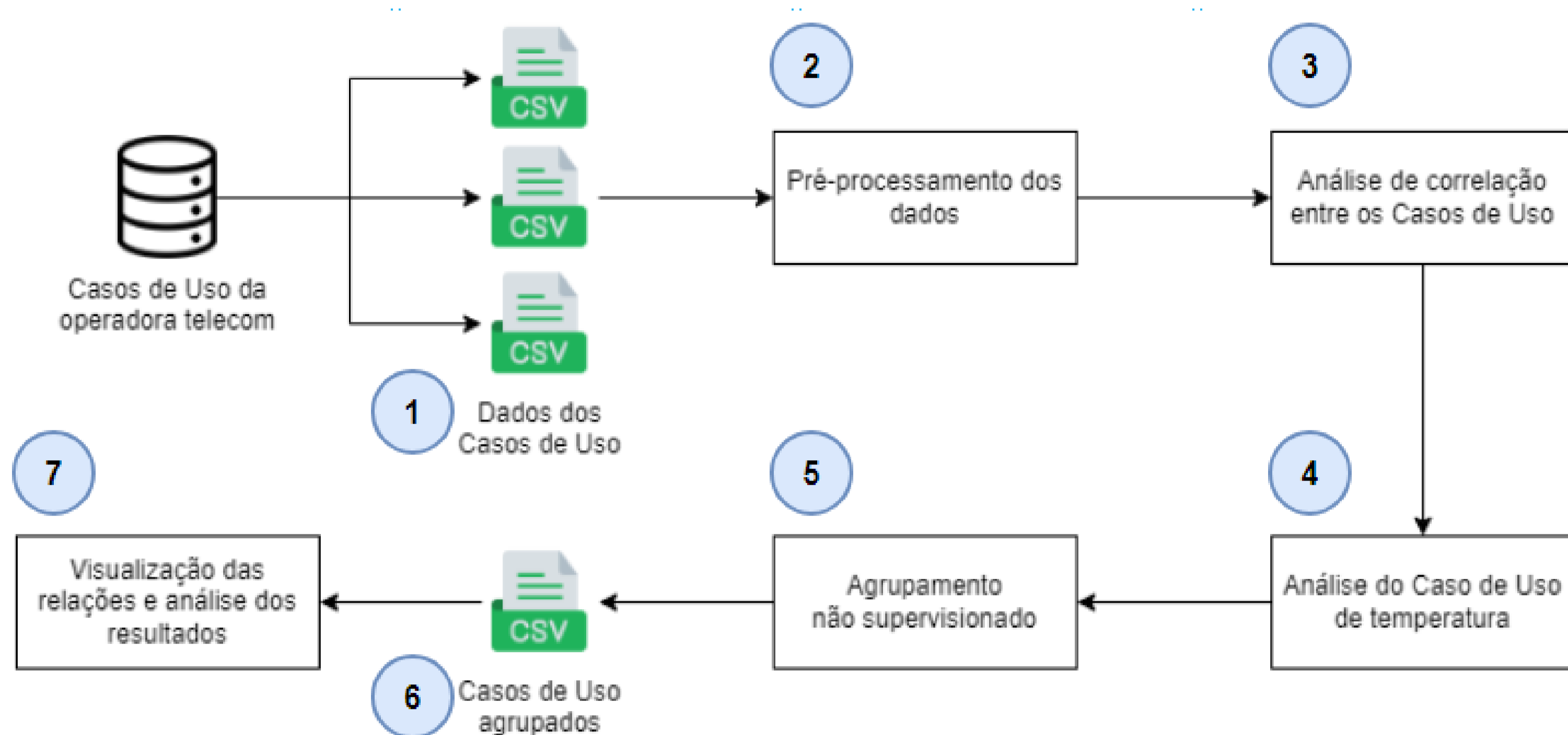




Ciência de Dados Aplicada a Redes Telecom 4G e 5G: Um Estudo de Caso Utilizando Algoritmos Não Supervisionados de Agrupamento Para Identificação de Padrões de Eventos em Antenas

PROBLEMA E OBJETIVO: As operadoras de telecomunicações recebem grandes volumes de dados de suas antenas de telefonia móvel (eNodeB), que representam erros ocorridos nas mesmas, chamados de "Casos de Uso". No entanto, a vasta quantidade de dados dificulta as análises e tomadas de decisão a partir desses eventos. Este projeto tem por objetivo utilizar algoritmos de aprendizado de máquina não supervisionados para identificar padrões nos eventos das antenas 4G e 5G, com dados disponibilizados por uma empresa de tecnologia a partir de um banco de dados relacional.

