Increasing the Detection Performance of Genuine Separation Minima Infringements with XGBOOST

Christian Verdonk Gallego, Chen Xia, Miguel García Martínez, Diego Piva Álvarez,
CRIDA
Madrid, Spain
{ceverdonk, cxia, mgmartinez, dpiva}@e-crida.enaire.es

Francisco Javier Pérez Heras
Safety Monitoring Team
ENAIRE
Madrid, Spain
fpheras@enaire.es

***Abstract*— The Automatic Safety Monitoring Tool (ASMT) is a key component of the Safety Management System of an air navigation service provider. Its main function is to monitor and record separation minima infringements and runway incursions. The recording of these events helps monitoring key safety performance indicators. The ASMT provides candidate events which are suitable to be considered as genuine infringements. These candidate events are often non-genuine due to different circumstances, such as noise in the vertical profile of the aircraft trajectories. The process of filtering genuine events from the initial population of candidate is an important step for accurately reporting on the safety performance indicators. This paper presents the workflow followed for deploying XGBoost for discriminating between genuine and non-genuine events within PERSEO, which works as the Automatic Safety Monitoring Tool for ENAIRE. Results show a significant increase on the detection performance of the separation minima infringements, which can facilitate larger analysis due to an increment of the detection precision while maintaining its sensitivity.**

Keywords- safety performance, xgboost, asmt, separation minima infringement, PERSEO

I. INTRODUCTION

Safety is paramount in Air Traffic Management. As such, Air Navigation Service Providers (ANSPs) are subject to the provisions under the Commission Implementing Regulation (EU) 2019/317 of the European Commission [1]. These provide the legal framework for the *'Performance and Cost Scheme'* under the Reference Period 3. The framework establishes the basis for the performance targets for the Key Performance Areas of Safety and Capacity, among others.

Such regulation establishes the indicators for monitoring safety at ANSP level, and among them, the rate of separation minima infringements (SMI) within the airspace where the air navigation service provider provides air traffic services, calculated as the total number of separation minima infringements with any contribution from air traffic services, or Communication, Navigation and Surveillance (CNS) services with a safety impact divided by the total number of controlled flight hours within that airspace.

The reporting of this indicator is facilitated through the identification of SMIs automatically. This is detailed in the Acceptable Means of Compliance (AMC) and Guidance Material (GM) for the implementation and measurement of Safety Key Performance Indicators (SKPIs) [2]. The AMC established the basic components of automated safety data recording systems, which are commonly known as *Automatic Safety Monitoring Tools* **(ASMT)** [3]**.**

PERSEO is a multipurpose tool developed by CRIDA whose objective is to facilitate the data-driven decision-making processes related to Air Traffic Management in ENAIRE through the exploitation of multiple conformed data sources.

Among other capabilities, PERSEO includes an ASMT which currently fulfils the regulatory requirements for ENAIRE. The PERSEO ASMT is capable of exploiting surveillance and flight data in order to detect SMIs, and provides a user-friendly interface in order to select genuine events and discard non-relevant ones.

PERSEO ASMT core processes analyse all potential interactions between aircraft within the airspace controlled by ENAIRE to detect separation minima infringements. The most common use cases are the detection of SMIs in *En-route* and T*erminal* airspaces. In these airspaces, a SMI occurs if the horizontal and vertical distances of a pair of aircraft fall below given thresholds. These thresholds are 5NM and 3NM respectively for the horizontal distance minima, and 1000ft for the vertical distance.

As it has been said, the core process of the PERSEO ASMT computes the horizontal and vertical distances for all potential pairs of aircraft that have flown in the airspace controlled by ENAIRE following a pure geometric approach. The track *timestamp* acts as a common variable to link the trajectories.

The horizontal accuracy of the surveillance systems allows the identification of horizontal infringements with high confidence levels. However, the altitude track data is provided by the surveillance with a granularity of 100ft (as given by the aircraft Mode-C), and in addition, anomalies may appear that can cause "fluctuations" in the recorded pressure altitude.

Consequently, many interactions fulfil the geometric conditions of being a SMI when actually they are not actual SMIs. These kinds of situations are usually induced by radar track's anomalies, as previously mentioned.

The PERSEO ASMT core algorithm implements different filtering steps (similar to those presented in [4]) in order to mitigate the detection of false SMIs. These filters allow to make the tool functional, but the analyses still required a post-filtering of false detections. The safety analyst can always access to the entire population of candidate SMIs.

For the sake of continuous improvement of the tool and the facilitation of larger safety analysis [5] without the need of human intervention for SMI post-processing, the PERSEO ASMT pursued the introduction of a supervised classifier to discriminate between genuine and non-genuine automatically.

This paper presents the methodology, results and the architecture in production of the supervised classifier.

II. Application of the CRISP-DM Methodology

The challenge is addressed as a typical data science process, following the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology [6]. The methodology comprises the following stages: *Problem Statement*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, and Deployment.

