



UNIVERSIDADE FEDERAL DO PARANÁ

VINICIUS PAULO RAIMUNDO

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

CURITIBA

2022

VINICIUS PAULO RAIMUNDO

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

Dissertação apresentada ao Programa de Pós-Graduação em Gestão de Organizações, Liderança e Decisão, Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Gestão de Organizações, Liderança e Decisão.

Orientador: Prof. Dr. Thiago André Guimarães

CURITIBA

2022

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIAS SOCIAIS APLICADAS

Raimundo, Vinicius Paulo

People analytics com aplicação de análise preditiva : estudo de caso em rescisões de contratos de estágio / Vinicius Paulo Raimundo. – Curitiba, 2022.

1 recurso on-line : PDF.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Gestão de Organizações, Liderança e Decisão. Orientador: Prof. Dr. Thiago André Guimarães.

1. Gestão de pessoas. 2. Estágios. 3. Predição. 4. Análise de regressão logística. I. Guimarães, Thiago André. II. Universidade Federal do Paraná. Programa de Pós-Graduação em Gestão de Organizações, Liderança e Decisão. III. Título.

Bibliotecária: Maria Lidiane Herculano Graciosa CRB-9/2008



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS SOCIAIS E APLICADAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO DE
ORGANIZAÇÕES, LIDERANÇA E DECISÃO - 40001016172P9

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GESTÃO DE ORGANIZAÇÕES, LIDERANÇA E DECISÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de VINICIUS PAULO RAIMUNDO intitulada: **PEOPLE ANALYTICS COM APLICAÇÃO DE ANÁLISE PREDITIVA: ESTUDO DE CASO EM RESCISÕES DE CONTRATOS DE ESTÁGIO**., sob orientação do Prof. Dr. THIAGO ANDRÉ GUIMARÃES, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 04 de Novembro de 2022.

Assinatura Eletrônica

07/11/2022 08:10:12.0

THIAGO ANDRÉ GUIMARÃES

Presidente da Banca Examinadora

Assinatura Eletrônica

07/11/2022 08:35:21.0

MAIKO FERNANDES BUZZI

Avaliador Externo (55002145)

Assinatura Eletrônica

07/11/2022 09:39:38.0

SIMONE CRISTINA RAMOS

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Avenida Prefeito Lothario Meissner, 632 - CURITIBA - Paraná - Brasil

CEP 80210-170 - Tel: (41) 3360-4464 - E-mail: ppgold@ufpr.br

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.

Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 233657

Para autenticar este documento/assinatura, acesse <https://www.prppg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp> e insira o código 233657

Dedico a caminhada até o mestrado aos meus pais, que sempre me incentivaram e vibraram com minhas conquistas e também à minha esposa e filhas, que compreenderam minha ausência durante os estudos e me apoiaram até o final.

AGRADECIMENTOS

Agradeço especialmente ao meu orientador Prof. Dr. Thiago André Guimarães, que desde o início dessa jornada acreditou no meu potencial e durante todo o período se dedicou à orientação de forma decisiva para que eu chegasse até aqui.

Meus agradecimentos aos professores e equipe do PPGOLD, que mesmo com os obstáculos impostos pela pandemia, cumpriram o importante papel de nos guiar nessa jornada, em especial a Profa. Dra. Simone Cristina Ramos, Prof. Dr. Cassius Tadeu Scarpin e Prof. Dr. Maiko Fernandes Buzzi, que participaram e contribuíram com este projeto durante as bancas.

Estendo também o agradecimento aos colegas de mestrado e também aos membros do GTAQ, contribuindo para constantes trocas de conhecimento, destacando o amigo e mestre Cássio de Araújo Silva, grande incentivador para minha entrada no mestrado.

Por fim, agradeço a empresa e equipe onde trabalho, local de realização deste estudo de caso.

Quanto mais aumenta nosso conhecimento, mais evidente fica nossa
ignorância. (John F. Kennedy)

RESUMO

Um desafio atual da área de Gestão de Pessoas é o desenvolvimento de métricas estruturadas sobre a gestão do capital humano em organizações (Qamar e Samad, 2021). O tema *People Analytics*, emerge como uma possibilidade de desenvolvimento de valor para a organização através das métricas de gestão de talentos, porém carece de estudos empíricos que demonstrem como tal valor pode ser gerado (Huselid, 2018). 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 das técnicas de Regressão Logística e Máquina de Vetores de Suporte, como esforço 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 cerca de 25000 rescisões de contratos ocorridas no ano de 2021. Os resultados apontaram que os índices gerais de acurácia dos modelos desenvolvidos situam-se entre 75% a 86%, de acordo com as variáveis explicativas consideradas. Os principais fatores determinantes decorrem das informações relacionadas ao período do estágio, como data de término, duração e data de rescisão. 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 associados à 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 por esta metodologia.

Palavras-chave: *People Analytics*. Estágios. Predição. Regressão Logística. Máquina de Vetores de Suporte.

ABSTRACT

A current challenge in people management is the development of the area of dimensioning human structures on capital management in organizations (Qamar and Samad, 2021). The People Analytics theme emerges as a possibility of development for an organization through talent management measurements, but it lacks empirical studies that demonstrate how such value can be generated (Huselid, 2018). Faced with this, the scenario of this research is the practice of People Analytics in a case of practical application of internship logistics management, performing a predictive analysis of regression techniques and machine application of support techniques, in an effort to identify the factors determinants of early terminations of internship offers. To point out, an analysis index of a case study result in an internship agent from Paraná, with analysis results from a case study in which general agents from Paraná, 75% according to the explanatory variables. The main determinants of correctness of information related to the internship period, such as term data, duration and termination data. The conclusion is that the results can help as parties involved in the internship process (companies, students, teaching agents and integration agents), in order to settle in advance the impacts associated with proactive actions. In addition, this work brings theoretical and empirical contributions. From the point of view, the understanding of the subject can support the best processes of recruitment, selection and management of interns. By the theoretical, the theory of application thought in the process of expanding the thought of the value generated by this methodology.

Keywords: People Analytics. Internship. Prediction. Logistic Regression. Support Vector Machine.

LISTA DE FIGURAS

FIGURA 1 - ESTRUTURA DO PEOPLE ANALYTICS (INTERSECÇÃO)	22
FIGURA 2 - FORMA DA RELAÇÃO LOGÍSTICA ENTRE VARIÁVEIS DEPENDENTES E INDEPENDENTES	23
FIGURA 3 - GRÁFICO DE $p = fZ$	25
FIGURA 4 - CURVA DE SENSIBILIDADE PARA TRÊS VALORES DE CUTOFF	28
FIGURA 5 - CRITÉRIO DE ESCOLHA DO MODELO COM MAIOR ÁREA DA CURVA ROC	29
FIGURA 6 - UM HIPERPLANO DE SEPARAÇÃO (w,b) PARA UM CONJUNTO DE TREINAMENTO BIDIMENSIONAL	32
FIGURA 7 - UM HIPERPLANO DE MARGEM ÓTIMA COM SEUS VETORES DE SUPORTES DESTACADOS	33
FIGURA 8 - UM MAPA DE CARACTERÍSTICA PODE SIMPLIFICAR A TAREFA DE CLASSIFICAÇÃO	34
FIGURA 9 - (a) CONJUNTO DE DADOS NÃO LINEAR; (b) FRONTEIRA NÃO LINEAR NO ESPAÇO DE ENTRADAS; (c) FRONTEIRA LINEAR NO ESPAÇO DE CARACTERÍSTICAS	35
FIGURA 10 - ANÁLISE DE COMPONENTES PRINCIPAIS - GRÁFICO DE VARIÁVEIS	42
FIGURA 11 - RESULTADO DA CURVA ROC PARA O MODELO SEM O MÊS DE RESCISÃO	46
FIGURA 12 - IMPORTÂNCIA DAS VARIÁVEIS - SEM MÊS DE RESCISÃO..	47
FIGURA 13 - RESULTADO DA CURVA ROC PARA O MODELO COM O MÊS DE RESCISÃO	49
FIGURA 14 - IMPORTÂNCIA DAS VARIÁVEIS - COM MÊS DE RESCISÃO .	50
FIGURA 15 - RESULTADO CURVA ROC – DADOS DE TREINO	52
FIGURA 16 - RESULTADO CURVA ROC	53

LISTA DE QUADROS

QUADRO 1 - MATRIZ DE CLASSIFICAÇÃO DO MODELO.....	27
QUADRO 2 - RESUMO DOS ESTUDOS CONSIDERADOS.....	40
QUADRO 3 - VARIÁVEIS UTILIZADAS.....	43
QUADRO 4 - SÍNTESE DAS AÇÕES.....	57

LISTA DE TABELAS

TABELA 1 - FUNÇÕES KERNEL MAIS COMUNS	35
TABELA 2 - DADOS DISTRIBUÍDOS - TREINO E TESTE.....	44
TABELA 3 - RESÍDUOS DE DESVIO PARA O MODELO SEM O MÊS DE RESCISÃO	45
TABELA 4 - COEFICIENTES PARA O MODELO SEM O MÊS DE RESCISÃO	45
TABELA 5 - RESULTADO DA MATRIZ DE CONFUSÃO DOS DADOS DE TREINO E TESTE.....	46
TABELA 6 - RESÍDUOS DE DESVIO PARA O MODELO COM O MÊS DE RESCISÃO	48
TABELA 7 - COEFICIENTES PARA O MODELO COM O MÊS DE RESCISÃO	48
TABELA 8 - RESULTADO DA MATRIZ DE CONFUSÃO DOS DADOS DE TREINO E TESTE.....	49
TABELA 9 - RESULTADO DA MATRIZ DE CONFUSÃO DOS DADOS DE TREINO.....	52
TABELA 10 - RESULTADO DA MATRIZ DE CONFUSÃO DOS DADOS DE TREINO.....	53
TABELA 11- RESUMO DOS RESULTADOS.....	54

LISTAS DE ABREVIATURAS E SIGLAS

AI	- Agente de Integração
AUC	- <i>Area Under the Curve</i>
EAD	- Educação à Distância
HRA	- <i>Human Resources Analytics</i>
IE	- Instituição de Ensino
LDA	- Análise Discriminante Linear
ML	- <i>Machine Learning</i>
MLE	- <i>Maximum Likelihood Estimation</i>
PA	- <i>People Analytics</i>
PCE	- Parte Concedente de Estágio
PPC	- Projeto Pedagógico do Curso
TCE	- Termo de Compromisso de Estágio
TRA	- Teoria da Ação Racional
SVM	- <i>Support Vector Machine</i>
RBF	- <i>Radial-Basis Function</i>
RH	- Recursos Humanos
RL	- Regressão Logística
RLB	- Regressão Logística Binária
ROC	- <i>Receiver Operating Characteristic</i>
VIF	- <i>Variance Inflation Factor</i>

SUMÁRIO

1	INTRODUÇÃO	15
2	REFERENCIAL TEÓRICO	18
2.1	LEI DE ESTÁGIO	18
2.2	PEOPLE ANALYTICS	19
2.3	TÉCNICAS DE PREDIÇÃO	22
2.3.1	Regressão logística binária	22
2.3.2	Máquina de vetores de suporte	30
2.4	TRABALHOS CORRELATOS	35
3	DESENVOLVIMENTO	40
3.1	COLETA, PREPARAÇÃO E DEFINIÇÃO DAS VARIÁVEIS	41
3.2	APLICAÇÕES DAS TÉCNICAS E RESULTADOS	44
3.2.1	Regressão logística	44
3.2.1.1	Modelo sem mês de rescisão	44
3.2.1.2	Modelo com o mês de rescisão	47
3.2.2	Máquina de vetores de suporte	51
3.2.2.1	Modelo sem mês de rescisão	51
3.2.2.2	Modelo com o mês de rescisão	52
3.3	INTERPRETAÇÃO DA PESQUISA	54
4	CONCLUSÃO	58
	REFERÊNCIAS	60

1 INTRODUÇÃO

O programa de estágio é de grande importância para a inclusão de jovens no mercado de trabalho, pois esta atividade não exige experiência laboral prévia, além de possibilitar ao estudante aplicar conhecimentos teóricos ao contexto prático. Conforme argumentado por Oliveira *et al.* (2010), o processo de estágio exerce papel central na vida do estudante, visto que esta experiência ocorre, em geral, na transição da juventude para a fase adulta, período este onde aportam-se sensíveis mudanças biológicas, psicológicas e de integração social.

No âmbito operacional, a viabilização do estágio enseja o cumprimento de alguns requisitos básicos 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). Este contrato detalha as responsabilidades das partes envolvidas, conforme normas especificadas na Lei de Estágio (BRASIL, 2008). O vínculo e a observação de tais normas são válidos até que se ocorra a rescisão do TCE, cuja vigência máxima é de dois anos¹.

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 remunere o período trabalhado e eventuais dias de recesso, além de entregar o termo de realização do estágio para o estudante. Todavia, essa flexibilidade acaba contribuindo para a rotatividade dos estagiários nas PCEs, e prever esta rescisão antecipada é um grande desafio.

Uma das alternativas para que a organização disponha de maior compreensão sobre o processo de *turnover* de seus funcionários é através do *People Analytics* (PA), cujo interesse de pesquisadores e de organizações no âmbito de Recursos Humanos (RH) vem sendo ampliado nos últimos anos (Tursunbayeva *et al.*, 2018; Huselid, 2018). Em PA a análise de dados por métodos científicos é usada para auxiliar nas decisões de gerenciamento, seja em aquisição de talentos, retenção, colocação, promoção, remuneração ou força de trabalho e planejamento de sucessão. A análise de pessoas e o *big data*

¹ Quando se tratar de portador de deficiência, o contrato de estágio pode ser redigido com vigência indeterminada.

passam a ser, então, recursos valiosos em RH. A expectativa é que a *PA* não apenas fará o RH avançar em termos de análise e *insights*, mas também criará a capacidade de previsão e tecnologia sugestiva (Walford-Wright e Scott-Jackson, 2018).

