Applicability of Traditional and AI Models in Predictive Maintenance in Aviation

Dimitri Reiswich
Lufthansa Technik AG, AVIATAR, Hamburg, Germany *dimitri.reiswich@lht.dlh.de*

ABSTRACT

The optimization of technical operations in the aviation industry is critical to reduce both, costs and operational incidences. Aircraft and maintenance data as well as advanced algorithms are the key ingredients for such an optimization. The rise of AI algorithms combined with an increase in the available data to feed those algorithms provides a challenge to the traditional approaches. A question of interest is which model type should be used for a specific problem. In order to illustrate the difference in modelling approaches, we will introduce a hydraulic leakage case as a basis for our discussion. The application of traditional modelling approaches is compared to that of an AI based time series prediction. We will show that traditional use cases are better suited for the introduced problem but also discuss situations where AI would be advantageous. We will then turn our attention to the question when and how such models can be used in a strictly regulated industry like aviation and what the potential future governance could look like.

1. HYDRAULIC LEAKAGE USE CASE

An Airbus A320 has 3 hydraulic systems (blue, green, yellow) which are all responsible for different parts of the aircraft. Figure 1 shows a representative time series of a green system. The hydraulic oil quantity in liters is displayed on the y-axis. The x-axis represents the corresponding flight index which is an integer representing consecutive flights. The quantity was measured at a specific time during the flight. The figure shows characteristic changes in the time series, we'll call each of the identified phases cycles. We notice an overall downward trend of the time series which demonstrates that the hydraulic oil quantity is typically decreasing after each flight. The explanation for this is that the hydraulic system is not 100% isolated and some oil naturally leaves the system during standard flight operations. Cycle 1 is a standard cycle followed by an abrupt jump after which cycle 2 starts. The abrupt jump was caused by a maintenance event where hydraulic oil was refilled. Such maintenance actions

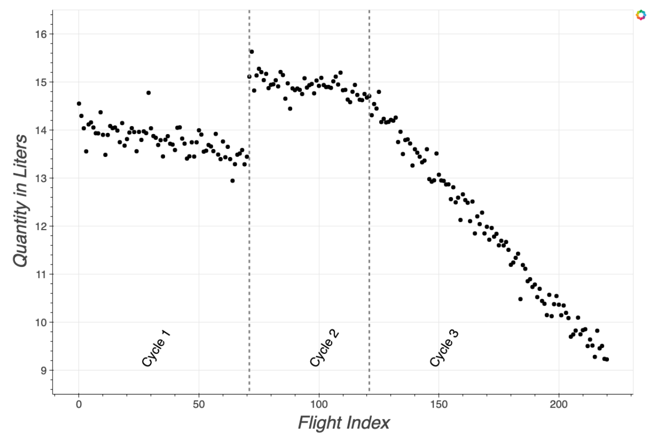

Figure 1. Airbus A320 hydraulic quantity time series for the green system. The flight index is an integer representing consecutive flights. Each point is a snapshot measurement of the quantity within a flight. Vertical dashed lines represent different phases of the system, called cycles.

are done on a regular basis to prevent hydraulic oil over- and underfills. After index 120 we observe a much larger negative slope in the time series, which is the beginning of the leakage cycle 3. Leakages can happen due to cracks in the hydraulic cylinders or deterioration of system components. We are interested in a model which helps to identify and alert when leakages occur such that maintenance actions can be taken as soon as they are identified.

2. TRADITIONAL MODELLING

A clear separation between a traditional model compared to an AI model is difficult. Linear regression has been used in the 19th century (Wikipedia, 2023, Linear Regression), but is also part of the standard literature about machine learning, see (Hastie, Tibshirani, & Friedman, 2009). For the scope of this paper, traditional models are loosely defined as models developed before the recent 21st century advances in AI model development[1].

[1]Note that scientific work on neural networks has been successfully performed in the 20th century, see (Macukow, 2016) for a more detailed coverage of the history of neural nets. However, many of the current state-of-the-art models were developed in the 21st century.

