source allies

Anomaly Detection with IoT Data

SUMMARY

IoT Data sensors are not perfect and can send bad data. Odometer data is used to create preventative maintenance plans, so when data is incorrect, it can cause trucks to not go in when they should, leading to money and time lost, or get maintenance before they should, leading to unnecessary spending. We showcase how ML can be used to highlight these abnormal readings, allowing our partner to get preventative maintenance plans back on track.

CHALLENGE

Semi-Automated Pipeline for Evaluating Odometer Data
- Replace the existing manual process with a automated alert for incorrect odometer data coming in from third-party sensors

Accessible Results
- Make results of models easily accessible for technical and non-technical users

Example Project for Future Analytics Projects
- Implement best practices and documentation for this project to stand as a repeatable example for future projects

SOLUTION

Third-party sensor data comes into the AWS environment via AWS Database Migration Service, landing in an S3 bucket, then is ingested into Amazon SageMaker. Scripts are set up in this pipeline to clean and normalize the data before training occurs. An Amazon SageMaker endpoint is hosted to allow model interaction and prediction. AWS S3 is used to land predictive results where they will be evaluated and acted on. The model consists of an Isolation forest algorithm that goes through a hyperparameter optimization stage when training. F1, Precision, Recall, Specificity, and False Anomaly Rate are used to evaluate model performance.

RESULTS

Our model and application reduces reaction time in fixing bad odometer data within the source systems. A pipeline was established to train, test, and validate the model and determine bad odometer data. This pipeline will make next iterations 10x faster, reducing time needed for future development. Best practices were set up, including monitoring, logging, and playbooks to act as a sample analytics project template for the future .