

## **USO DE MINERAÇÃO DE DADOS EM UM PROCESSO DE SELEÇÃO DE PESSOAS**

Ítalo Alberto Do Nascimento Sousa - Universidade de São Paulo

Antonio Geraldo Da Rocha Vidal - FEA/USP

Edison Spina

### **Resumo**

Este relato tecnológico consiste em um estudo de caso da aplicação de Mineração de dados em uma Organização da Sociedade Civil. Ele se baseia em uma exploração de dados de um processo de seleção de pessoas. Tem-se como objetivo, tornar a seleção mais rápida e eficiente. Ou seja, determinar os principais critérios de avaliação e diminuir a 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. Os resultados indicam que alguns critérios de avaliação influenciam mais na tomada de decisão, como os critérios Comprometimento e Trabalho em equipe. Esses critérios podem ser utilizados para suporte na tomada de decisão de candidatos incertos, pois indicam quais características pesam mais na seleção. O resultado de quais critérios são os mais importantes em uma tomada de decisão são implantados hoje na organização e tornam o processo de seleção mais rápido e assertivo.

## Resumo

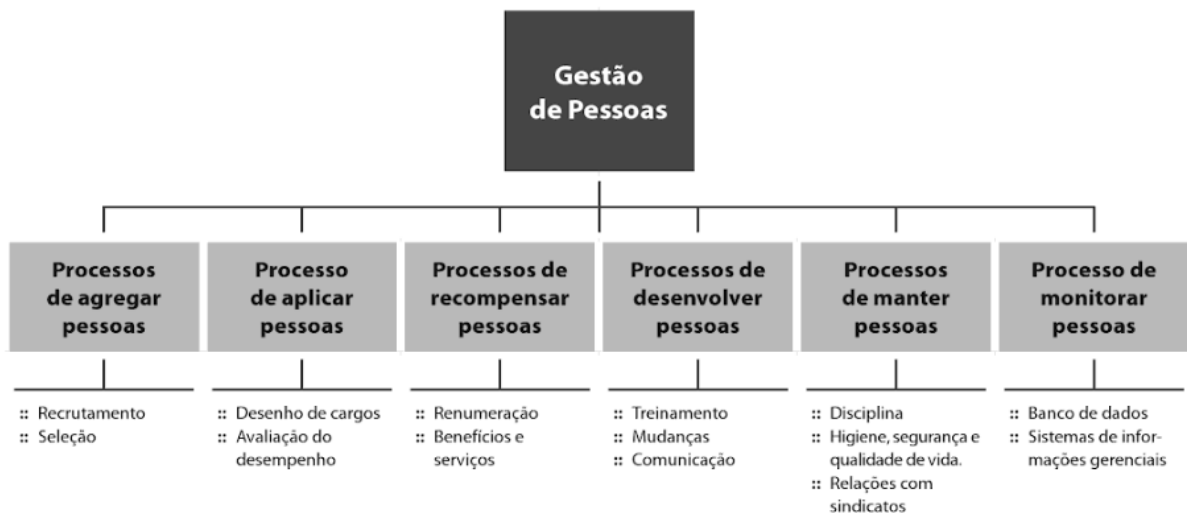
Este relato tecnológico consiste em uma investigação do uso de técnicas analíticas avançadas para Gestão de Recursos Humanos. Ele é desenvolvido por meio de um estudo de caso da aplicação de Mineração de dados em uma Organização da Sociedade Civil. Esse estudo de caso se baseia na exploração de dados de um processo de Seleção de pessoas. Tem-se como objetivo, tornar a seleção mais rápida e eficiente. Ou seja, determinar os principais critérios de avaliação e diminuir a 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. Os resultados indicam que alguns critérios de avaliação influenciam mais fortemente 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 os mais importantes em uma tomada de decisão são implantados hoje na organização e tornam o processo de seleção mais rápido e assertivo.

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

## 1 INTRODUÇÃO

É notável um investimento crescente nos últimos anos das organizações na área da gestão de RH (Recursos Humanos), independente do porte de cada organização. Ela é reconhecida cada vez mais como um recurso estratégico na busca por vantagem competitiva (BAYLÃO; ROCHA, 2014). Esta área, que é responsável pela gestão das pessoas em uma organização, pode ser dividida em seis processos, como expresso pela figura 1 Caxito (2012).

Figura 1 – Divisões da área de gestão de pessoas.