A. Problem Statement

1) Problem requirements

The goal of the PERSEO ASMT is to provide to the safety practitioner a list of genuine SMIs that occur within a given period and airspace, minimising the number of non-genuine ones.

The scope of the problem is limited to SMI that occurred within **Spanish** En-route airspace, i.e., above 24,500ft (FL245). This limitation was included to introduce in a controlled manner the new technical component to the process.

In addition, the SMIs between military flights were excluded as the performance of military differs from the commercial aircraft. For example, a sharp dive is a very common manoeuvre for military aircraft, but if it appears in a commercial aircraft vertical profile, it implicates either a fluctuation or an accident.

The problem was addressed as a *classification* problem, where the output of the model is to determine whether a given SMI is **genuine** or not. In this regard, ENAIRE Safety Monitoring Team provided a list of SMI for the considered period that was treated as the **ground truth**.

The model should work in a daily basis, embedded within the extraction, transformation and loading processes that feed CRIDA's Data warehouse (DWH). An initial description of the data model which supports the ASMT function is provided in [7].

In addition, the model should be capable of working with the data already stored, and these data (from 2013) shall be retrofitted with the output of this model. This poses a limitation, as no new data could be used, and data transformations feeding the model should be kept to a minimum.

2) Evaluation of the Baseline Performance

The first step is to evaluate the current performance of the filters that are applied to the output of the PERSEO ASMT core processes, considering the ground truth provided by ENAIRE. This dataset is called from now on ***baseline***. The output of the PERSEO core processes is denominated **interaction**.

The evaluation of a classification problem relies on a reliable identification of True Positive (TP), i.e., genuine cases to be shown. The genuine cases are known from the **ground truth** provided by ENAIRE.

Considering the variables for the evaluation the following:

- **TP**: true positive, interaction correctly marked as genuine SMI.
- **FP**: false positive, interaction wrongly marked as genuine SMI.
- **TN**: true negative, interaction correctly marked as non-genuine.
- **FN**: true negative, interaction wrongly marked as non-genuine.

The evaluation metrics are the followings:

Sensitivity or true positive rate (TPR): indicates the proportion of genuine detected SMIs from the whole population of genuine SMIs.

$$TPR = \frac{TP}{TP + FN}$$

Specificity or true negative rate (TNR): indicates the proportion of cases that are correctly marked as negative (non-genuine SMI) from the non-genuine baseline sample.

$$TNR = \frac{TN}{TN + FP}$$

False positive rate (FPR): indicates the proportion of cases that are wrongly marked as positive from the non-genuine baseline detections.

$$FPR = \frac{FP}{TN + FP}$$

Precision: indicates the proportion of correctly detected positives from the population marked as genuine.

$$Precision = \frac{TP}{TP + FP}$$

F1 Score: is a measure of a classification accuracy, by combining the precision and sensitivity scores.

$$F1 = 2\frac{Precision * TPR}{Precision + TPR}$$

TABLE I. BASELINE CLASSIFIER PERFORMANCE

	Sensitivity	*Specificity*	*FP Rate*	*Precision*	*F1*
Sample	86.6%	69.4%	30.6%	71.4%	78.1%

It can be seen from the results reported in TABLE I that the baseline classifier had an 30.6% of FPR, and a F1 score if

78.1%. These figures made that the Precision score fell to levels that required a lot of post-analysis to detect genuine SMI, as the practitioner could always access to all the population of **interactions** (candidate SMIs).

B. Data understanding – Taxonomy of Cases

The second step was to understand the underlying cases that might cause the appearance of non-genuine SMIs among the detected interactions.

The team carried out an analysis on a large sample of interactions for this purpose. The analysis resulted in a taxonomy of the different cases that characterises an interaction in its vertical layer. It is summed up in TABLE II, where it is also shown whether those interactions should be initially considered as genuine or not.

The following subsections provide more details for the categories A-E.

TABLE II. TAXONOMY OF CASES

Category	Definition	**Genuine?**
A	Tracks affected by fluctuations or other kind of anomalies	0
B1	Flight established at FL XX100 o XX900 ft during a short period of time: between cruising flights (including holding pattern)	0
B2	Flight established at FL XX100 o XX900 ft during a short period of time: at least one of them is climbing or descending (including holding pattern)	1
C	Flight established at XX100 o XX900 ft	0
D	At least one of the two flights is climbing or descending	1
E	Small infringement and short duration	1
M	Between military flights	Out of the scope
V	Between VFR (visual flight rule) flights.	Out of the scope
Level Bust	Level Bust	1

1) Case A

In this case, the *geometric* SMI is produced due to anomalies in the track. Therefore, it is not a genuine SMI.

Most of the anomalies appears in the vertical plane, this is, a fluctuation in the flight level. They could be small fluctuations of (100ft-200ft) or larger ones (200ft to 1000ft), for example.