Marler e Boudreau (2017) ao realizarem uma revisão de literatura sobre *Human Resources Analytics* (HRA), destacam que apesar do crescente interesse pelo tema, ainda são poucas e limitadas as evidências científicas sobre a adoção de HRA na tomada de decisão, com conclusão da necessidade de maior investigação e estudos empíricos. Conclusão semelhante chegaram os autores Qamar e Samad (2021) em sua pesquisa, com limitados trabalhos na literatura que adotaram estudos empíricos.

No âmbito da análise de *turnover*, Abbassi e Hollman (2000) o descrevem como sendo a rotação de trabalhadores no mercado de trabalho, entre empresas, empregos e profissões, bem como entre os estados de emprego e desemprego. Também destacam que o *turnover* excessivo muitas vezes gera consequências negativas para a organização, como o aumento de custos, dada a queda de produtividade inerente ao processo de substituição de pessoas. Considerando este cenário de rotatividade, Gao *et al.* (2019) afirmam que se a empresa puder prever quais funcionários irão deixar a organização no futuro, poderão se antecipar ao fato e tomar medidas para reduzir esta probabilidade.

Segundo Ongori (2007) as organizações investem em seus funcionários em termos de admissão e treinamento, desenvolvendo, mantendo e retendo-os em seus quadros. Há, portanto, a necessidade de se ter maior compreensão dos fatores determinantes à rotatividade de funcionários, para que as ações sejam precisamente endereçadas à contenção de suas fontes causadoras. O autor complementa que vários pesquisadores tentaram identificar os fatores que levam as pessoas a deixarem as organizações, mas que até aquele momento, havia pouca consistência nos achados.

Neste sentido, Fallucchi *et al.* (2020), apontam que a inteligência artificial pode ser uma alternativa no auxílio ao entendimento das fontes geradoras de *turnover*, haja vista sua amplitude de uso no auxílio à tomada de decisão, incluindo a área de Gestão de Pessoas. Uma das técnicas de predição amplamente utilizadas por pesquisadores é a Regressão Logística (RL). Peng *et al.* (2002) indicam que a RF pode ser uma ferramenta poderosa quando a variável

resposta é dicotômica. Os autores afirmam ainda que a RL ganhou popularidade, por ocasião da ampla disponibilidade de softwares disponíveis no mercado.

Saradhi e Palshikar (2011) demonstram em sua pesquisa que as técnicas de aprendizado de máquina, por exemplo a Máquina de Vetores de Suporte (SVM), podem ser usadas para construir modelos preditivos confiáveis e precisos para a rotatividade de funcionários.

Embora a utilização de técnicas de predição seja uma prática exitosa no âmbito do suporte decisório nas organizações, a gestão de contrato de estágios ainda é um nicho muito pouco explorado. 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. Todavia, este contexto possui paralelo no âmbito da Gestão de Pessoas, com estudos orientados à predição de desligamentos de funcionários nas empresas (*turnover*). 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.

Considerando o referencial teórico e os estudos efetivados nas técnicas de RL, o objetivo dessa pesquisa é a aplicação de *People Analytics* para identificar os fatores determinantes às rescisões antecipadas de contratos de estágio, e nesta esteira desenvolver uma metodologia para a predição de rescisões que podem ocorrer antes da finalização do contrato. Para tal, propõe-se o estudo das técnicas de classificação Regressão Logística (RL) e Máquina de Vetores de Suporte (SVM) e posterior análise de caso em um Agente de Integração de Estágio (AI) no estado do Paraná.

A estrutura do trabalho é descrita como segue: na segunda Seção o documento apresenta o referencial teórico, que abordará: 1. aspectos legais sobre a prática laboral do estágio, 2. definições de *People Analytics*, 3. conceitos teóricos sobre as técnicas de classificação e predição utilizadas e 4. um levantamento bibliográfico sobre a aplicação de técnicas de predição em *turnover*. A Seção 3, apresenta as aplicações das técnicas e reporta e interpreta os resultados do estudo de caso, enquanto a Seção 4 tece as considerações finais do trabalho e alguns direcionamentos futuros.

2 REFERENCIAL TEÓRICO

Este referencial teórico aborda os aspectos fundamentais que suportam este estudo. Primeiramente, dar-se-á a dimensão legal do estágio e definições de *People Analytics*. Na sequência, aspectos pertinentes acerca da Regressão Logística e Máquina de Vetores de Suporte serão abordados. Finalizando, será apresentado levantamento bibliográfico sobre o emprego da utilização das técnicas de predição em *turnover*.

2.1 LEI DE ESTÁGIO

Em 15 de setembro de 2008 foi sancionada a Lei n. 11.788, conhecida como Lei do Estágio. Em seu artigo 1, define-se:

Estágio é ato educativo escolar supervisionado, desenvolvido no ambiente de trabalho, que visa à preparação para o trabalho produtivo de educandos que estejam frequentando o ensino regular em instituições de educação superior, de educação profissional, de ensino médio, da educação especial e dos anos finais do ensino fundamental, na modalidade profissional da educação de jovens e adultos. (Brasil, 2008).

Ainda segundo a mesma Lei, o estágio visa o aprendizado de competências próprias da atividade profissional e à contextualização curricular, objetivando o desenvolvimento do educando para a vida cidadã e para o trabalho.

O programa de estágio não deve estar condicionado à experiência prévia do candidato, mas necessita compor o Projeto Pedagógico do Curso (PPC), seguindo um itinerário de formação do estudante. Assim, as atividades práticas devem estar alinhadas com as diretrizes pedagógicas do curso ao qual o estágio se destina. O mercado brasileiro de estagiários, segundo as últimas estatísticas da Abres (Estatísticas, Abres, 2021) é de 900000 estagiários, sendo 686000 no ensino superior e 214000 no ensino médio e técnico. As seis maiores áreas em número de vagas de estágio são: Administração (16,8%), Direito (7,3%), Comunicação Social (6,2%), Informática (5,2%), Engenharias (5,1%) e Pedagogia (4,2%).

Para viabilizar o estágio é obrigatório cumprimento de requisitos básicos, envolvendo Instituição de Ensino (IE), Estudante e Parte Concedente de Estágio (PCE). Também é possível a participação do Agente de Integração (AI) nesse processo, conforme definido na Lei de Estágio. Dentre as atribuições desse AI estão gerenciar uma base de dados de estudantes, identificar oportunidades de estágio, ajustar as condições de realização do estágio e realizar o acompanhamento administrativo.

No processo de contratação de estagiários através de um AI é necessário que a IE e a PCE celebrem um convênio tríplice. Neste âmbito, as PCEs ofertam as vagas de estágio e o AI realiza o processo de divulgação com os estudantes cadastrados em seu banco de dados, cabendo também ao AI o encaminhando de candidatos que atendam aos requisitos legais e operacionais da vaga oportunizada. A PCE então realiza o processo de seleção desses candidatos e, após definir o aprovado, efetiva através do AI o Termo de Compromisso de Estágio (TCE), detalhando as responsabilidades das partes envolvidas conforme normas especificadas na Lei de Estágio. O vínculo e a observação de tais normas têm vigência até que se ocorra a rescisão do contrato de estágio, que pode ser motivado pelo término da data estabelecida no TCE, ou a pedido de qualquer uma das partes envolvidas. Segundo a Lei de Estágio (BRASIL, 2008), o vínculo empregatício se configura apenas quando o estágio está em desconformidade com esta lei, e nesse caso deve ser aplicado a legislação trabalhista e previdenciária vigente.

O AI onde o estudo será conduzido atua no mercado desde 1967 e gerencia atualmente mais de 27000 contratos de estágios no Paraná, com aproximadamente 6300 empresas (PCEs) contratantes, em diferentes segmentos de mercado.

2.2 PEOPLE ANALYTICS

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, são poucas evidências que a *Human Resource Analytics (HRA)* 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 lutado 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.

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 e Boudreau (2017); Qamar e 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.

Fitz-enz (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).

2.3 TÉCNICAS DE PREDIÇÃO

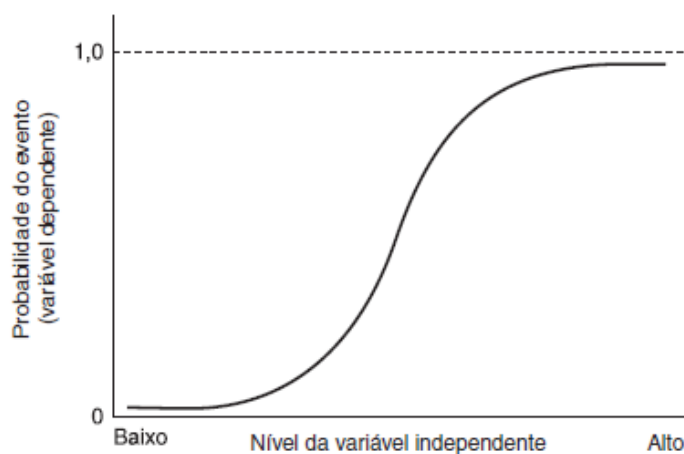
2.3.1 Regressão logística binária

Uma das técnicas de predição utilizadas nessa pesquisa é a Regressão Logística, que será explorado nessa seção. Também serão descritos conceitos importantes para compreensão dos resultados do estudo de caso, como matriz de classificação, desempenho dos modelos, acurácia e curva *ROC*.

Hair *et al.* (2009) descrevem a regressão logística como uma forma especializada de regressão que é formulada para prever e explicar uma variável categórica binária (dois grupos), e não uma medida dependente métrica. A forma da variável estatística de regressão logística é semelhante à da variável

estatística da regressão múltipla. Na FIGURA 2 tem-se a representação gráfica entre as variáveis independentes e dependente, utilizando a curva logística. À medida que a probabilidade de ocorrência de uma variável independente reduz a curva, se aproxima de 0, porém nunca alcança este valor (curva assintótica). Analogamente, quando a probabilidade de ocorrência do evento aumenta, os valores previstos aumentam a taxas decrescentes na curva se aproximando de 1,0 também de maneira assintótica.

FIGURA 2- FORMA DA RELAÇÃO LOGÍSTICA ENTRE VARIÁVEIS DEPENDENTES E INDEPENDENTES



FONTE: HAIR *et al.* (2009)

Um exemplo de aplicação da regressão logística binária em *turnover*, enseja a quantificação da probabilidade de um funcionário ser desligado da empresa, com base nas seguintes variáveis independentes: nas suas médias de avaliações de desempenho, seu salário, e no número de faltas por mês. A variável dependente nesse caso seria binária, isto é, ser desligado ou não. Nesse caso, a variável dependente é do tipo binária ou *dummy*, assumindo o valor zero se o funcionário for desligado e o valor um, caso o funcionário permaneça.

Fávero e Belfiore (2017) descrevem que quando o fenômeno em estudo puder ser representado por meio de apenas duas categorias, aplica-se a técnica de regressão logística binária. Por outro lado, se o fenômeno em estudo apresentar mais de duas categorias como possibilidade de ocorrência, será preciso definir uma categoria como referência e depois elaborar a técnica de regressão logística multinomial.

O objetivo da regressão logística binária é quantificar a probabilidade de ocorrência de um evento, representado de forma qualitativa dicotômica por uma variável Y , assumindo 0 ou 1, considerando o comportamento de variáveis explicativas X . De forma a exemplificar, define-se em (1) um vetor de variáveis explicativas Z , chamado de logito, contendo uma constante α , e a soma do produto das variáveis explicativas X com os parâmetros estimados β . O conceito de chance de ocorrência de um evento é conhecido por *odds*, e está representado em (2), considerando a probabilidade p_i de ocorrência do evento de interesse para cada observação, isto é, em função dos parâmetros estimados para cada variável explicativa. Em (3) a regressão logística binária define o logito Z como o logarítmico natural da chance. Dessa forma, tem-se em (4) a probabilidade de ocorrência do evento e a probabilidade de não ocorrência em (5).

$$Z_i = \alpha + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_k \cdot X_{ki} \quad (1)$$

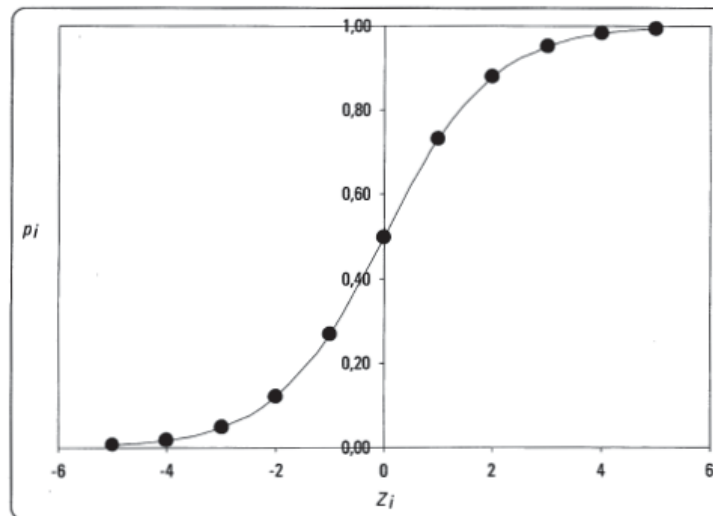
$$P(odds)_{Y_i=1} = \frac{p_i}{1 - p_i} \quad (2)$$

$$\ln\left(\frac{p_i}{1 - p_i}\right) = Z_i \quad (3)$$

$$p_i = \frac{e^{Z_i}}{1 + e^{Z_i}} = \frac{1}{1 + e^{-Z_i}} \quad (4)$$

$$1 - p_i = 1 - \frac{e^{Z_i}}{1 + e^{Z_i}} = \frac{1}{1 + e^{Z_i}} \quad (5)$$

Na FIGURA 3 ilustra-se a dinâmica acumulativa para as probabilidades estimadas, em função de vários valores assumidos por Z , situando entre 0 e 1, garantido quando se impôs que o logito fosse igual ao logaritmo natural da chance.