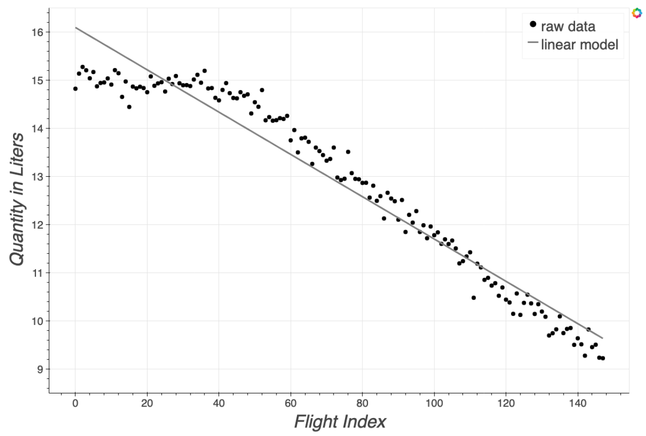

Figure 2. Linear fit to all the data.

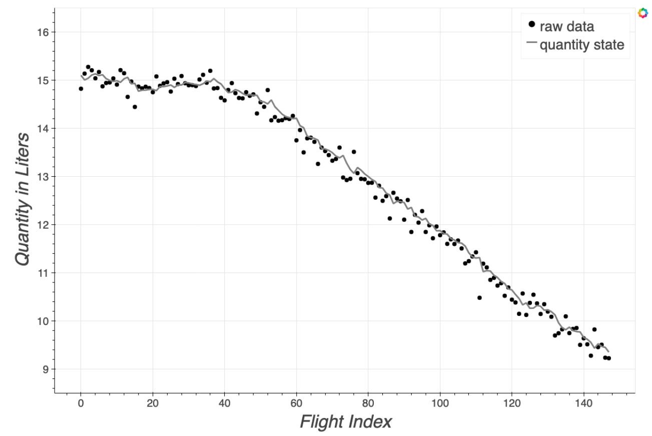

Figure 3. Quantity estimate for the state space model as defined in Equation 3.

For simplicity, we will focus on modelling the time series covering cycle 2 and 3 to avoid dealing with the jump. The resulting model can then be applied to individual cycles. The following traditional models could be chosen to model the system:

1. A linear regression model.

2. A discrete linear state space model.

We will develop a model architecture for both of them.

2.1. Linear Regression

In a linear setup the data is modelled as

$$y_i = \beta_0 + \beta_1 t_i + \epsilon_i \tag{1}$$

where y_i is the hydraulic quantity at flight cycle t_i. Modelling the quantity via a linear regression has two challenges. Firstly, we observe some large outliers which typically occur when the hydraulic system is in active use at the time of the measurement. A standard linear regression is sensitive with respect to outliers. Robust linear regression models should be used instead, see (Yu & Yao, 2017). However, a more difficult challenge to the linear model is the sudden change of slope initiating the 3rd cycle. A linear fit to all the data will clearly not perform well as the slope changes during the time series, see Figure 2. The time series is piece-wise linear. Change point detection algorithms could be used instead, see (Killick & Eckley, 2014). The output of such an analysis would be parameters β_1^k which are slopes identified for cycle k. Most interestingly, β_1^k would represent the actual leakage rate which can be monitored. However, note that a typical monitoring application requires an **online** prediction where leakage rates are estimated and predicted on each new incoming data point. Most change point algorithms work in an offline setup, which needs to be addressed.

2.2. Discrete Linear State Space Model

State space models are popular in control engineering, see (Bar-Shalom, Li, & Kirubarajan, 2001) for a comprehensive introduction. In such a setup, we distinguish between the observed variable y_i and the true unobserved oil quantity q_i and leakage rate l_i which are both hidden. The model assumes that the observed quantity is the hidden quantity polluted with noise. The observed hydraulic quantity y at flight cycle i can then be modeled as

$$\begin{aligned} y_i &= q_i + \epsilon_i^y && (2)\\ q_i &= q_{i-1} + l_{i-1}(t_i - t_{i-1}) + \epsilon_i^q && (3)\\ l_i &= l_{i-1} + \epsilon_i^l && (4) \end{aligned}$$

