**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**

**NÚCLEO DE EDUCAÇÃO A DISTÂNCIA**

**Pós-graduação Lato Sensu em Ciência de Dados e Big Data**

**José Nunes da Silva Junior**

**ANÁLISE CHURN RATE TELECOM**

**Belo Horizonte**

**2023**

José Nunes da Silva Junior

ANÁLISE DE CHURN RATE TELECOM

Trabalho de Conclusão de Curso apresentado

ao Curso de Especialização em Ciência de

Dados e Big Data como requisito parcial à

obtenção do título de especialista.

Belo Horizonte

2023

**SUMÁRIO**

**1. Introdução**

**1.1. Contextualização**

Nos últimos anos, o setor de telecomunicações tem observado um enorme avanço tecnológico, avanço este, que possibilitou novas oportunidades de negócios para as empresas. A facilidade de contratar determinados serviços com poucos cliques também é encontrada ao se tentar cancelar este mesmo serviço. Dada a crescente competição no mercado bem como o aumento da demanda dos consumidores por serviços mais personalizados e acessíveis, um elemento vem se tornando o foco das atenções: o churn rate.

O “churn rate”, ou taxa de rotatividade de clientes, é uma métrica crucial para empresas no ramo de telecomunicações, diferentemente de um modelo de compra e venda onde se obtêm o lucro no momento em que o processo se concretiza, o modelo de negócios de telecomunicações se caracterizam por um investimento inicial feito pela empresa, seja em infraestrutura, equipamentos e/ou instalação na residencia do cliente. Nesse modelo de negócio, o cliente só passa a dar lucro meses ou ate mesmo anos depois de contratar o serviço.

Diante disso, tornou-se fundamental para as empresas o foco nesse indicador, uma vez que a retenção de clientes é crucial para o sucesso e a sustentabilidade dos negócios.

**1.2. O problema proposto**

Devido as características de investimento inicial alto e lucro futuro, a presente analise priorizara no RECALL, uma vez que devemos buscar minimizar o maximo possivel de falsos negativos, onde o modelo determina que o cliente não vai cancelar porem o cliente cancela

**2. Coleta de Dados**

O conjunto de dados *Telecom Customer Churn Prediction,*objeto de análise deste projeto pode ser encontrado através do sítio da *Maven Analytics*, uma plataforma de desafios relacionados a analises de dados.

Figura 01 - Fonte da base de dados que ser*á utilizada. Disponível em: <https://www.mavenanalytics.io/data-playground?search=telecom>*

Na Figura NUMERO é possível observar os nomes e propriedades dos arquivos que serão utilizados. O conjunto em questão consiste em 3 (três) *datasets*  que possuem informações e finalidades distintas: *telecom\_customer\_churn*, que contêm a base principal dos dados que serão analisados neste artigo; *telecom\_zipcode\_population*, que possui dados adicionais relacionados a população por código postal e; *telecom\_data\_dictionary*, que como o próprio nome sugere, trata-se de um dicionário que contém as informações do que cada coluna representa em nossa base principal.

**3. Tratamento de Dados**

Figura 02 - Datasets que serão utilizados

Para a execução do projeto, foi utilizada a linguagem de programação interpretada Python, na sua versão 3.9.13, além do *Jupyter Notebook*, ambiente de desenvolvimento interativo baseado na Web

Nesta etapa do projeto, será utilizada majoritariamente a biblioteca *Pandas* (Figura XXXX) para manejo dos dados bem como de eventuais ajustes que possam ser necessários para deixar a base de dados melhor estruturada, e com isso, buscar melhores interpretações dos algoritmos de *Machine Learning*.

Figura 03 - Importação da biblioteca Pandas

BIBLIOGRAFIA