FIGURA 3- GRÁFICO DE $p = f(Z)$ 

FONTE: FÁVERO; BELFIORE (2017)

Fávero e Belfiore (2017), alertam que se a proporção de observações definida como evento e não evento da amostra deve ser idêntica à correspondente em relação a população. Caso se conheça esta proporção é necessário aplicar uma correção no valor do intercepto estimado no modelo amostral.

Na regressão logística binária, a estimação dos parâmetros da equação da probabilidade é realizada pelo uso da função de verossimilhança (*likelihood function*), a partir da qual será elaborada em (6) a estimação por máxima verossimilhança (*maximum likelihood estimation - MLE*). Segundo Hair *et al.* (2009), a estimação de máxima verossimilhança (MLE) é comumente empregada em modelos de equações estruturais. Uma alternativa aos usuais mínimos quadrado sem regressão múltipla, a MLE é um procedimento que melhora por iterações as estimativas de parâmetros para minimizar uma função de ajuste especificada.

$$LL = \sum_{i=1}^n \left\{ \left[Y_i \cdot \ln \left(\frac{e^{Z_i}}{1 + e^{Z_i}} \right) \right] + \left[(1 - Y_i) \cdot \ln \left(\frac{1}{1 + e^{Z_i}} \right) \right] \right\} \quad (6)$$

Para se avaliar a significância estatística geral do modelo na regressão logística binária e em alguns casos como indicador de desempenho do modelo

escolhido, pode ser utilizado o *pseudo R² de McFadden*, apresentado em (7), que possibilita comparar dois ou mais modelos distintos, considerando como critério positivo o maior valor.

$$pseudo R^2 = \frac{-2.LL_0 - (-2.LL_{m\acute{a}x})}{-2.LL_0} \quad (7)$$

Segundo Fávero e Belfiore (2017), o teste χ^2 (qui-quadrado) permite uma verificação inicial sobre a existência do modelo que está sendo proposto, uma vez que, se todos os parâmetros estimados $\beta_j (j = 1, 2, \dots, k)$ forem estatisticamente iguais a 0, o comportamento de alteração de cada uma das variáveis X não influenciará a probabilidade de ocorrência do evento em estudo. A estatística χ^2 segue a expressão (8).

$$\chi^2 = -2.(LL_0 - LL_{m\acute{a}x}) \quad (8)$$

O resultado do teste é em graus de liberdade e ao nível de significância de 5%, o *p-valor* for menor que 0,05, rejeita a hipótese nula, sendo possível dar continuidade à análise de regressão. Este teste avalia a significância conjunta das variáveis explicativas, não definindo quais destas variáveis consideradas no modelo são estatisticamente significativas para influenciar a probabilidade de ocorrência do evento.

Desta forma, é preciso avaliar se cada um dos parâmetros do modelo de regressão logística binária é estatisticamente significativo. Para isso, a estatística *z de Wald* fornece o teste para cada parâmetro do modelo. As expressões para o cálculo das estatísticas *z de Wald* de cada parâmetro α e β são dadas, respectivamente em (9) e (10), em que *SE* mede o erro-padrão (*standard error*) de cada parâmetro em análise.

$$z_\alpha = \frac{\alpha}{SE(\alpha)} \quad (9)$$

$$z_{\beta_j} = \frac{\beta_j}{SE(\beta_j)} \quad (10)$$

Após o cálculo das estatísticas *z de Wald*, o pesquisador pode utilizar a tabela de distribuição da curva normal padrão para obtenção dos valores críticos a um dado nível de significância e verificar se tais testes rejeitam ou não a hipótese nula (H_0). Caso seja rejeitada, conclui-se que há associação entre as variáveis.

Como na regressão logística binária a ocorrência de um evento é definida por um ponto de corte (*cutoff*). Para um *cutoff*=0,5, qualquer valor de $p > 0,5$ será considerado a incidência do evento (FIGURA 4). Definido o *cutoff*, constrói-se uma matriz de classificação do modelo (QUADRO 1) e define-se a sensibilidade (11), especificidade (12) e eficiência global ou acurácia (13).

QUADRO 1- MATRIZ DE CLASSIFICAÇÃO DO MODELO

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

FONTE: O autor (2022)

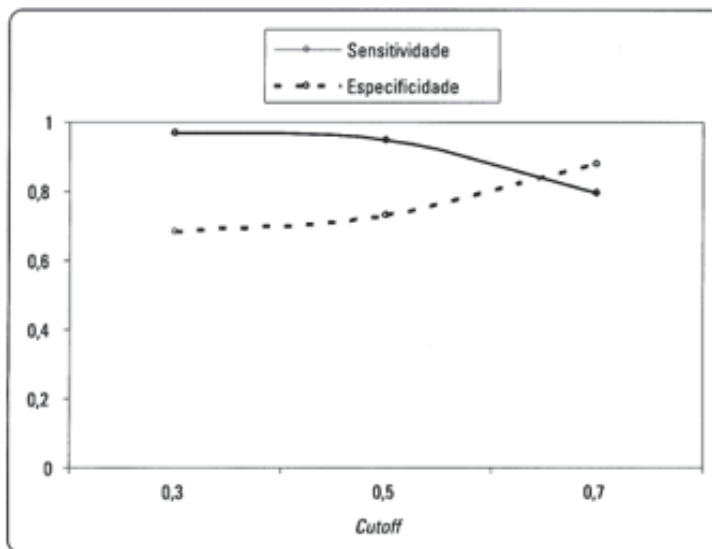
$$sensibilidade = \frac{VP}{VP + FN} \quad (11)$$

$$especificidade = \frac{VN}{FP + VN} \quad (12)$$

$$eficiência\ global = \frac{VP + VN}{VP + FN + FP + VN} \quad (13)$$

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 4, é calculada com diferentes *cutoff*, sendo possível analisar em qual valor a sensibilidade se iguala a especificidade.

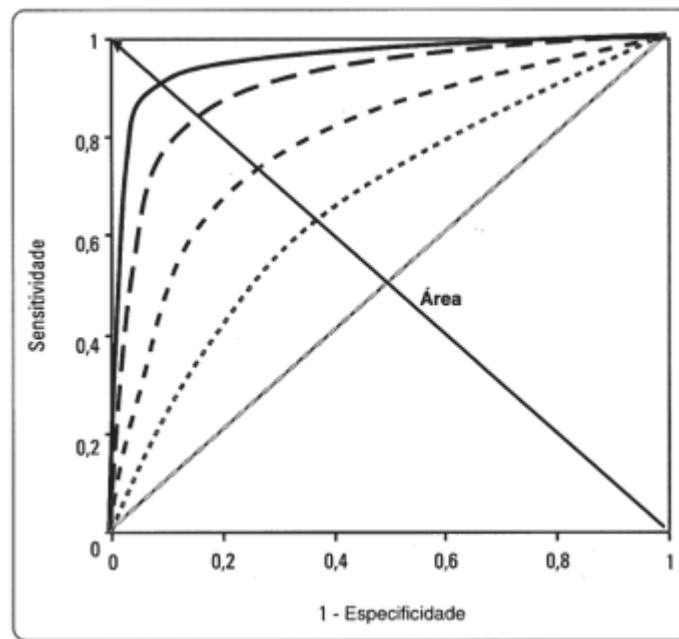
FIGURA 4 - CURVA DE SENSIBILIDADE PARA TRÊS VALORES DE *CUTOFF*



FONTE: FÁVERO; BELFIORE (2017)

A curva ROC (FIGURA 5), apresenta a sensibilidade e especificidade de forma convexa, indicando que o modelo com maior área abaixo da curva (*area under the curve - AUC*) apresenta maior eficiência global de previsão.

FIGURA 5 - CRITÉRIO DE ESCOLHA DO MODELO COM MAIOR ÁREA DA CURVA ROC



FONTE: FÁVERO; BELFIORE (2017)

Ainda de acordo com Fávero e Belfiore (2017), um dos principais problemas de predição reside na multicolinearidade das variáveis explicativas. Esta anomalia ocorre quando há correlações muito elevadas entre tais variáveis, o que impossibilita a identificação de qual delas efetivamente influenciam a variável dependente.

Neste sentido, Hair *et al.* (2009) relatam que as duas medidas mais comuns para se avaliar a multicolinearidade são a tolerância (14), onde R_k^2 é o coeficiente de determinação para a previsão da variável k pelas outras variáveis independentes na variável estatística de regressão, e seu inverso, o fator de inflação de variância – *Variance Inflation Factor* (VIF) (15). À medida que o valor da tolerância se torna menor, a variável é melhor prevista pelas outras variáveis independentes (colinearidade). Também indicam que uma referência comum entre os pesquisadores é quando o VIF for maior que 10, existem problemas de multicolinearidade e devem ser observados com atenção e executar ações corretivas, como por exemplo, omitir esta variável do modelo.

$$tolerância = 1 - R_k^2 \quad (14)$$

$$VIF = \frac{1}{\text{tolerância}} \quad (15)$$

2.3.2 Máquina de vetores de suporte

Outra técnica de predição utilizada nessa pesquisa é a Máquina de Vetores de Suporte - *Support Vector Machines* (SVMs), que segundo Cristianini e Shawe-Taylor (2000) são sistemas de aprendizagem que utilizam um espaço de hipótese de funções lineares em um espaço de recursos de alta dimensão, treinado com um algoritmo de aprendizagem da teoria da otimização, que possuem viés de aprendizagem derivado da teoria de aprendizagem estatística. Esta teoria foi introduzida por Vapnik (1995) sendo considerado um dos principais métodos de estratégia de aprendizagem.

Steinwart e Christmann (2008) destacam que o SVM se aplica em várias situações. Os autores identificam três razões para o sucesso desta técnica: 1. sua capacidade de aprender bem com somente um pequeno número de parâmetros, 2. sua robustez contra vários tipos de violações de modelos e *outliers* e 3. sua eficiência computacional comparada com outros métodos.

Ma e Guo (2014) também descrevem que SVM foi desenvolvido com base na teoria de aprendizagem estatística e mais tarde adotado em aprendizagem de máquina, com aplicações em robótica, visão computacional, reconhecimento de padrões, segurança da computação, análise de imagens médicas, entre outras.

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 supervisionadas envolve construir um modelo estatístico para predição, ou estimativa, com a saída baseada em uma ou mais entradas. Com a aprendizagem estatística não supervisionada, existem entradas, mas as saídas não são supervisionadas e nem é possível ter a previsão de uma resposta, pois não sabemos a resposta verdadeira. Utilizada muitas vezes como parte de uma análise exploratória de dados, sendo possível aprender os relacionamentos e estrutura dos dados (JAMES *et al.*, 2013).

Considerando a aprendizagem supervisionada e a abordagem de aprendizagem de máquina (*Machine Learning*), uma quantidade limitada de valores de entrada com valores de saídas conhecidas são utilizadas para “aprender” o presumido, mas desconhecido o 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).

De acordo com Cristianini e Shawe-Taylor (2000), dentro da aprendizagem supervisionada, é fornecido um conjunto de dados treinamento com exemplos de entradas e as respectivas saídas. Usualmente os exemplos estão na forma de vetores de atributos, sendo os espaços de entrada um subconjunto de R^n . Uma vez os vetores de atributos estão disponíveis, um número de conjunto de hipóteses pode ser escolhido para o problema, e entre estes, as funções lineares estão entre as mais simples e de fácil entendimento de aplicação.

Nos casos onde os atributos podem ser linearmente separáveis, é possível explicar a ideia chave do SVM. Dado um conjunto de treinamento S que contém pontos em duas classes, um SVM separa as classes por meio de um hiperplano determinado por certos pontos de S , denominado como vetores de suporte. No caso separável, este hiperplano maximiza a margem e todos os vetores de suporte ficam na posição na mesma distância mínima a partir do hiperplano, que chamados vetores margem (PONTIL; VERRI, 1998).

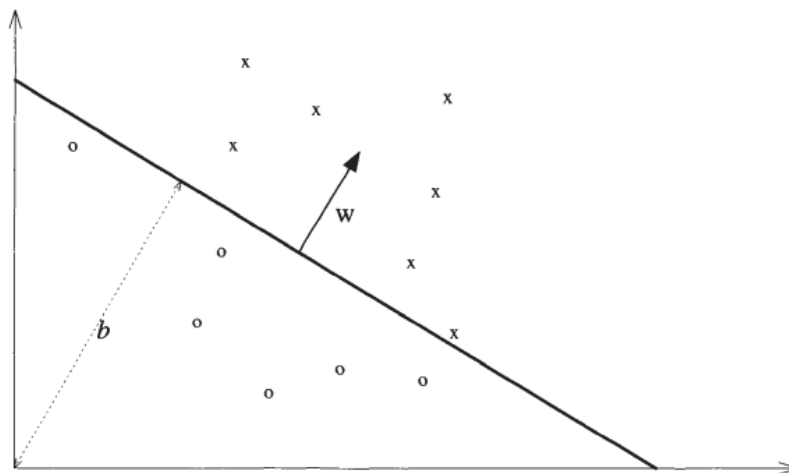
Cristianini e Shawe-Taylor (2000) apresentam que a classificação binária é frequentemente implementada utilizando uma função de valor real $f : X \subseteq R^n \rightarrow R$ na seguinte forma: a entrada $x = (x_1, \dots, x_n)$ é atribuída a uma classe positiva se $f(x) \geq 0$ e a uma classe negativa se $f(x) \leq 0$. Considera-se o caso onde $f(x)$ é uma função linear de $x \in X$, podendo ser escrito conforme (16), onde $(w, b) \in R^n \times R$ são parâmetros que controlam a função e regra de decisão é dada por $sgn(f(x))$, onde será utilizada a convenção que $sgn(0) = 1$, sendo a função sinal sgn representando o hiperplano que separa os dados com maior margem. O vetor w define uma direção perpendicular ao hiperplano, enquanto varia o valor de b movendo o hiperplano paralelo a si mesmo.

$$f(x) = (w \cdot x) + b \quad (16)$$

$$= \sum_{i=1}^n w_i x_i + b$$

Cristianini e Shawe-Taylor (2000) complementam que a interpretação geométrica desse tipo de hipótese é que o espaço de entrada X é separado em duas partes pelo hiperplano definido pela equação $(w \cdot x) + b = 0$ conforme FIGURA 6. Um hiperplano é um subespaço a fim de dimensão $n - 1$ que divide o espaço em duas metades que correspondem as entradas das duas classes distintas, como exemplo da FIGURA 6, onde o hiperplano é a linha mais espessa, com a região positiva acima e a região negativa abaixo. O vetor w define uma direção perpendicular ao hiperplano, ao variar o valor de b , move o hiperplano paralelo a si mesmo. Portanto, uma representação envolvendo $n + 1$ parâmetros livres é necessário, se alguém deseja representar todos os possíveis hiperplanos em R^n .

