



Congresso Internacional
de Administração
ADM 2022

**24 a 28
de outubro**

**SOBREVIVÊNCIA DAS ORGANIZAÇÕES
EM TEMPOS INCERTOS:**

O papel dos gestores e do ambiente externo
no **sucesso** e no **fracasso** organizacional.

PEOPLE ANALYTICS COM APLICAÇÃO DE ANÁLISE PREDITIVA: ESTUDO DE CASO EM RESCISÕES DE CONTRATOS DE ESTÁGIO

PEOPLE ANALYTICS WITH PREDICTIVE ANALYSIS APPLICATION: CASE STUDY ON TERMINATION OF INTERNSHIP CONTRACTS

ÁREA TEMÁTICA: GESTÃO DE PESSOAS

Vinicius Paulo Raimundo, Universidade Federal do Paraná (UFPR), Brasil, viniciusraimundo@hotmail.com

Hellen Cristina Spengler, Universidade Federal do Paraná (UFPR), Brasil, hspengler.mat@gmail.com

Thiago André Guimarães, Universidade Federal do Paraná (UFPR) / Instituto Federal do Paraná (IFPR) - Campus Curitiba, Brasil, thiagoandre@ufpr.br

Simone Cristina Ramos, Universidade Federal do Paraná (UFPR), Brasil, simone.crisr@gmail.com

Resumo

Um desafio atual da área de Gestão de Pessoas é o desenvolvimento de métricas estruturadas sobre a gestão do capital humano em organizações e o tema *People Analytics* coloca-se como uma possibilidade de desenvolvimento de valor, porém carece de estudos empíricos que demonstrem como tal valor pode ser gerado. Frente a este cenário, o objetivo dessa pesquisa é a aplicação de *People Analytics* em um caso prático de gestão de contratos de estágio, realizando a análise preditiva por meio da técnica de *Random Forest* para identificar os fatores determinantes às rescisões antecipadas de contratos de estágio. Para tal, foi conduzido um estudo de caso em um Agente de Integração de Estágio do estado do Paraná, com análise de 24976 rescisões de contratos ocorridas em 2021. Os resultados apontam que os índices gerais de acurácia do modelo desenvolvido foram de 76% de acordo com as variáveis consideradas. A conclusão é que os resultados podem auxiliar as partes envolvidas no processo de estágio (empresas, estudantes, instituições de ensino e agentes de integração) de forma a se dirimir os impactos decorrentes da rescisão antecipada por meio de ações proativas. Além disso, este trabalho traz contribuições teóricas e empíricas. Do ponto de vista gerencial, o entendimento da matéria pode subsidiar melhores processos de recrutamento, seleção e retenção de estagiários. Pelo ângulo teórico, a aplicação de *People Analytics* ao processo estudado amplia o entendimento do valor gerado pela metodologia.

Palavras-chave: *people analytics; random forest; machine learning; estágios; gestão de pessoas*

Abstract

A current challenge in the People Management area is the development of structured metrics on the management of human capital in organizations and the People Analytics theme stands as a possibility for value development, but it lacks empirical studies that demonstrate how such value can be generated. Faced with this scenario, the objective of this research is to apply People Analytics in a practical case of management of internship contracts, performing predictive analysis using the Random Forest technique to identify the determining factors for early termination of internship contracts. To this end, a case study was carried out in an internship integration agent in the state of Paraná, with an analysis of 24976 contract terminations that occurred in 2021. The results indicate that the general accuracy rates of the model developed were 76% according to with the variables considered. The conclusion is that the results can help the parties involved in the internship process (companies, students, educational institutions and integration agents) in order to resolve the impacts resulting from early termination

through proactive actions. In addition, this work brings theoretical and empirical contributions. From a managerial point of view, understanding the subject can support better recruitment, selection and retention processes for interns. From the theoretical angle, the application of People Analytics to the studied process broadens the understanding of the value generated by the methodology.

Keywords: *people analytics; random forest; machine learning; stages; people management*

1. INTRODUÇÃO

People Analytics vem ampliando o interesse de pesquisadores e de organizações no âmbito de Recursos Humanos (RH) nos últimos anos (Tursunbayeva *et al*, 2018; Huselid, 2018). Em *People Analytics* a análise de dados por métodos científicos é usada para auxiliar nas decisões de gerenciamento, seja em aquisição de talentos, retenção, colocação, promoção, remuneração, e planejamento de sucessão. A análise de pessoas e o *big data* passam a ser, então, recursos valiosos em RH. A promessa da *People Analytics* é que não apenas fará o RH avançar em termos de análise e *insights*, mas também criará a capacidade de previsão e tecnologia sugestiva (Walford-Wright & Scott-Jackson, 2018).

Conforme Angrave *et al* (2016), a área de RH apresenta certa defasagem em relação a outras áreas funcionais de gerenciamento na adoção de tecnologias analíticas e na análise de *big data*. Ainda segundo os autores, apesar de relatos otimistas de profissionais da área, poucas são as evidências que a *Human Resources Analytics (HR Analytics)* está se transformando em uma capacidade obrigatória que poderá garantir no futuro o RH como uma função estratégica. Minbaeva (2018) também afirma que apesar do enorme interesse no *Human Capital Analytics*, as organizações têm demonstrado esforço para efetivamente utilizar análises estatísticas de dados no RH, superando a fase tradicional de relatórios operacionais.