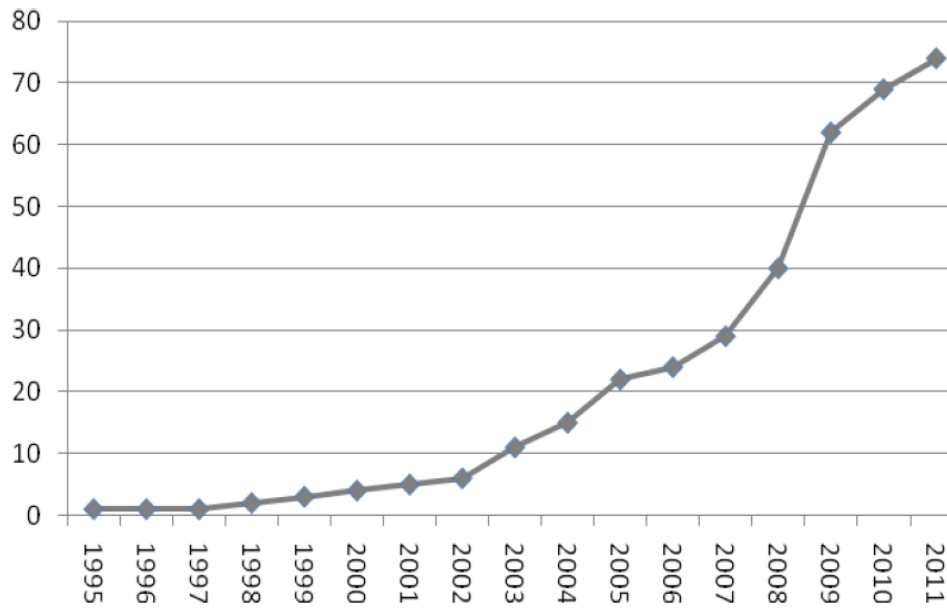


Fonte: Caxito (2012).

O Processo de agregar pessoas é o responsável pelo Recrutamento e Seleção de Pessoas. O Processo de aplicar pessoas realiza a incorporação das pessoas à empresa, ou seja, 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. E por fim, o Processo de monitorar pessoas é responsável pela elaboração e manutenção de um banco de dados que permita aos profissionais de RH e gestores o suporte nas tomadas de decisão no que se ao pessoal da organização.

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 – Publicação de artigos em Mineração de Dados para Recursos Humanos.



Fonte: Strohmeier e Piazza (2012)

Dentre todos os processos de Recursos Humanos citados, destaca-se o Processo de agregar pessoas, responsável pelo Recrutamento e Seleção. Ele é o responsável por agregar competências e talentos necessários para colaborar 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 requerida 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 (BAYLÃO; ROCHA, 2014) (CAXITO, 2012).

A partir disso, esta pesquisa tem como objetivo utilizar mineração de dados de forma exploratória em um processo de seleção de pessoas de uma organização para tornar a seleção mais rápida e eficiente.

## 2 CONTEXTO DA REALIDADE INVESTIGADA

Este relato tecnológico é baseado em dados reais de uma Organização da Sociedade Civil que atua com assistência social. Ela em sua missão, parte da idéia 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 dos chamados soft skills, que são compostos por habilidades socioemocionais, de comunicação, autoconhecimento, carreira e inteligência emocional para estudantes de baixa renda. Isso se dá para que o pleno desenvolvimento dessas habilidades possa tornar os profissionais mais competitivos no mercado de trabalho, e assim, terem mais condições de crescer profissionalmente.

Para viabilizar sua missão, a organização conta com profissionais líderes do mercado nas diferentes linhas de soft skills que contribuem voluntariamente na realização de um programa de formação gratuito 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 básicas individuais, como estudantes que trabalham bem em equipe, saibam conversar entre si, organizados, motivados e comprometidos com sua formação tanto pessoal quanto profissional. Tais características são almejadas pois durante o programa eles realizam projetos sociais reais que exigem um bom rendimento para um pleno retorno social. Baseado nisso, a organização estabelece os objetivos principais do processo de seleção, pelos quais busca-se descobrir sobre os candidatos:

- O quão interessados eles estão em aprimorar seus soft skills;
- O quão excelentes estão nas habilidades básicas individuais;
- O quão engajados eles estão para participar do programa de formação.

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.

