



ORGANIZATION



COUNTRY/REGION

Spain

INDUSTRY

Transportation

BUSINESS NEED

Internal Fraud Detection based on
Machine Learning and Data Science Techniques

THE CUSTOMER

Alsa is one of the biggest bus operating companies in the world. Alsa was founded in 1923 in Spain, and became the leading road passenger transporter in the 1980's. By 2000's was become one of the leading companies in Western Europe and expanded activities to Asia and North Africa. In 2007 was acquired by the British National Express.

Alsa has revenues above \$700mio per year, manage more than 3400 buses and transports more than 350mio passengers per year. Since 2017 Alsa is investing heavily in technology in order to improve business and make it more efficient.

THE NEED

The customer suspected some fraudulent behaviors were taking place in the transport operations in certain routes. The challenge in the project was the fraud had "no face"; i.e., it was a customer's suspicion, therefore the fraudulent behaviors had to be profiled and confirmed, and from then, once there were solid fraudulent patterns, the AI model had to detect them in the daily operations and be ready to profile and learn new fraudulent patterns.

THE SOLUTION

In December 2018 both companies signed a four-month agreement to apply artificial intelligence techniques to the before mention problem.

The project was developed in several phases:

1. *Data loading*: Defining the data structure and format, names, and features.
2. *Data wrangling*: Data processing to make them adequate for the AI models developed.
3. *Data analysis*: Jupyter Notebooks generation which summarized the problem, the several approaches followed and tested, the data available, the volume of them, etc.
4. *Feature engineering*: From the existing variables, generating new features to feed the AI models to enrich the information model.

5. *Machine Learning*: Due to the lack of labeled historical data hAltta team conducted an unsupervised learning approach based on anomaly detection techniques. Several analysis have been created in order to measure different ways to create underlying datasets and apply different algorithms. Based on the study of model outputs several configuration have been selected as candidates to use in real life. Those selected models have been "consulted" in order to extract highlights in fresh data and based on several manual analysis performed the whole team (business and technical sides) validated the approach.

With this process the customer was able to create a tool which runs periodically and generate a list with all cases ranked by "anomaly level".

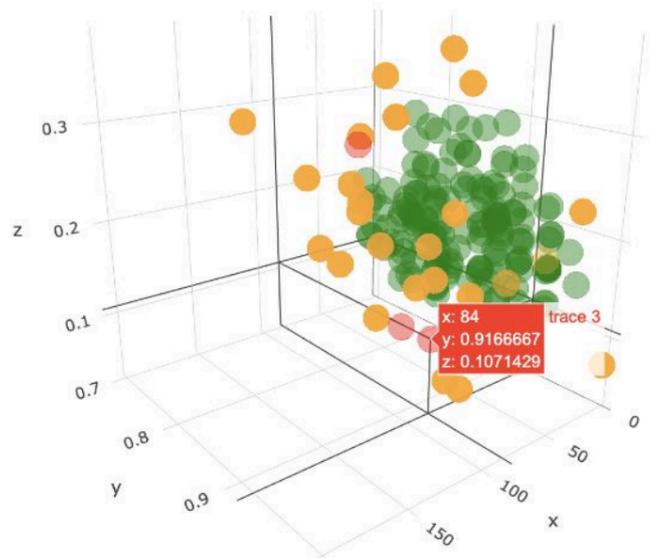


Fig. 1.- Anomalies 3D visualizations.

THE BENEFITS

The customer has harvested several benefits from this AI-based tool development; namely:

- ✓ Through the combination of business knowledge with results delivered by the tool ,daily operations can be strategically planned.
- ✓ The impact of actions conducted with the information created increases customer economic incomes.
- ✓ Several situations existing in data have been reported as new knowledge for different areas in the company.

TECHNICAL INFORMATION



hAltta used the so-called ‘Python Data Science Toolkit’, which is formed by open source tools and libraries such as:

- Numpy
- SciPy
- Pandas
- ScikitLearn
- Bokeh
- MatPlotLib
- SeaBorn
- ...

Experiments and visualizations were shown to the client in early stages, using Jupyter Notebooks.



For source code versioning purposes, hAltta made use of private Git repositories.

There was neither special needs for infrastructure in order to run experiments or to operationalize the model in this project.

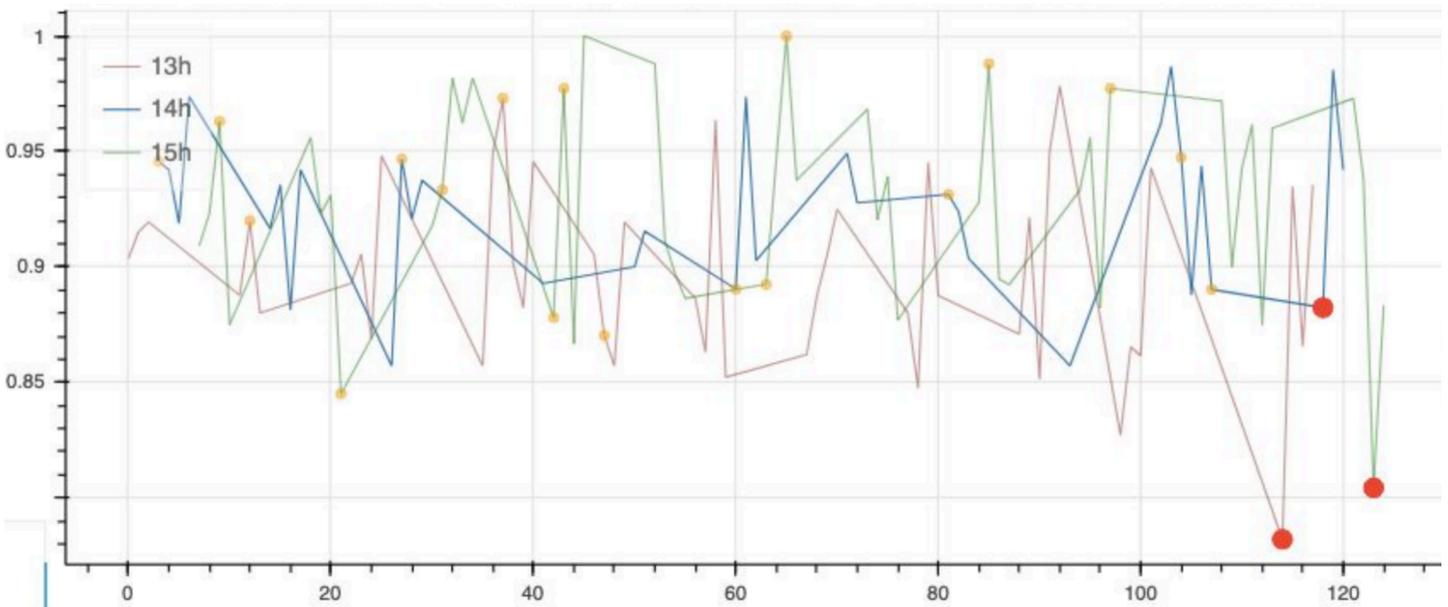


Fig. 2.- Analysis of several cases over time based on model predictions.