De acordo com Hamilton e Sodeman (2020), o emprego da análise de dados em RH é direcionado com maior ênfase à triagem de candidatos para contratação. Todavia, como perspectiva geral da empresa, a análise dos dados relacionados ao RH deve ter como objetivo capturar o alinhamento estratégico entre capital humano e lucratividade, gerando vantagem competitiva e melhorando o desempenho da organização.

Marler e Boudreau (2017) ao realizarem uma revisão de literatura sobre *HR Analytics*, destacam que apesar do crescente interesse pelo tema, ainda são poucas e limitadas as evidências científicas sobre a adoção desta prática na tomada de decisão, e concluem sobre a necessidade de mais pesquisas científicas e estudos empíricos. Conclusão semelhante sobre a escassez de estudos empíricos é reportada também por Qamar e Samad (2021).

No melhor do nosso conhecimento não foram encontrados estudos semelhantes na literatura que aplicam técnicas para a predição de rescisão antecipada de contratos de estágio. Com a intenção de contribuir para a redução deste *gap*, este trabalho endereça a aplicação de *People Analytics* em um Agente de Integração de Estágio (AI) no estado do Paraná, realizando a análise preditiva com aplicação da técnica de *Random Forest* para identificar os fatores determinantes às rescisões antecipadas de contratos de estágio, e neste âmbito desenvolver uma metodologia para a predição de rescisões que podem ocorrer da antes da finalização do contrato.

Entre as contribuições diretas deste trabalho, objetiva-se municiar o AI estudado na tomada de decisão, possibilitando atuar de maneira proativa junto aos grupos mais propensos à rescisão antecipada, visando a diminuição desses números ou a preparação operacional e financeira caso a rescisão não possa ser evitada. A estrutura do trabalho é descrita como segue: na segunda Seção o documento apresenta o referencial teórico, que abordará as definições de *People Analytics* e os conceitos teóricos sobre a técnica de classificação e predição utilizada.

A Seção 3, apresenta as aplicações das técnicas e os resultados preliminares do estudo de caso, enquanto a Seção 4 tece as considerações e os direcionamentos futuros do trabalho.

2. REFERENCIAL TEÓRICO

Neste referencial teórico são abordados os aspectos fundamentais que suportam este estudo. Primeiramente, dar-se-á aos conceitos de *People Analytics*. Na sequência, aspectos pertinentes sobre da técnica de *Random Forest* são detalhados.

2.1 People Analytics

People Analytics é um tema emergente que geralmente apresenta-se com terminologias equivalentes, entre eles: *Human Resources Analytics*, *Human Capital Analytics*, *Workforce Analytics*, *Human Resources Metrics* e *Talent Analytics*. (Marler & Boudreau, 2017; Qamar & Samad, 2021; Tursunbayeva *et al*, 2018).

Segundo Larsson e Edwards (2021), *People Analytics* se caracteriza como uma prática organizacional que emprega a análise apoiada em dados, por meio de verificações empíricas, que comensuram e qualificam a performance dos funcionários em processos de gestão de pessoas, conectando-a aos resultados de negócios, propiciando tomadas de decisões respaldadas em evidências.

Marler e Boudreau (2017) descrevem *Human Resource Analytics* como um processo da área de gestão de pessoas que viabiliza, através da tecnologia da informação, o uso de análises estatísticas, descritivas e visuais de dados, aplicadas aos processos da área, aos dados de capital humano, performance da organização e benchmarks externos, a fim de mensurar o efeito comercial e apoiar um processo decisório baseado em dados.

Jac (2010) considera o *Human Resource Analytics* em primeiro lugar como uma ferramenta de comunicação, que reúne dados de diversas fontes, como pesquisas e operações de diferentes unidades ou níveis, para desenhar uma imagem coesa e acionável das condições atuais e prováveis futuros. Ainda segundo o autor, o nível das análises que as empresas devem se importar depende do propósito organizacional, sugerindo cinco etapas de análises: 1. Registrar o trabalho (contratar, pagar, treinar, apoiar e reter), 2. Relacionar aos objetivos da organização (qualidade, inovação, produtividade, serviços), 3. Comparar os resultados com outros (*benchmarking*), 4. Compreender o comportamento e os resultados passados (análise descritiva) e 5. Prever probabilidades futuras (análise prescritiva).

Seguindo esta linha, Fitz-Enz e Mattox II (2014) complementam que *Human Resource Analytics* é uma abordagem baseada em evidências para tomar melhores decisões e dividem em três níveis:

- Descritivo: métricas tradicionais do RH (taxa de rotatividade, tempo de preenchimento de vagas, custo de contratação, funcionários treinados, etc.) com foco em redução de custos e melhoria de processos, que revela e descreve relacionamentos de dados atuais e históricos, disponibilizados por exemplo por meio de *dashboards* e relatórios periódicos.
- Preditivo: abrange uma variedade de técnicas (estatísticas, modelagem, mineração de dados) que usam fatos atuais e históricos para fazer previsões sobre o futuro. Trata-se de probabilidade e impacto potencial, envolvendo por exemplo, modelos utilizados para aumentar a probabilidade de selecionar as pessoas certas para contratar, treinar e promover.