Os dados utilizados na análise se baseiam nos critérios adotadas no processo de seleção para avaliar os candidatos. São realizadas duas etapas para avaliação e geração de pontuação no processo de seleção, a dinâmica em grupo e a entrevista individual. Eles são coletados por meio de formulários preenchidos pelos avaliadores no decorrer das etapas e ao fim do processo são transferidos para planilhas do Excel.

Na dinâmica em grupo são utilizadas os seguintes critérios e pontuações:

Figura 3 – Critérios e pontuações da dinâmica em grupo.

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

Na dinâmica em grupo, é colocada uma métrica 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:

Figura 4 – Critérios e pontuações da entrevista individual.

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

Os dois passos ocorrem de forma independente, ou seja, os avaliadores não interagem entre si e não influenciam na tomada de decisões um do outro. Cada etapa conta com status 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 do status de aprovação do candidato para o fim do processo.

Figura 5 – Denominação do estado de aprovação.

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%

Na entrevista, também há uma métrica diferente, que corresponde com a métrica Assiduidade, porém ela é isolada e não contribui na geração da pontuação para seleção dos candidatos, sim para peso de decisão nos desempates ou verificação. A métrica Assiduidade é gerada a partir de uma pergunta simples e direta sobre a participação de cada candidato no programa durante a entrevista. Ela é definida de forma qualitativa pelo avaliador com as opções de resposta:

Figura 6 – Índices de Assiduidade e sua descrição.

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

Os dados fornecidos do processo de seleção 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, foi realizado um tratamento de dados, como limpeza de dados faltantes, estimação de dados baseado em dados anteriores ou posteriores, e também são excluídas as linhas de dados quando quase

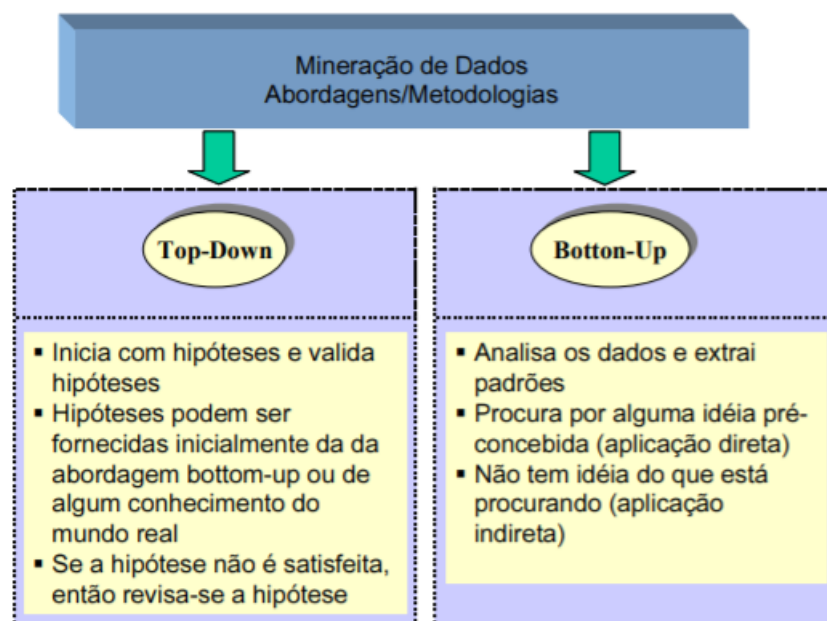
todos os dados da linha se encontram faltantes. Foi realizado um tratamento e preparação desses dados, trazendo uma nova planilha de dados analíticos que pode ser utilizada agora em um processo de mineração de dados.

## 2.1 DIAGNÓSTICO DA 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 a seleção mais rápida e eficiente. Ou seja, determinar os principais critérios de avaliação e diminuir a 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.

Figura 7 – Tipos de abordagens de Mineração de Dados.



Fonte: Côrtes, Porcaro e Lifschitz (2002)



As abordagens, ou metodologias, de mineração de dados descrevem como o processo é conduzido. Essas abordagens podem ser do tipo top-down ou bottom-up, como pode ser visto na figura 7. A abordagem top-down funciona como um teste de hipótese, pois ela é utilizada quando há uma ideia pré-concebida que se deseja confirmá-la ou refutá-la. 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 híbrida que funciona como a combinação das duas Côrtes, Porcaro e Lifschitz (2002).