An example of this case is shown in Figure 1. It illustrates the visual output of the PERSEO ASMT HMI. On the left-hand side of the figure, it can be observed the geographical evolution of the trajectories involved in the interaction. The right-hand side shows the altitude of both trajectories, the horizontal separation at the centre, and the vertical separation at the bottom. It can be seen that vertical minima suffers a jump from

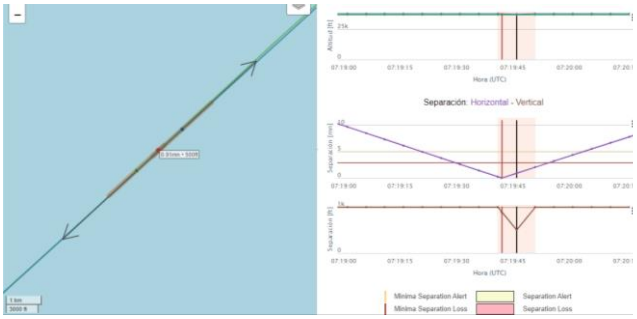

Figure 1. Case A – Vertical track fluctuations

1000ft (no separation infringement) to 500ft (separation infringement) in just one track. This is considered a *non-genuine* SMI and a fluctuation.

2) Case B

As it has been said previously, the Mode C of the aircraft reports the aircraft pressure altitude at intervals of the closest 100ft. Therefore, established flights can cruise with a 100ft deviation with respect its cleared level, it could be caused by various reasons, and it is hypothesised in this paper that this might be due to two different causes:

a) Case B1 – Mode C lack of precision

In the attempt to have the criterion as clear as possible, the category B1 is further divided into 3 groups. Figure 2 presents a simple sketch of them.

The red line represents the track with anomalies, the blue one is the other flight participating in the interaction, and the light red rectangle indicates the period of separation minima infringement. The different cases presented in the aforementioned figure are described as follows:

