



Artificial Intelligence Serving Fraud
and Default Detection in the Airline Industry

ORGANIZATION



IATA
www.iata.org

ORGANIZATION SIZE

1,600 employees

COUNTRY/REGION

Worldwide

INDUSTRY

Aviation

BUSINESS NEED

Fraud and default detection on
the flight ticket sale activity

THE CUSTOMER

Founded in Havana, Cuba, in April 1945, the International Air Transport Association (IATA) is the trade association for the world's airlines, representing some 275 airlines or 83% of total air traffic. IATA supports many areas of aviation activity and help formulate industry policy on critical aviation issues. Advocating for the interests of airlines across the globe, IATA challenges unreasonable rules and charges, holds regulators and governments to account, and strives for sensible regulation. For over 70 years, it has developed global commercial standards upon which the air transport industry is built. It helps airlines operate safely, securely, efficiently, and economically under clearly defined rules.

THE NEED

IATA Settlement Systems (ISS) are the backbone of the global air transport industry. They enable the swift, reliable, and efficient movement of funds between airlines and their travel partners. In 2016, the ISS settled \$401.4 billion.

Inside the ISS, the Billing and Settlement Plan (BSP) is a system designed to facilitate and simplify the selling, reporting and remitting procedures of IATA accredited passenger sales agents, as well as improve financial control and cash flow for BSP airlines. There are BSP operations in some 180 countries and territories. The system currently serves approximately 400 participating airlines. In 2016 IATA's BSP processed \$219.0 billion.

Despite these fine-tuned procedures and processes, the fraud and default episodes by accredited passenger sales agents have been an ongoing problem for the organization for the last years, reaching the tipping point in 2013 when the unrecovered debt amounted to its maximum level. Since then several strategies -mainly based on predictive analytics and statistical methods- have been implemented to reduce these economic losses; there have been some improvements, the problem and its related economic losses have kept on occurring though.

THE SOLUTION

In August 2016 IATA and hAlta signed a one-year project to use artificial intelligence in order to tackle the fraud and default problem by the accredited passenger sales agents. The project had a two-fold scope:

1. Testing the suitability of using artificial intelligence to cope with IATA default problem.

2. The development of a machine learning-based model -from hAlta Yoken solution- to detect frauds with as much accuracy as possible.

The first semester of the project was devoted to verify if artificial intelligence could be employed to deal with this ongoing problem. hAlta dealt with the data of Madrid Global Delivery Center Europe. The dimensions of the dataset were around 1.5 million agents' billing periods reviewed -from January, 2014 to July 2016, and approximately 25,000 agents' behaviors analyzed.

During the feature engineering stage of the machine learning project, hAlta created more than 100 features, being some of them Billing Fluctuation, Short Payment Number, CPI Index, Financial Guarantees, Country Market Share Size, Monthly Billing Fluctuation, etc.

With respect to the particular technical approaches followed during the whole length of the project, the algorithms Random Forest, Logistic Regression, Xgboost, Decision Trees, SVC, amongst others, were employed.

From the results achieved hAlta did prove that artificial intelligence could be applicable to this problem set and it could add significant value. Nonetheless, on the one hand owing to the random nature of the factors behind defaults, It was impossible to design algorithms since no differences existed between right and wrong flight ticket sale behaviors. On the other hand due to the meaningless economic impact of defaults versus frauds, hAlta recommended IATA to concentrate on frauds -the second stage of this project.

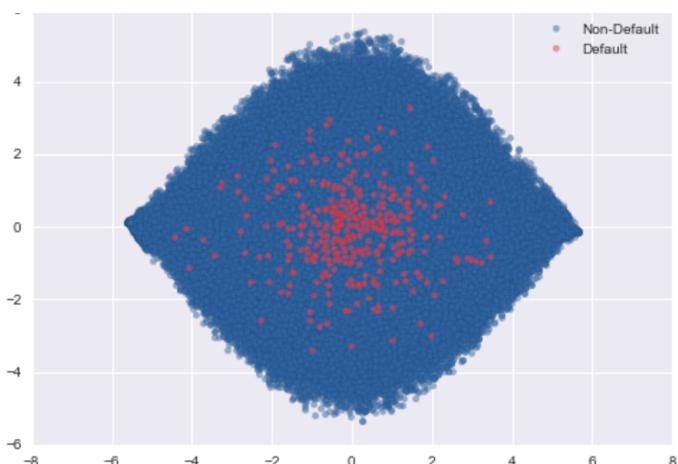


Fig. 1. Clustering non-defaulting and defaulting sales agents. As it can be noticed, there was no clear decision boundary between both groups mainly owing to the random nature of defaults.

The second semester of work was devoted to create a machine learning-based model to detect frauds with as much accuracy as possible. In order to perform this task, hAltta focussed on France and United Kingdom -as the countries in which the episodes of frauds had had the highest incidence.

The dimensions of the dataset this time were as follows: in UK, around 40 million of sale operations reviewed and close to 2,000 accredited passenger sales agents analyzed; in France around 40 million of sale operations reviewed and close to 3,000 accredited passenger sales agents analyzed. The dates encompassed were from October 2015 to July 2017.

The final results of this second part of the project confirmed hAltta had successfully developed a machine learning-based model -from hAltta Yoken methodology- to detect frauds in the ticket sale activity performed by accredited passenger sales agents, with an accuracy over 90%.

THE BENEFITS

From the work performed by hAltta, IATA could extract several positive lessons being the three most important:

1. Apprehending certain guidelines regarding the pre-processing and processing of raw data, upon facing an artificial intelligence project.
2. Detecting -through the exploratory analysis performed- different and new anomalous sales agents' behaviors, which had never been found out previously by the standard analysis procedures implemented in IATA; e.g., the so-called by hAltta "zombie sales agents".
3. And the last and most significant, obtaining a machine learning-based model to detect frauds committed by accredited passenger sales agents with an accuracy over 90%. With this model spotting fraudulent behaviors by accredited passenger sales agents would take place as soon as the HOT files - raw data- were received by the artificial intelligence classifier, avoiding thus the economic losses in which IATA used to incur due to certain delays and gaps in the standard procedure.

hAltta's dedication and focus on the problem surpassed IATA's expectations, by not only fulfilling requirements but also by going the extra mile that separates excellent from exceptional. As the model continues to keep on learning new fraudulent patterns and improving its performance, the next steps would be ideally to include more external information sources to enhance the detection and prediction capabilities of the model. In sum, IATA is benefiting from a top-notch fraud detection artificial intelligence model, enabling it to continue shaping the industry.

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
- Keras
- 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 BitBucket repositories.

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

"hAltta did an amazing job throughout the year devoted to develop the machine learning-based model. Thanks to its know-how and experience in artificial intelligence, and its strong commitment to the project, the machine learning model achieved an impressively unexpected high performance, reaching a level over 90% accuracy upon detecting fraudulent travel agencies. This first successful experience with artificial intelligence applied to a particular problem in the airline industry must definitely show IATA the path to be followed."

Giorgio Moretti.

Assistant Director Remittance & Settlement,
ITS Business Solution at IATA