- Prescritivo: a análise prescritiva vai além das previsões e descreve opções de decisão e otimização da força de trabalho. É usado para analisar dados complexos para prever resultados e fornecer opções de decisão e mostrar impactos comerciais alternativos. Envolve por exemplo, modelos usados para entender como os investimentos alternativos em aprendizado impactam o resultado final.

Outra definição foi realizada por West (2020), *People Analytics* consiste aplicar evidência nas decisões de gestão de pessoas e o coloca na interseção de quatro estruturas, conforme Figura 1.



Figura 1. Estrutura do *People Analytics* (Intersecção)
Fonte: West (2020)

Um case de sucesso que deu visibilidade ao termo *People Analytics* foi a divulgação em 2011 do *Project Oxygen*, que por meio da análise dos dados conseguiram identificar oito comportamentos dos bons gerentes, e com essa informação pode incorporar aos programas de treinamentos e desenvolvimento gerencial na organização (Shrivastava *et al.*, 2018).

2.2 *Machine Learning e Random Forest*

Aprendizagem estatística refere-se a um vasto conjunto de ferramentas para entendimento dos dados. Estas ferramentas podem ser classificadas com supervisionadas ou não supervisionadas. Aprendizagem estatística supervisionada envolve construir um modelo estatístico para predição, com a saída baseada em uma ou mais entradas. Com a aprendizagem estatística não supervisionada existem entradas, mas as saídas não são supervisionadas, no entanto é possível aprender os relacionamentos e estrutura dos dados (James *et al.*, 2013).

Considerando a aprendizagem supervisionada uma quantidade limitada de valores de entrada com valores de saídas conhecidas são utilizadas para “aprender” o presumido, mas desconhecido relacionamento funcional entre os valores de entrada e os valores de saída por

um algoritmo, na qual torna-se possível prever o valor de saída para futuros valores de entrada (Steinwart & Christmann, 2008).

Um tipo de aprendizagem supervisionada é a aprendizagem de máquina (*Machine Learning*), que segundo Sullivan (2017), pode ser definida como um processo de entrada de dados nos sistemas de computadores de forma que o computador aprenda a capacidade de processar e executar no futuro sem ser explicitamente programado ou alimentado com dados semelhantes ou extras. Esta forma de aprendizagem pode possuir ampla aplicação e em diferentes áreas, como no processamento de linguagem natural, jogos digitais, diagnósticos médicos, análise de comportamento, detecção de fraudes, entre outras.

Proposta por Breiman (2001), a Floresta Aleatória (*Random Forest*) é um método já consolidado de aprendizagem de máquina. Sullivan (2017) a define como um versátil e inteligente método de *Machine Learning* que pode realizar tarefas tanto de classificação quanto de regressão.

Segundo Liaw e Wiener (2002) é possível determinar a importância de uma variável observando o aumento do erro de previsão quando os dados (*out-of-bag - OOB*) para essa variável são permutados, enquanto todos os outros são deixados inalterados. Os cálculos necessários são realizados árvore por árvore à medida que a floresta aleatória é construída.

Random Forest faz parte dos métodos *ensemble*, que são capazes de combinar diferentes modelos de árvores de decisões. Neste método as árvores são criadas aleatoriamente e treinadas por métodos *bagging* (*Bootstrap Aggregating*), que busca aprimorar as combinações entre as árvores e aumentar a acurácia geral, utilizando a combinação de outras duas técnicas: a agregação e o *bootstrap*. Embora este método seja mais custoso computacionalmente, uma vez que gerará grande número de comparações, apresentará resultados com maior aleatoriedade e facilmente adaptáveis aos problemas propostos.

Já seguindo para a análise dos resultados com as técnicas estatísticas que trabalham com classificação, isto é, o modelo com base nas variáveis irá apontar ou não a incidência de um evento, utiliza-se o conceito do *cutoff*, que é um ponto de corte no valor da probabilidade (p). Na Figura 2, caso o *cutoff* fosse 0.5, qualquer valor de $p > 0,5$ será considerado a incidência do evento, e $p < 0.5$ a não incidência. Definido o *cutoff*, é possível elaborar uma tabela com a classificação do modelo, chamada matriz de confusão, definindo a sensibilidade (1), especificidade (2) e eficiência global ou acurácia (3), conforme mostrado na Tabela 1.

		Observação do Evento	
		Incidência	Não incidência
Previsão do Evento	Incidência	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Não incidência	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Tabela 1. Matriz de Classificação do Modelo
Fonte: Os autores (2022)

$$sensibilidade = \frac{VP}{VP + FN} \quad (1)$$

$$\text{especificidade} = \frac{VN}{FP + VN} \quad (2)$$

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

Com a definição da sensibilidade e especificidade é possível criar a curva de sensibilidade e a curva ROC (*Receiver Operating Characteristic*). A curva de sensibilidade, mostrada na Figura 2 é calculada com diferentes *cutoff*, sendo possível analisar em qual valor de *cutoff* a sensibilidade se iguala a especificidade.

