



UNIVERSIDADE PRESBITERIANA MACKENZIE
FACULDADE DE COMPUTAÇÃO E INFORMÁTICA
TECNOLOGIA EM CIÊNCIAS DE DADOS

PROJETO APLICADO IV

Projeto Anatel: Acompanhamento de Reclamações Registradas na ANATEL

PROFESSOR: GUSTAVO SCALABRINI SAMPAIO

GRUPO:

GUSTAVO CASSIMIRO – 10415853 – 10415853@mackenzista.com.br

MAIKI SOARES – 10415481 – 10415481@mackenzista.com.br

VANESSA CORDEIRO – 10415118 – 10415118@mackenzista.com.br

MARIA FERNANDA – 10424791 – 10424791@mackenzista.com.br

NATHAN SAMPAIO – 10408439 – 10408439@mackenzista.com.br

São Paulo
2025

SUMÁRIO

INTRODUÇÃO.....	3
OBJETIVO	3
Metas específicas	3
MOTIVAÇÕES E JUSTIFICATIVAS	4
Benefícios esperados da solução.....	4
DESCRIÇÃO DA BASE DE DADOS	5
Metadados principais	5
Período de coleta.....	5
Fonte	5
REFERENCIAL TEÓRICO	5
PIPELINE DE SOLUÇÃO	6
DISCUSSÃO DA PROPOSTA	9
CRONOGRAMA	10
REPOSITÓRIO NO GITHUB.....	11
REFERÊNCIAS	11

INTRODUÇÃO

A área de conhecimento envolvida neste projeto é o **ODS 9 — Indústria, Inovação e Infraestrutura**, que orienta iniciativas voltadas à melhoria e modernização de sistemas produtivos e de serviços essenciais. Ao relacionar esse objetivo às telecomunicações, evidencia-se a necessidade de desenvolver mecanismos que possibilitem monitorar e fortalecer a infraestrutura de rede, estimulando a inovação e a qualidade dos serviços prestados à população.

O problema selecionado tem como foco a análise das manifestações registradas junto à **Agência Nacional de Telecomunicações (Anatel)** a respeito das operadoras que atuam no Brasil. A base de dados utilizada compreende informações sobre reclamações, denúncias, pedidos de informação e sugestões, organizadas em formato de série temporal.

O desafio consiste em transformar esse volume de dados em um produto analítico capaz de apoiar gestores públicos e privados na tomada de decisão, permitindo a antecipação de picos de reclamações, a identificação de padrões sazonais e a avaliação do impacto de intervenções em infraestrutura e inovação tecnológica.

Ao detalhar esse problema, observa-se que a ausência de monitoramento preditivo dificulta a resposta ágil a falhas sistêmicas e a elaboração de estratégias preventivas. Além disso, a falta de indicadores padronizados pode comprometer a capacidade de avaliar a efetividade de políticas públicas e investimentos em inovação.

Nesse sentido, o projeto propõe o desenvolvimento de uma solução analítica que, a partir de séries temporais, permitirá acompanhar a evolução das manifestações da população, detectar variações inesperadas e apoiar decisões estratégicas que contribuam para a melhoria da infraestrutura de telecomunicações no país.

OBJETIVO

O objetivo deste projeto é desenvolver um **produto analítico baseado em séries temporais** capaz de monitorar, analisar e prever a evolução das manifestações registradas pelos consumidores junto à Anatel em relação aos serviços de telecomunicações.

Metas específicas

- Identificar padrões históricos das manifestações, considerando **sazonalidade, tendências e variações** por UF, operadora, serviço e tipo de registro.
- Construir um **pipeline de dados** para coleta, organização e tratamento das séries temporais de manifestações.
- Aplicar **métodos estatísticos e de modelagem preditiva** para gerar previsões de curto e médio prazo sobre o volume de registros.
- Implementar **técnicas de detecção de anomalias** que auxiliem na identificação de eventos inesperados ou problemas emergentes.

MOTIVAÇÕES E JUSTIFICATIVAS

A escolha do tema é motivada pela importância crescente das **telecomunicações** como um pilar essencial para a vida cotidiana, a competitividade das empresas e a inclusão digital da população.

Segundo dados da **Anatel (2023)**, foram registradas aproximadamente **3,2 milhões de reclamações** de consumidores contra operadoras de telecomunicações no Brasil em 2022, com destaque para serviços de banda larga fixa e telefonia móvel.

