

HR Analytics e Mineração de Dados para Seleção de Pessoas

Applications Track

I. A. N. Sousa¹, E. Spina²

Universidade de São Paulo, Brazil
italoengenharia@usp.br

Abstract. A área de Analytics tem crescido de maneira expressiva tanto em escopo como em variedade de aplicações. Organizações têm investido de maneira crescente nos últimos anos na área da gestão de Recursos Humanos. Dentre as principais aplicações de Analytics, destaca-se a área de HR Analytics, uma abordagem que se baseia em dados para tomar melhores decisões sobre pessoas em uma organização. A mineração de dados destaca-se como um dos principais meios de obtenção de resultados para viabilizar essa prática. Estudos que envolvem HR Analytics têm crescido nos últimos anos, mas poucas dessas investigações realizam trabalhos quantitativos ou experimentais. Nesta pesquisa, aplica-se uma metodologia de mineração de dados para otimização de um processo de seleção de pessoas em uma Organização da Sociedade Civil. Os resultados indicam que as métricas de avaliação desse processo podem ser melhoradas e que o uso dessa metodologia de mineração de dados pode contribuir com a otimização do processo de seleção.

Categories and Subject Descriptors: H.2.8 [Data Mining]: Data Mining Applications

Keywords: Gestão de Recursos Humanos, HR Analytics, Mineração de dados, Seleção de pessoas

1. INTRODUÇÃO

Organizações têm investido de maneira crescente nos últimos anos na área da gestão de RH (Recursos Humanos). Esta área é reconhecida cada vez mais como um recurso estratégico na busca por vantagem competitiva [da Silva Baylão and de Sousa Rocha 2014] [Dulebohn and Johnson 2013].

Com o avanço da computação, pesquisadores têm desenvolvido sistemas de suporte à decisão, sistemas especialistas e usado análise de dados para melhorar os resultados da gestão de RH. A área de análise de dados, ou em inglês *Analytics*, têm crescido de maneira expressiva tanto em escopo como em variedade de aplicações [Shmueli et al. 2017] [Chien and Che 2006].

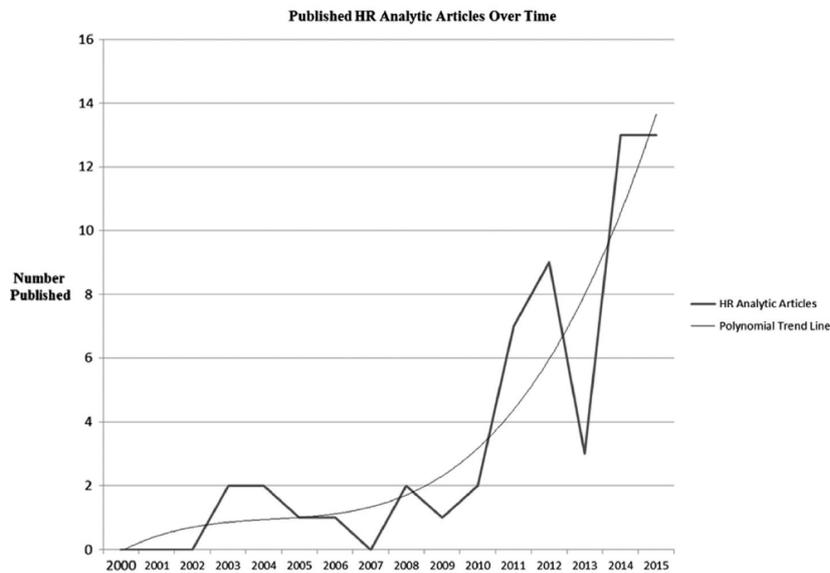
Dentre as principais aplicações de *Analytics*, destaca-se a área de *HR (Human Resources) Analytics*. *HR Analytics* é uma abordagem que se baseia em evidência para tomar melhores decisões sobre pessoas em uma organização. Esse conceito consiste em ferramentas e tecnologias que varia desde relatórios simples de métricas até a elaboração de modelos preditivos [Marler and Boudreau 2016].