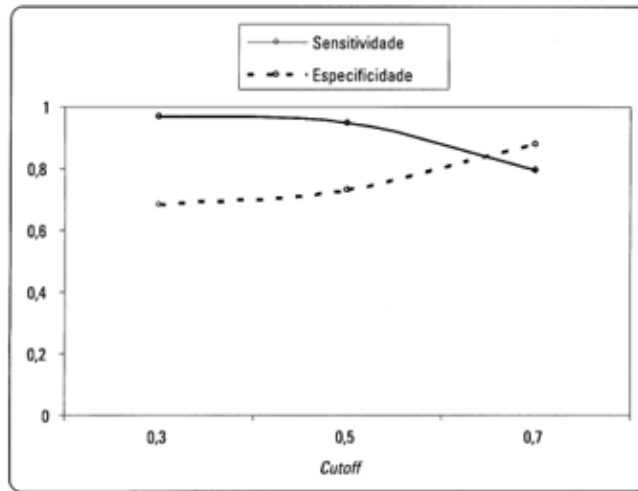


Figura 2. Curva de sensibilidade para três valores de *cutoff*
Fonte: Fávero; Belfiore (2017)

A curva ROC representado na Figura 3, apresenta a sensibilidade e especificidade em forma de formato convexo indicando que o modelo com maior área abaixo da curva (*area under the curve - AUC*) apresenta maior eficiência global de previsão.

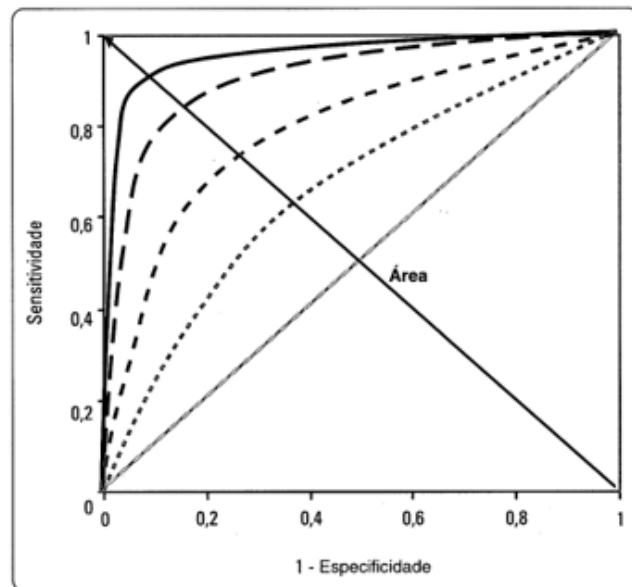


Figura 3. Critério de escolha do modelo com maior área da curva ROC
Fonte: Fávero; Belfiore (2017)

3. DESENVOLVIMENTO

O estudo de caso para aplicação de *People Analytics* foi realizado em uma organização que atua como Agente de Integração de Estágio do estado do Paraná/Brasil, que atualmente realiza a gestão de cerca de 30000 contratos de estágio e 6300 empresas contratantes.

No Brasil, o estágio é regulamentado por meio da Lei de Estágio (BRASIL, 2008) sendo necessário cumprir os requisitos que envolvem a instituição de ensino (IE), o estudante, e a parte concedente de estágio (PCE), cujas atribuições individuais são consolidadas em um termo de compromisso de estágio (TCE), que detalha as responsabilidades das partes envolvidas conforme normas especificadas. O vínculo e a observação de tais normas são válidos até que se ocorra a rescisão do TCE, com vigência máxima de dois anos, exceto quando se tratar de portador de deficiência, que não possui limite de estágio.

Uma característica de destaque no programa de estágio é a pouca burocracia e facilidade na rescisão antecipada do TCE, que pode ser realizada a pedido de qualquer uma das partes, sem necessidade de aviso prévio. Para tal, basta que a PCE pague o período trabalhado e eventuais dias de recesso, além de entregar o termo de realização do estágio para o estudante. Esta característica contribui para a rotatividade dos estagiários nas PCEs, dificultando a predição dessa rescisão.

3.1 Coleta e Preparação dos Dados

Os dados utilizados na pesquisa são compostos de registros das rescisões (antecipadas ou não) ocorridas no período de 01/01/2021 a 31/12/2021. Foram consideradas as seguintes regras para obtenção dos dados:

- contratos de estágio de estudantes que estavam cursando ensino médio ou ensino superior;
- estágios da modalidade não obrigatório;
- excluídas rescisões por motivos compulsórios: uma vez que o término de contrato não depende das variáveis disponibilizadas, como por exemplo a transferência de curso ou instituição de ensino, mudança de cidade ou país, término de convênio entre AI e PCE, efetivação, entre outros.

Também foi realizado o tratamento dos dados, eliminando *outliers* e registros que não tinham todas as informações completas. Todos os dados que identificavam o estudante ou qualquer parte do contrato foram desconsiderados, disponibilizando apenas dados anônimos.

A Tabela 2 descreve as variáveis utilizadas, sendo que a variável resposta (dependente) na aplicação da técnica é a ocorrência da rescisão antecipada do contrato de estágio. Já as variáveis explicativas (independentes) foram selecionadas de acordo com a disponibilidade dos dados disponíveis nos sistemas da empresa e que fazem parte das informações obrigatórias no cadastro do estudante para compor o contrato de estágio.

