

Garantindo a qualidade de dados em estações meteorológicas

Kauê Guirado, Rafael Gonçalves

Prof. Dr. Fábio Silva Lopes

Motivação e Objetivo

Ao decorrer dos anos a ciência conseguiu grandes avanços na previsão climática e do tempo, estudando diversos fatores e parâmetros atmosféricos.

Diversas decisões são tomadas com a utilização de dados gerados a partir de sensores que estão sujeitos a interferência de fenômenos externos e erros do próprio hardware, muitas vezes distorcendo ou prejudicando uma análise precisa.

Pensando nisso, buscamos a implementação de um sistema capaz de garantir a qualidade destes dados, proporcionando uma análise mais eficaz, facilidade de manutenção e redução no tempo de tomada de decisão.

Metodologia

Para a aplicação, os indicadores de qualidade foram utilizados dados providos pela Estação meteorológica da Universidade Presbiteriana Mackenzie, que após tratamento e consolidação foram enviados a um servidor.

Foi utilizada a linguagem de programação Python para o desenvolvimento do servidor, incluindo os algoritmos de validação de precisão e completude de dados. Esses algoritmos utilizam basicamente os princípios da estatística de variância, media e desvio padrão.

Para uma análise mais detalhada rica em elementos visuais apresentamos nossos dados em gráficos feitos com a API HighCharts que nos possibilita ver os valores finais do indicadores de qualidade bem como os valores supostamente errados.

Resultados

Após a utilização de uma massa de testes o algoritmo de precisão detectou períodos que se encontram fora do desvio padrão e os classificou como inválidos como é possível observar no gráfico de temperatura (1) abaixo. Já o algoritmo de completude concluiu que 95.92% destes dados são válidos, ou seja, os valores de temperatura e umidade estão concisos conforme é apresentado no gráfico de completude (2).

Como trabalho futuro será implementado uma chamada no servidor capaz de receber novos dados em tempo real para calcular sua precisão e completude possibilitando gráficos atualizados, além de uma API de notificação capaz de enviar um e-mail informando uma possível falha nos dados.

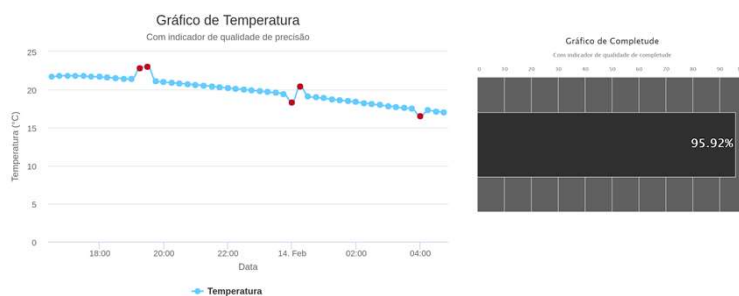


Gráfico de Temperatura(1) e Gráfico de Completude(2) – 13/03/2018 16:30 à 14/03/2018 04:45

Conclusões

Concluimos que com a implementação da qualidade de dados é possível obter um sistema muito mais confiável, onde determinados erros possam ser encontrados e solucionados com maior facilidade e rapidez a fim de proporcionar dados mais consistentes e confiáveis.

Com a aplicação de apenas 2 indicadores de qualidade já foi possível ter uma análise muito mais detalhada dos dados, garantindo assim sua confiabilidade para utilização muito mais precisa.

Referências

GAMBA; Implementação de uma arquitetura para aquisição de dados em fluxo incorporando atributos de qualidade, 2016..

BATINI, C.; SCANNAPIECO, M. Data and Information Quality: Dimensions, Principles and Techniques. Springer, 2016



Faculdade de Computação e Informática
Mackenzie