Nesse contexto, a mineração de dados destaca-se como um dos principais meios de obtenção de novos resultados para estudos analíticos [Chien and Che 2006]. Mineração de dados se refere aos métodos utilizados para viabilizar a prática de *Analytics* [Shmueli et al. 2017]. Ela consiste no processo de identificação de padrões originais e válidos baseado no uso de dados [Strohmeier and Piazza 2012].

Estudos que envolvem essas abordagens de *HR Analytics* e suas aplicações têm crescido de maneira expressiva nos últimos anos, como pode ser visto na figura 1 [Marler and Boudreau 2016].

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/>) pela disponibilização dos dados da pesquisa. Agradecimento ao apoio do CEST (Centro de Estudos Sociedade e Tecnologia). Copyright©2018 Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

Fig. 1. Publicações em HR Analytics.



Porém, apesar do interesse e atuação na área de *HR Analytics* estar em crescimento, poucas dessas investigações realizam trabalhos quantitativos ou experimentais [Marler and Boudreau 2016].

Baseado nisso, esta pesquisa tem como intuito contribuir para a área de *HR Analytics* com uma abordagem analítica por meio de um estudo exploratório de análise de dados em uma das principais frentes da gestão de RH, a seleção de pessoas. Nesse contexto, aplica-se uma metodologia de mineração de dados com o objetivo de otimizar o processo de seleção de pessoas em um caso de estudo investigado.

2. HR ANALYTICS

2.1 Seleção de pessoas

Seleção de pessoas possui um papel fundamental na gestão de RH, pois determina a qualidade de entrada do pessoal em uma organização. Ou seja, essa área é responsável por agregar competências e talentos necessários para colaborar na busca por vantagem competitiva diante de um contexto altamente complexo [da Silva Baylão and de Sousa Rocha 2014] [de Andrade Caxito 2012].

O processo de seleção consiste na tomada de decisão do candidato a ser escolhido para atuar em determinada função da organização. Ele tem função realizar uma comparação do padrão de características dos candidatos com um referencial estabelecido para o processo de seleção. Esse rigor de escolha ocorre com o intuito de manter ou aumentar o desempenho do pessoal, para com isso colaborar com a eficácia da organização [da Silva Baylão and de Sousa Rocha 2014].

2.2 Mineração de dados

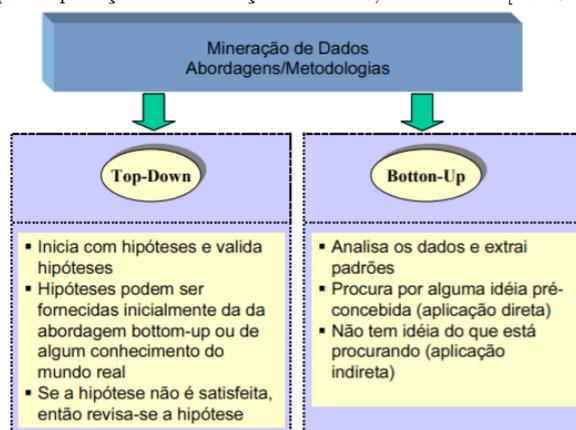
A mineração de dados consiste em uma série de técnicas que possuem diferentes abordagens para solucionar problemas. De maneira geral, são métodos avançados de análise de dados utilizados para trabalhar com variáveis em busca de padrões razoáveis [Han et al. 2012][Shmueli et al. 2016].

Os principais métodos utilizados para mineração de dados são a classificação, predição, regras de associação, análise preditiva, redução de dados, visualização e exploração de dados, aprendizado supervisionado e não supervisionado [Shmueli et al. 2016] [Shmueli et al. 2017].

2.3 Abordagens da mineração de dados

Diante da variedade de tarefas de mineração de dados, cada aplicação busca utilizar uma tarefa específica ou uma combinação de tarefas que possa melhor se adequar para resolução de sua questão levantada na definição da proposta de mineração de dados. As abordagens de mineração de dados descrevem como o processo é conduzido. Ele pode ser conduzido por meio das abordagens *top-down* e *bottom-up*, como pode ser visto na figura 2 extraída de [da Costa Côrtes et al. 2002].