Neste projeto de análise, utiliza-se a condução segundo uma abordagem bottom-up. Ou seja, busca-se analisar os dados sobre a seleção, 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 o padrão de distribuição dos candidatos aprovados e dos sujeitos à Verificação é bastante semelhante, podendo até tornar-se o mesmo parâmetro.

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

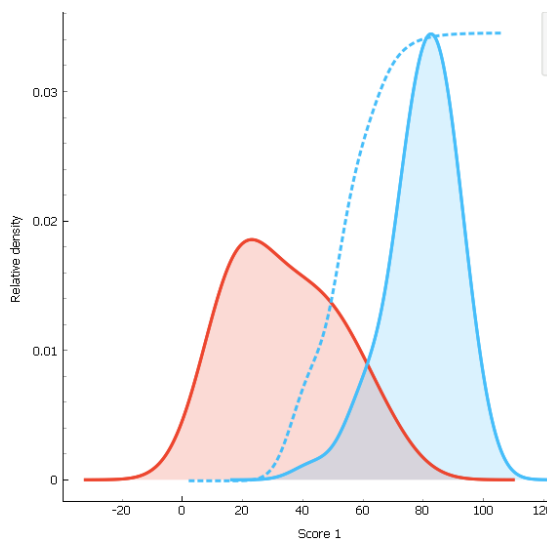
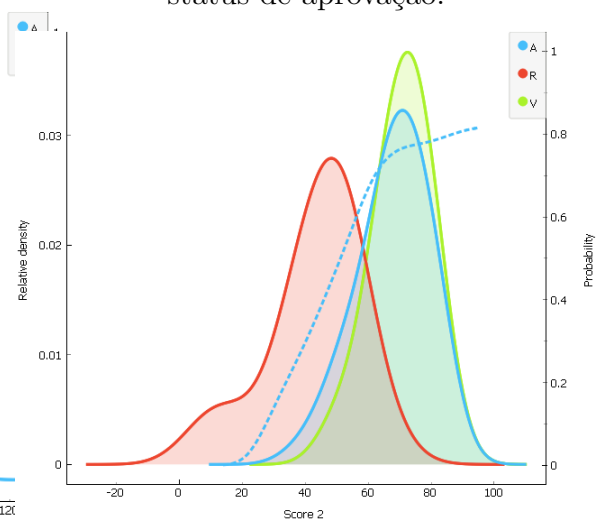


Figura 9 – Distribuição da pontuação da Entrevista Individual em relação ao status de aprovação.

