Project Synopsis/Project Concept Document

Project Number	37
Project Title	MLOps enabling anomaly detection in real-time sensor data streams
Document	Dass Project Concept Document
Creation Date	28.01.2022
Created By	Tanmay Goyal, Soveet Nayak, Lakshmi Girija, Imami Syed
Client	Smartterra

Description

- Sensors are used to monitor and get data at a high velocity from a water distribution network. These data streams are analysed using DeepAR or TFT for anomalous behaviour.
- The overall objective of this project is to identify the anomalies and predict what and where the problem could have been.
- Analyse data with large granularity in real-time, pre-process it and clean the data.
- To eliminate any errors before running machine learning algorithms to get an idea of different models to predict different kinds of losses, including revenue and water losses.
- Analyse the data over time and come to an accurate conclusion.

Profile of Users

To be updated in accordance with progress in the project and client meetings. We currently only have a high level overview of the project at hand.

Enter the user's profile to display the user's access to the website based on the user's profile. Allow/restrict access to various fields on the website. After uploading, relevant document, the user can check graphs and data in human readable form.

Feature Highlights

- We need to build a front-end application using -> Javascript, React to take user input for the MLOps(a file, mostly a csv containing data from water pipelines and metres over a period of time) and visualise the results. We need to process (smooth or clean) incoming data at a high velocity and store all this evaluated data in MongoDB on the basis of time indexing.

- Next, we need to build a backend application using -> Python, Frappe to orchestrate a data pipeline (primarily Azure Data Factory), a database (MongoDB) and an ML engine (Kubeflow). The ML data is arranged in Kubeflow clusters. A backend orchestrator should regularly kickoff an ML engine. The goal of the orchestration layer is to optimise and streamline systematic, repeatable processes.

- The backend should retrieve the latest processed data streams and ML insights.

- The front end should display conclusions drawn from this processed information(accessed from the backend) using graphs, charts, etc.

Usage Model and Diagrams (if any)

To be updated in accordance with progress in the project and client meetings. We currently only have a high level overview of the project at hand.

A description of how the system will be potentially used by the various users. This should be captured in a diagram. As a flow of screens or operations or a set of flow charts.