

Leticia da Luz



[LinkedIn](#) | [GitHub](#) | [E-mail](#)

Olá, este é o meu portfólio de projetos de Ciência de Dados.

Neste espaço busco apresentar um pouco sobre as minhas habilidades para resolver problemas de negócios por meio de Análise de Dados, Estatística e Machine Learning utilizando as linguagens R e Python.

Portfólio

Sobre mim

Me chamo Letícia da Luz. Atualmente, estou em formação em Ciência de Dados pela Data Science Academy e recentemente concluí o Bootcamp de Data Science do programa Santander Coders, uma parceria entre a Ada Tech e a Santander Open Academy.

Sou farmacêutica graduada pela Universidade Federal do Paraná e atuei em diversos setores da área da saúde. Meu interesse por dados surgiu durante minha experiência na área de epidemiologia, onde percebi o impacto significativo da análise de dados na tomada de decisões e na melhoria de processos.

Meu objetivo é ajudar as empresas a superar desafios na tomada de decisões estratégicas, fornecendo soluções baseadas em dados e transformando informações em ideias que impulsionem o crescimento e a inovação.

Linguagens de Programação:

- R: readr, dplyr, caret, randomforest, ggplot2, xgboost e catboost.
- Python: pandas, scikit-learn, matplotlib e seaborn.
- SQL: DQL (Data Query Language).

Ferramentas:

- Microsoft Azure Machine Learning
- PostgreSQL

Projetos

Previsão de Cancelamento de Serviço de Internet e Segmentação de Clientes



Contexto: Com a crescente competição entre os provedores de internet, manter os clientes existentes tornou-se uma prioridade vital para aumentar a receita. Diante

desse cenário, entender o fenômeno do churn, ou seja, a taxa de cancelamento de serviços, é crucial. Os provedores buscam identificar quais clientes apresentam maior probabilidade de cancelar seus serviços, pois isso lhes permite antecipar tais movimentos e implementar estratégias eficazes de retenção.

Objetivos:

- O primeiro objetivo é criar um modelo capaz de prever quando um cliente está prestes a cancelar seu serviço de internet.
- O segundo objetivo é agrupar os clientes que cancelaram o serviço de internet usando técnicas de clusterização. Ao identificar grupos de clientes com características e comportamentos semelhantes, conseguimos desenvolver estratégias de marketing específicas para reter esses clientes.

Resultados:

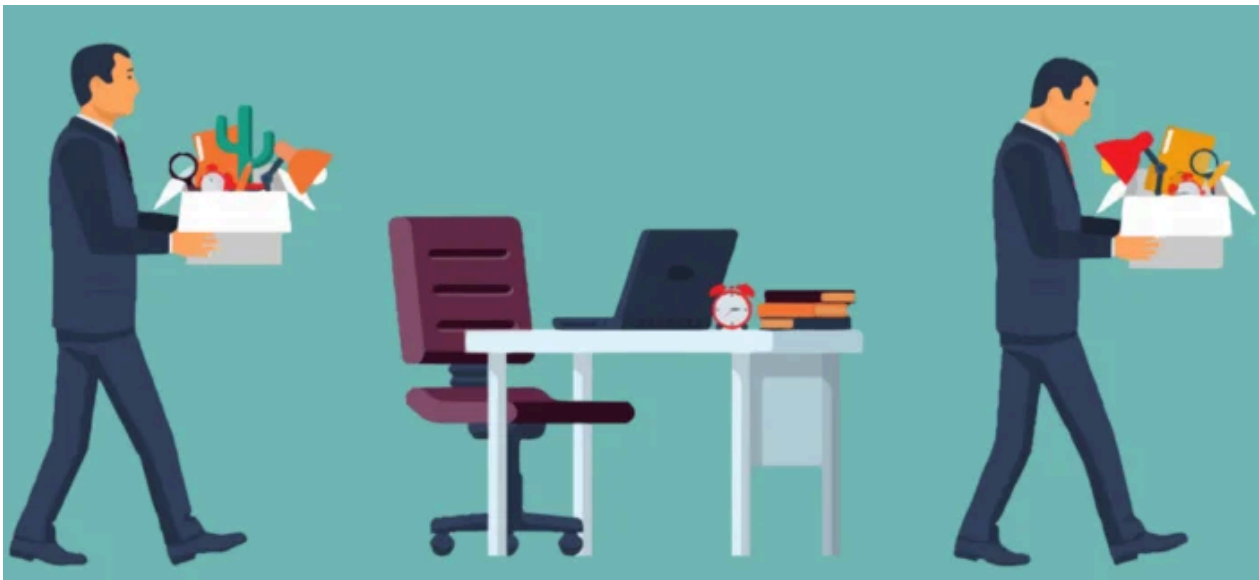
- Construímos seis modelos de Machine Learning utilizando três estimadores: KNN, SVC e XGBoost. Realizamos pré-processamento das variáveis e otimização de hiperparâmetros. A métrica escolhida foi o recall, visando minimizar falsos negativos e identificar o maior número de clientes em risco de churn. O modelo final, utilizando o estimador XGBoost, alcançou um recall de 0.94 nos dados de teste, indicando sua eficácia em identificar casos de churn.
- Realizamos também uma análise de clusterização com os clientes que cancelaram o serviço, usando os algoritmos K-Means e DBSCAN. Embora ambos tenham produzido scores semelhantes, optamos pelo K-Means devido à sua eficiência computacional e capacidade de identificar uma segmentação mais detalhada dos dados, otimizando a diferenciação entre perfis de clientes.