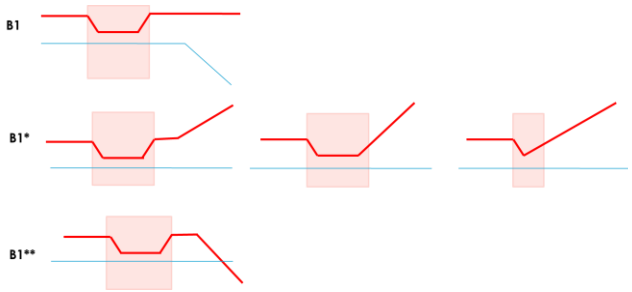

Figure 2. Case B1 – Instrumental Lack of precision

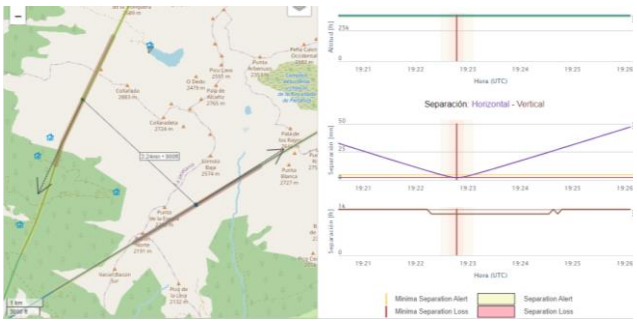

Figure 3. Example of Case B1

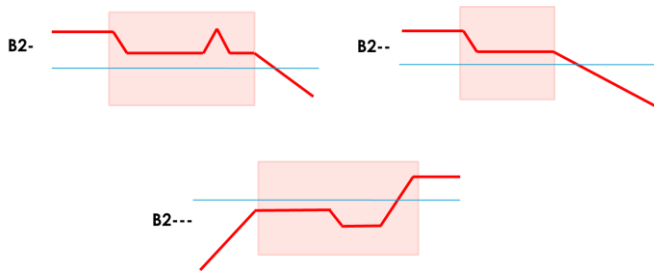

Figure 4. Case B2 – Potential Pilot Actions

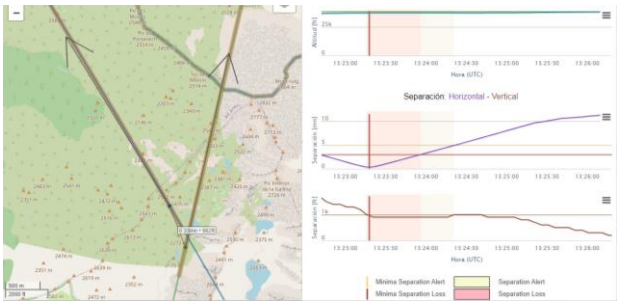

Figure 5. Example of Case B2---

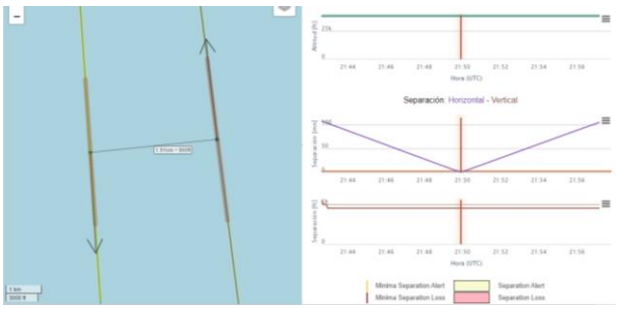

Figure 6. Example of Case C

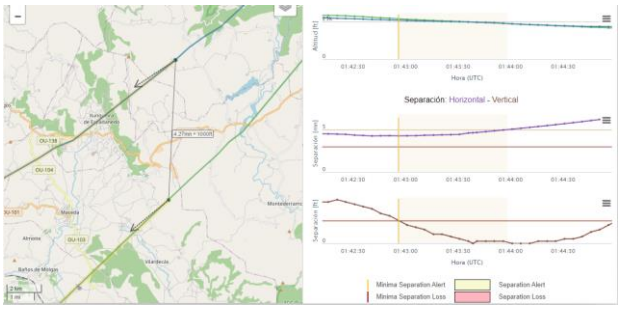

Figure 7. Example of Case D

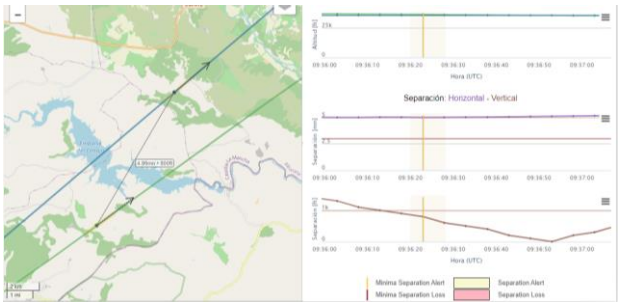

Figure 8. Example of Case E

- **B1**: the "classic" one, with a valley in the middle of the cleared level (in cruise phase). This is illustrated with an example in Figure 3.
- **B1***: shortly after the valley, the flight starts moving in the opposite direction. The second B1* example in Figure 2 shows a flight (red line) which has descended 100ft and immediately after the valley, it starts climbing. The reason of considering this case as non-genuine is because when an aircraft is planned to climb, it is unlikely that the pilot would decide to descend first. The first B1* example is actually the same as B1, but it is included in this group due to the global form it has, which could induce some doubts.
- **B1****: it is similar to B1* but for descents.

b) Case B2 – Potential Pilot or ATC Actions

The category B2 is also split in three subcategories. The SMI always appears in the first segment of a descent or the last segment of a climb. These cases are considered as genuine SMIs because the infringed segment in all 3 groups seems to be the product of a very slow climb or descent.

The reason of their appearance could be intentional, in order to wait until the minimum horizontal separation is reached before clearing further a descent or a climb (in order to avoid further reducing the vertical separation while the horizontal one is already infringed). The safety practitioners considered that these might be genuine, and that should be assessed in a case by case, so this should be shown by the classifier.

- **B2-**: it refers to a "classic" B2 but with a short fluctuation in between. The fluctuation is irrelevant because it is not the cause of infringement.
- **B2--**: "Classic" B2.
- **B2---**: Similar to B2-, the only difference is instead of a peak it is a valley. An example of this subcase is illustrated in Figure 5.

3) Case C-E

These three cases are simpler in their conception, so they described in this subsection together.

- **Case C:** A flight could be established during the whole or most of cruising phase at XX1 o XX9 ft, (example FL311, FL309 respectively). It is similar to the Case B1. An example is illustrated in Figure 6.
- **Case D:** Typical SMI when an aircraft or both change their flight levels. An example is illustrated in Figure 7.
- **Case E:** The infringement is slight and with a short duration, and therefore, its degree of safety relevance is also small. Usually, **D** cases usually also presents the characteristics of an **E** case. An example is illustrated in Figure 8.

The other remaining cases are between **military flights**, flights under VFR and Level Busts. The first and the second are not discussed in this paper as they fall out of the scope, whereas the latter has been widely discussed in different papers and safety notices [8].

C. Data Preparation

1) Selection of Sample for Training and Testing

The data preparation phase should abide to the constraints that were detailed in Section II.A.1). In this regard, the dataset for training the supervised classifier (see next section) was selected including the genuine dataset provided by ENAIRE.

