

## Estudos de modelos de predição de cancelamentos de serviços de telefonia utilizando técnicas de aprendizado de máquina

**Aluno:** João Victor Serralha, Guilherme Martins ([joaoserralha@hotmail.com](mailto:joaoserralha@hotmail.com), [guilherme.mg.oliveira@outlook.com](mailto:guilherme.mg.oliveira@outlook.com))

**Orientador:** Flavio Tonidandel ([flaviot@fei.edu.br](mailto:flaviot@fei.edu.br))

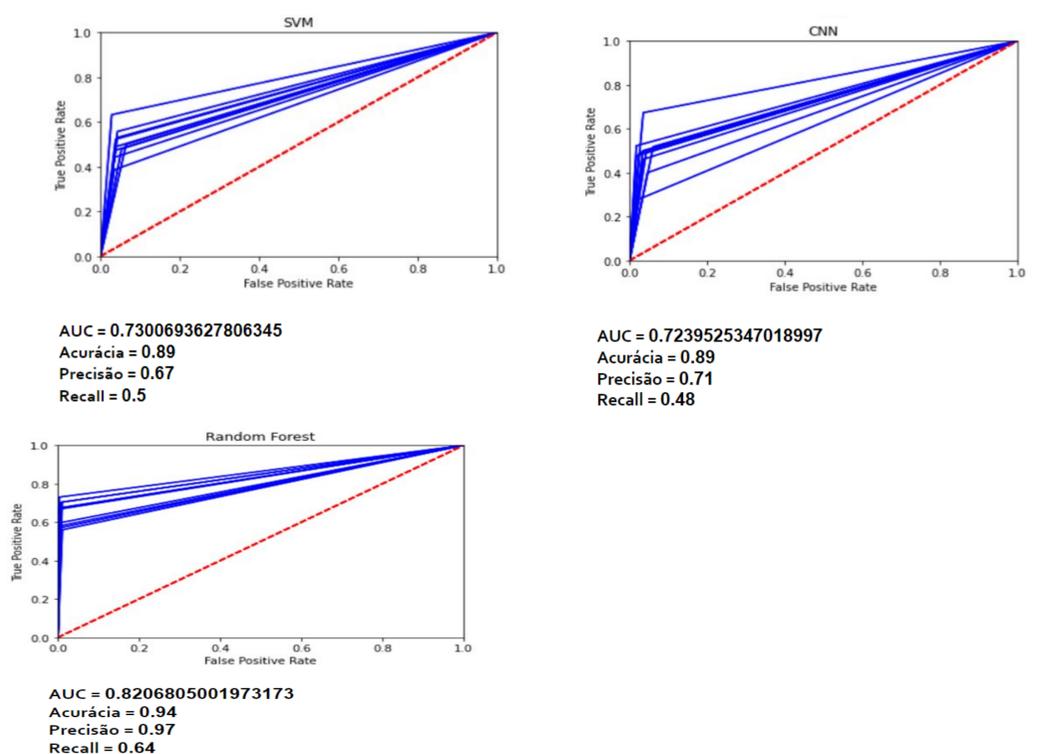
Existem diversas pesquisas afirmando que é mais barato para uma empresa manter o cliente atual do que conquistar novos clientes. Baseado nessa premissa, houve um crescente investimento em tecnologias com o intuito de prevenir o churn, ou seja, cancelamento do serviço por parte do cliente. Iremos introduzir nesse contexto, tecnologias de aprendizado de máquina, e através dessas, podemos treinar modelos classificadores que irão identificar quais dos clientes atuais se encaixam no mesmo padrão de clientes que cancelaram o serviço anteriormente, ou seja, quais clientes provavelmente estão prestes a cancelar o serviço.

O objetivo de reter o cliente é especialmente mais difícil para as empresas de telefonia, já que a grande quantidade de clientes acaba facilmente fugindo do controle dos analistas de pós vendas. Há diversas técnicas de aprendizado de máquina disponíveis, e todas elas conseguem produzir um modelo classificador de churn. Inicialmente, foi realizado uma pesquisa para descobrir os quais são os três algoritmos mais utilizados nesse contexto, que são: SVM, Random Forest e CNN. Nesse trabalho, foi construído três modelos a partir dos algoritmos citados acima. E cada modelo foi aplicado em uma base de dados contendo registros reais de clientes de uma empresa de telefonia francesa chamada Orange Telecom. Os resultados foram utilizados para a construção de suas determinadas métricas, tendo como objetivo, avaliar o desempenho dos classificadores. Após o treinamento e validação dos modelos, foi possível chegar em uma conclusão de qual algoritmo produziu o melhor modelo nesse contexto e as possíveis razões para tal resultado.

A metodologia, como mostra abaixo, é composta de três estágios principais, cujo objetivo é, a partir de um data set pré-processado, gerar três modelos classificadores, cada um criado a partir de uma das seguintes técnicas de Machine Learning: SVM, CNN e Random Forest. Após essa etapa, serão comparados os resultados através das métricas: matriz de confusão e curva ROC.



Após a aplicação dos classificadores, conseguimos obter a matriz de confusão e a curva ROC. Analisando essas informações, é possível calcular as métricas que nos mostram o desempenho dos nossos classificadores. São eles: AUC (area under the ROC curve), a acurácia, precisão e o recall, conforme demonstrado na imagem abaixo:



De acordo com os resultados mostrados anteriormente, o Random Forest obteve resultados melhores e mais constantes se comparados com os outros dois algoritmos. O algoritmo é provavelmente a solução mais pragmática para predição de cancelamento de serviços, principalmente com a volatilidade apresentada no mercado atualmente.