De acordo com a pesquisa **TIC Domicílios (Cetic.br, 2023)**, **83% dos domicílios brasileiros possuem acesso à internet**, reforçando a relevância dos serviços de telecomunicações para inclusão social e desenvolvimento econômico.

Nesse cenário, o registro de manifestações junto à Anatel constitui **uma fonte estratégica de dados** para compreender a qualidade dos serviços prestados e os principais pontos de insatisfação dos usuários.

A relevância do tema está no potencial de transformar essas informações em **inteligência prática** para gestores públicos e privados.

Benefícios esperados da solução:

- Antecipar problemas, reduzindo impactos negativos à população.
- Detectar padrões sazonais e eventos inesperados, apoiando a **gestão preventiva**.
- Avaliar o impacto de **investimentos em infraestrutura e inovação tecnológica**.
- Gerar **indicadores confiáveis** para órgãos reguladores.

Além disso, os **atores envolvidos** incluem:

- **Consumidores**: com serviços mais estáveis e atendimento mais eficiente.
- **Operadoras**: com insights sobre falhas recorrentes e oportunidades de melhoria.
- **Órgãos reguladores**: com indicadores confiáveis para monitorar o setor.
- **Sociedade em geral**: favorecida pela inclusão digital e modernização da infraestrutura.

Por fim, a solução pode ser aplicada em:

- **Dashboards internos** das operadoras.
- **Painéis de acompanhamento** de órgãos reguladores.
- **Relatórios de gestão de qualidade**.

DESCRIÇÃO DA BASE DE DADOS

O conjunto de dados utilizado refere-se às solicitações registradas pelos consumidores junto à **Anatel**, incluindo reclamações, denúncias, pedidos de informação e sugestões.

A base está disponível no **Portal de Dados Abertos do Governo Federal**.

Metadados principais

- **Título:** Solicitações Registradas na Anatel
- **Descrição:** Quantidade de registros de reclamações, denúncias, pedidos de informação e sugestões em relação às operadoras e à Anatel.
- **Etiquetas:** ANATEL, Banda Larga Fixa, Celular Pós-Pago, Celular Pré-Pago, Móvel Pessoal, SCM, Serviço Móvel, STFC, Telefone Fixo, Telefonia Móvel, TV por Assinatura, reclamação, denúncia, pedido de informação, sugestão
- **Autor:** Agência Nacional de Telecomunicações – Anatel
- **E-mail do autor:** rcts@anatel.gov.br
- **Mantenedor:** Gerência de Tratamento de Solicitações de Consumidores – RCTS
- **E-mail do mantenedor:** rcts@anatel.gov.br
- **Granularidade temporal:** mensal

Período de coleta

- Janeiro de 2015 até dezembro de 2023

Fonte

- Portal de Dados Abertos do Governo Federal:
<https://dados.gov.br/dados/conjuntos-dados/solicitacoesregistradasnaanatel>
- Agência Nacional de Telecomunicações (Anatel)

REFERENCIAL TEÓRICO

O estudo de **séries temporais** é amplamente aplicado em diferentes setores, incluindo economia, saúde e telecomunicações, permitindo a previsão de tendências e a detecção de padrões sazonais (Box, Jenkins & Reinsel, 2015).

No contexto das telecomunicações, essas técnicas são fundamentais para compreender o comportamento da demanda e o impacto de políticas de melhoria de infraestrutura.

Pesquisas recentes destacam o uso de modelos estatísticos como **ARIMA** e **Prophet** para previsão de séries temporais em bases públicas de reclamações (Silva & Andrade, 2021).

Outros estudos apontam o potencial de modelos baseados em **aprendizado de máquina**, como **LSTM (Long Short-Term Memory)**, que oferecem maior capacidade de capturar padrões não lineares e complexos (Zhang et al., 2020).

Comparação de abordagens

- **Modelos estatísticos (ARIMA):**
 - Vantagens: simplicidade e interpretabilidade.
 - Limitações: desempenho inferior em cenários de alta variabilidade.
- **Modelos de redes neurais (LSTM):**
 - Vantagens: maior precisão em grandes volumes de dados.
 - Limitações: alta demanda computacional e menor transparência nos resultados.

No Brasil, a análise de dados da Anatel é utilizada por órgãos reguladores para acompanhar a evolução da qualidade do serviço e orientar políticas públicas (Anatel, 2023).