VARIÁVEL	TIPO DE VARIÁVEL	VALORES POSSÍVEIS
ANTECIPADO	Dependente – qualitativa - que indica se ocorreu a rescisão antecipada.	Sim Não
BOLSA	Independente – contínua - representando o valor da bolsa auxílio.	
IDADE	Independente – contínua - representando a idade do estudante.	
TIPO_IE	Independente – qualitativa - representando o tipo da Instituição de Ensino.	Privada Pública
TIPO_EMPRESA	Independente – qualitativa - representando o tipo da Empresa Contratante.	Privada Pública
NIVEL_CURSO	Independente – qualitativa - representando o nível do curso do estudante.	Médio Superior
SEXO_ESTUDANTE	Independente – qualitativa - representando o sexo do estudante.	Masculino Feminino
TURNO_AULA_ESTUDANTE	Independente – qualitativa - representando o turno da aula do estudante.	Manhã Tarde Noite Integral EAD
PORTE_CIDADE	Independente – qualitativa - representando o porte da cidade do local do estágio.	Pequeno Médio Grande Rural Micro
DIST_RESIDENCIA_EMPRESA	Independente – contínua - representando o valor da distância em metros da residência do estudante até o local de estágio.	
MES_INICIO	Independente – contínua – representando a mês de início do contrato	
DATA_FINAL_ORIGINAL	Independente – contínua – representando a data final prevista do contrato	
DURACAO_CONTRATO_ORIGINAL	Independente – contínua – representando a quantidade de meses de duração prevista do estágio	
PORTE_EMPRESA	Independente – qualitativa - representando o porte da empresa considerando o número de funcionários	Muito Pequena Pequena Média Grande Muito Grande

Tabela 2. Variáveis utilizadas
Fonte: Os autores (2022)

3.2 Aplicação de *Machine Learning* com *Random Forest* e Resultados

No desenvolvimento dessa pesquisa foi utilizada a técnica de *Random Forest*, empregando a linguagem R, por meio do ambiente *RStudio*.

No modelo o valor da variável dependente ANTECIPADO teve como fator positivo a ocorrência da rescisão. Separou-se os dados em dois conjuntos: treino e teste, com 75% dos

dados para treino, representando 18732 registros e 25% para testes, representando 6244, mantendo a mesma proporção de contratos antecipados dos dados originais.

Conforme a Figura 4, as rescisões antecipadas de contrato de estágio realizadas por este Agente de Integração possuem um alto índice, sendo que dos 24976 contratos rescindidos em 2021, 15509 (62%) resultaram em rescisão antecipada e 9467 (38%) contratos chegaram ao final da vigência do contrato.

Na Figura 4 são apresentadas as distribuições da população em relação as rescisões, considerando as variáveis qualitativas de porte da cidade, tipo de IE, tipo de empresa, nível do curso, sexo do estudante, turno da aula e porte da empresa. Ao observar o grupo majoritário das rescisões antecipadas, os estudantes são do sexo feminino e cursando ensino superior no turno da noite, provindo de IEs privadas e tendo contrato de estágio em empresas privadas de porte muito pequeno a médio, sendo localizadas em cidades de porte médio a grande.

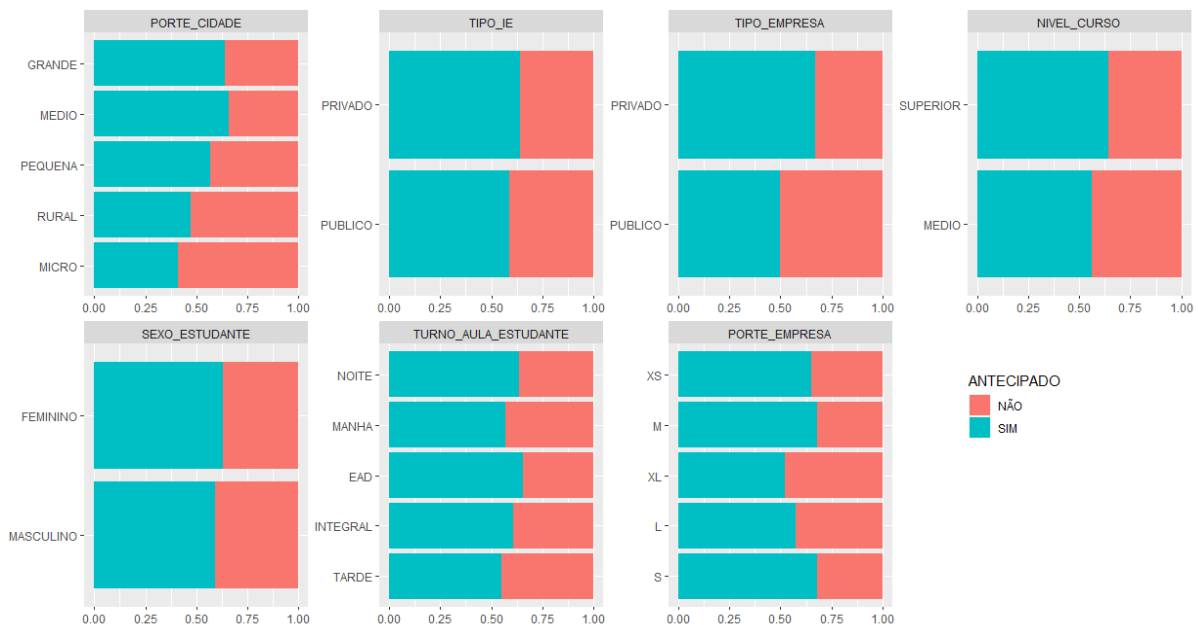


Figura 4- Distribuição da população para variáveis qualitativas
Fonte: Os autores (2022)

Tomando as variáveis contínuas, conforme a Figura 5, tem-se que em média os estudantes possuem idades entre 15 e 25 anos, recebendo bolsas com valores de 400 a 900 reais. Os contratos em grande maioria possuem duração de 1 ou 2 anos, iniciando entre os meses de janeiro a março.