Fig. 2. Abordagens para aplicação da mineração de dados, extraída de [da Costa Côrtes et al. 2002].



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 [da Costa Côrtes et al. 2002].

3. APLICAÇÃO EM UM CASO REAL

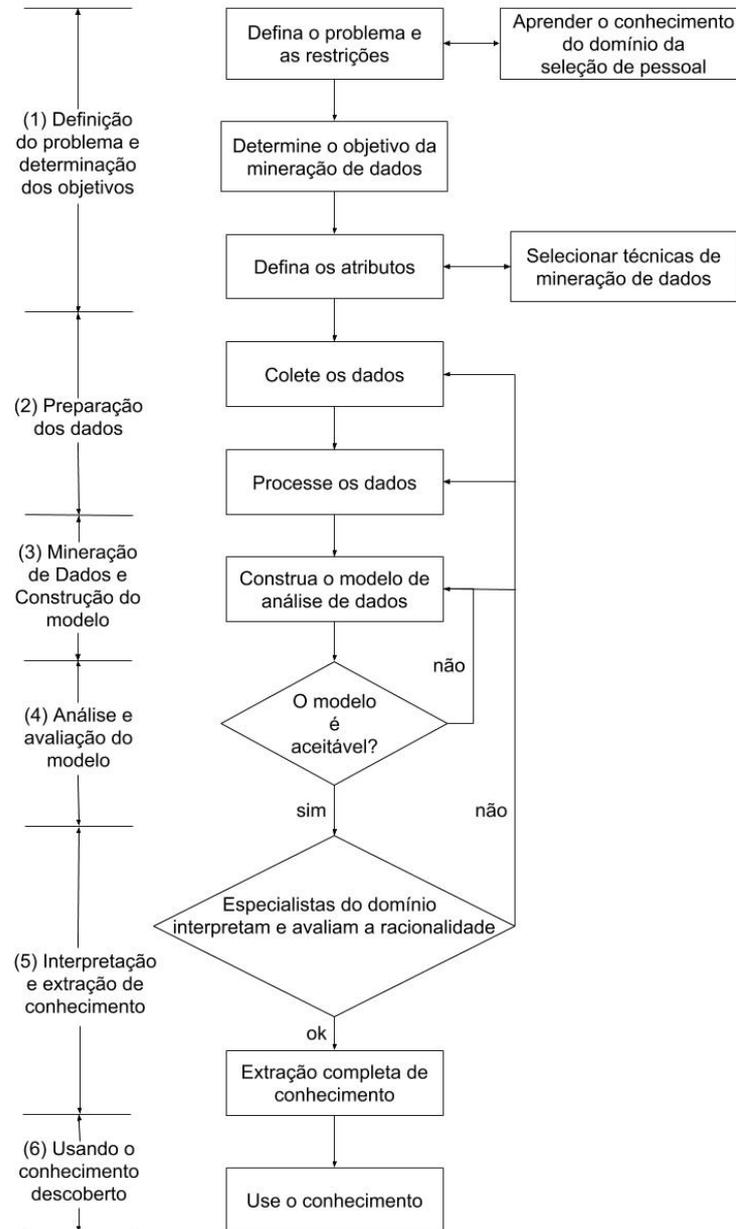
Esta pesquisa utiliza uma estrutura de mineração de dados para recursos humanos baseada no modelo apresentado pela figura 3, que é voltado para processos de seleção de pessoas [Chien and Che 2006]. Seu uso permite a extração de padrões sobre o conjunto de dados dos candidatos de um processo para um melhor entendimento dos dados e suporte na tomada de decisão da seleção de pessoas.

3.1 Definição do problema e determinação dos objetivos

Este experimento é baseado em dados reais de uma OSC (Organização da Sociedade Civil) que atua com ensino e assistência social. Segundo os membros da organização, a cada processo há mais inscritos e fica mais difícil decidir na seleção apenas com as pontuações coletadas na dinâmica e entrevista. Como os recursos são limitados, eles só podem trabalhar com um número máximo de 55 estudantes para suas ações sociais. Ou seja, a seleção dos candidatos a serem escolhidos no processo deve ser bem elaborada. Se essa elaboração não ocorrer de uma forma bem articulada, a tendência é tanto de aumentar o número de evasão do programa, como também de uma má realização de suas atividades.

