

IES Final Project

Martim Gouveia Sousa - nº 85672

University of Aveiro, DETI

January 4, 2023

Project definition

Assumptions

- Generate a synthetic stream of visitors, where each visitor is associated with a pageID;
- Each pageID represents a time series, aggregated in an hourly frequency for Prediction/Machine Learning purposes;
- We have a total of 10 time series (Page IDs).

Goals

- Display the data through graphs and charts in a dashboard refreshed roughly in real time;
- Predict the incoming number of visitors for the next H hours with probabilistic forecasting methods;
- Retrain a new model whenever distribution shifts occur (model starts performing terribly);
- Monitor the model performance in real time.

Architecture

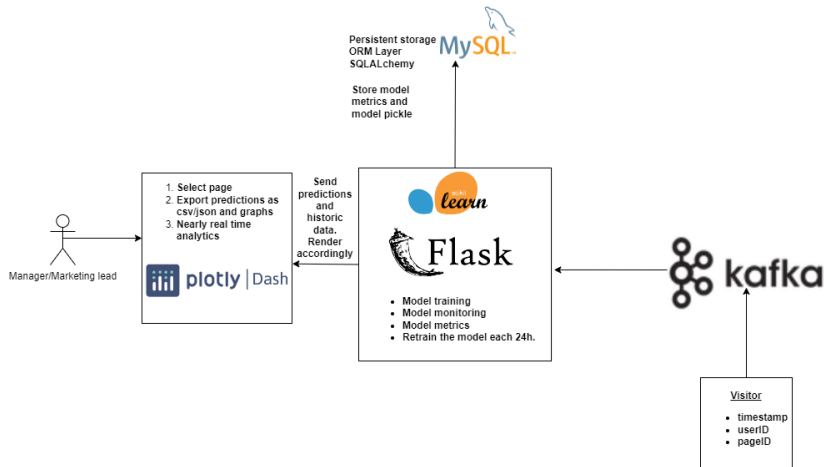


Figure: Project architecture diagram.

Tech Stack

- Docker;
- Confluent Kafka;
- Flask;
- MySQL;
- Plotly Dash.

Prospect work/Not Implemented

- Schedule a retraining scheme, e.g., retrain each model every week;
- Monitor model performance in real time;
- Nginx (static content) to act as a reverse proxy for Gunicorn (dynamic content) with HTTPS requests;
- Celery or other technology to run intensive tasks in the background.