Ferramentas Utilizadas:

- Google Colab
- Python

Para acessar o projeto basta clicar em:  [View on GitHub](#)

Ciência de Dados em Recursos Humanos



Contexto: Este projeto é dedicado à análise e compreensão das altas taxas de rotatividade de funcionários enfrentadas por uma empresa de tecnologia. A empresa tem experimentado uma rotatividade significativa de funcionários, chegando a 30% em alguns semestres. Essa alta taxa de rotatividade tem custos substanciais e impactos no negócio.

Objetivos:

- Utilizar técnicas de análise exploratória, modelos preditivos e clusterização para desenvolver estratégias mais eficazes de retenção e identificar fatores críticos que contribuem para o turnover.

Resultados:

- Na Análise Exploratória dos Dados, identificamos que as variáveis 'Pontuação de Desempenho' e 'Horas Extras' mostram uma relação significativa com a taxa de desligamento. Funcionários com pontuações de desempenho mais altas e aqueles que realizam mais horas extras têm uma maior probabilidade de se desligarem da empresa.
- Construímos modelos de Machine Learning, a métrica escolhida para otimização foi o recall. Ao final, utilizamos uma abordagem de ensemble que combina as forças de dois algoritmos de ML para melhorar o desempenho da classificação através de um VotingClassifier. O resultado final foi um recall de 0.62 nos dados de teste, indicando um desempenho razoável na identificação de funcionários propensos a sair.
- Realizamos uma análise de clusterização dos funcionários desligados utilizando o algoritmo K-Means, a análise de clusterização dos funcionários desligados revelou que o Cluster 1 é composto por funcionários de nível sênior que fazem mais horas extras e recebem um salário maior. Os padrões encontrados neste projeto podem ajudar a empresa a desenvolver estratégias específicas para diminuir o turnover.

- Este projeto me rendeu o 1º lugar entre mais de 500 inscritos no desafio.

Ferramentas Utilizadas:

- Google Colab
- Python

Para acessar o projeto basta clicar em:  [View on GitHub](#)

Mini projetos

- [Mini projeto 1 - Análise combinatória e Probabilidades em Python](#)
- [Mini projeto 2 - Variáveis discretas e suas distribuições em R](#)
- [Mini projeto 3 - Variáveis contínuas e suas distribuições em R](#)
- [Mini projeto 4 - Estatística Descritiva em R](#)
- [Mini projeto 5 - Testes Paramétricos em R](#)
- [Mini projeto 6 - Testes Não Paramétricas em R](#)
- [Mini projeto 7 - Teste de Qui-quadrado em R](#)
- [Mini projeto 8 - Análise Exploratória de Dados utilizando SQL e Python](#)
- [Mini projeto 9 - Construção de um algoritmo baseado em k-NN em Python](#)

Certificados

Para acessar os certificados, basta clicar [aqui](#)

Entre em contato comigo por meio de:



© 2024 Leticia da Luz.