The variable y_i is the unobserved (hidden) state quantity q_i plus the observation noise ϵ_i^y at flight cycle t_i. The hidden quantity is driven by its previous value plus the leakage rate l multiplied by the cycle difference $t_i - t_{i-1}$. The hidden quantity is also driven by a white noise term ϵ_i^q. The leakage rate itself is also a hidden state variable and modeled as a sum of the previous value plus its own noise term ϵ_i^l. This is a more complex setup compared to the linear regression system and the estimation can be performed using the Kalman filter, see (Chui, Chen, et al., 2017). Covariance matrices for the noise terms need to be defined but the discussion of the parameter tuning is outside the scope of this paper. Despite the more complex setup, we have the advantage to explicitly model both, the quantity and the leakage rate as individual processes.

The results of the estimate for the given problem are presented in Figure 3 and 4. Figure 4 shows the estimated leakage rate and demonstrates how the model can cope with the abrupt changes of the leakage and adapts accordingly. The leakage rate l_i is explicitly modelled as a stochastic variable which is expected to change as implied by the variance of the noise term ϵ_i^l. Imagine that we would actually create an

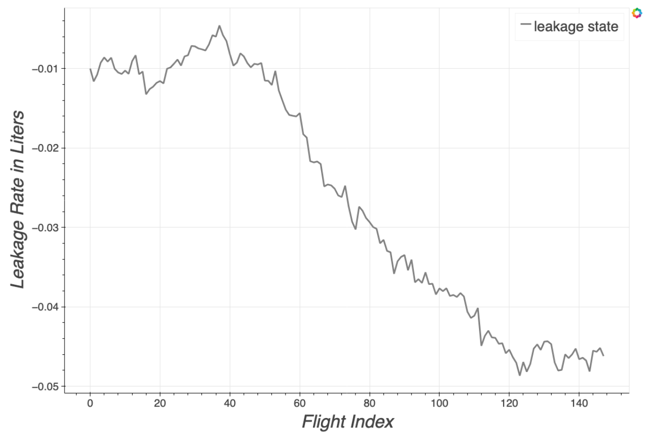

Figure 4. Leakage rate state estimate for the state space model as defined in Equation 4.

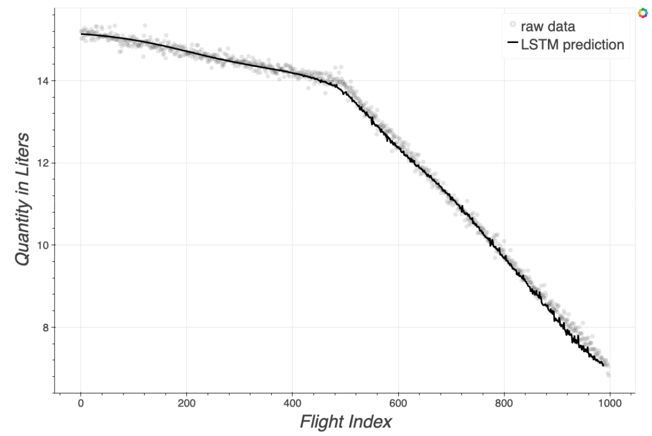

Figure 5. Quantity prediction by a LSTM.

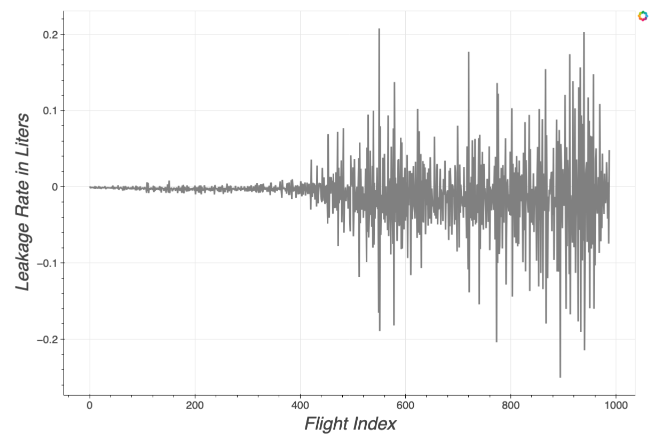

Figure 6. Leakage prediction by an LSTM.

alert based on the latest estimate of the leakage rate. Such an alert should be robust regarding data noise to avoid creating too many false positives. While the linear estimator in Equation 1 can produce any leakage estimate, the range of leakage values for the state space models is determined by the noise distribution in Equation 4. By choosing a suitable variance, we can control how large a typical update can be which leads to a more robust estimator. Finally, the state space model is working natively in an online fashion which can be used for real time alerting.