A subset of genuine cases was easily identifiable from the dataset provided by ENAIRE, but the sample was deemed too small for training. In addition, it should be noted that only SMIs that are categorised as A (Alpha) or B (Bravo) by applying the Risk Assessment Methodology of EUROCONTROL should be reported within the SKPI. In this sense, this dataset was biased towards these types of incidents, so it should be completed with SMIs which are considered as non-safety relevant. Examples of these could be when the SMI is infringed when the aircraft are already diverging (i.e. the horizontal distance has already reached its minimum).

On the other hand, the interactions population as provided by the PERSEO ASMT core processes was much larger. A randomly selected and balanced sample, discarding the genuine SMIs provided by ENAIRE was selected and labelled following the previous taxonomy.

1) Selection of Parameters

The current PERSEO ASMT core process is embedded in a larger Extract, Transform and Load (ETL) process, which is run in a daily basis and supports the entire pipeline for CRIDA's DWH [9].

The inclusion of additional transformation processes to those already existent was a potential option, but should be considered very carefully. These processes went through quality verification checks before their deployment in production, and their modifications are a last-resource option.

As a result, only accessible data during this process are considered. In this sense, the PERSEO ASMT core processes access to the surveillance information, and therefore, the transformation of this information was considered key.

TABLE III presents the series of features that were finally selected for training the model. Some of them (marked as ***Initial PERSEO ASMT Core Process***) are an output of the core process for detecting the interactions between aircraft.

Among them, the most relevant are *Convergency* and *Track Anomaly*. These are the result of transformation processes that try to find the convergence of two aircraft before their closest point of approach (CPA); and whether *any* of the tracks that compose the potential SMI present a fluctuation.

TABLE III PARAMETERS CONSIDERED FOR THE MODEL (THE LAST COLUMN INDICATES THE ORIGIN OF THE DATA)

Parameter's name (In Spanish)	**Parameter**	**Type**	
FLTTYP1	Flight type of flight 1	1: M or S 0: rest	Initial PERSEO ASMT Core Process
FLTTYP2	Flight type of flight 2	1: M or S 0: rest	
%H (x_H)	Horizontal separation/ Horizontal Separation minima	Percentage	
%V (x_V)	Vertical separation / Horizontal Separation Minima	Percentage	
Alt. (Alt_)	Altitude	Numeric values (discrete FL)	
Anomalías en traza (Anomal_asEn Traza)	Original track anomaly identified by PERSEO.	1: no anomaly 2: fluctuation (100) 3: fluctuation (500) 4: garbling	
Convergente	Convergency state when the SMI starts.	1: convergent 0: divergent	
Duración (s) (Duracion_s_)	Duration of the SMI	Numerical values	
Duración crítica (s) (Duracion_critica_s_)	Duration of from instant of min. separation until the end of the SMI	Numerical values	
Mean_alt_A	Mean of the modulus operation of the FL and 10, of Flight 1 during t_{CPA} ± 30s	Numerical values	Transformed from Surveillance after the core PERSEO AST Process
Mean_alt_B	Mean of the modulus operation of the FL and 10 for Flight 2 during t_{CPA} ± 30s	Numerical values	
std_vSep	Standard deviation of the vertical separation during t_{CPA} ± 30s	Numerical values	
mean_vSep	Mean vertical separation during t_{CPA} ± 30s	Numerical values	
std_hSep	Standard deviation of the horizontal separation during t_{CPA} ± 30s	Numerical values	
mean_hSep	Mean horizontal separation during t_{CPA} ±	Numerical values	

The rest of the variables are transformations that are easily obtainable if the SMI period is available by using aggregation functions, powered by the database engines. These are marked as ***Transformed from Surveillance after the core PERSEO ASTM Process***.

Looking at the different cases of the taxonomy, the following considerations were made for the selection and generation of the new features:

- The percentages of horizontal and vertical separation are useful to detect the presence of conflict induced by

fluctuation. For example, fluctuations have typical values between as 100 ft and 500 ft in the vertical plane.

- The mean altitude of a flight within the t_{CPA} ± 30s can provide the unit number of the mean flight level. It detects the vertical changes.
- The mean and standard deviation of the two separation measurements can detect whether the flights have experienced any kind of fluctuations during the 60 s range around t_{CPA}.

These are illustrated by means of Figure 9.

III. MODELLING

The algorithm used for the classification is XGBoost (eXtreme Gradient Boosting), a supervised machine learning algorithm built within the gradient boosting framework [10].

The selection of XGBoost was driven by its performance in classification tasks, but also to the ease which it can be deployed within JAVA-based ETL pipelines (see Section V).

As a boosting algorithm, it benefits from working with a sequence of trees, one after another. This way of training enables the improvement of the model by correcting the predictions from the previous tree in each step of the training. And it reduces the error by reducing the bias, this is, the result will be closer to the real output but more disperse [11].

However, XGBoost presents many differences, in some case, enhancements compared to the original gradient boosting, such as:

- the possibility of setting different objective function
- the presence of regularization term to prevent overfitting
- the missing values handling
- the use of parallel boosting: the parallelisation does not refer to the train of independent tree in parallel like Random Forest. Instead, it happens within each tree of the model.