FIGURA 6 – UM HIPERPLANO DE SEPARAÇÃO (w,b) PARA UM CONJUNTO DE TREINAMENTO BIDIMENSIONAL

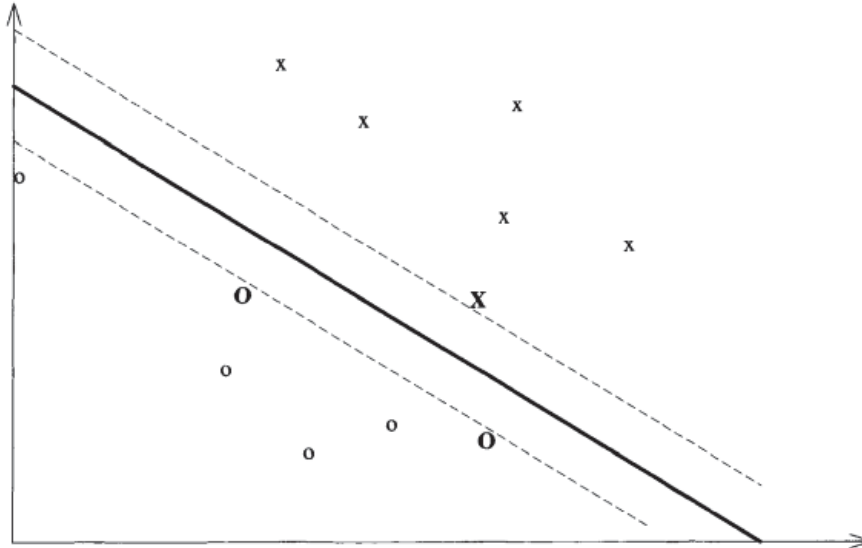


FONTE: CRISTIANINI; SHAWE-TAYLOR (2000)

Segundo Cristianini e Shawe-Taylor (2000), o mais simples modelo de SVM é chamado classificador de margem ótima, que trabalha apenas com dados linearmente separáveis no espaço de característica e, portanto, não pode ser utilizado em muitas situações da vida real. Este modelo procura o hiperplano de

separação com maior margem, conforme mostrado na FIGURA 7, e sendo formulado utilizando as restrições em (17).

FIGURA 7 - UM HIPERPLANO DE MARGEM ÓTIMA COM SEUS VETORES DE SUPORTES DESTACADOS



FONTE: CRISTIANINI; SHAWE-TAYLOR (2000)

$$(w \cdot x^+) + b = +1 \quad (17)$$

$$(w \cdot x^-) + b = -1$$

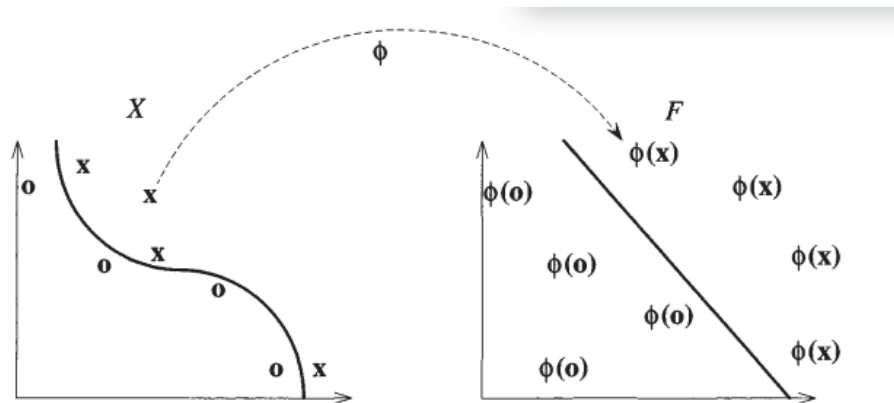
Nos casos onde os dados não podem ser separados linearmente no espaço de característica, uma extensão do SVM pode ser utilizada a otimização de margens suaves, introduzindo variáveis não negativas $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ em (18).

$$y_1(w \cdot x_i + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, \forall i = 1, \dots, n \quad (18)$$

Conforme Cristianini e Shawe-Taylor (2000), dentro do sistema de aprendizagem SVM, para os casos nos quais os dados não são linearmente separáveis, é requerido a utilização de hipóteses de espaços mais expressivos que funções lineares. Representações por *kernels* oferecem uma solução alternativa, projetando os dados dentro de um espaço de característica de alta dimensão para incrementar o poder computacional da aprendizagem de máquina linear. A FIGURA 8 mostra um exemplo de um mapeamento de

recursos de um espaço bidimensional de entrada para um espaço bidimensional de característica, onde os dados não podem ser separados por uma função linear em um espaço de entrada, mas podem ser separados no espaço de característica.

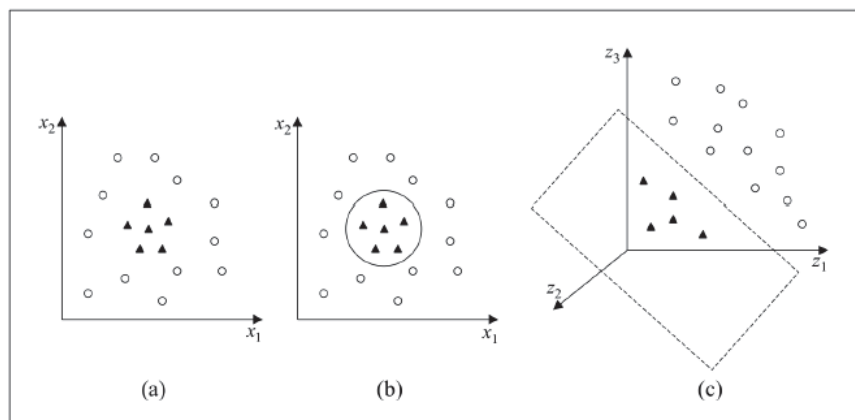
FIGURA 8 - UM MAPA DE CARACTERÍSTICA PODE SIMPLIFICAR A TAREFA DE CLASSIFICAÇÃO



FONTE: CRISTIANINI; SHAWE-TAYLOR (2000)

Em Lorena e De Carvalho (2007), a visualização da separação dos dados por meio do espaço de características em contraste ao espaço de entradas é apresentada na FIGURA 9. As funções *kernel* seguem as condições estabelecidas pelo teorema de Mercer. Alguns *kernels* mais utilizados na prática são os Polinomiais, Os Gaussianos ou RBF (*Radial-Basis Function*) e os Sigmoidais, listados na TABELA 1.

FIGURA 9 - (a) CONJUNTO DE DADOS NÃO LINEAR; (b)FRONTEIRA NÃO LINEAR NO ESPAÇO DE ENTRADAS; (c) FRONTEIRA LINEAR NO ESPAÇO DE CARACTERÍSTICAS



FONTE: LORENA; DE CARVALHO (2007)

TABELA 1 - FUNÇÕES KERNEL MAIS COMUNS

Tipo de Kernel	Função $K(x_i, x_j)$	Parâmetros
Polinomial	$(\delta(x_i \cdot x_j) + k)^d$	δ, k e d
Gaussiano	$\exp(-\sigma \ x_i - x_j\ ^2)$	σ
Sigmoidal	$\tanh(\delta(x_i \cdot x_j) + k)$	δ e k

FONTE: LORENA; DE CARVALHO (2007)

2.4 TRABALHOS CORRELATOS

No melhor do nosso conhecimento, não foram encontradas na literatura científica estudos sobre predições em rescisões antecipadas de contratos de estágio, todavia, serão descritos nessa seção artigos que remetem a aplicação de técnicas de predição focados na rotatividade de funcionários (*turnover*) e em aplicações que fazem menção aos estágios, com objetivo de auxiliar no desenvolvimento dessa pesquisa e possibilitar a realização de análises comparativas das técnicas, metodologias e resultados.

Nadler e Kros (2014) realizaram a coleta de informações por meio de uma pesquisa com motoristas considerando a Teoria da Ação Racional (TRA), e com as variáveis obtidas nessa pesquisa, aplicou-se a técnica de RLB para identificar os fatores que levam à rotatividade de motoristas de uma transportadora que possuem uma alta taxa de rotatividade. A variável

dependente utilizada foi teve como pergunta: “*Qual a probabilidade de você mudar de emprego nos próximos 12 meses?*”, resultando nas respostas alta ou baixa probabilidade. Os resultados concluem que baixa qualidade de gestão da empresa, motoristas que tiveram mais empregos anteriores, baixa idade e os que dirigem menos quilômetros por ano tem maior probabilidade de rotatividade. Os autores relatam que a pesquisa fornece uma contribuição gerencial importante que permite identificar os motoristas com maior risco de desligamento e tomar ações proativas.

Quinn *et al.* (2002) aplicaram os modelos de RLB e uma ferramenta de redes neurais (*Multi-Layered Perceptron*), com objetivo de comparar estas duas abordagens para prever a rotatividade de funcionários em uma empresa que atua como agência de serviços de proteção à criança, definindo como variável dependente da pesquisa se o funcionário deixou a agência ou não. Em relação a comparação das técnicas, os resultados encontrados nessa comparação foi que a taxa de previsão geral dos modelos foi equivalente, contrariando a expectativa que previam resultados superiores das redes neurais em comparação a RL. Os resultados de sucesso do modelo com os dados de testes variaram entre 60% a 70%. A melhor variável preditora apontada na pesquisa foi a etnia (afro-americana ou não), porém os resultados do estudo, segundo os autores, não foram suficientes para dar aos gestores previsões de rotatividade aceitáveis.

Em Rombaut e Guerry (2018), a rotatividade voluntária de funcionários foi abordada aplicando a técnica de RLB na filial Belga de uma empresa privada especializada em soluções de RH, com objetivo de desenvolver uma abordagem metodológica para investigar e prever esta rotatividade, apenas com dados disponíveis no banco de dados da empresa, totalizando 4041 funcionários ao longo dos anos de 2006 a 2016. O modelo de RLB alcançou uma taxa de *AUC* de 0.7432 e demonstrou que algumas variáveis causam diminuição na rotatividade, como gênero feminino, mais tempo na empresa, mais idade e estar casado. Estas conclusões permitem aos gestores realizarem ações proativas, como entrevistas com os grupos mais propensos, revisão nos processos de recrutamento e seleção ou ainda rever as medidas de gestão dos recursos humanos.

Construir e comparar modelos preditivos para a rotatividade voluntária foi um dos temas de pesquisa no estudo de caso de Saradhi e Palshikar (2011), realizado em uma grande organização. Dados coletados durante um período de dois anos para construir um modelo preditivo de rotatividade de funcionários, considerando 20 atributos para construção dos modelos preditivos em três grupos: I. demitidas, II. liberadas e III. retidas. Foram utilizadas as técnicas de *SVM*, *Random Forest* e *Naïve Bayesian*, alcançando índices de acuraria entre 73.97 (*Naïve Bayesian* no grupo I) a 92.28 (*Random Forest* no grupo II). Os autores relataram que a principal contribuição do artigo é a demonstração de que técnicas de aprendizado de máquina podem ser utilizadas para construir modelos preditivos confiáveis e precisos para rotatividade de funcionários.

D.Aiao (2013) analisou dados de 309 funcionários desligados de uma instituição superior de ensino na Nigéria, por meio de modelos preditivos utilizando três tipos de classificadores da técnica de Árvores de Decisão, com objetivo de prever o desligamento desses funcionários. O estudo mostrou que o salário e tempo de serviços foram os fatores determinantes no desligamento, sendo os funcionários com maiores salários e mais tempo de empresa tem menor probabilidade de rotatividade. A acuraria dos modelos variaram entre 61% a 74% e o índice AUC entre 0.748 a 0.784.

Punnoose e Ajit (2016) estudaram a rotatividade de funcionários nos dados disponíveis nos sistemas de recursos humanos em uma equipe de liderança de uma varejista americana global, com objetivo de comparar a técnica *Extreme Gradient Boosting (XGBoost)* com os preditores Regressão Logística Binária, *Naïve Bayesian*, Floresta Aleatória, K-Vizinho mais Próximo (KNN), Análise Discriminante Linear (LDA) e Máquina de Vetores de Suporte (SVM). Considerando o indicador *AUC*, a técnica *XGBoost* apresentou melhor resultado nesse estudo, alcançando índice de 0.88 com dados de treino.

Em Ponnuru (2020), a rotatividade de funcionários em uma empresa de programação de computadores foi abordada, utilizando como estudo de caso informação fornecidas pela área de recursos humanos da IBM, com dados de 1470 funcionários. Foi utilizado a técnica de RLB, considerando algumas variáveis como idade, satisfação no trabalho, renda mensal e anos na empresa. O resultado de eficiência do modelo foi de 85% no índice de exatidão na previsão

de desgaste voluntário do funcionário, que resulta no desligamento e índice *AUC* de 0.552.