3. AI MODELS

There are various AI models which can be applied to time series forecasting, see for example (Masini, Medeiros, & Mendes, 2023), (Ahmed, Atiya, Gayar, & El-Shishiny, 2010), (Zhao, Lu, Chen, Liu, & Wu, 2017). In such a setup, the hydraulic quantity y is modeled as a dependent variable

$$y = f(x) + \epsilon \tag{5}$$

where x is an independent variable, f is some fixed but unknown function and ϵ is an error term, see (Hastie et al., 2009). The independent variable x can contain any input features, in particular the flight cycles t_i but also previous values of the quantity y. The unknown function f is approximated by a neural net class, for example as a recurrent or long short-term memory network (LSTM) (Hochreiter & Schmidhuber, 1997). In practice, a neural net type and architecture are chosen and optimized such that, for example, the fitted function $\hat{f}$ minimizes the mean squared error

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2 \tag{6}$$

given n observations. The observed quantities are again y_i and x_i are the observed independent variables (features) at time t_i. The quantity at cycle $i+1$ can be predicted by $\hat{f}(x_{i+1})$, the leakage rate itself as $\hat{f}(x_{i+1}) - \hat{f}(x_i)$. We will use an LSTM network to perform our analysis. LSTM networks, often employed for time series analysis and sequential data processing, have the strength of learning long-term dependencies. As features, we will take both the current cycle and the previous quantity, so the network can potentially incorporate autoregressive features similar to Equation 3.

The fitted time series for the quantity is shown in Figure 5. A first look at the quantity figure looks promising, the LSTM seems to have fitted a curve which does reasonably well at approximating the input data. The corresponding leakage rate is shown in Figure 6. The leakage is very noisy as the LSTM fit is not smooth enough which impacts the first order derivative. Leakages as high as 200ml are reported, and the estimation couldn't be used for practical purposes without applying some smoothing to the original curve. However, even if the fit to the quantity looks ok, one should ask: What has the network actually learned? As it is a black box, it's hard to tell until one looks at the prediction of the model on unseen data. This is illustrated in Figure 7 where we ask the model to predict on a time series which doesn't have a leakage. If the

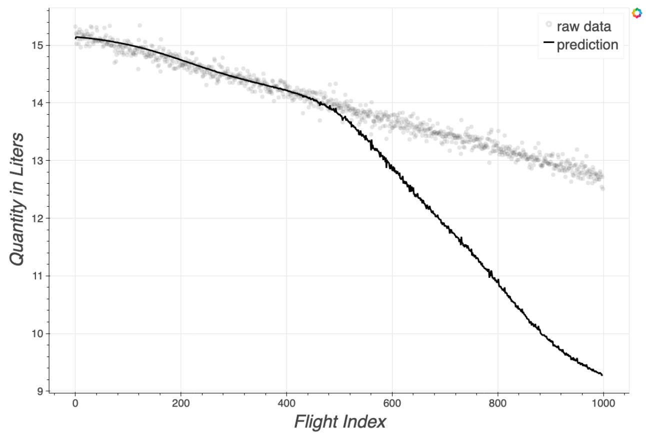

Figure 7. Quantity prediction by an LSTM on unseen data which doesn't include a leakage.

model has learned autoregressive features, it would be able to predict correctly by adding a constant leakage rate to the previous value. Looking at the figure shows that the model has learned the time dependence with the exact point of the previous leakage. It applies this to the new data where no leakage is present. Training the model on non leakage data and applying it to leakage data shows the same problem. Note, that the bad out of sample performance is not driven by extrapolation as the out of sample data range is well within the domain of the training data set.

This example shows the dangers of blindly applying AI models to any problem, focusing too much on finding the best fit without questioning and challenging the model. Testing AI models is critical to assess their reliability, robustness, and applicability.