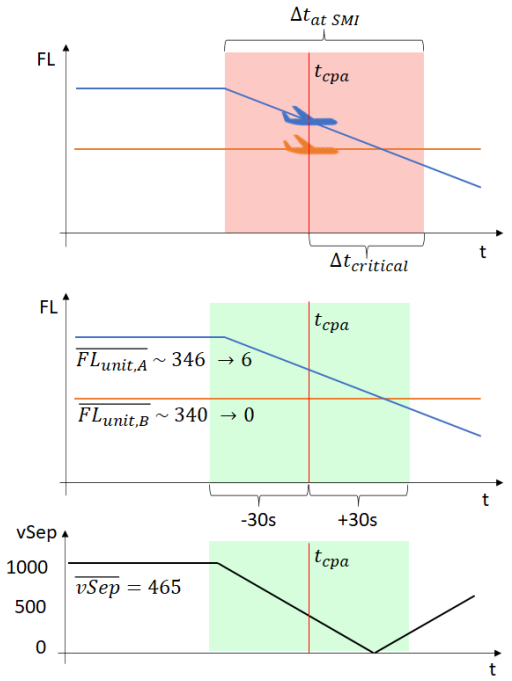

Figure 9. Graphical illustrations of the features

The XGBoost algorithm also provides a large number of hyperparameters that characterise the training process. They have to be tuned to acquire an acceptable performance. For that purpose, the Python Scikit-learn package [12] was used (see GridSearchCV) for the hyperparameter tuning. The model was trained for all possible combination for the defined hyperparameters' values, and the associated scores stored, selecting at the end the hyperparameters which provided the best one. The parameters and the selected range for analyses were:

- **Objective function: "binary: logistic"**. This function is prepared for binary classification problem. The model trained with this function gives the predicted probability as the variable to determine the predicted output, depending on the threshold that is considered.
- **max_depth**: it is the maximum depth that a tree can have. The range was [3, 5, 7, **9**]
- **min_child_weight**: minimum number of cases in a node to stop further split the sample. The range was [**1**, 3, 5]
- **eta**: it is the learning rate of the model. Adjustment step size. The range was [**0.05**, 0.1, 0.15, 0.2]
- **Gamma**: overfitting control parameter. It prevents the trees from fitting to noise. The range was [0, 0.1, 0.2, **0.3**, 0.4, 0.5]
- **Subsample**: sample size used in each tree. It also controls overfitting. The range was [0.6, 0.7, 0.8, **0.9**, 1]
- **colsample_bytree**: percentage of features (inputs) used in each tree. The range was [**0.5**,0.6, 0.7, 0.8, 0.9, 1]

The rest of hyperparameters were set to their default values.

As the dataset is balanced, it was not necessary to add the weight hyperparameter. The hyperparameters that provided the best performance score were selected (marked in bold in the previous enumeration).

IV. RESULTS EVALUATION

For testing the classifier with a new sample, a sample of 300 interactions classified with the new model were analysed. This sample was not used during the phase of training and initial testing.

Figure 10 and Figure 11 present the output of the XGBoost model for the analysed sample, which is a *probability* between 0 and 1 of the case being a genuine SMI. Figure 10 presents the case for those interactions that were not marked initially as anomalous, whereas Figure 11 illustrates those that were marked as anomalous by the PERSEO ASMT core process.

The blue dots are cases that the team considered to be a genuine SMI, whereas the red dots are non-genuine SMIs. It can be directly observed that the sharpest improvement is in the

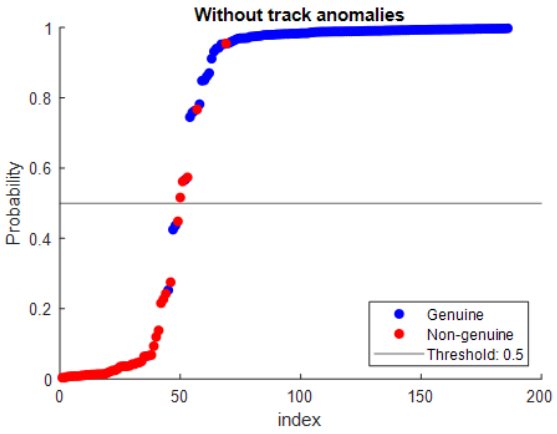

Figure 10. Output from the model for interactions not previously characterised as anomalous.