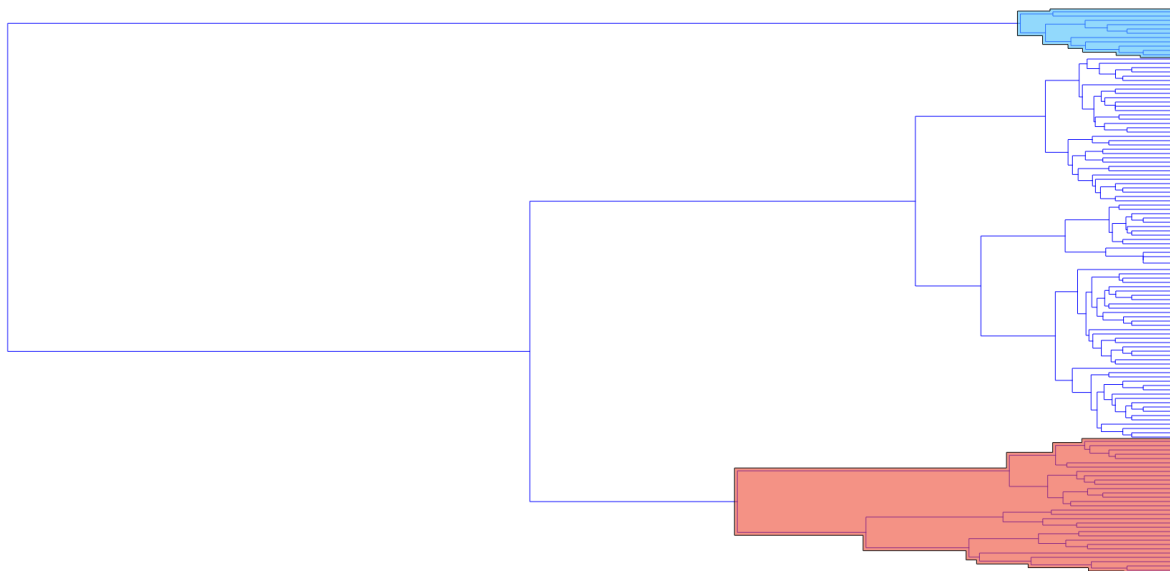


### 3 PROPOSTA DE INTERVENÇÃO

Para explorar os dados do processo de seleção e conseguir melhorá-lo, utilizam-se técnicas de mineração de dados que se encaixam nesse problema, como a classificação. Por meio da técnica de classificação chamada de Agrupamento, candidatos com pontuação de critérios semelhantes são agrupados e grupos com pontuações diferentes são divididos. Baseado nos dados da seleção, tem-se o agrupamento representado pela figura 10. Os grupos não-destacados são compostos totalmente pelos candidatos aprovados. Os grupos destacados em azul e vermelho representam o agrupamento dos candidatos amplamente rejeitados.

A partir disso, verifica-se que o grupo azul representa todos os candidatos reprovados, e já o grupo vermelho representa os candidatos que possuem características semelhantes de reprovação, mas que possuem alguns candidatos bem sucedidos. Ou seja, o grupo vermelho possui no geral candidatos com pontuações semelhantes nos critérios 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 para a análise de aprovação. Os resultados obtidos trazem a resposta de quais critérios possuem o maior peso na avaliação.

Na dinâmica em grupo, o melhor ganho de informação vem do Trabalho em equipe, depois da Comunicação e Iniciativa. Na entrevista individual, ele vem do Comprometimento, seguido pela Motivação e Trabalho em Equipe, visto pela figura 11.

Figura 11 – Ranking do ganho de informação dos critérios.

	#	Info. gain
<b>N</b> Commitment 3		0.333
<b>N</b> Motivation 4		0.296
<b>N</b> Team work 4		0.289

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, apesar de terem características predominantemente semelhantes aos reprovados, possuem maior probabilidade de serem aprovados na seleção.

Como a métrica Comprometimento é a 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, conseguiram aprovação.

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 Assiduidade nessa pergunta final.

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

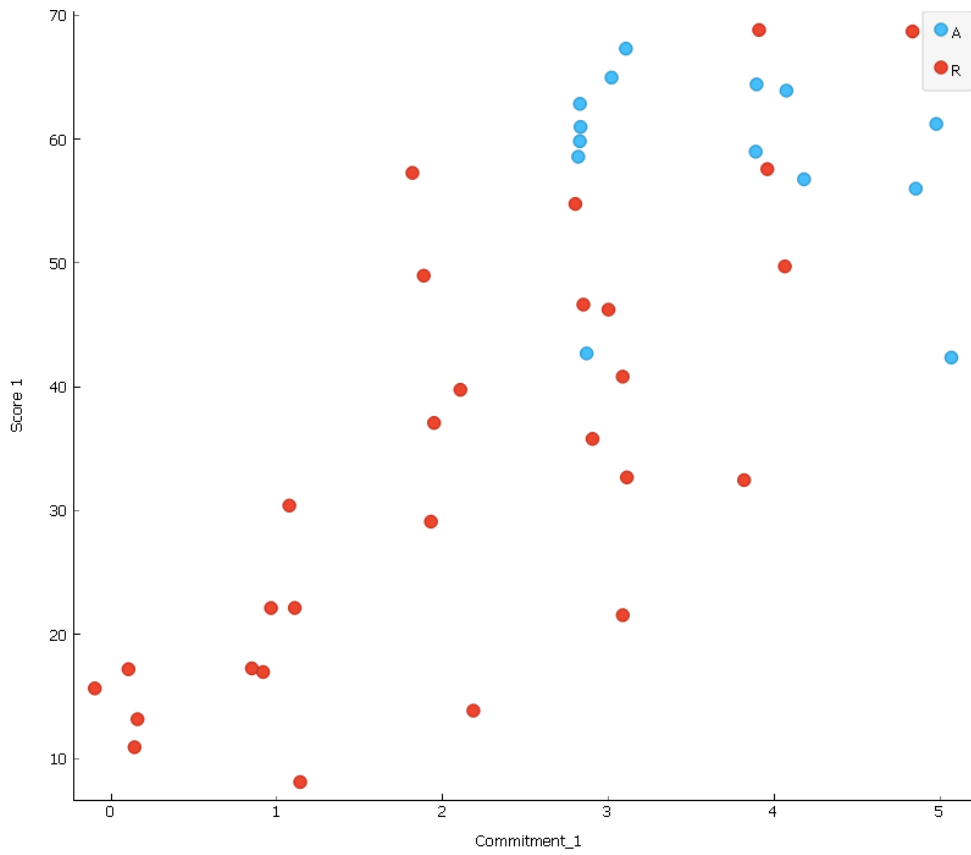
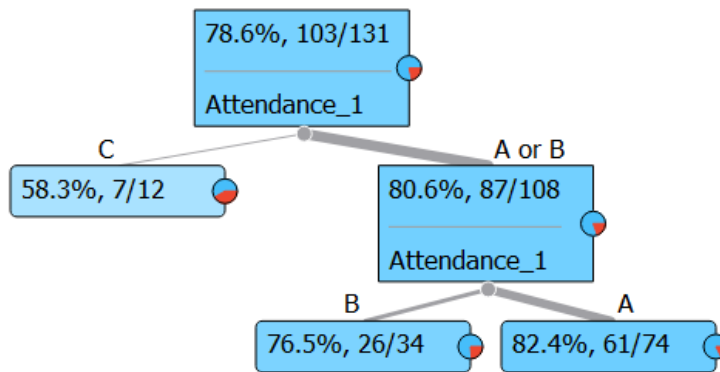


