



Aprendizado de Máquina na Gestão Estratégica de Pessoas. Aplicando Modelo Preditivo de Turnover

Vinicius de Oliveira Boen,¹ Luis Gustavo Nonato²
ICMC-USP

1 Introdução

Gerenciar a taxa de turnover é fundamental para a saúde e o sucesso de qualquer empresa a longo prazo. Pesquisadores organizacionais mostraram que o turnover voluntário de funcionários qualificados trás uma grande perda para a empresa, começando com todo o valor substancial que é perdido, isso também afeta os resultados relacionados à produtividade e reduz o desempenho financeiro, impactando diretamente na vantagem competitiva. Inclusive, se a causa não for identificada a tempo, pode evoluir para um turnover coletivo causando ainda mais danos [1].

Estudos apontam vários fatores que podem afetar direta ou indiretamente o turnover. Entre eles, relatos dos resignados costumam julgar negativamente tanto líderes, como, autoritários e imprudentes), quanto as condições do ambiente de trabalho, como pagamento, trabalho por turnos, avaliações de desempenho e capacidade subutilizada. Existem também as influências não profissionais como conflito trabalho-família, quando os funcionários optam por resignação para cuidar dos filhos ou quando pedem demissão por melhores benefícios corporativos. Por outro lado, empresas conseguiram reduzir a taxa de turnover após levantar um programa que aumentava a satisfação no trabalho e com isso, elaborar uma estrutura de como a insatisfação evolui para turnover [2].

A favor de minimizar essa perda, algumas organizações têm conduzido pesquisas e técnicas de mineração de dados para avaliar o nível de satisfação e indicar potenciais razões por trás da insatisfação dos funcionários [3], auxiliando gestores e responsáveis na identificação da motivação certa de cada funcionário, possibilitando a elaboração de medidas para o planejamento de retenção de funcionários qualificados e de sucessão, aumentando a produtividade e a lucratividade da organização.

Diante do contexto da importância do controle da taxa de turnover nas empresas, este trabalho tem como objetivo aplicar o uso de técnicas de mineração de dados com aprendizado de máquina para o estudo de como o envolvimento de tecnologias pode auxiliar na tomada de decisão, através da coleta e manipulação dos dados disponíveis nos sistemas de RH de uma empresa.

¹boen@usp.br

²gnonato@icmc.usp.br

2 Material e Metodologia

2.1 Mineração de Dados

O conjunto de dados utilizado no trabalho foi extraído via interface de programação (API) diretamente da plataforma de RH. Em seguida, através da utilização de técnicas de mineração de dados foram corrigidos problemas considerados comuns em aplicações reais de gestão de pessoas que são a grande quantidade de dados ausentes e os dados inconsistentes. Para a adequação do conjunto aos algoritmos de classificação, também foi necessário remover informações contendo muita variação textual, e remover informações muito semelhantes, ou correlacionadas entre si. Ao final desse processo o conjunto de dados ficou com 4517 amostras, 3016 (66%) pertencem a classe de funcionários ativos e 1501 (34%) pertencem a classe de funcionários desligados voluntariamente e um total de 17 atributos sendo Turnover o atributo de interesse.

2.2 Pré Processamento

Como pré processamento foram aplicadas as seguintes técnicas: *One Hot Encoding* do atributo Gênero do Funcionário, Normalização dos atributos pelo método *z-score*, Re-amostragem dos dados com *under-sampling* e *over-sampling* devido ao desbalanceamento do conjunto de dados e por fim, o Particionamento do conjunto para Treino, Teste e Validação.

2.3 Aprendizado Supervisionado

Os classificadores utilizados como função para a predição desse evento, foram: *Naïve Bayes*, *Regressão Logística*, *Decision Tree*, *Random Forest*. E foram avaliados através das medidas clássicas de desempenho calculadas a partir da matriz de confusão, onde no caso de classificação binária, possui duas colunas indicando a quantidade de amostras preditas para cada classe e duas linhas indicando a quantidade de amostras verdadeiras para cada classe. Partindo dessa matriz tem-se as medidas de avaliação usadas nesse trabalho: Acurácia, Precisão, Sensibilidade, Especificidade, F1-Score, Curva ROC, Área da Curva ROC [5].

3 Resultados

A figura a 1 ilustra os resultados de acurácia de todos os classificadores através de gráficos boxplots separados por treino e validação.

Após concluído o treino das técnicas propostas, utilizou-se os 30% restantes das amostras para avaliação dos classificadores. Os resultados estão descritos na tabela 1, onde ao analisá-la, é possível concluir que o classificador de melhor desempenho foi o *Random Forest* e em segundo lugar a *Regressão Logística*. Porém para as análises na etapa de pós processamento foi escolhido o classificador *Regressão logística* por ter regras mais simples de interpretação.