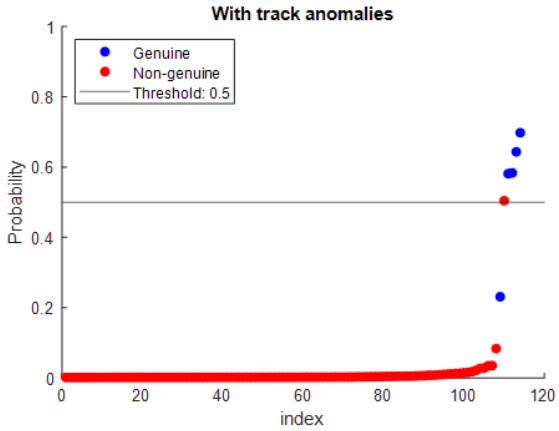

Figure 11. Output from the model for interactions characterised as anomalous

correct detection of non-genuine SMIs from those that passed the first filter.

TABLE IV presents the performance metrics for the baseline filters versus the XGBoost, should we define a **0.5** threshold for discriminating between genuine and non-genuine SMIs. It can be observed the sharp increment in performance provided by the XGBoost layer.

TABLE IV. PERFORMANCE OF XGBOOST VERSUS THE BASELINE, CONSIDERING A THRESHOLD OF 0.5

	Sensitivity	***Specificity***	***FP Rate***	***Precision***	***F1***
Filters Baseline	96.5%	69.5%	30.5%	73.1%	83.2%
XGBOOST	96.5%	96.2%	3.8%	95.8	96.1%

The FP rate decreased a 27%, whereas the precision increased 22 percentual points. In practical terms, this meant that, for this specific sample, the safety analyst only had to discard 1 false positive per every 25 candidates, whereas for the previous system, this was approximately between 1 per 3 and 1 per 4. This performance was deemed acceptable by the safety team and therefore, the development was stopped at this point without further iterations on the model. The SMIs shown in the previous figures include all types of SMI. The majority did not have any safety impact.

V. DEPLOYMENT

The last step of the methodology presented in the introduction of Section II is the deployment of the product in production. As it has been said, XGBoost was partially selected considering the advantages that it presented for its deployment. CRIDA DWH's data ingestion process is supported by data integration tools that are JAVA-based.

In order to accelerate the deployment of the model in CRIDA's data pipelines, the team followed the process presented in Figure 12.

The training and testing of the model were conducted in a development environment, powered by Python. In the meantime, the specifications for wrapping the model within the pipelines were provided to the data engineering team. As soon as the model was validated and its performance accepted, the model was deployed in the data pipeline and the orchestration process (i.e., updating prior registers and launching it in a daily basis) was triggered.

Also, in parallel, the new visualisation requirements were elicited and tested. A key requirement was that all the information should be available for the safety practitioner. The practitioner can select the threshold model for showing or hiding candidate SMIs. It can be set to a higher threshold, which will guarantee fewer false positives. On the other hand, the threshold could be relaxed in order to identify more genuine SMI, at the cost of more false positives.

In addition, it has been defined an automated labelling through PERSEO ASMT Core, in order to update the model once there is a sufficient sample of new labelled data, which will contribute to a continuous improvement of the model

VI. CONCLUSIONS

Safety is paramount in Air Traffic Management. In this regard, ANSPs in Europe should report in a yearly basis the ratio between Separation Minima Infringements with a safety impact divided by the total number of controlled flight hours within their controlled airspaces.

For this purpose, the Automatic Safety Monitoring Tools were developed to detect automatically SMIs. PERSEO ASMT provides a functionality for detecting potential SMIs between flights, including a filter for showing only genuine SMIs.

This paper presented the workflow followed for deploying a Machine Learning model (XGBoost) in production, also retrofitting for all the data available (since 2013) to increase the detection performance of SMIs.

The paper presented a taxonomy of vertical anomalies in the tracks, features that were selected for the model as well as the hyperparameters selected.