There are some situations, where an AI model can be advantageous. If we have many observed features which all have an influence on the system, it might be hard to explicitly model the dependence in a state space model as the system dynamics need to be known. As a concrete example, the hydraulic quantity is highly dependent on the outside temperature at the time of the measurement. The volume is higher when temperatures are high. The quantity dynamics are consequently not only driven by the flight cycles and the current leakage rate but also the outside temperature which should be included as a factor in the estimation. This is fairly easy to do in both, Equation 5 and 1 by simply introducing the temperature as an independent variable. Explicit modelling work has to be done in the equation system 3 though, but a standard physical law can be incorporated easily into the equation. When more variables are involved and the relationship is not clear, more complex models could provide a valid alternative.

4. Model Governance in Predictive Maintenance in the Aviation Industry

In aviation, the role of predictive maintenance algorithms has been historically limited by regulatory driven safety requirements. For example, consider the case of a maintenance interval for a specific aircraft component, like the hydraulic system introduced earlier. Such an interval is typically determined conservatively via an approval process involving the aviation authorities and the original equipment manufacturer. Both, the manufacturer and the authorities have little visibility of the component lifecycle as the corresponding data is owned and acquired by the airlines or companies maintaining those components. If data is available, it is natural to start applying analytics and models to monitor actual failures with the goal of calculating an optimal maintenance interval which is in line with the data. This helps to evaluate the suitability of the originally proposed maintenance intervals. The outcome of this could point to shorter or longer maintenance interval times. Interval times which are calculated to be shorter than originally assumed are problematic as failures happen more often than expected. Longer interval times are in the interest of airlines, as they would save maintenance costs. However, even if regulators agree to the analytical approach, letting algorithms determine maintenance intervals raises follow-up questions related to responsibility should a safety related incidence happen. This topic involves complex ethical discussions and will not be the focus of this section. Instead, we will assume that algorithms will be approved to do certain tasks and discuss how the setup could look like including authorities.

A data driven approach always includes some form of analytics which generate insights from the data. This can range from simple statistics (aggregations, histograms) to more advanced models. In case of a more complex model, the following questions need to be answered:

1. Is the chosen model suitable for the given problem?
2. Is enough data available to fit the model?
3. What is the uncertainty of the model outcome?
4. How sensitive are the model parameters, i.e. how robust is the model?
5. Can model predictions be explained?
6. What is the model risk, i.e. the risk of choosing the wrong model?

Answering those questions would involve debates in a scientific environment as scientists have different opinions on each of the mentioned points. This will not be different when authorities are involved who eventually have to agree and approve the proposed methodology. While aviation authorities seem to have started developing governance around AI models (EASA, 2021), this process is still in an early stage. Another example where authorities cover analytical models is given in (FAA, 2018). The document explicitly mentions

Weibull and Pareto analysis as potential analytical tools without being explicit on how such an analysis should be performed in order to be accepted by the regulators. Sometimes, it helps to look into other industries to find examples of an existing process, and we will consider the financial industry as an example. While aviation and finance are different and safety regulation is not comparable with the regulation of financial markets, financial authorities have a very mature and well established model governance process. The financial industry has developed sophisticated models to price and risk manage financial instruments starting as early as in the 1970s when the famous Black Scholes pricing model was published (Black & Scholes, 1973). Since then, very sophisticated models were developed to price complex financial instruments. Financial instrument prices are calculated by using stochastic calculus, Monte Carlo simulations and partial differential equations. Given the importance of the financial industry and their large role in the stability of economy, regulators need to guarantee that the pricing and risk methodology is robust and accurately reports risk and balance sheets. How is this done in practice? Example references and discussions are provided in (Federal Deposit Insurance Corporation, 2005), (Federal Reserve, 2022). Firstly, a split in responsibilities is imposed on the financial institutions where the role of a model developer is separated from the role of a model validator. Both of the roles have different incentives. The model developer is primarily interested to develop a model to solve a business problem and get it into the production. The model validator independently validates that the model is reflecting the true financial risk by running validations, stress tests and model reviews. The validator is not subject to the pressure from the business side, the organizations and reporting lines are separated. The regulators require all big financial institutions to have such a setup. In addition, regulators have their own teams of quantitative modelers whose role is to audit, validate and challenge the models. All production models need to be extensively documented and regulators validate them in order to assess their quality and applicability. This process is well established and shows how such a cooperation between the industry and authorities could potentially look like in a world where complex models are in use.