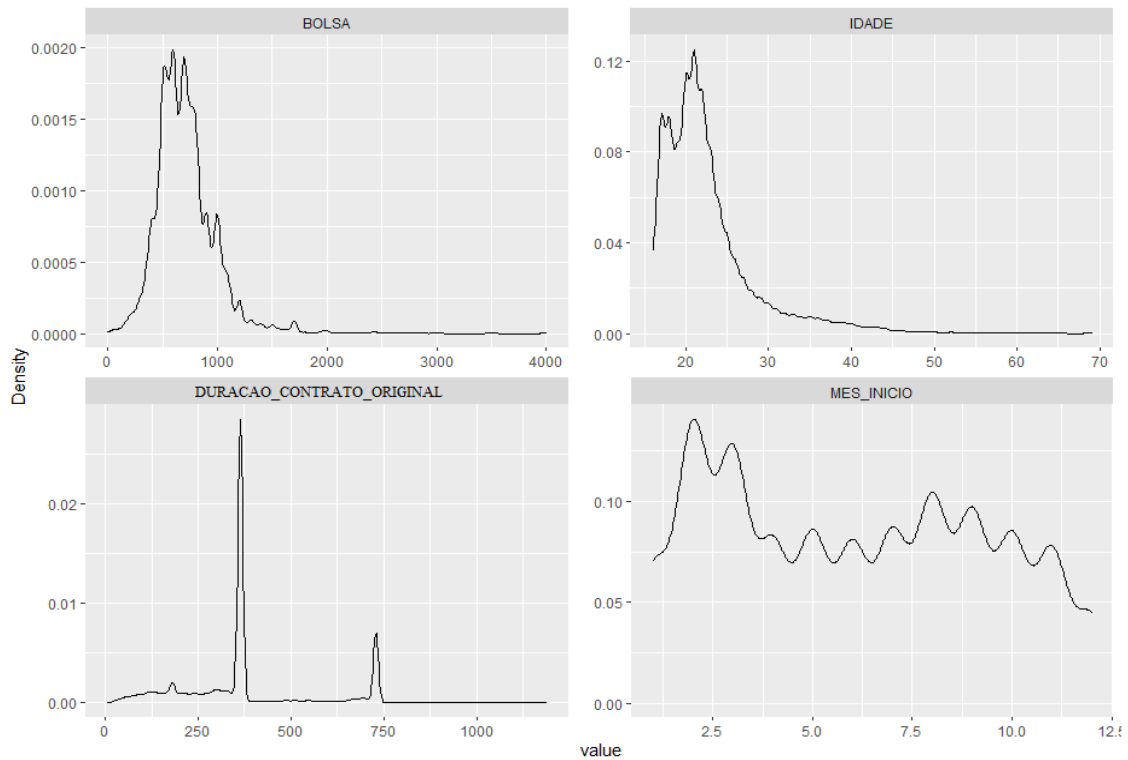


Figura 5- Distribuição da população para variáveis quantitativas
Fonte: Os autores (2022)

Os principais resultados da aplicação do modelo de *Random Forest* estão apresentados na Figura 6, que traz a curva *ROC* e na Tabela 3 apresentando as matrizes de confusões dos resultados referentes aos dados de treino e testes. Observa-se que a acurácia geral obtida na referência de treino foi de 76%, enquanto a sensibilidade foi de 80% e a especificidade de 69%. Assim tendo valores aproximados entre as medidas, embora a sensibilidade tenha se destacado, tendo uma propensão para avaliação de ocorrências positivas do evento.

PREDIÇÃO	REFERÊNCIA TREINO		REFERÊNCIA TESTE	
	ANTECIPADO	FINAL	ANTECIPADO	FINAL
ANTECIPADO	9312	2160	3068	732
FINAL	2365	4895	764	1680
ACURÁCIA	0.7584		0.7604	
SENSIBILIDADE	0.7975		0.8006	
ESPECIFICIDADE	0.6938		0.6965	

Tabela 3- Resultado da Matrix de Confusão dos dados de Treino e Teste (corte 0.5)
Fonte: Os autores (2022)

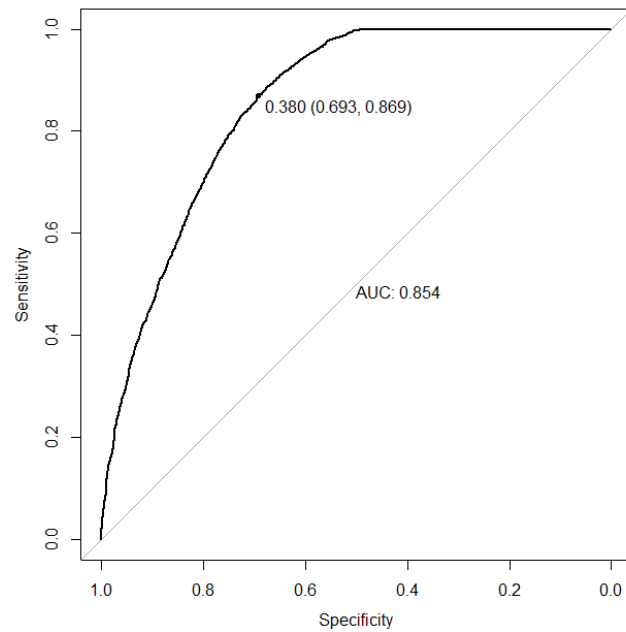


Figura 6- Resultado Curva ROC
Fonte: Os autores (2022)