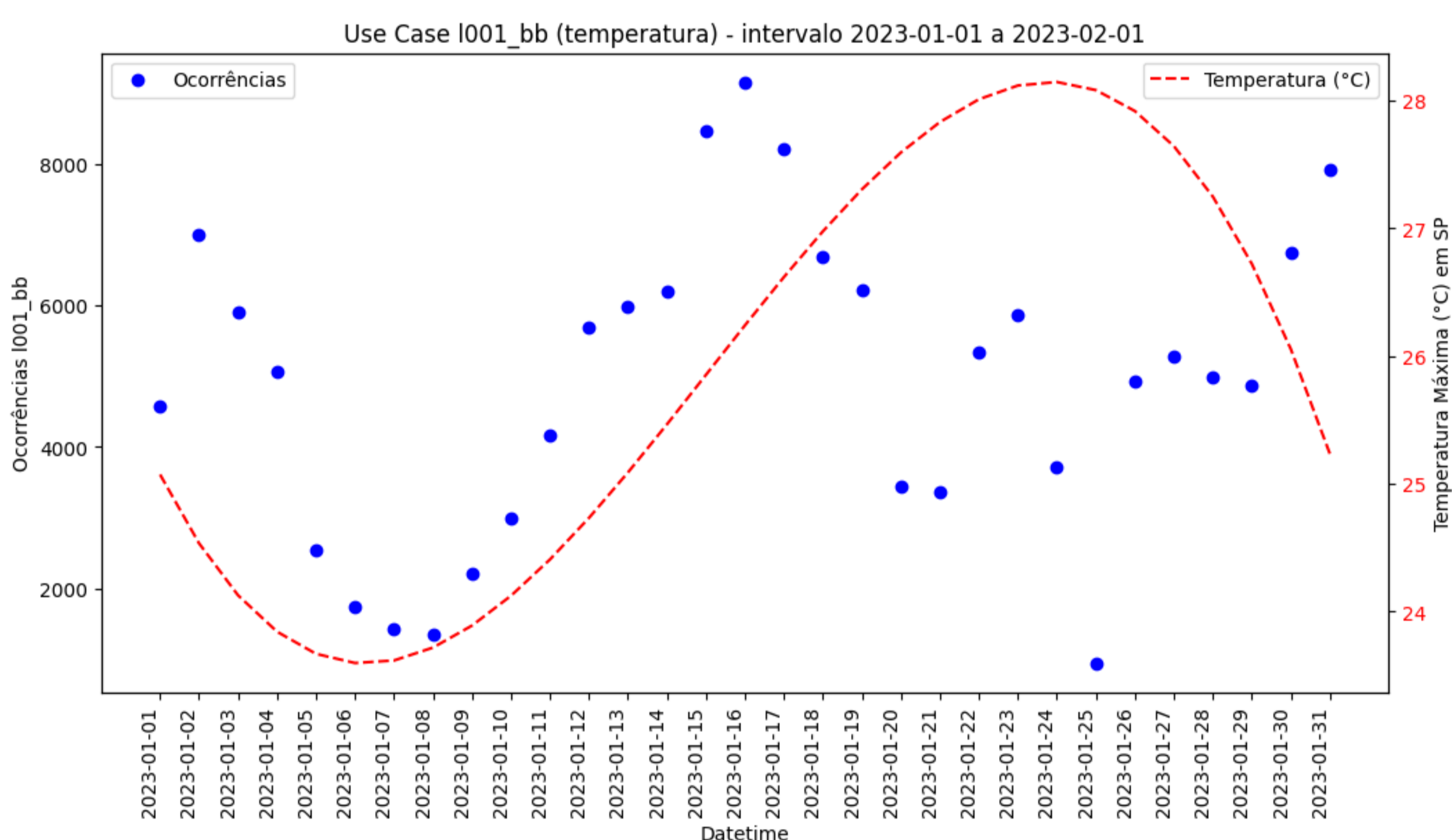
METODOLOGIA:



- 1 Obtenção dos dados referentes aos eventos ocorridos na rede móvel entre dezembro 2022 e maio 2023.
- 2 Pré-processamento dos dados, inclui a união de todos os arquivos em um único arquivo, filtragens e formatações.
- 3 Geração de um mapa de calor ou matriz de correlação, entre todos os 22 Casos de Uso extraídos da base de dados.
- 4 Separação do Caso de Uso de temperatura do arquivo unificado, a partir de onde é obtido o histórico de temperaturas por meio de uma API.
- 5 Os Casos de Uso são agrupados por meio da utilização de um algoritmo de aprendizado de máquina não supervisionado.
- 6 Obtenção do arquivo de saída resultante da etapa anterior, com cada Caso de Uso pertencendo ao seu respectivo agrupamento.
- 7 Visualização e análise das relações entre os agrupamentos formados, a fim de classificar a efetividade do modelo em alcançar o objetivo definido.

