DataFIX: Automatic Tools for Data Review and Substitution

CeADAR — Centre for Applied Data Analytics Research

MARKET NEED

Novel business applications use raw, real-life or online data that contains various types of errors.

Existing data cleaning platforms lack automation and require high-levels of expert interaction.

Hence, identifying and correcting erroneous data is expensive, requiring analysts' expertise and ad-hoc solutions.

There is a real need for solutions that help to automate data quality investigation and where appropriate automated correction of data.
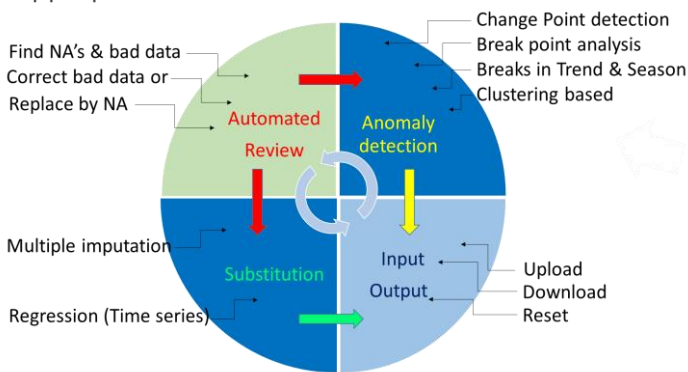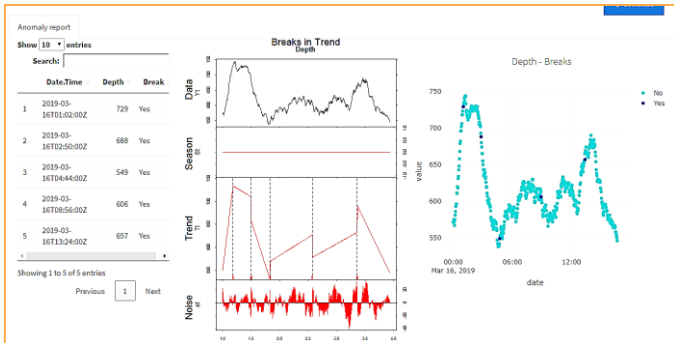

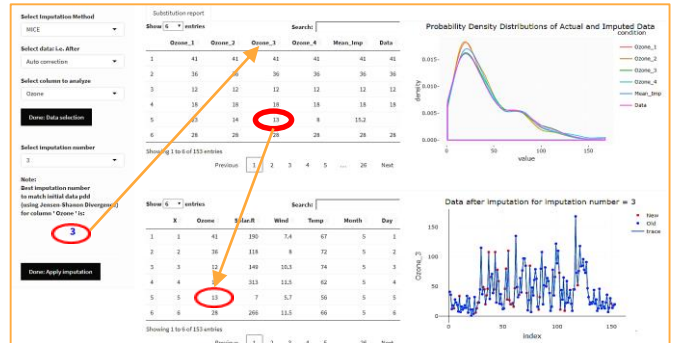*DataFIX workflow and methods.*

TECHNOLOGY SOLUTION

Our solution automates the data correction process with an easy to use interface based on state of the art algorithms for:

➢ Data cleaning
➢ Anomaly detection for time-series data, and
➢ Advanced methods for data substitution.

DataFIX is developed as an interactive web application that provides automatic data cleaning and substitution *without prior workflow design.* A range of data visualisation and data correction options are provided for a variety of data types.

DataFix provides a range of domain-independent, unsupervised techniques:

➢ Original automatic data cleaning workflow;
➢ Statistical and ML based methods for anomaly detection (trend, change & break point analysis, or clustering methods);
➢ Data substitution based on multiple imputation by chained equations with best choice based on density distribution match.

APPLICABILITY

The DataFIX platform provides automated data review & substitution for various business challenges::
➢ Raw, online or sensor data, historical data (Numerical, categorical, time-series);
➢ Monitoring sensor data (as batch);
➢ E-banking, E-commerce, fraud detection, manufacturing (fault detection).

DataFIX visualizations illustrating detection of break/change points (left) and data substitution (right).

RESEARCH TEAM

Dr. Tamara Matthews, TU Dublin

Sandesh Gangadhar, TU Dublin

Dr. Robert John Ross, TU Dublin

www.ceadar.ie

TECHNOLOGY CENTRE
ENTERPRISE IRELAND
IDA IRELAND SUPPORTED

OLLSCOIL TEICNEOLAÍOCHTA BHAILE ÁTHA CLIATH
TU DUBLIN
TECHNOLOGICAL UNIVERSITY DUBLIN