A partir disso, tem-se como objetivo do negócio, tornar a seleção mais rápida e eficiente. Ou seja, determinar as principais métricas 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.

Fig. 3. Estrutura de mineração de dados para seleção de pessoas, baseado em [Chien and Che 2006].



3.2 Estrutura de mineração de dados para seleção de pessoas

Os atributos utilizados na análise se baseiam nas métricas adotadas no processo de seleção para avaliar os candidatos. Esses atributos correspondem com os resultados das etapas da dinâmica em grupo e da entrevista individual. Eles são obtidos por meio da pontuação gerada nessas etapas do processo de seleção. A partir disso, são obtidos 252 dados por meio da coleta advinda dos formulários preenchidos pelos avaliadores no decorrer das etapas, e então ao fim do processo são transferidos para planilhas.

Na etapa da dinâmica em grupo, cada candidato recebe uma nota, de zero a cinco, e conta com as métricas de avaliação: Iniciativa (*Initiative*), Organização (*Organization*), Trabalho em equipe (*Team*

work), Motivação (*Motivation*), Comprometimento (*Commitment*) e Comunicação (*Communication*). Nessa etapa, ele pode ser A (Aprovado) ou R (Reprovado) pelo avaliador. Na etapa da entrevista individual são praticamente as mesmas métricas, com exceção da Iniciativa. E então, pode ser A (Aprovado), R (Reprovado), ou sujeito a V (Verificação) para deixar a decisão do status de aprovação do candidato para o fim do processo.

Na entrevista, há uma métrica singular da Assiduidade (*Attendance*), porém ela é isolada e não contribui na geração da pontuação para seleção dos candidatos, apenas contribui como peso de decisão nos desempates. Essa métrica é gerada por meio de uma pergunta simples e direta sobre a participação de cada candidato no programa. Ela é categorizada pelo avaliador, com as opções de resposta:

- A: candidatos que têm certeza de que irão para o final e absolutamente não faltarão;
- B: candidatos que pensam que vão até o fim, mas talvez venham a faltar algum dia;
- C: candidatos que não têm muita certeza se irão até o final.

A tabela do processo seletivo utilizada para a pesquisa não estava em uma formatação adequada. Portanto, realiza-se uma estruturação desses dados para transformação em dados prontos para análise. Nisso, são aplicadas técnicas de limpeza para dados faltantes, como a estimação pela média de valores dos dados que devem estar presente em cada local baseado nos dados anteriores e posteriores. Também são excluídas as linhas de dados quando quase todos os dados da linha se encontram faltantes.

A construção do modelo da investigação é conduzido segundo uma abordagem *bottom-up*. Ou seja, busca-se analisar os dados sobre a seleção, explorar seu comportamento e extrair padrões que possam melhorar o processo. A fase de análise e avaliação do modelo é a responsável pela revisão e avaliação do modelo elaborado para confirmação da sua validade para o momento de extração de conhecimento. Nesta investigação, são realizadas análises à respeito de dois objetivos para obtenção de conhecimento: busca por padrões não óbvios e a confirmação de hipóteses existentes. A fase da interpretação e extração de conhecimento consiste na extração de conhecimento dos modelos analisados. Os resultados da mineração de dados são avaliados e interpretados para viabilizar a confirmação do conhecimento extraído. Por fim, finaliza-se a aplicação da estrutura de mineração com o uso do conhecimento adquirido. Esse conhecimento pode ser a base de suporte para tomadas de decisão da organização e ser usado para gerar estratégias para o processo de seleção.

4. MINERAÇÃO DE DADOS E CONSTRUÇÃO DO MODELO

Na construção do modelo, busca-se primeiramente 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 das figuras 4 e 5, o status de aprovação tende a ser maior em valores de pontuação mais altos, mas apesar disso, alguns candidatos com pontuações mais baixas ainda são aprovados. Nota-se 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.