No âmbito do processo de estágio, em Veludo-De-Oliveira *et al.* (2013) um dos temas tratados no estudo foi a empregabilidade do estudante universitário por meio do estágio e teve como objetivo de pesquisa identificar quais fatores são determinantes da conquista de uma vaga de estágio por universitários. As variáveis, obtidas por meio de questionário aplicado com 302 candidatos no estado de São Paulo em um agente de integração de estágio, foram analisadas e aplicada técnica de regressão logística binária e regressão múltipla, prevendo se o estudante tinha ou não conseguido uma vaga de estágio. A variável satisfação com a instituição de ensino e o curso foram considerados determinante para conseguir a vaga.

Um estudo para prever a satisfação dos estagiários observando três fatores amplos, característica do trabalho, característica do ambiente e fatores contextuais foi realizado por D'Abate *et al.* (2009), que destacaram a importância do estágio para a formação dos estudantes e para as empresas. Após coletar dados de alunos de graduação do Departamento de Administração e Negócios em uma faculdade americana, aplicou-se análise de regressão logística multinomial, indicando as variáveis que influenciavam na satisfação dos estagiários, observando as características do trabalho (variedade de habilidades, identidade e significado da tarefa, autonomia e feedback), do ambiente do estágio (oportunidade de aprendizagem, apoio da supervisão e dos colegas de trabalhos e satisfação com a empresa) e de fatores contextuais (trabalho com horas flexíveis, deslocamento razoável, satisfação com o pagamento e localização desejável). Os resultados demonstraram que os melhores fatores preditores da satisfação do estágio foram o significado da tarefa, *feedback*, oportunidade de aprendizagem, suporte da supervisão e satisfação com a empresa.

Estudo conduzido por Seyitoğlu e Yirik (2015), na Akdeniz University Faculty of Tourism, localizada na Turquia, pesquisou em que grau a satisfação com o estágio e educação no turismo vocacional desses alunos preveem seu desenvolvimento profissional e percepção industrial. Utilizando como técnica a regressão múltipla, os resultados segundo os pesquisados realmente sugerem que estagiários satisfeitos tem impacto significativo no desenvolvimento

profissional e na percepção industrial, permitindo que ações sejam realizadas para que as empresas criem melhores condições no ambiente de estágio, trazendo mais estudantes nessa área, conseqüentemente possibilitando futuros profissionais no setor do turismo.

Em Hergert (2009) foi avaliado as percepções dos alunos sobre o valor dos estágios, e por meio da técnica de regressão multivariada, utilizando uma série de regressões *stepwise*, determinando quais os fatores são mais significativos e com mais impacto para os estudantes. As variáveis foram distribuídas em grupos de demografia (idade, gênero, ano escolar e rendimento no curso), formato/pedagogia (pagamento remunerado, horas trabalhadas e integração do estágio com o curso) e orientação profissional (proximidade com o emprego ideal, interesse nas atividades e a utilidade do estágio para encontrar um emprego) e obtidas com a aplicação de questionário com estudantes de administração de uma universidade americana. Os resultados encontrados foram que estudantes mais velhos, em maiores anos escolares e com alto rendimento escolar estão significativamente relacionados ao valor do estágio. Fatores relacionados a remuneração e ao desenvolvimento de carreira também são valorizados pelos estagiários.

De forma a organizar os estudos supracitados sobre a aplicação de técnicas de predição em gestão de pessoas, o QUADRO 2 sintetiza as referências, classificando-as conforme a área, tema e técnica empregada.

QUADRO 2 – RESUMO DOS ESTUDOS CONSIDERADOS

TEMA	TÉCNICA	TRABALHO
Rotatividade de motoristas em uma transportadora	Regressão Logística Binária	Nadler; Kros (2014)
Rotatividade de funcionários em uma agência de serviços humanos	Regressão Logística Binária Redes Neurais (Multi-Layered Perceptron)	Quinn et al. (2002)
Rotatividade de funcionários	Regressão Logística Binária	Rombaut; Guerry (2018)
Rotatividade de funcionários	Máquina de Vetores de Suporte (SVM) Floresta Aleatória Naïve Bayesian	Saradhi; Palshikar (2011)
Rotatividade de funcionários em Instituição Superior na Nigéria	Árvores de Decisão	D.Alao (2013)
Rotatividade de funcionários em um varejista global	Extreme Gradient Boosting (XGBoost) Regressão Logística Binária Naïve Bayesian Floresta Aleatória K-Vizinho maisPróximo (KNN) Análise Discriminante Linear (LDA) Máquina de Vetores de Suporte (SVM).	Punnoose; Ajit (2016)
Rotatividade funcionários em uma empresa de programação	Regressão Logística Binária	Ponnuru (2020)
Empregabilidade do candidato para vaga de estágio	Regressão Logística Binária Regressão Múltipla	Veludo-De-Oliveira et al. (2013)
Satisfação dos Estagiários	Regressão Logística Multinomial	D'Abate et al. (2009)
Satisfação dos Estagiários	Regressão Múltipla	Seyitoğlu; Yirik (2015)
Experiência de Estágio mais Valorosa	Regressão Multivariada	Hergert (2009)

FONTE: O autor (2022)

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.

As técnicas de RL e SVM foram aplicadas utilizando algoritmos implementados em linguagem R por meio do ambiente *Rstudio*.

3.1 COLETA, PREPARAÇÃO E DEFINIÇÃO DAS VARIÁVEIS

Os dados foram extraídos da base do AI, por meio de consultas utilizando a ferramenta SQL Server, exportando em formato “.csv”. Os dados utilizados na pesquisa foram compostos de registros das rescisões ocorridas no período de 01/01/2021 a 31/12/2021, abrangendo 24974 contratos. Foram consideradas algumas regras para obtenção dos dados, como utilizar apenas contratos de estágio de estudantes que estavam cursando ensino médio ou ensino superior, estágios da modalidade não obrigatórios e foram excluídas rescisões por motivos compulsórios, isto é, que o término de contrato não depende das variáveis disponibilizadas, como: transferência de curso ou instituição de ensino, mudança de cidade ou país, término de convênio entre AI e PCE, efetivação, entre outros.

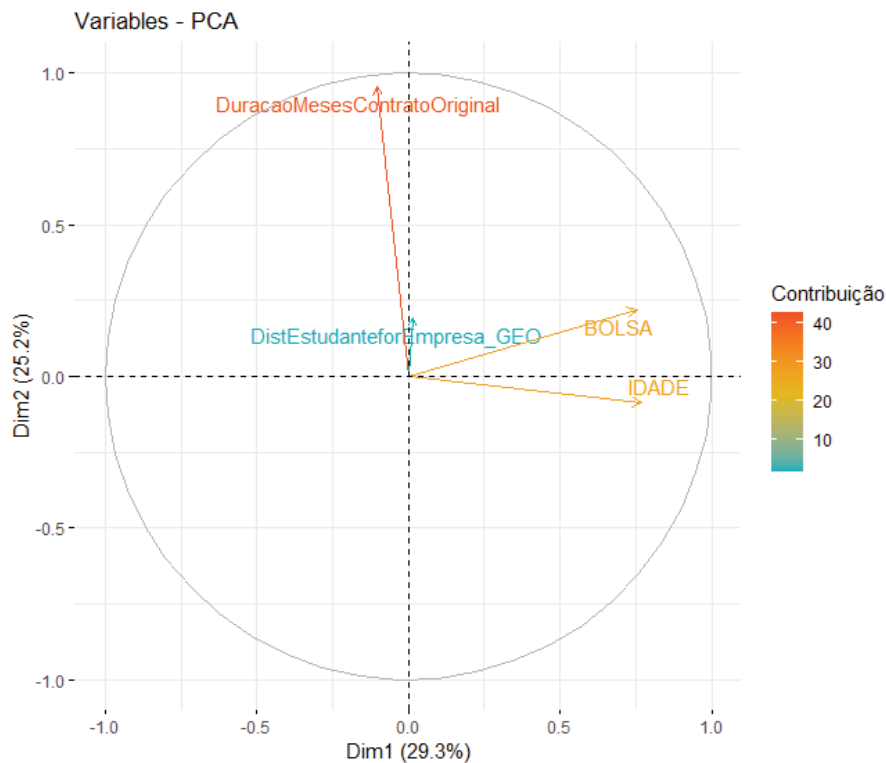
Dos 24974 contratos rescindidos, 15509 (62%) resultaram em rescisões antecipadas e 9467 (38%) contratos chegaram ao final da vigência do contrato.

Foi realizada a limpeza nos 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.

Outra ação efetivada antes da definição das variáveis utilizadas foi a aplicação da técnica exploratória de análise fatorial por meio do método componentes principais, que segundo Fávero e Belfiore (2017) permite, a partir de um conjunto de variáveis originais correlacionadas entre si, seja determinado outro conjunto de variáveis (fatores) resultantes da combinação linear do primeiro conjunto, possibilitando, dentro dos objetivos do método, a redução de variáveis que serão utilizadas na etapa de predição. Este método foi aplicado com as variáveis quantitativas (BOLSA, IDADE, DIST_RESIDENCIA_EMPRESA, DURACAO_CONTRATO_ORIGINAL), porém

o resultado apresentou ausência de multicolinearidade entre as variáveis (FIGURA 10), portanto não sendo necessário desconsiderar alguma delas na predição.

FIGURA 10- ANÁLISE DE COMPONENTES PRINCIPAIS - GRÁFICO DE VARIÁVEIS



FONTE: O autor (2022)

O QUADRO 3 descreve as variáveis utilizadas, onde 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 do AI e que fazem parte das informações obrigatórias no cadastro do estudante para compor o contrato de estágio.

QUADRO 3 - VARIÁVEIS UTILIZADAS

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
DIST_RESIDENCIA_EMPRESA	Independente – contínua - representando o valor da distância em metros da residência do estudante até o local de estágio.	
MES_INICIO	Independente – qualitativa – representando a mês de início do contrato	01 a 12
DATA_FINAL_ORIGINAL	Independente – qualitativa – representando a data final prevista do contrato	01/01/2021 a 31/12/2021
MES_RESCISAO	Independente – qualitativa – representando o mês efetivado da rescisão	01 a 12
DURACAO_CONTRATO_ORIGINAL	Independente – contínua – representando a quantidade de meses de duração prevista do estágio	01 a 24
PORTE_EMPRESA	Independente – qualitativa - representando o porte da empresa considerando o número de funcionários	MuitoPequena Pequena Média Grande MuitoGrande

FONTE: O autor (2022)

3.2 APLICAÇÕES DAS TÉCNICAS E RESULTADOS

Para aplicação da técnica, foram consideradas duas versões dos dados, diferindo pela inclusão ou não do mês de rescisão do contrato de estágio (MÊS_RESCISAO, ver QUADRO 3)

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 e 25% para testes, mantendo a mesma proporção de contratos antecipados dos dados originais, conforme TABELA 2.

TABELA 2 - DADOS DISTRIBUÍDOS - TREINO E TESTE

Dados Totais		Dados Treino		Dados Testes	
ANTECIPADO	FINAL	ANTECIPADO	FINAL	ANTECIPADO	FINAL
15509	9467	11677	7055	3832	2412

FONTE: O autor (2022)

3.2.1 Regressão logística

Para aplicação da técnica de RL, foi utilizado a linguagem R, no pacote *stats*, na família binomial (*link = 'logit'*).

3.2.1.1 Modelo sem mês de rescisão

Os principais resultados da aplicação do modelo de RL sem considerar a variável de mês rescisão estão apresentados nas tabelas extraídas do software *RStudio*. Na TABELA 3 são apresentados os resíduos de desvio (*Deviance Residuals*) e na TABELA 4 os coeficientes, obtido por meio do comando *summary*. A FIGURA 11 traz a curva ROC e em TABELA as matrizes de confusões dos resultados referentes aos dados de treino e testes. Para este modelo, após análise da curva ROC, definiu o *cutoff* em 0.35, ponto em que é maximizado as curvas de sensibilidade e especificidade. Executando o modelo com *cutoff* de 0.5, que tradicionalmente é padronizado, a acurácia se manteve equivalente, no valor de 0.7346 na base de treino e 0.7333 na base de testes. Também foi executado o modelo retirando as variáveis

(TURNO_AULA_ESTUDANTE, PORTE_CIDADE e DIST_RESIDENCIA_EMPRESA) com p -valor > 0.05 , que não são estatisticamente significativas para influenciar a probabilidade de ocorrência do evento, porém a acurácia se manteve equivalente, no valor de 0.7362 na base de treino e 0.7434 na base de testes.

A acurácia total do modelo aplicado aos dados de treinos e testes foram semelhantes, sendo 74% com treino e 75% com testes, representando a capacidade do modelo de prever satisfatoriamente este cenário de rescisão.