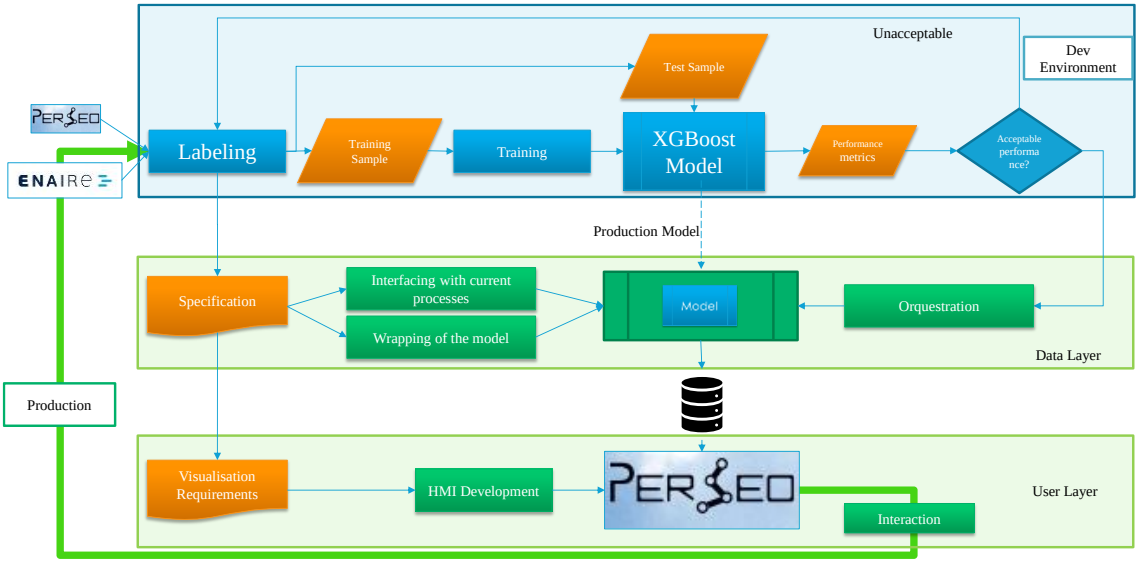

Figure 12. Deployment of the model

The performance of the model was verified against the baseline performance, with a sharp increment on the precision of the model, gaining full acceptability by the teams in charge of its development and its exploitation. To illustrate the gains, the FPR decreased by a 27%, meanwhile the precision increased by a 22%.

The full deployment of these technologies can pave the way for an effective and trustable deployment of *Safety Intelligence*. For example, the identification of trends for eliciting recommendations in a large basis can only be supported by data in the origin with a very high level of integrity. In addition, these studies support SESAR funded project such as FARO [13].

ACKNOWLEDGMENT

The authors would like to thank the entire past and present PERSEO team at CRIDA, and especially Jurgen Treml. The authors would like to thank the Safety Monitoring Team of ENAIRE for their continuous support, especially Mónica García Medina and Leticia Sánchez-Palomo Bermúdez.

REFERENCES

[1] European Commission, *Commision Impementing Regulation (EU) 2019/317 of 11 February 2019 laying down a performance and charging scheme in the single European sky and repealing Implementing Regulations (EU) No 390/2013 and (EU) No 391/2013*, no. OJ L 56, 25.2.2019 1–67. .

[2] European Aviation Safety Agency (EASA), "Acceptable Means of Compliance and Guidance Material for the implementation and measurement of Safety Key Performance Indicators (SKPIs) (ATM performance IR)," no. 2. Cologne, 2014, [Online]. Available: https://www.easa.europa.eu/sites/default/files/dfu/Annex to ED Decision 2014-035-R.pdf.

[3] S. Pozzi *et al.*, "Safety monitoring in the age of big data: From description to intervention," *Proc. 9th USA/Europe Air Traffic Manag. Res. Dev. Semin. ATM 2011*, pp. 391–399, 2011.

[4] R. Palacios and R. J. Hansman, "Filtering Enhanced Traffic Management System (ETMS) altitude data," *Metrol. Meas. Syst.*, vol. 20, no. 3, pp. 453–464, 2013, doi: 10.2478/mms-2013-0039.

[5] I. Buselli *et al.*, "Natural language processing for aviation safety: extracting knowledge from publicly-available loss of separation reports," *Open Res. Eur.*, vol. 1, p. 110, 2021, doi: 10.12688/openreseurope.14040.1.

[6] R. Wirth and J. Hipp, "CRISP-DM: towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39," *Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min.*, no. 24959, pp. 29–39, 2000, [Online]. Available: https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining.

[7] C. E. Verdonk Gallego, V. F. Gómez Comendador, F. J. Sáez Nieto, G. Orenga Imaz, and R. M. Arnaldo Valdés, "Analysis of air traffic control operational impact on aircraft vertical profiles supported by machine learning," *Transp. Res. Part C Emerg. Technol.*, vol. 95, no. July 2017, pp. 883–903, 2018, doi: 10.1016/j.trc.2018.03.017.

[8] European Organisation for the Safety of Air Navigation (EUROCONTROL), "European Action plan for the Prevention of Level Bust." Brussels, 2004, [Online]. Available: https://www.skybrary.aero/bookshelf/books/244.pdf.

[9] M. G. Martínez, J. G. Moreno, and G. Sendino, "Resilient Arrival Runway Occupancy Time prediction for decision-making tool in Barcelona (LEBL) airport .," 2020.

[10] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, vol. 42, no. 8, pp. 785–794, doi: 10.1145/2939672.2939785.

[11] D. Nielsen, "Tree Boosting With XGBoost Why Does XGBoost Win "Every" Machine Learning Competition?," *Tree Boost. With XGBoost - Why Does XGBoost Win "Every" Mach. Learn. Compet.*, no. December, 2016, [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2433761.

[12] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 127, no. 9, pp. 2825–2830, Jan. 2012, doi: 10.1289/EHP4713.

[13] FARO Consortium, "FARO Results - saFety And Resilience guidelines for aviatiOn," 2020. https://cordis.europa.eu/project/id/892542/results.