Para explorar o porquê alguns candidatos com notas mais baixas são aprovados, utiliza-se a técnica do Agrupamento. Por meio dessa técnica, candidatos com pontuações de métricas semelhantes são agrupados e grupos com pontuações diferentes são divididos. A divisão desses agrupamentos das pontuações das características dos candidatos resulta nos grupos representados pela figura 6. Os grupos não-destacados (em branco) são compostos em sua totalidade pelos candidatos aprovados. O grupo destacado em azul representa os candidatos reprovados em sua totalidade. Já o grupo destacado em vermelho representa os candidatos amplamente reprovados, mas com algumas aprovações em sua composição. A partir disso, verifica-se que o grupo vermelho representa candidatos que possuem características semelhantes de reprovação, mas que apesar disso, tem alguns candidatos bem sucedidos. Ou seja, o grupo vermelho possui no geral candidatos com pontuações semelhantes de reprovação nas métricas contabilizadas mas que por algum motivo são aprovados pelos avaliadores.

Fig. 4. Distribuição da pontuação da Dinâmica em grupo em relação ao status.

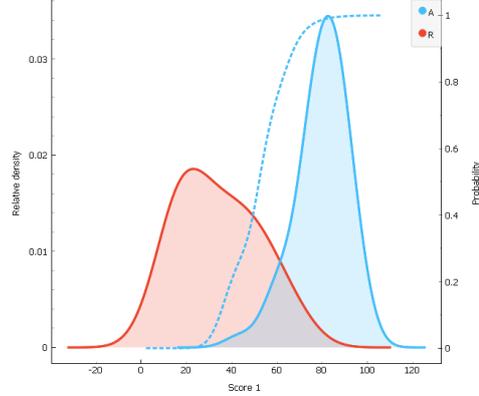


Fig. 5. Distribuição da pontuação da Entrevista Individual em relação ao status.

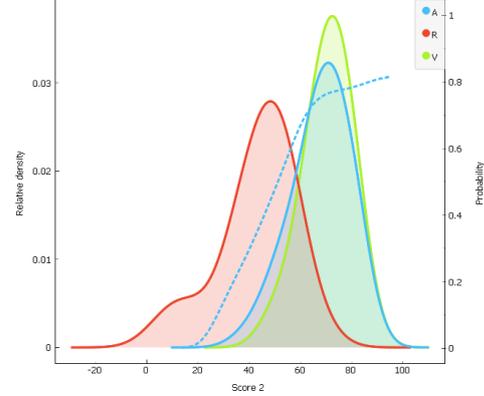
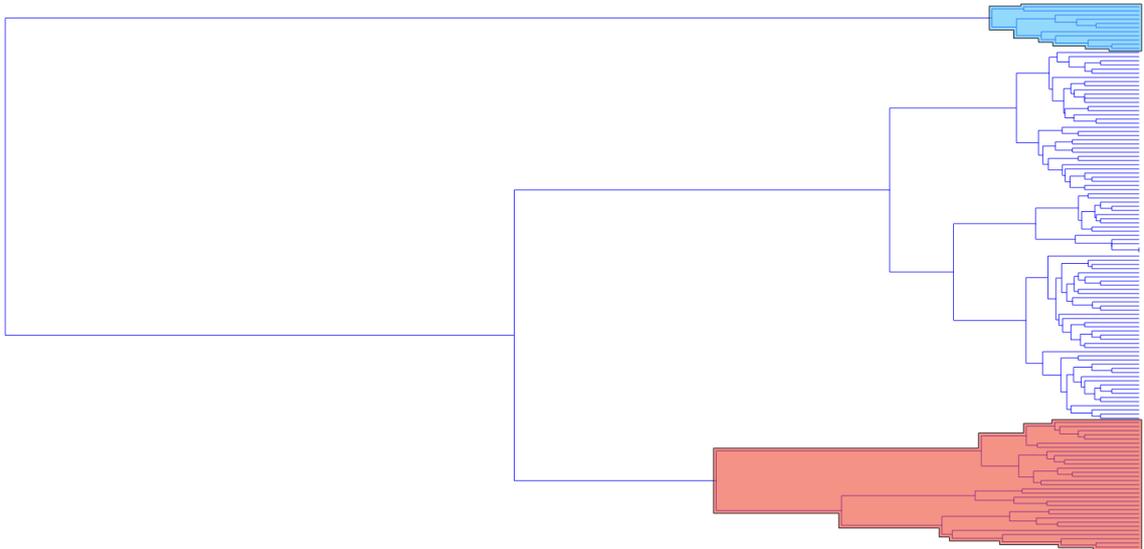