TABELA 3 - RESÍDUOS DE DESVIO PARA O MODELO SEM O MÊS DE RESCISÃO

Resíduos de Desvio (<i>Deviance Residuals</i>)				
<i>Min</i>	1Q	<i>Median</i>	3Q	<i>Max</i>
-2.8065	-0.7772	-0.4516	0.9282	2.7733

FONTE: O autor (2022)

TABELA 4 – COEFICIENTES PARA O MODELO SEM O MÊS DE RESCISÃO

(continua)

Coeficientes (<i>Coefficients</i>)				
	Estimate	Std Error	Z Value	Pr(> z)
<i>(Intercept)</i>	8.245e+01	1.788e+00	46.118	< 2e-16 ***
BOLSA	2.946e-04	6.932e-05	4.251	2.13e-05 ***
IDADE	1.759e-02	3.585e-03	4.907	9.26e-07 ***
TIPO_IE Publico	-8.889e-02	4.800e-02	-1.852	0.064038
TIPO_EMPRESA Publico	2.703e-01	4.401e-02	6.142	8.16e-10 ***
NIVEL_CURSO Superior	-5.916e-01	5.450e-02	-10.856	< 2e-16 ***
SEXO_ESTUDANTE Masculino	2.094e-01	3.817e-02	5.485	4.15e-08 ***
TURNO_AULA_ESTUDANTE Integral	1.402e-01	1.218e-01	1.151	0.249627
TURNO_AULA_ESTUDANTE Manhã	3.655e-01	6.313e-02	5.789	7.06e-09 ***
TURNO_AULA_ESTUDANTE Noite	3.079e-02	5.313e-02	0.579	0.562319
TURNO_AULA_ESTUDANTE Tarde	3.916e-01	1.190e-01	3.291	0.000997 ***
PORTE_CIDADE Medio	1.492e-03	5.123e-02	0.029	0.976762
PORTE_CIDADE Micro	9.444e-01	9.371e-02	10.078	< 2e-16 ***
PORTE_CIDADE Pequena	1.791e-01	6.644e-02	2.695	0.007033 **
PORTE_CIDADE Rural	7.628e-01	8.390e-02	9.091	< 2e-16 ***
DIST_RESIDENCIA_EMPRESA	2.734e-08	1.012e-07	0.270	0.787021
MES_INICIO	1.343e-02	5.335e-03	2.517	0.011850 *

TABELA 4 – COEFICIENTES PARA O MODELO SEM O MÊS DE RESCISÃO

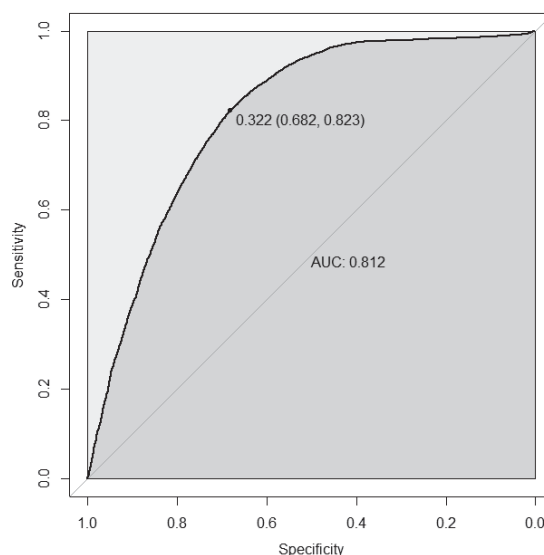
(conclusão)

Coefficientes (Coefficients)				
	Estimate	Std Error	Z Value	Pr(> z)
DATA_FINAL_ORIGINAL	-5.003e-08	1.081e-09	-46.301	< 2e-16 ***
DURACAO_CONTRATO_ORIGINAL	-1.836e-01	5.053e-03	-36.332	< 2e-16 ***
PORTE_EMPRESA Pequena	-4.698e-01	1.109e-01	-4.236	2.27e-05 ***
PORTE_EMPRESA Media	-1.560e-01	7.068e-02	-2.208	0.027257 *
PORTE_EMPRESA Grande	-1.334e-01	5.410e-02	-2.465	0.013702 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

FONTE: O autor (2022)

FIGURA 11 - RESULTADO DA CURVA ROC PARA O MODELO SEM O MÊS DE RESCISÃO



FONTE: O autor (2022)

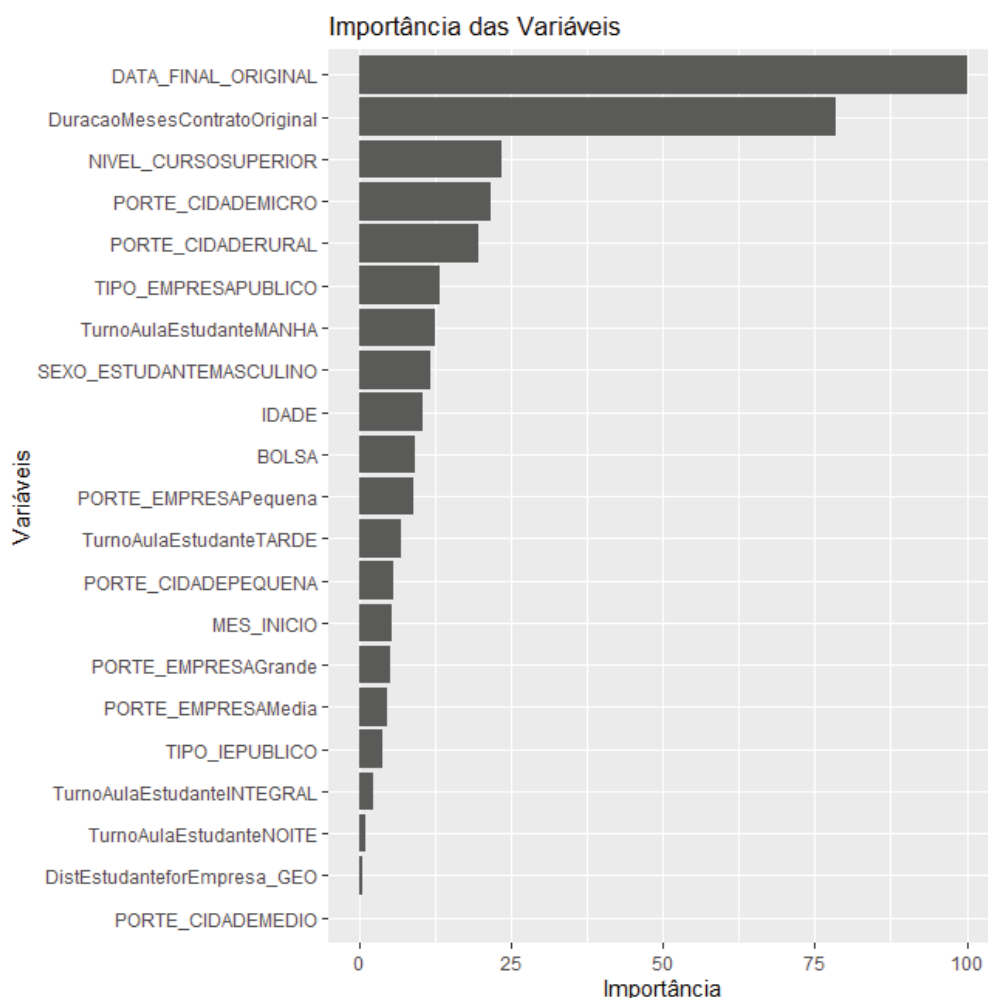
TABELA 5 - RESULTADO DA MATRIZ DE CONFUSÃO DOS DADOS DE TREINO E TESTE

Predição	Referência Treino		Referência Teste	
	ANTECIPADO	FINAL	ANTECIPADO	FINAL
ANTECIPADO	8288 (71%)	1516 (22%)	2758 (72%)	493 (20%)
FINAL	3389 (29%)	5539 (78%)	1074 (28%)	1919 (80%)
TOTAL	11677 (100%)	7055 (100%)	3832 (100%)	2412 (100%)
ACURÁCIA	0.7381		0.7490	
SENSIBILIDADE	0.7098		0.7197	
ESPECIFICIDADE	0.7851		0.7956	

FONTE: O autor (2022)

A FIGURA 12 representa a classificação da importância das variáveis desse modelo. Nota-se que as variáveis relacionadas ao período do contrato de estágio se destacam entre os fatores determinantes a predição antecipada.

FIGURA 12 - IMPORTÂNCIA DAS VARIÁVEIS - SEM MÊS DE RESCISÃO



FONTE: O autor (2022)

3.2.1.2 Modelo com o mês de rescisão

Com objetivo de conseguir um modelo com melhor ajuste, foi incluído variável mês de rescisão. Na TABELA 6 são apresentados os resíduos de desvio (*Deviance Residuals*) e na TABELA 7 os coeficientes, obtido por meio do comando *summary*. A FIGURA 13 traz a curva ROC e na TABELA 8 as matrizes de confusões dos resultados referentes aos dados de treino e testes. Para este modelo, após análise da curva ROC, definiu o *cutoff* em 0.35, ponto em que é maximizado as curvas de sensibilidade e especificidade.

Seguindo a dinâmica a aplicada sem a variável mês de rescisão, o modelo foi executado com *cutoff* de 0.5, que tradicionalmente é padronizado, a acuraria se manteve equivalente, no valor de 0.8156 na base de treino e 0.8155 na base de testes. Também foi executado o modelo retirando as variáveis (BOLSA, TIPO_IE, TIPO_EMPRESA e DIST_RESIDENCIA_EMPRESA) com *p-valor* > 0.05, que não são estatisticamente significativas para influenciar a probabilidade de ocorrência do evento, porém a acurácia se manteve equivalente, no valor de 0.814 na base de treino e 0.8193 na base de testes.

A acurácia total do modelo aplicado aos dados de treinos e testes foram semelhantes, com taxa de 82%, representando um ótimo resultado na capacidade do modelo de prever corretamente a rescisão antecipada.

TABELA 6 - RESÍDUOS DE DESVIO PARA O MODELO COM O MÊS DE RESCISÃO

Resíduos de Desvio (<i>Deviance Residuals</i>)				
<i>Min</i>	1Q	<i>Median</i>	3Q	<i>Max</i>
-3.6737	-0.5614	-0.2484	0.7222	2.5759

FONTE: O autor (2022)

TABELA 7 – COEFICIENTES PARA O MODELO COM O MÊS DE RESCISÃO

(continua)

Coeficientes (<i>Coefficients</i>)				
	<i>Estimate</i>	<i>Std Error</i>	<i>Z Value</i>	<i>Pr(> z)</i>
<i>(Intercept)</i>	1.692e+02	2.876e+00	58.822	< 2e-16 ***
BOLSA	6.873e-05	7.890e-05	0.871	0.38368
IDADE	1.759e-02	4.146e-03	4.243	2.21e-05 ***
TIPO_IE Publico	-5.891e-02	5.530e-02	-1.065	0.28672
TIPO_EMPRESA Publico	-7.812e-02	5.111e-02	-1.529	0.12638
NIVEL_CURSO Superior	-4.588e-01	6.229e-02	-7.365	1.77e-13 ***
SEXO_ESTUDANTE Masculino	2.391e-01	4.406e-02	5.428	5.70e-08 ***
TURNO_AULA_ESTUDANTE Integral	3.425e-01	1.300e-01	2.634	0.00845 **
TURNO_AULA_ESTUDANTE Manha	8.723e-02	7.277e-02	1.199	0.23060
TURNO_AULA_ESTUDANTE Noite	-1.178e-01	6.085e-02	-1.935	0.05296
TURNO_AULA_ESTUDANTE Tarde	2.879e-01	1.366e-01	2.108	0.03507 *
PORTE_CIDADE Medio	1.498e-02	5.885e-02	0.254	0.79913
PORTE_CIDADE Micro	9.453e-01	1.083e-01	8.728	< 2e-16 ***

TABELA 7 – COEFICIENTES PARA O MODELO COM O MÊS DE RESCISÃO

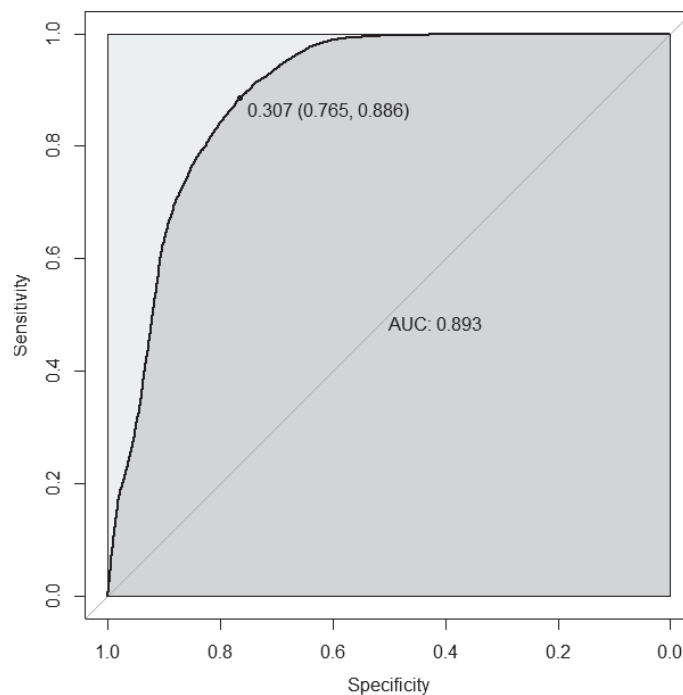
(conclusão)

Coeficientes (Coefficients)				
	<i>Estimate</i>	<i>Std Error</i>	<i>Z Value</i>	<i>Pr(> z)</i>
PORTE_CIDADE Pequena	8.957e-02	7.634e-02	1.173	0.24065
PORTE_CIDADE Rural	7.338e-01	9.632e-02	7.618	2.58e-14 ***
DIST_RESIDENCIA_EMPRESA	-5.698e-08	1.042e-07	-0.547	0.58446
MES_INICIO	-3.304e-02	6.323e-03	-5.226	1.73e-07 ***
DATA_FINAL_ORIGINAL	-1.056e-07	1.781e-09	-59.267	< 2e-16 ***
DURACAO_CONTRATO_ORIGINAL	-9.087e-02	5.336e-03	-17.031	< 2e-16 ***
PORTE_EMPRESA Pequena	-3.948e-01	1.266e-01	-3.118	0.00182 **
PORTE_EMPRESA Media	-1.860e-01	8.149e-02	-2.282	0.02248 *
PORTE_EMPRESA Grande	-1.581e-01	6.300e-02	-2.510	0.01207 *
MES_RESCISAO	4.561e-01	8.405e-03	54.261	< 2e-16 ***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

FONTE: O autor (2022)

FIGURA 13 - RESULTADO DA CURVA ROC PARA O MODELO COM O MÊS DE RESCISÃO



FONTE: O autor (2022)