In aviation, we could think about a slow start where algorithms are applied to non-safety critical components and approved through an established process by regulators. In a first step, modelling related organizational structures should be created in both, the authorities and the industry. A governance process related to the performance, suitability and deployment of the algorithm should then be established. Once the process is established and accepted by both parties, the component coverage could be extended. Eventually, one might think about a similar setup to the financial industry where model development and model validation are split and authorities run own analytical teams who audit the process.

5. Conclusion

After introducing a predictive maintenance case for an aircraft, we have presented traditional and AI modelling approaches to solve the problem. We concluded that traditional models with explicit modelling of the dynamics are advantageous and that blindly applying AI models isn't recommended. We also discussed that a more complex multivariate system where explicit modelling is challenging could require AI. Finally, we briefly discussed a potential future model governance structure for the discussed models by looking into the setup in the financial industry.

Acknowledgment

The author would like to thank Francesco Marino and Marius Schneider for the fruitful discussions and valuable feedback.

References

Ahmed, N., Atiya, A., Gayar, N., & El-Shishiny, H. (2010, 08). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, *29*, 594-621. doi: 10.1080/07474938.2010.481556

Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, *81*(3), 637–654.

Chui, C. K., Chen, G., et al. (2017). *Kalman filtering*. Springer.

EASA. (2021). *Easa concept paper: First usable guidance for level 1 machine learning applications* (Tech. Rep.). Author. https://www.easa.europa.eu/en/easa-concept-paper-first-usable-guidance-level-1-machine-learning-applications-proposed-issue-01pdf.

FAA. (2018). *Reliability program methods— standards for determining time limitations* (Tech. Rep.). U.S. Department of Transportation Federal Aviation Administration. https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/1035253.

Federal Deposit Insurance Corporation. (2005). *Model governance.* https://www.fdic.gov/regulations/examinations/supervisory/insights/siwin05/siwinter05-article1.pdf. ([Online; accessed 08-June-2023])

Federal Reserve. (2022). *Approach to supervisory model development and validation.*

https://www.federalreserve.gov/publications/2022-supervisory-stress-test-methodology-approach-supervisory-model.htm. ([Online; accessed 08-June-2023])

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Killick, R., & Eckley, I. (2014). changepoint: An r package for changepoint analysis. *Journal of statistical software*, *58*(3), 1–19.

Macukow, B. (2016). Neural networks–state of art, brief history, basic models and architecture. In *Computer information systems and industrial management: 15th ifip tc8 international conference, cisim 2016, vilnius, lithuania, september 14-16, 2016, proceedings 15* (pp. 3–14).

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of economic surveys*, *37*(1), 76–111.

Wikipedia. (2023). *Linear regression — Wikipedia, the free encyclopedia.* http://en.wikipedia.org/w/index.php?title=Linear\%20regression&oldid=1153729683. ([Online; accessed 08-June-2023])

Yu, C., & Yao, W. (2017). Robust linear regression: A review and comparison. *Communications in Statistics-Simulation and Computation*, *46*(8), 6261–6282.

Zhao, B., Lu, H., Chen, S., Liu, J., & Wu, D. (2017). Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, *28*(1), 162–169.

BIOGRAPHIES

Dimitri Reiswich is the Head of Data Science at Lufthansa Technik's AVIATAR. He holds a Diploma in Business Mathematics from the University of Hamburg as well as a PhD in Financial Mathematics from the Frankfurt School of Finance and Management. Dimitri has substantial experience in advanced mathematical and statistical modelling. He spent the first years of his career in the financial industry where his role was in the area of financial instruments and markets modelling. He then joined Lufthansa Technik to lead the Data Science team which develops advanced predictive maintenance models with the goal of optimizing scalable technical operations of AVIATAR's clients.