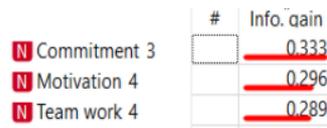
Fig. 6. Agrupamento pelas pontuações das métricas.



Para explorar o porquê dessas aprovações, busca-se saber no agrupamento vermelho o que difere os candidatos aprovados dos demais. Para isso, realiza-se um ranking das métricas que trazem o maior ganho de informação (*information gain*) para a análise de aprovação, sendo que a medida usada é a quantidade de ganho de informação. Os resultados obtidos trazem a resposta de quais métricas possuem o maior peso na avaliação. Na dinâmica em grupo, o maior 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, expresso pela figura 7.

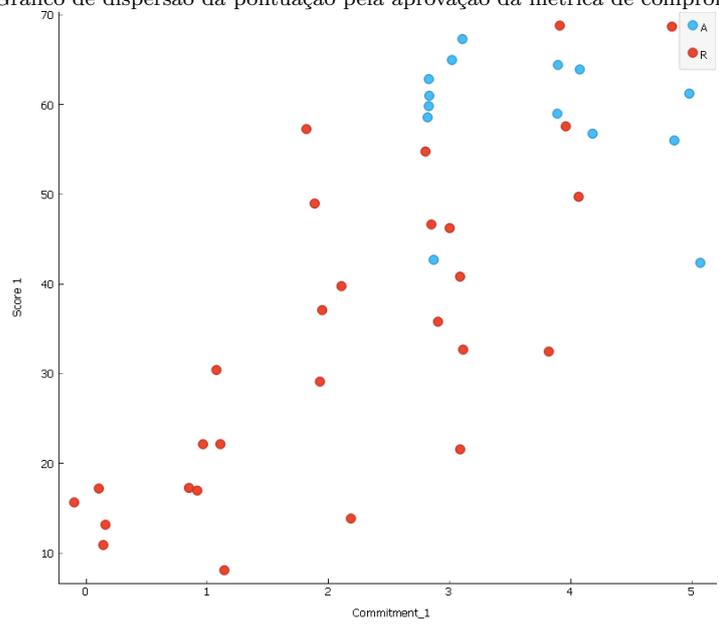
Na análise do quanto esses candidatos bem sucedidos diferem dos demais, o gráfico de dispersão é utilizado para verificar qual é a relação entre esses candidatos e as pontuações de cada uma das métricas avaliadas. Ao verificar cada item, nota-se que aqueles que oferecem maiores valores de Comprometimento 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

Fig. 7. Ranqueamento do ganho de informação das métricas.



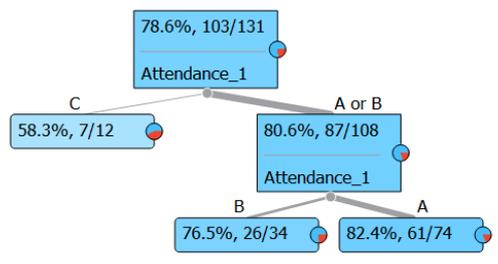
vermelho, o gráfico da dispersão dos candidatos aprovados em relação às suas pontuações, apresentado pela figura 8. Este gráfico indica que candidatos com maiores pontuações de Comprometimento, mesmo tendo várias semelhanças com outros reprovados, conseguiram aprovação.

Fig. 8. Gráfico de dispersão da pontuação pela aprovação da métrica de comprometimento.