3.1 Pós Processamento

Com os resultados das predições dos classificadores chegando ao usuário da ponta através da plataforma de *people analytics*, o interesse passa a ser, não nos casos onde o classificador acertou

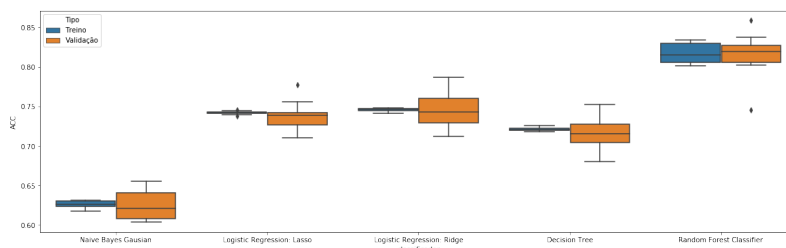


Figura 1: Boxplot das validações cruzadas dos classificadores

Classificador	Acurácia	Precisão	Sensibilidade	F1-Score	AUC
Naive Nayes	0.60	0.64	0.62	0.59	0.76
Regressão Logística	0.72	0.72	0.72	0.72	0.81
Decision Tree	0.68	0.69	0.70	0.69	0.79
Random Forest	0.82	0.82	0.82	0.82	0.90

Tabela 1: Resultado dos Classificadores no Conjunto de Teste

que o funcionário saiu, mas sim, nos casos onde o classificador deu uma probabilidade alta de turnover para um funcionário que ainda está ativo. Assim, os responsáveis por aquele funcionário podem analisar os atributos e trabalhar em planos de ação para tentar reverter o possível turnover. Adicionalmente, outra análise de suma importância é a relação entre probabilidade do turnover e tempo até o turnover, de fato.

Para isso, após 17 meses da extração do primeiro conjunto de dados, extraiu-se novamente as informações da aplicação de RH, porém agora foram filtrados os funcionários que foram classificados com probabilidade de turnover maior que 50%, a tabela 2 resume essa relação.

Probabilidade de Turnover	Situação Atual	Quantidade	%
50 a 60%	A / D / T	43 / 23 / 19	52% / 26% / 21%
61 a 70%	A / D / T	25 / 16 / 12	47% / 30% / 22%
71 a 80%	A / D / T	18 / 2 / 9	62% / 7% / 31%
81 a 90%	A / D / T	2 / 2 / 3	28% / 28% / 42%
91 a 100%	A / D / T	0 / 1 / 1	0% / 50% / 50%

Tabela 2: Situação atual dos funcionários para cada faixa de probabilidade de turnover. A = Ativo, D = Demitido, T = Turnover

Percebe-se que a proporção de turnover em cada faixa aumenta com a probabilidade, inclusive, a relação do tempo até o turnover ser confirmado pode ser observado na figura 2, para os 44 casos onde o turnover foi confirmado.

Os pontos próximos do dia zero (eixo y) representam os casos em que o turnover ocorreu logo depois da Extração das informações. Além disso, a reta mostra que existe uma leve tendência de casos com maior probabilidade de turnover demorem menos para serem confirmados, isso também pode ser interpretado através da correlação entre essas duas variáveis que foi de -0.17 .

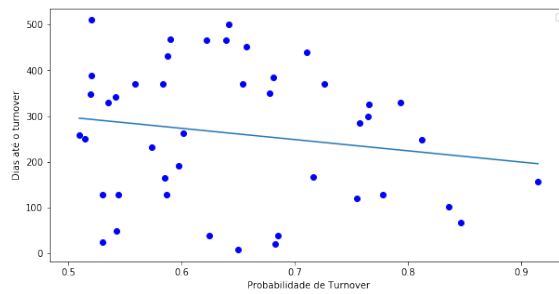


Figura 2: Scatter plot da probabilidade de turnover com dias até o turnover.

4 Conclusões

O principal desafio deste trabalho foi o de aplicar técnicas de mineração de dados nas informações da aplicação de RH. Durante esse processo foi possível concluir o quão importante é o domínio da área de ciência de dados e o domínio das regras de negócio da empresa, regras essas que podem impactar em como deve ser o tratamento dos dados e que podem direcionar as hipóteses de exploração dos mesmos.

O classificador escolhido no pós processamento obteve resultados interessantes e que se adéqua as regras de negócio da empresa, além disso, relação de probabilidade de turnover com tempo até o turnover também são análises que confirmam a relação e a importância de se tomar as decisões baseadas em dados, mesmo que os usuários finais (líderes e gestores) muitas vezes não possuem o conhecimento analítico para interpretação dos classificadores.

Referências

- [1] Hausknecht, John P., Trevor, Charlie O., Howard, Michael J. Causes and consequences of collective turnover: A meta-analytic review, *Journal of Applied Psychology*, 2013. DOI: <http://dx.doi.org/10.1037/a0032380>.
- [2] Hom, Peter W and Lee, Thomas W and Shaw, Jason D and Hausknecht, John P. One hundred years of employee turnover theory and research. *Journal of applied psychology*. American Psychological Association. 2017.
- [3] Rohit Punnoose and Pankaj Ajit. Prediction of Employee Turnover in Organizations using Machine Learning Algorithms. A case for Extreme Gradient Boosting. *International Journal of Advanced Research in Artificial Intelligence*. 2016,
- [4] Jaynes, Edwin T. Probability theory: The logic of science. *Cambridge university press*. 2003.
- [5] Rodrigo Fernandes de Mello and Moacir Antonelli Pontintor. Machine Learning: A Practical Approach on the Statistical Learning. *Springer International Publishing AG, part of Springer Nature*. 2003, 2020.