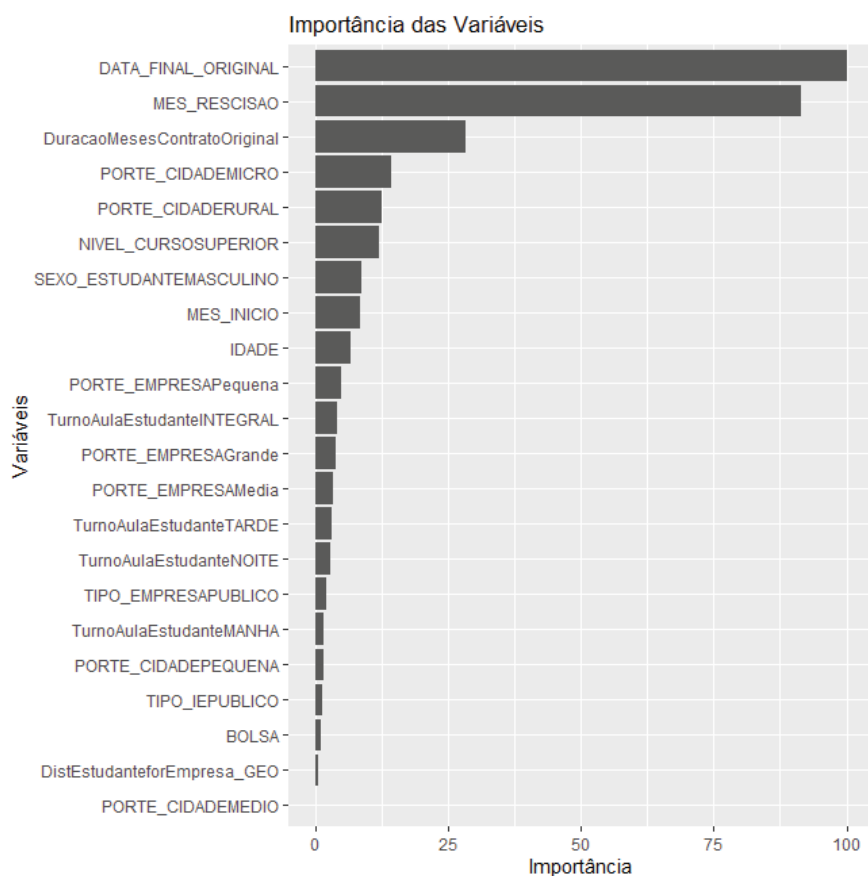
TABELA 8 - RESULTADO DA MATRIZ DE CONFUSÃO DOS DADOS DE TREINO E TESTE

Predição	Referência Treino		Referência Teste	
	ANTECIPADO	FINAL	ANTECIPADO	FINAL
ANTECIPADO	9217 (79%)	1022 (15%)	3055 (80%)	356 (15%)
FINAL	2460 (21%)	6033 (85%)	777 (20%)	2056 (85%)
TOTAL	11677 (100%)	7055 (100%)	3832 (100%)	2412 (100%)
ACURÁCIA	0.8141		0.8185	
SENSIBILIDADE	0.7893		0.7972	
ESPECIFICIDADE	0.8551		0.8524	

FONTE: O autor (2022)

A FIGURA 14 representa a classificação da importância das variáveis desse modelo. Nota-se que as variáveis relacionadas ao período do contrato de estágio se destacam entre os fatores determinantes a predição antecipada.

FIGURA 14- IMPORTÂNCIA DAS VARIÁVEIS - COM MÊS DE RESCISÃO



FONTE: O autor (2022)

3.2.2 Máquina de vetores de suporte

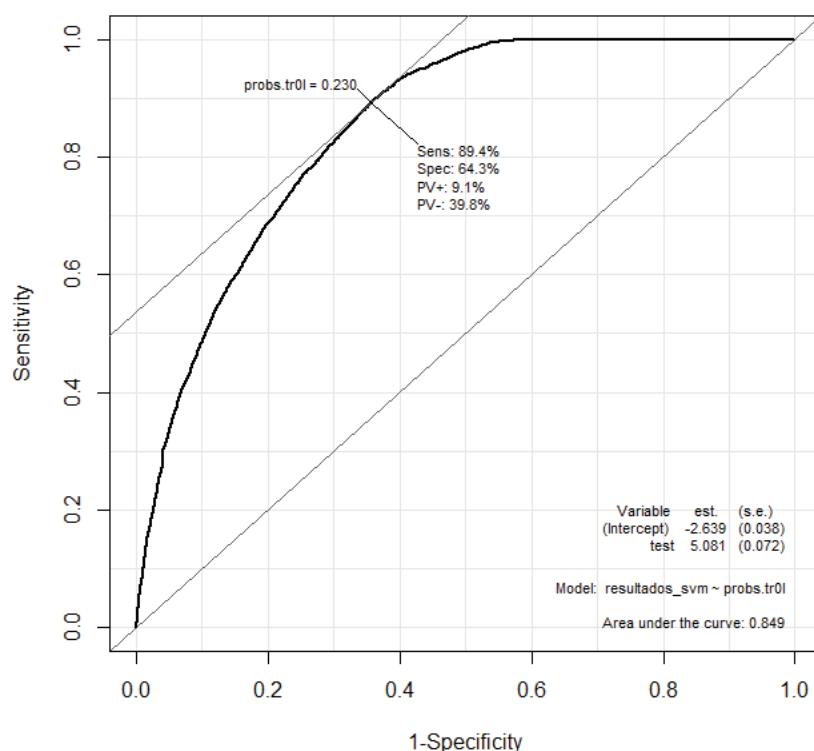
Outra ferramenta utilizada para desenvolver um modelo de predição de contratos de estágio é a Máquina de Vetores de Suporte (SVM). Este modelo também foi desenvolvido utilizando a linguagem R, por meio do pacote *e1071*, utilizando os parâmetros: *SVM-Type = C-classification*, *SVM-Kernel = radial* e *cost = 1*. Assim como utilizado com a Regressão Logística, foram analisados dois modelos, sem considerar o mês de rescisão e outra considerando esta variável.

3.2.2.1 Modelo sem mês de rescisão

Os principais resultados da aplicação do modelo de SVM sem considerar a variável de mês rescisão estão apresentados em FIGURA 15 que traz a curva ROC dos dados de treino e na TABELA 9 as matrizes de confusões dos resultados referentes aos dados de treino e testes.

O resultado da curva ROC foi 0.849 e a acurácia total do modelo aplicado aos dados de treinos e testes foram semelhantes, sendo 75% para treino e testes, representando a capacidade do modelo de prever satisfatoriamente este cenário de rescisão.

FIGURA 15 - RESULTADO CURVA ROC – DADOS DE TREINO



FONTE: O autor (2022)

TABELA 9 - RESULTADO DA MATRIZ DE CONFUSÃO DOS DADOS DE TREINO

Predição	Referência Treino		Referência Teste	
	ANTECIPADO	FINAL	ANTECIPADO	FINAL
ANTECIPADO	9184 (79%)	2068 (29%)	3043 (79%)	712 (30%)
FINAL	2493 (21%)	4987 (71%)	789 (21%)	1700 (70%)
TOTAL	11677 (100%)	7055 (100%)	3832 (100%)	2412 (100%)
ACURÁCIA	0.7565		0.7596	
SENSIBILIDADE	0.7865		0.7941	
ESPECIFICIDADE	0.7069		0.7048	

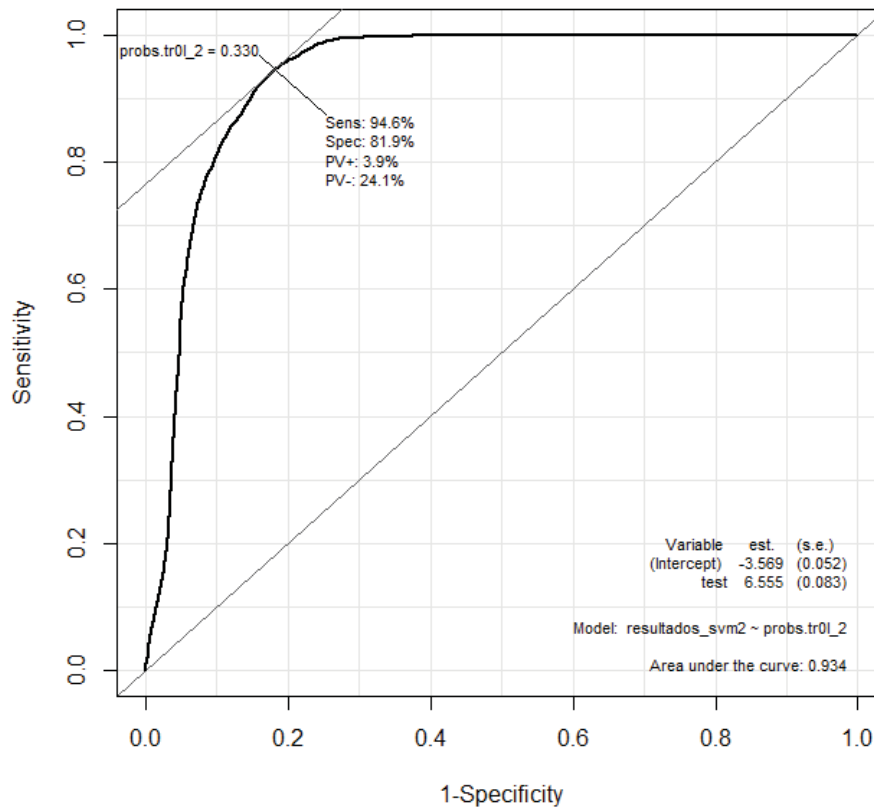
FONTE: O autor (2022)

3.2.2.2 Modelo com o mês de rescisão

Assim como na aplicação da RL, com objetivo de conseguir um modelo com melhor ajuste, foi incluído variável mês de rescisão. A FIGURA 16 traz a curva *ROC* e na TABELA 10 as matrizes de confusões dos resultados referentes aos dados de treino e testes.

A acurácia total do modelo aplicado aos dados de treinos e testes foram semelhantes, com taxa de 86%, representando um ótimo resultado na capacidade do modelo de prever corretamente a rescisão antecipada.

FIGURA 16- RESULTADO CURVA ROC



FONTE: O autor (2022)

TABELA 10 - RESULTADO DA MATRIZ DE CONFUSÃO DOS DADOS DE TREINO

Predição	Referência Treino		Referência Teste	
	ANTECIPADO	FINAL	ANTECIPADO	FINAL
ANTECIPADO	10060 (86%)	825 (12%)	3314 (86%)	316 (13%)
FINAL	1617 (14%)	6230 (88%)	518 (14%)	2096 (87%)
TOTAL	11677 (100%)	7055 (100%)	3832 (100%)	2412 (100%)
ACURÁCIA	0.8696		0.86664	
SENSIBILIDADE	0.8615		0.8648	
ESPECIFICIDADE	0.8831		0.8690	

FONTE: O autor (2022)

3.3 INTERPRETAÇÃO DA PESQUISA

Comparando as duas técnicas utilizadas, o SVM apresentou índices de acurácia ligeiramente superiores em relação à RL. Dessa forma, entende-se que o SVM auferiu maior assertividade na predição das rescisões antecipadas. Nos casos considerando a data de rescisão como variável de entrada, o modelo tem a capacidade de prever corretamente 87% estas rescisões. Conforme apresentado nas referências teóricas desse trabalho, a acurácia é calculada por meio dos números de Verdadeiros Positivos e Verdadeiros Negativos, isto é, são os acertos que o modelo consegue prever. Para chegar ao resultado de 87%, o modelo consegue prever 86% das rescisões que serão antecipadas e 87% das rescisões que chegarão ao final do contrato. O restante das predições (13%) representa os casos na qual o modelo errou na predição. A TABELA 11 apresenta o resumo dos resultados dos índices de acurácia e acertos dos modelos, com os dados de testes.

TABELA 11 - RESUMO DOS RESULTADOS

Técnica	Sem variável data de rescisão			Com variável data de rescisão		
	Acurácia	VP	VN	Acurácia	VP	VN
RL	0.7490	72%	80%	0.8185	80%	85%
SVM	0.7596	79%	70%	0.8690	86%	87%

FONTE: O autor (2022)

Considerando que não se encontrou na literatura científica estudos relacionados à predição de contratos de estágio, conduzimos uma análise comparativa com modelos de predição de *turnover* de funcionários. Nesse cenário, os índices de acertos do modelo desse trabalho estão condizentes com os resultados obtidos nos trabalhos descritos na seção de Trabalhos Correlatos.

Outro ponto relevante a ser considerado são as variáveis que mais influenciam nos modelos de predição, demonstrando que alterações dessas variáveis são determinantes no processo de rescisão de contratos de estágio. Durante a aplicação das técnicas, especificamente no modelo sem a variável de entrada mês de rescisão, as variáveis relacionadas ao período do estágio, como data final do contrato e duração do estágio, foram determinantes para predição da rescisão. O nível do curso, porte da cidade, tipo de empresa, turno de aula,

sexo, idade e bolsa são as mais importantes, nesta sequência, para a predição do modelo. A comparação das variáveis mais importantes no modelo com outros estudos na literatura científica de *turnover* é prejudicada, pois o contrato de estágio tem como uma de suas características a duração pré-definida, o que não ocorre com as contratações de funcionários efetivos das empresas.

Quando incluímos a variável mês de rescisão nos modelos, os resultados têm significativa melhoria nos índices de acertos, fortalecendo a relevância das variáveis relacionadas ao tempo do contrato nas predições.

Considerando estes índices de acertos dos modelos e a importância das variáveis, é possível o direcionamento de ações proativas que possam diminuir a probabilidade de ocorrência das rescisões antecipadas ou ainda se preparar operacionalmente e financeiramente caso a rescisão não possa ser evitada.