Iniciativas acadêmicas também reforçam que o uso de séries temporais aplicadas às reclamações pode apoiar a antecipação de falhas sistêmicas e otimizar o atendimento ao consumidor (Ferreira & Oliveira, 2021).

Este projeto se apoia em abordagens já consolidadas na literatura, ao mesmo tempo em que propõe um diferencial: a **integração de diferentes métodos de modelagem** em um pipeline analítico voltado especificamente para a base de manifestações da Anatel.

PIPELINE DA SOLUÇÃO

O pipeline proposto para o desenvolvimento do produto analítico está estruturado em cinco etapas principais, desde a obtenção dos dados brutos até a apresentação dos insights em dashboards interativos.

Abaixo detalhamos cada uma dessas etapas:

1. Coleta de Dados

Descrição:

Coleta dos dados de reclamações registradas no portal de dados abertos da Anatel. Download do arquivo CSV atualizado.

Ferramentas:

- Python (*requests*, *pandas*)
- Agendamento via cron ou ferramenta de orchestration (ex: Apache Airflow)

Saída:

Base histórica consolidada (desde 2015) e atualizações incrementais.

2. Pré-processamento e Limpeza

Descrição:

Nesta etapa, o dataset bruto passa por um rigoroso processo de limpeza e padronização para garantir sua qualidade e consistência. As tarefas incluem o tratamento de valores ausentes, a unificação de categorias (ex: nomes de operadoras), codificação de variáveis categóricas e a agregação dos dados na granularidade temporal desejada (mensal por UF, serviço, operadora).

Tarefas incluem:

- Tratamento de registros com UF não informada
- Unificação de nomes de operadoras (ex: "CLARO", "CLARO SA" → "CLARO")
- Criação de variáveis temporais (mês, ano, trimestre)
- Agregação por quantidade de solicitações por mês, UF, serviço e operadora.

Saída:

Dataset limpo e agregado, pronto para análise.

Ferramentas:

- pandas
- numpy
- matplotlib
- seaborn
- plotly
- scikit-learn

3. Análise Exploratória e Feature Engineering

Descrição:

O objetivo desta fase é aprofundar o entendimento sobre os dados, identificando tendências, padrões de sazonalidade e correlações. Com base nesses insights, são criadas novas variáveis (features, ex: médias móveis, lags, indicadores de sazonalidade) que enriquecem o dataset e melhoram o desempenho dos modelos preditivos.

Tarefas:

- Visualização de séries temporais por operadora/UF
- Identificação de sazonalidade (ex: aumento de reclamações no fim do ano)

Criação de variáveis derivadas:

- Lags (valores de períodos anteriores): *lag_1*, *lag_12*.
- Médias móveis: *rolling_mean_3*, *rolling_mean_12*.
- Indicadores sazonais e de eventos: *month*, *year*, *is_peak*.

Saída:

Um dataset final enriquecido com features temporais relevantes para a modelagem.

4. Modelagem Preditiva e Detecção de Anomalias

Descrição:

Esta é a etapa central do projeto, onde são aplicados algoritmos para prever o volume futuro de reclamações e para detectar picos anômalos que possam indicar problemas sistêmicos. A abordagem combina modelos clássicos e de Machine Learning.

Abordagem dupla:

a) Previsão:

- ARIMA/SARIMA: Para capturar tendências e sazonalidade lineares.
- LSTM (Long Short-Term Memory): Para aprender padrões complexos e não lineares em séries temporais longas.

b) Detecção de Anomalias:

- Métodos estatísticos: Z-score e IQR para identificar outliers.
- Modelos de Machine Learning: Isolation Forest ou Local Outlier Factor (LOF) para anomalias mais complexas.
- Análise Contextual: As anomalias serão avaliadas considerando o contexto específico de cada UF e operadora.

Validação e Métricas:

Será utilizada uma divisão temporal dos dados (ex: treino até 2021, teste para 2022-2023) e validação cruzada para séries temporais. As métricas de avaliação incluem MAE, RMSE e MAPE para previsão, e Precisão/Recall para a detecção de anomalias.

Ferramentas:

- statsmodels
- pmdarima
- tensorflow
- scikit-learn
- scipy

5. Visualização e Interpretação

Descrição:

A última etapa do pipeline consiste na criação de dashboards interativos para comunicar os resultados da análise de forma clara e acionável.

