



MLOps In Practice

SUMMARY

Our client, a leading professional services company, wanted to forecast how a Key Performance Indicator (KPI) will behave for the next year, given current trends. High levels of automation needed to be demonstrated, requiring minimal maintenance. To get automated predictions of an important business KPI, we built an MLOps system in AWS to automatically retrain, evaluate, deploy, and monitor an ML time-series model. These results are used within an internal tool called 'Business Performance Monitor' to visualize a projection of the KPI.

SOLUTION

Using an AWS Sagemaker pre-trained model with the Deep-AR algorithm, we built a time-series model using our information about previous months' KPIs to predict the next twelve months' KPI values. We hosted the model on a Sagemaker Realtime Endpoint and exposed it to clients through a Lambda function, which does data processing to simplify the client-facing API.

A service was set up to pull new KPI data monthly, add it to the training set, and train a new model using the updated data to keep the models up to date. Then the two models are squared off in a contest - whichever more accurately predicts the latest and most relevant data is considered better performing. The superior model is then automatically deployed using a CodePipeline and a report is sent to a Slack channel for observability.

Without actual values for our KPI in future months, model accuracy is measured retroactively, limiting its use as a health indicator. Instead, model drift is used as a leading indicator, measuring how different the model's predictions look from historical data, using the KS-Test and Population Stability Index. If drift surpasses a threshold, a slack channel is alerted so manual intervention can occur.

RESULTS

100%

Automated process
requiring no manual
intervention

8 Hours

8 pair hours saved
every retraining

60 Hours

60 hours saved for each
KPI report out event