Uma das ações proativas possíveis para a mitigação do risco de rescisão antecipada é analisar com mais profundidade as variáveis período de duração dos contratos de estágio e a data final. Como essas variáveis são geralmente definidos pelas IEs no momento da emissão do TCE, pode-se rever as definições de períodos e datas em conjunto com as IEs. Outra ação possível a ser realizada analisando os resultados dessa pesquisa é aplicar novos modelos com dados específicos para cada empresa contratante e observar quais as características das variáveis influenciam na rescisão antecipada, considerando as variáveis possíveis de serem previstas no momento do recrutamento e seleção dos candidatos, como valor de bolsa auxílio, turno do estudante, entre outras. Conhecendo estas informações antecipadas é possível diminuir a probabilidade de rescisão.

Nos casos que não são possíveis ações proativas que possam diminuir a probabilidade de rescisão antecipada, os resultados dessa pesquisa permitem que as partes interessadas se preparem com antecedência para a rescisão, operacionalmente e financeiramente. No âmbito operacional, a PCE, com informação de uma provável rescisão, tem condições de realizar estratégias de reposição desse estagiário ou ainda analisar previamente uma possível efetivação, valorizando o bom desempenho caso tenha ocorrido durante o período do estágio. Como estratégias de reposição, é possível antecipar o processo de recrutamento e seleção do novo estagiário e administrar o tempo final do estágio para repasse de informações entre a equipe ou adequações das

atividades no setor que este estagiário está envolvido. Esta previsibilidade permite uma gestão dos recursos humanos mais eficiente. A previsibilidade desta reposição também é importante para o AI, que poderá ser proativo no relacionamento com a PCE, antecipando os trâmites de nova abertura de vagas, agilizando o processo de divulgação para novos estudantes e emissão de novo contrato de reposição. Para as IEs que acompanham ativamente o contrato de estágio de seus alunos, não sendo possível evitar a rescisão antecipada, é interessante realizar estudos dos impactos do estudante no rendimento acadêmico e estratégias de recolocação.

No âmbito financeiro, compreender os motivos de rescisão antecipada e ter capacidade de predição nesse cenário, possibilita ao AI previsibilidade na arrecadação das taxas administrativas que são pagas pelas PCEs durante a administração dos contratos de estágio que estão vigentes. Para as PCEs, é possível prever até quando as bolsas serão pagas aos estagiários e as taxas pagas ao AI, bem como prever os encargos decorrentes do encerramento do contrato. Conforme alertado por Abbassi e Hollman (2000), alta taxa de *turnover* ocasiona altos custos para a empresa, portanto, manter o estagiário até o final do contrato previsto tem impactos financeiros positivos para as PCEs. É possível considerar que a bolsa financeira recebida pelos estagiários pode ter papel importante em custear as despesas nesse período de vínculos com as IEs, como locomoção, alimentação e também nas mensalidades escolares, sendo que o término antecipado do contrato de estágio poderá acarretar dificuldades de os estudantes continuar estudando, resultando na evasão escolar.

O impacto direto que a rescisão antecipada para o estudante não foi foco dessa pesquisa, porém o estudante, como papel central no processo de estágio, pode-se beneficiar caso as oportunidades de melhorias citadas acima se concretizarem.

O QUADRO 4 fornece a síntese das oportunidades de melhorias para os problemas identificados relacionados a cada parte interessada do processo de estágio.

QUADRO 4- SÍNTESE DAS AÇÕES

PARTE INTERESSADA	PROBLEMA IDENTIFICADO	OPORTUNIDADES DE MELHORIAS
AI / IEs	Influência da duração e datas dos contratos	Revisão das estratégias e prazos
AI / PCEs	Influência das variáveis	Avaliação no momento do recrutamento e seleção
PCEs	Âmbito operacional - provável rescisão	Estratégias de reposição – antecipar processo de recrutamento e seleção Estratégias de reposição – repasse de informações do estagiário para a equipe Efetivação
AI	Âmbito operacional - provável rescisão	Antecipação dos trâmites para novas vagas Agilidade na divulgação das vagas
IEs	Âmbito operacional - provável rescisão	Impactos no rendimento acadêmico.
PCEs	Âmbito financeiro – compreendendo os motivos de rescisão	Prever encargos decorrentes do encerramento do contrato Alto <i>turnover</i> – alto custos envolvidos
AI	Âmbito financeiro – compreendendo os motivos de rescisão	Previsibilidade na arrecadação das taxas administrativas
IEs	Âmbito financeiro – compreendendo os motivos de rescisão	Evasão escolar
AI / IEs / PCEs	Poucas variáveis explicativas	Estudos de novas variáveis para identificar rescisão antecipada

FONTE: O autor (2022)

Os resultados dessa pesquisa corroboram com Walford-Wright e Scott-Jackson (2018) que destacam a importância de *People Analytics* no gerenciamento de pessoas por meio da análise de dados por métodos científicos.

4 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. Ainda que a literatura reporte estudos afetos à Gestão de Pessoas, particularmente no âmbito dos fatores determinantes ao *turnover* de funcionários, 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*.

Com relação aos modelos avaliados, os resultados mostram uma acuraria geral de 75% utilizando a técnica de RL e 76% com a técnica SVM nos dados de testes, desconsiderando o mês de rescisão. Estes números foram obtidos mesmo considerando as características particulares do programa de estágio, como a facilidade de rescisão e variáveis envolvendo as partes: IE, PCE e estudante. Quando o mês de rescisão é incorporado ao modelo de predição, os resultados mostram uma acuraria geral de 82% com a RL e 86% com o SVM nos dados de testes. Nesse cenário, apesar de não ser possível no início estágio prever se haverá ou não rescisão antecipada, pode-se analisar mês a mês a influência das variáveis e o perfil do estágio caso a rescisão ocorra em determinado período.

Os principais fatores determinantes foram as variáveis relacionadas ao período do estágio, como data de término, duração e data de rescisão. Outras variáveis também são relevantes, como nível do curso do estudante porte da cidade onde o estágio é realizado.

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.

Um grande legado dessa pesquisa para este AI é o conhecimento gerado na compreensão dos dados e das técnicas estatísticas, com possibilidades de aplicar a predição em outros processos igualmente relevantes. Também contribui para a comunidade científica na aplicação de modelos de predição em Gestão de Pessoas.

Com relação a direcionamentos futuros, acredita-se que os desempenhos das técnicas possam ser aprimorados 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, carga de atividades no estágio, satisfação na empresa, gestão do supervisor, entre outras. 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. Também é possível aprimorar o desempenho dessas técnicas utilizando aprimoramentos estatísticos disponíveis na linguagem R. Vislumbra-se ainda a avaliação comparativa de outras técnicas de predição.

REFERÊNCIAS

- ABBASSI, S.; HOLLMAN, K. TURNOVER The Real Bottom Line Consequences of Turnover. **Public Personnel Management**, v. 29, n. 3, p. 333–342, 2000.
- ANGRAVE, D.; CHARLWOOD, A.; KIRKPATRICK, I.; LAWRENCE, M.; STUART, M. HR and analytics: why HR is set to fail the big data challenge. **Human Resource Management Journal**, v. 26, n. 1, p. 1–11, 2016. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/1748-8583.12090>>.
- BRASIL. Lei nº 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/l11788.htm. Acesso em: 07 fev. 2022.
- CRISTIANINI, N.; SHAWE-TAYLOR, J. **An Introduction to Support Vector Machines and Other Kernel-based Learning Methods**. Cambridge University Press, 2000.
- D'ABATE, C.; YOUNDT, M.; WENZEL, K. Making the most of an internship: An empirical study of internship satisfaction. **Academy of Management Learning and Education**, v. 8, n. 4, p. 527–539, 2009.
- D.ALAO, A. B. A. Analyzing employee attrition using decision tree algorithms. **Information Systems & Development Informatics**, v. 4, n. 1, p. 17–28, 2013.
- ESTATÍSTICAS, ABRES. Disponível em <https://abres.org.br/estatisticas>. Acesso em: 10 fev. 2021.
- FALLUCCHI, F.; COLADANGELO, M.; GIULIANO, R.; DE LUCA, E. W. Predicting employee attrition using machine learning techniques. **Computers**, v. 9, n. 4, p. 1–17, 2020.
- FÁVERO, L. P.; BELFIORE, P. Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. , , n. 2, 2017.
- FITZ-ENZ, J.; MATTOX, J. R. (ORGS.). **Predictive Analytics for Human Resources**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2014.
- GAO, X.; WEN, J.; ZHANG, C. An Improved Random Forest Algorithm for Predicting Employee Turnover. **Mathematical Problems in Engineering**, v. 2019, 2019.
- HAIR, J. F.; BLACK, W. C.; BADIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. 6º ed. 2009.
- HAMILTON, R. H.; SODEMAN, W. A. The questions we ask: Opportunities and challenges for using big data analytics to strategically manage human capital

resources. **Business Horizons**, v. 63, n. 1, p. 85–95, 2020. Elsevier Ltd. Disponível em: <<https://doi.org/10.1016/j.bushor.2019.10.001>>.

HERGERT, M. Student Perceptions Of The Value Of Internships. **Business Education**, v. 2, n. 8, p. 9–14, 2009.

FITZ-ENZ, J. The new HR analytics: predicting the economic value of your company's human capital investments. **American Management Association**, 2010.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning**. New York, NY: Springer New York, 2013.

HUSELID, M. A. The science and practice of workforce analytics: Introduction to the HRM special issue. **Human Resource Management**, v. 57, n. 3, p. 679–684, 2018. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/hrm.21916>>.

LARSSON, A.-S.; EDWARDS, M. R. Insider econometrics meets people analytics and strategic human resource management. **The International Journal of Human Resource Management**, v. 33, n. 12, p. 2373–2419, 2022. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/09585192.2020.1847166>>.

LORENA, A. C.; DE CARVALHO, A. C. P. L. F. Uma Introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007.

MA, Y.; GUO, G. **Support vector machines applications**. 2014.

MARLER, J. H.; BOUDREAU, J. W. An evidence-based review of HR Analytics. **The International Journal of Human Resource Management**, v. 28, n. 1, p. 3–26, 2017. Routledge. Disponível em: <<http://dx.doi.org/10.1080/09585192.2016.1244699>>.

MINBAEVA, D. B. Building credible human capital analytics for organizational competitive advantage. **Human Resource Management**, v. 57, n. 3, p. 701–713, 2018.

NADLER, S. S.; KROS, J. F. Logistic Regression Approach to Predicting Truck Driver Turnover. **International Journal of Applied Logistics**, v. 5, n. 1, p. 15–32, 2014.

OLIVEIRA, S. R. DE; PICCININI, V. C.; SILVEIRA, C. D. S. O trabalho como representação: a visão dos jovens universitários. **Revista de Ciências da Administração**, v. 12, n. 28, p. 171–197, 2010.

ONGORI, H. A review of the literature on employee turnover. **African Journal of Business Management**, Vol. 1, No. 3, pp. 049-054, 2007

PENG, C. Y. J.; LEE, K. L.; INGERSOLL, G. M. An introduction to logistic regression analysis and reporting. **Journal of Educational Research**, v. 96, n. 1, p. 3–14, 2002.

PONNURU, S. R. Employee Attrition Prediction using Logistic Regression. **International Journal for Research in Applied Science and Engineering Technology**, v. 8, n. 5, p. 2871–2875, 2020.

PONTIL, M.; VERRI, A. Properties of Support Vector Machines. **Neural Computation**, v. 10, n. 4, p. 955–974, 1998.

PUNNOOSE, R.; AJIT, P. Prediction of Employee Turnover in Organizations using Machine Learning Algorithms. **International Journal of Advanced Research in Artificial Intelligence**, v. 5, n. 9, p. 22–26, 2016.

QAMAR, Y.; SAMAD, T. A. Human resource analytics: a review and bibliometric analysis. **Personnel Review**, v. 51, n. 1, p. 251–283, 2022. Disponível em: <<https://www.emerald.com/insight/content/doi/10.1108/PR-04-2020-0247/full/html>>.

QUINN, A.; RYCRAFT, J. R.; SCHOECH, D. Building a Model to Predict Caseworker and Supervisor Turnover Using a Neural Network and Logistic Regression. **Journal of Technology in Human Services**, v. 19, n. 4, p. 65–85, 2002.

ROMBAUT, E.; GUERRY, M.-A. Predicting voluntary turnover through human resources database analysis. **Management Research Review**, v. 41, n. 1, p. 96–112, 2018.

SARADHI, V. V.; PALSHIKAR, G. K. Employee churn prediction. **Expert Systems with Applications**, v. 38, n. 3, p. 1999–2006, 2011.

SEYITOĞLU, F.; YIRIK, S. Internship Satisfaction of Students of Hospitality and Impact of Internship on the Professional Development and Industrial Perception. **Asia Pacific Journal of Tourism Research**, v. 20, n. April 2015, p. 1414–1429, 2015.

STEINWART, I.; CHRISTMANN, A. Support Vector Machines. 1st ed. New York, NY: **Springer**, New York, NY, 2008.

SHRIVASTAVA, S.; NAGDEV, K.; RAJESH, A. Redefining HR using people analytics: the case of Google. **Human Resource Management International Digest**, v. 26, n. 2, p. 3–6, 2018.

TURSUNBAYEVA, A.; DI LAURO, S.; PAGLIARI, C. People analytics—A scoping review of conceptual boundaries and value propositions. **International Journal of Information Management**, v. 43, p. 224–247, 2018. Elsevier. Disponível em: <<https://doi.org/10.1016/j.ijinfomgt.2018.08.002>>.

VAPNIK, V. N. *The Nature of Statistical Learning Theory*. New York, NY: **Springer**, New York, 1995.

VELUDO-DE-OLIVEIRA, T. M.; BALLESTERO-ALVAREZ, M. E.; OLIVEIRA, M. J.; VIDIGAL, P. R. Empregabilidade e seus antecedentes para conquista da vaga de estágio por universitários. **Revista Brasileira de Orientação Profissional**, v. 14, n. 1, p. 47–59, 2013.

WEST, M. **People Analytics para Leigos**. Alta Books, 2020.