Ferramentas:

- Streamlit ou Power BI
- Gráficos de séries temporais com previsões
- Alertas visuais para anomalias detectadas

Exemplo de visualizações:

- Série histórica com intervalo de confiança da previsão
- Heatmap de reclamações por UF/mês
- Top 5 operadoras com mais reclamações no período

DISCUSSÃO DA PROPOSTA

Este pipeline foi desenhado para ser reproduzível, escalável e orientado a negócio. A combinação de modelos clássicos (ARIMA) e modernos (LSTM) permite balancear interpretabilidade e precisão. A detecção de anomalias contextualizada por UF/operadora ajudará a identificar problemas localizados com agilidade.

A escolha de ferramentas open source (Python, Streamlit) visa garantir acessibilidade e baixo custo de implantação, alinhando-se ao caráter extensionista do projeto.

CRONOGRAMA

O cronograma abaixo detalha as principais atividades de desenvolvimento do projeto, alinhadas ao calendário da disciplina e às etapas descritas no pipeline da solução.

Período	Atividade	Entregáveis
10/09 – 20/09	Pesquisa bibliográfica e revisão de trabalhos relacionados	Referencial teórico inicial
21/09 – 25/09	Redação das seções de introdução, objetivos, motivação, descrição da base e pipeline	Documento preliminar (versão A2)
26/09	Entrega A2	Referencial teórico, pipeline da solução e cronograma inicial
27/09 – 02/10	Coleta de dados da Anatel, tratamento inicial (limpeza e padronização)	Base consolidada em CSV
03/10 – 06/10	Análise exploratória inicial (EDA) e visualizações preliminares	Gráficos e relatório descritivo
07/10	Implementação do modelo base (ARIMA ou Prophet)	Notebook com primeiros resultados
08/10	Entrega A3	Notebook atualizado com EDA, pré-processamento e modelo base
09/10 – 20/10	Refinamento do pré-processamento e feature engineering	Dataset enriquecido
21/10 – 05/11	Treinamento de modelos avançados (LSTM, SARIMA) e comparação de desempenho	Relatório de avaliação de modelos
06/11 – 15/11	Implementação de detecção de anomalias (Isolation Forest, Z-Score, LOF)	Notebook com resultados de anomalias
16/11 – 22/11	Construção de dashboards interativos (Streamlit/Power BI)	Dashboard funcional
23/11 – 26/11	Redação final do artigo, gravação do vídeo de apresentação e ajustes no GitHub	Artigo + vídeo
27/11	Revisão coletiva e checagem da reprodutibilidade no GitHub	Versão final
28/11	Entrega A4 (final)	GitHub completo, vídeo de apresentação, artigo e notebook final

REPOSITÓRIO NO GITHUB

Repositório: <https://github.com/Cassimirogustavo/Projeto-Aplicado-4>

Entrega 1: https://github.com/Cassimirogustavo/Projeto-Aplicado-4/blob/main/cd_projeto_aplicado_IV_entrega_1_vfinal.ipynb

Entrega 2: https://github.com/Cassimirogustavo/Projeto-Aplicado-4/blob/main/cd_projeto_aplicado_IV_entrega_2.ipynb

REFERÊNCIAS

ANATEL. Relatórios de Reclamações de Consumidores. Agência Nacional de Telecomunicações, 2023. Disponível em: <https://www.gov.br/anatel>

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. Time Series Analysis: Forecasting and Control. 5. ed. Hoboken: Wiley, 2015.

CETIC.BR. TIC Domicílios 2023: Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros. São Paulo: Comitê Gestor da Internet no Brasil, 2023.

CHUNG, J. et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv preprint arXiv:1412.3555, 2014.

FERREIRA, A. P.; OLIVEIRA, R. Análise de Séries Temporais Aplicada a Reclamações em Serviços Públicos. Revista Brasileira de Sistemas de Informação, v. 17, n. 2, p. 45–62, 2021.

SILVA, M. R.; ANDRADE, P. C. Previsão de Reclamações em Telecomunicações Utilizando Séries Temporais. Anais do Congresso Brasileiro de Informática, p. 221–233, 2021.

ZHANG, S. et al. Time Series Forecasting Using LSTM Networks: A Case Study in Telecommunications. Journal of Big Data, v. 7, n. 1, p. 1–15, 2020.