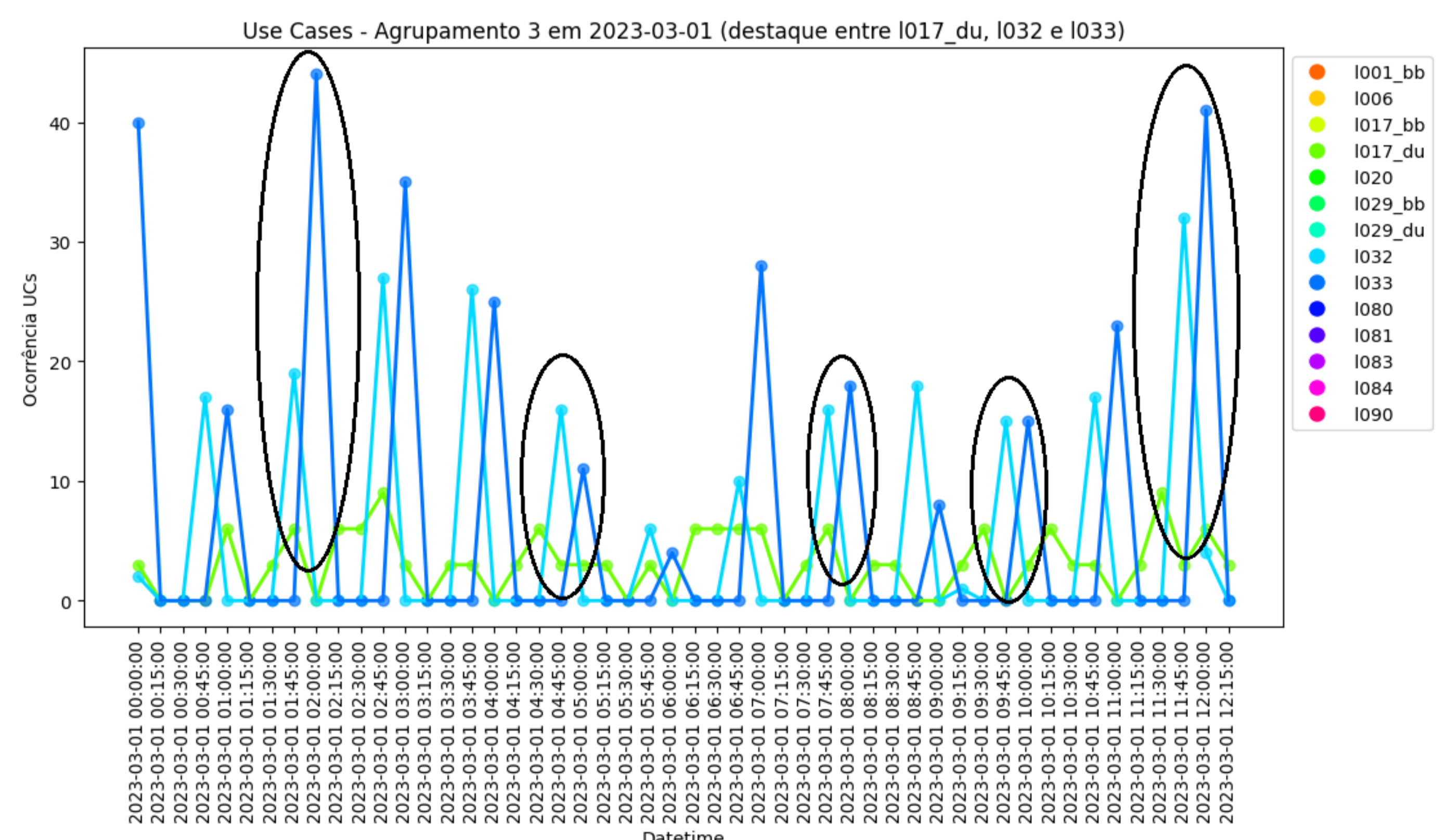
RESULTADOS:

EVENTO DE TEMPERATURA x TEMPERATURA MÁXIMA EM SP (°C)



A Figura demonstra o período entre os dias 01/01/2023 e 01/02/2023, com a curva do valor da temperatura e o número de ocorrências do evento sendo similares, indicando uma correlação.

SEQUÊNCIA DE EVENTOS



Na figura são exibidas sequências entre 3 eventos: **Bit Error Rate Alto** (número significativo de erros nos dados recebidos em comparação com os dados transmitidos), seguido de **contadores de disponibilidade da célula: tempo que o eNodeB está fora de serviço**, seguido de **distúrbio de energia no rádio**.

CONCLUSÃO: O modelo combinou técnicas estatísticas descritivas e aprendizado de máquina não supervisionado, sendo eficaz em demonstrar os eventos que ocorrem isoladamente, simultaneamente, alternadamente e/ou sequencialmente.