Figura 13 – Árvore de decisão da porcentagem de aprovação pelo nível de Assiduidade dos candidatos.



## 4 CONCLUSÕES E CONTRIBUIÇÕES

A ideia da utilização de mineração de dados para o processo de seleção é deixá-lo mais objetivo utilizando os fatos como suporte na tomada de decisão e não as impressões. A experiência dos membros da organização coloca a métrica Assiduidade como isolada e última métrica a ser analisada. Os resultados analíticos provam que essa métrica traz bastante assertividade ao processo, pois tem uma relação proporcional com a taxa de aprovação, mesmo sem utilizar essa métrica na geração da pontuação do processo.

Os resultados da mineração de dados indicam que a métrica Comprometimento tem maior influência na Aprovação e Reprovação dos candidatos. Pela sua importância, percebe-se que é possível colocar um peso maior para essa métrica na avaliação. Essa proposta faz com que os avaliadores sejam mais criteriosos ao julgar essa métrica e com isso podem aumentar sua representação na geração da pontuação final.

O conhecimento dos principais critérios que pesam 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 Assiduidade 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 dá as indicações do resultado, e também torna mais assertivo pois é elaborado baseado no histórico quantitativo de classificação dos candidatos. Assim sendo, essa pesquisa traz como contribuição para outros processos de seleção de pessoas que o fato de encontrar os critérios que mais pesam em uma seleção é uma forma mais ótima de tomar decisões quando os candidatos são incertos e de difícil processo de tomada de decisão.

Para trabalhos futuros, almeja-se elaborar um esquema de pesos nas variáveis de aprovação em conjunto com os critérios principais, pois com isso o processo de seleção fica mais assertivo e são descobertos como cada parâmetro influencia na decisão. Também pode ser realizado como trabalho futuro a partir de novas análises em mais processos de seleção, um refinamento do modelo que pode melhorar ainda mais a tomada de decisão.

## 5 Agradecimentos

Este trabalho foi realizado durante uma bolsa de estudos financiada pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). Agradecimento à ONG Instituto CEO do Futuro (<http://ceodofuturo.org.br/>) por proporcionar os dados e a interação necessária para realização da pesquisa. Agradecimento ao apoio do CEST (Centro de Estudos Sociedade e Tecnologia).

## Referências<sup>1</sup>

BAYLÃO, A. L. da S.; ROCHA, A. P. de S. A importância do processo de recrutamento e seleção de pessoal na organização empresarial. 2014. Citado 2 vezes nas páginas 2 e 3.

CAXITO, F. de A. *Recrutamento e Seleção de Pessoas*. 1. ed. [S.l.: s.n.], 2012. Citado 2 vezes nas páginas 2 e 3.

CÔRTEZ, S. da C.; PORCARO, R. M.; LIFSCHITZ, S. Mineração de dados: Funcionalidades, técnicas e abordagens. 2002. Citado 2 vezes nas páginas 7 e 8.

STROHMEIER, S.; PIAZZA, F. Domain driven data mining in human resource management: A review of current research. 2012. Citado na página 3.

---

<sup>1</sup> De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.