informação, analisa-se, por meio do agrupamento vermelho, o gráfico da dispersão dos candidatos aprovados em relação às suas pontuações, apresentado pela figura 12. Este gráfico indica que candidatos com maiores pontuações de Comprometimento, mesmo tendo várias semelhanças com outros reprovados, diferem-se no resultado final e conquistam aprovação.

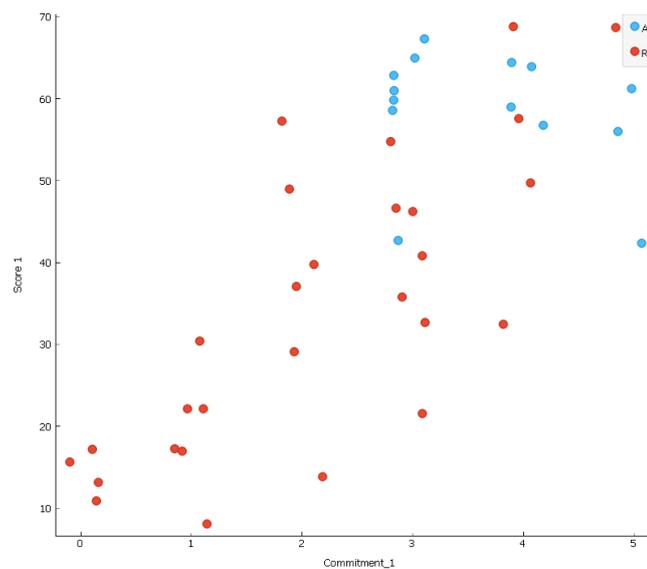


Figura 12: Gráfico de dispersão da pontuação pela aprovação da métrica de comprometimento.

Utiliza-se uma árvore de decisão para classificar as pessoas que respondem mais positivamente à pergunta simples e final sobre o desejo de comparecer de forma assídua ao programa de formação, como pode ser visto pela figura 13. Nota-se então, que os candidatos que tiveram a maior taxa de aprovação em todas as etapas do processo foram aqueles que tiveram o maior nível de disponibilidade nessa pergunta final.

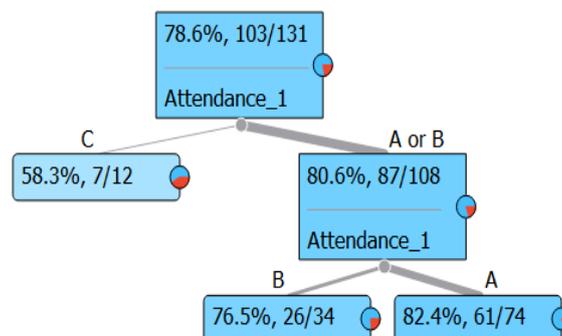


Figura 13: Árvore de decisão da porcentagem de aprovação pelo nível de Disponibilidade dos candidatos.

## CONCLUSÕES E RECOMENDAÇÕES FINAIS

A ideia da utilização de mineração de dados para o processo de seleção é deixá-lo mais objetivo utilizando fatos como suporte na tomada de decisão e não apenas as impressões. A experiência dos membros da organização coloca o critério Disponibilidade como isolado e como última opção para colaborar na avaliação. Porém, os resultados analíticos indicam que esse critério traz bastante assertividade ao processo, pois tem uma relação proporcional com a taxa de aprovação, mesmo sem utilização desse critério na geração da pontuação do processo.

Os resultados da mineração de dados indicam que o critério Comprometimento tem maior influência tanto na Aprovação quanto na Reprovação dos candidatos. Pela seu grau de importância, percebe-se que é viável colocar um peso maior para esse critério na avaliação. Essa proposta faz com que os avaliadores sejam mais criteriosos ao julgar esse critério, e com isso, podem aumentar sua representatividade na construção do resultado final.

O conhecimento dos critérios que mais influenciam na avaliação foi incorporado no processo de seleção dessa organização como fator decisivo para seleção de candidatos incertos. Ou seja, candidatos que são aprovados em uma etapa e reprovados na outra participam de um novo ranking de seleção, onde os fatores de tomada de decisão são baseados nas notas de Comprometimento e Trabalho em equipe, assim como também no nível de disponibilidade registrado para o candidato.

Esse novo método que leva em consideração os critérios principais de seleção para apoio na tomada de decisão torna o processo mais rápido, pois com as pontuações estabelecidas ele já gera as indicações do resultado, e também torna mais assertivo pois é elaborado baseado no histórico quantitativo de classificação dos candidatos.

Desta forma, essa pesquisa traz como contribuição para outros processos de seleção de pessoas o fato de que encontrar os critérios que mais pesam em uma seleção é uma forma útil de suportar o processo de avaliação quando se têm candidatos incertos para tomada de decisão.

Para trabalhos futuros, almeja-se aplicar uma metodologia *top-down* para testar possibilidades adquiridas por meio dos conhecimentos que esse estudo gerou. Propõe também elaborar um esquema de pesos nas variáveis de aprovação em conjunto com os critérios principais para testar se o processo de seleção fica mais assertivo e descobrir como cada parâmetro influencia na decisão. Também podem ser realizadas outras análises em novos processos de seleção para refinar o estudo que propõe melhorar a tomada de decisão.

### Agradecimentos

Este trabalho foi realizado durante uma bolsa de estudos financiada pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). Agradecimento especial à ONG Instituto CEO do Futuro (<http://ceodofuturo.org.br/>) pela colaboração com a pesquisa. Agradecimento ao apoio e suporte do CEST (Centro de Estudos Sociedade e Tecnologia).

## REFERÊNCIAS

Baylão, A. L. da S.; Rocha, A. P. de S. (2014). A importância do processo de recrutamento e seleção de pessoal na organização empresarial. *XI Simpósio de Excelência em Gestão e Tecnologia – SEGET 2014*.

Caxito, F. de A. (1012). *Recrutamento e Seleção de Pessoas*. IESDE BRASIL AS.

Côrtes, S. da C.; Porcaro, R. M.; Lifschitz, S. (2002). *Mineração de dados: Funcionalidades, técnicas e abordagens*. Rio de Janeiro. PUC.

Strohmeier, S.; Piazza, F. (2012). Domain driven data mining in human resource management: A review of current research. *Expert Systems with Applications*, 40(7), 2410-2420.