Por fim, utiliza-se uma árvore de decisão para classificar os candidatos que respondem mais positivamente à pergunta final sobre o desejo de comparecer de forma assídua ao programa de formação. 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, visto pela figura 9.

Fig. 9. Árvore de decisão da porcentagem de aprovação pelo nível de Assiduidade dos candidato.



5. CONCLUSÕES

Percebe-se na revisão teórica que apesar de haver um crescimento de pesquisas na área de *HR Analytics*, poucos dos estudos são empíricos. Nesta pesquisa, busca-se contribuir com uma investigação quantitativa da mineração de dados aplicada ao contexto de *HR Analytics*. Para isso, busca-se investigar uma das áreas mais prioritárias da gestão de RH, que é o processo de seleção de pessoas, pois ele quem define as qualidades e competências do pessoal que vai fazer parte da organização.

A experiência dos membros da organização coloca a métrica Assiduidade como isolada e última métrica a ser analisada de maneira categórica. Os resultados analíticos provam que essa métrica traz bastante assertividade ao processo, pois corresponde com 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 proporcionam a construção de um modelo que utiliza métricas como Comprometimento, Trabalho em equipe e Assiduidade como métricas de influência na Aprovação e Reprovação dos candidatos. Por sua importância, percebe-se que é possível colocar um peso maior para essas métricas na avaliação. Essa proposta faz com que os avaliadores sejam mais criteriosos ao julgar tais métricas e com isso podem aumentar sua representatividade na geração da pontuação final.

O conhecimento das principais métricas de seleção é incorporado ao processo dessa organização como fator de suporte na seleção dos candidatos. Desta forma, tem-se destacado quais métricas colaboram de maneira mais efetiva como apoio na tomada de decisão, já que foram obtidas baseado no histórico quantitativo de classificação dos candidatos. Assim sendo, essa pesquisa traz como contribuição para um processo de seleção, não necessariamente só de pessoas, que buscar encontrar as métricas que mais pesam em uma seleção é encontrar uma forma mais ótima de tomar decisões.

Para finalizar, por meio da aplicação da metodologia de mineração de dados no processo de seleção investigado, percebe-se que o uso de técnicas analíticas colabora com a otimização do processo por meio da descoberta de padrões não óbvios. Essa consolidação de conhecimento também pode ser usada para melhorar outras atividades de gestão em organizações. Para trabalhos futuros, podem ser realizadas novas análises em processos de seleção, não apenas de seleção de pessoas para entrada na organização, mas também de várias outras áreas da atuação do *HR Analytics*, como seleção para promoção ou demissão. Pode ser feito também como trabalho futuro o estudo em outros processos de seleção para refinamento do modelo de forma a melhorar ainda mais a tomada de decisão.

REFERENCES

- CHIEN, C.-F. AND CHE, L.-F. Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry, 2006.
- DA COSTA CÔRTEZ, S., PORCARO, R. M., AND LIFSCHITZ, S. Mineração de dados: Funcionalidades, técnicas e abordagens, 2002.
- DA SILVA BAYLÃO, A. L. AND DE SOUSA ROCHA, A. P. A importância do processo de recrutamento e seleção de pessoal na organização empresarial, 2014.
- DE ANDRADE CAXITO, F. *Recrutamento e Seleção de Pessoas*, 2012.
- DULEBOHN, J. H. AND JOHNSON, R. D. Human resource metrics and decision support: A classification framework, 2013.
- HAN, J., KAMBER, M., AND PEI, J. *Data mining: concepts and techniques*, 2012.
- MARLER, J. H. AND BOUDREAU, J. W. An evidence-based review of hr analytics, 2016.
- SHMUELI, G., BRUCE, P., YAHAV, I., PATEL, N., AND LICHTENDAHL, K. Data mining for business analytics: Concepts, techniques, and applications in r, 2017.
- SHMUELI, G., BRUCE, P. C., YAHAV, I., PATEL, N. R., AND KENNETH C. LICHTENDAHL, J. *Data Mining for Business Analytics: Concepts, Techniques, and Applications*, 2016.
- STROHMEIER, S. AND PIAZZA, F. Domain driven data mining in human resource management: A review of current research, 2012.