Na Figura 7 é apresentada o gráfico de influência das variáveis consideradas no modelo. A variável destacadamente com maior influência foi a DATA_FINAL_ORIGINAL. Outras variáveis com influência significativa foram de DuracaoMesesContratoOriginal, IDADE, MÊS_INICIO E TIPO_EMPRESA, PORTE_CIDADE, BOLSA e NÍVEL_CURSO. Enquanto dentre as variáveis menos influêntes estão o SEXO_ESTUDANTE, TIPO_IE, PORTE_EMPRESA, TurnoAulaEstudante e DistEstudenteforEmpresa_GEO.

Assim, observa-se um forte relacionamento com a época do ano em que o estágio é estabelecido, visto que as variáveis que envolvem a época e tempo do contrato apresentaram maior prevalência.

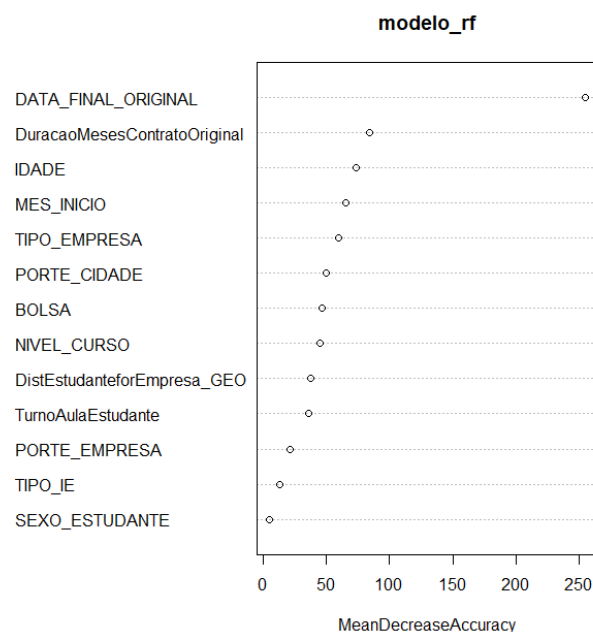


Figura 7- Influência das variáveis do modelo
Fonte: Os autores (2022)

Ainda se determinou os melhores pontos de corte, tanto para igualar os valores de sensibilidade e especificidade, quanto para melhorar a medida *F1-score*. Conforme a Figura 8, obteve-se o corte com valor de 0.5 para a medida de *F1-score* e o corte com valor de 0.44 para a sensibilidade e especificidade.

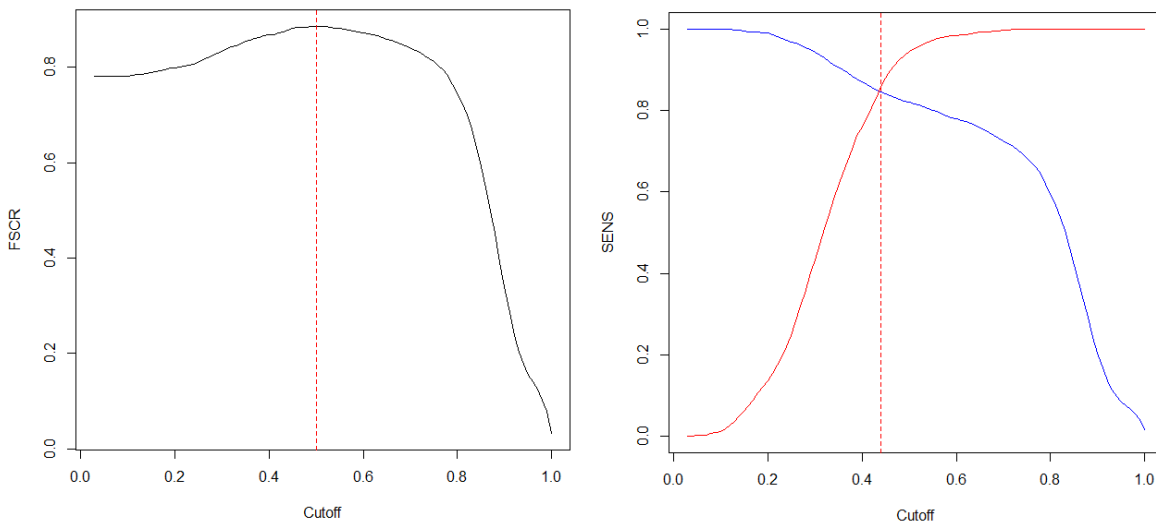


Figura 8- Pontos de corte para F1-score e para sensibilidade versus especificidade
Fonte: Os autores (2022)

Como o ponto de corte foi tradicionalmente estabelecido pelo valor de 0.5 nos testes anteriores, então na seguinte Tabela 4 são apresentados os valores do resultado da matriz de confusão para o corte de 0.44. Note-se que os valores de sensibilidade e especificidade se tornaram mais próximos se comparados com o corte tradicional, sem haver grandes diferenças na acurácia final.

PREDIÇÃO	REFERÊNCIA TREINO		REFERÊNCIA TESTE	
	ANTECIPADO	FINAL	ANTECIPADO	FINAL
ANTECIPADO	9285	2237	3007	741
FINAL	2248	4962	783	1713
ACURÁCIA	0.7605		0.7559	
SENSIBILIDADE	0.8050		0,7934	
ESPECIFICIDADE	0.6892		0.6980	

Tabela 4- Resultado da Matrix de Confusão dos dados de Treino e Teste (corte 0.44)
Fonte: Os autores (2022)

A acurácia total do modelo aplicado aos dados de treinos e testes foram semelhantes, representando a capacidade do modelo de prever satisfatoriamente este cenário de rescisão. Nota-se que utilizando o ponto de corte 0.44 obteve-se valores de sensibilidade e especificidade mais próximos, havendo declínio da acurácia em contrapartida.

CONCLUSÃO

Este trabalho apresentou um estudo de caso de aplicação de *People Analytics* para a predição do processo de rescisões antecipadas de contratos de estágio em um Agente de Integração de Estágio que atua no estado do Paraná. No melhor do nosso conhecimento, não foram

encontrados na literatura estudos endereçados à investigação dos mesmos fatores em contratos de estágio e ainda evidenciou poucos trabalhos empíricos envolvendo *People Analytics*.

Os resultados mostram uma acuraria geral de 76% nos dados de testes, mesmo considerando as características particulares do programa de estágio, como a facilidade de rescisão e variáveis envolvendo as partes: IE, PCE e estudante. Os resultados dos estudos na organização possibilitam que a tomada de decisão seja respaldada com base em evidências, permitindo que todas as partes envolvidas no contrato de estágio possam efetivar ações proativas para evitar a rescisão ou prever quando elas têm maior probabilidade de ocorrer, se preparando para eventuais impactos operacionais e financeiros como a variação do fluxo de rescisões.

Com relação a direcionamentos futuros, acredita-se que o que desempenho da *Random Forest* possa ser aprimorado com a avaliação de outras variáveis de representam características durante o período do estágio, como o meio de transporte utilizado, tempo de deslocamento, desempenho escolar e carga de atividades no estágio. Pela indisponibilidade momentânea de tais informações na base de dados do AI estudado, pesquisas futuras poderão ser conduzidas a partir da ampliação da base de dados. Vislumbra-se ainda a avaliação comparativa de outras técnicas de predição.

REFERÊNCIAS

- Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human resource management journal*, 26(1), 1-11.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.
- Fávero, L. P., & Belfiore, P. (2017). *Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®*. Elsevier Brasil.
- Fitz-Enz, J., & John Mattox, I. I. (2014). *Predictive analytics for human resources*. John Wiley & Sons.
- Jac, F. E. (2010). *The new HR analytics: predicting the economic value of your company's human capital investments*. Amacom.
- Hamilton, R. H., & Sodeman, W. A. (2020). The questions we ask: Opportunities and challenges for using big data analytics to strategically manage human capital resources. *Business Horizons*, 63(1), 85-95.
- Huselid, M. A. (2018). The science and practice of workforce analytics: Introduction to the HRM special issue. *Human Resource Management*, 57(3), 679-684.
- Larsson, A. S., & Edwards, M. R. (2022). Insider econometrics meets people analytics and strategic human resource management. *The International Journal of Human Resource Management*, 33(12), 2373-2419.
- Lei n. 11.788, de 25 de setembro de 2008. Dispõe sobre o estágio de estudantes e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 25 set. 2008. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/11788.htm. Acesso em: 07 fev. 2022.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR Analytics. *The International Journal of Human Resource Management*, 28(1), 3-26.

- Minbaeva, D. B. (2018). Building credible human capital analytics for organizational competitive advantage. *Human Resource Management*, 57(3), 701-713.
- Qamar, Y., & Samad, T. A. (2021). Human resource analytics: a review and bibliometric analysis. *Personnel Review*.
- Rosett, C. M., & Hagerty, A. (2021). *Introducing HR Analytics with Machine Learning*. Springer International Publishing.
- Shrivastava, S., Nagdev, K., & Rajesh, A. (2018). Redefining HR using people analytics: the case of Google. *Human Resource Management International Digest*.
- Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.
- Sullivan, W. (2017). 1: *Machine learning Beginners Guide Algorithms Supervised & Unsupervised learning, Decision Tree & Random Forest Introduction*. Create Space Independent Publishing Platform.
- Tursunbayeva, A., Di Lauro, S. and Pagliari, C. (2018), "People analytics – a scoping review of conceptual boundaries and value propositions", *International Journal of Information Management*, Vol. 43, pp. 224-247.
- Walford-Wright, G., & Scott-Jackson, W. (2018). Talent Rising; people analytics and technology driving talent acquisition strategy. *Strategic HR Review*.
- West, M. (2020). *People Analytics para Leigos*. Alta Books.