Increasing Operational Profitability through Better Flight Arrival Time Predictions

**Customer Profile**
Lufthansa Industry Solutions

**Location**
Germany

**Industry**
Airlines

**The Customer**
Lufthansa Industry Solutions is a consulting firm that works both for businesses within the Lufthansa family, and outside companies. Lufthansa Industry Solutions helps clients digitize and automate their business processes, and recently this has included using RapidMiner to apply data science to a variety of problems and opportunities.

**Executive Summary**
- Lufthansa needed a better prediction of when flights would arrive at their destination
- An accurate arrival prediction at the time the aircraft takes off allows time for the operations team to make adjustments at the destination to accommodate lateness
- Destination adjustments improve the passenger experience and reduce operating costs for Lufthansa
- RapidMiner enabled the Lufthansa Industry Solutions team, a consulting practice within Lufthansa, to leverage a large and diverse data set to build better arrival predictions than what was previously available

**About Lufthansa and the RapidMiner user group**
Lufthansa is the largest German airline, and when counting its subsidiaries, the largest airline in Europe, both in terms of fleet size and passengers. Headquartered in Cologne, and operating out of hubs in Frankfurt and Munich, Lufthansa flies worldwide. Its subsidiary airlines include Austrian Airlines, Swiss International Air Lines, Brussels Airlines, and Eurowings, including Germanwings.

The business is broader than just its airlines, though. Lufthansa also owns a number of aviation-related companies, such as: Lufthansa Technik (maintenance, completions, repair, and overhaul - MRO - services for aircraft), AirPlus (business travel payment solutions for corporations) and LGBS (finance, purchasing, human resources, and revenue accounting for a variety of companies).

The RapidMiner user group is part of Lufthansa Industry Solutions, yet another business within Lufthansa. Lufthansa Industry Solutions is a consulting firm that works both for businesses within the Lufthansa family, and outside companies. Lufthansa Industry Solutions helps clients digitize and automate their business processes, and recently this has included using RapidMiner to apply data science to a variety of problems and opportunities.
Lufthansa’s need: better predict arrival times of departing aircraft

When a Lufthansa plane arrives late at its destination, it does more than inconvenience passengers. In fact, what may seem like a tolerably small arrival delay for a passenger, say, less than thirty minutes, can have a severe operational impact on Lufthansa. A delayed arrival may impact catering service, flow of flight crews between aircraft, gate availability, connecting flights for passengers, and more. This can add up to significant costs for Lufthansa, especially in cases where a passenger misses a connecting flight.

On the other hand, the farther in advance that Lufthansa can tell that a flight will be delayed, and by how much, the more able it is to make adjustments to minimize costs - such as rescheduling catering, reassigning crew, changing gates to be closer to to departure gates for connecting passengers, pre-ordering shuttles between gates for passengers, or proactively rebooking them on later flights.

“We wanted to predict the actual arrival time of a flight to within a 5 minute window, at the very moment the plane took off,” according to Dr. Stanislaw Schmal, Analytics Project Manager for Lufthansa Industry Solutions. “This would typically give the Lufthansa operations team enough time to make any necessary adjustments at the arrival destination to minimize passenger inconvenience and costs for Lufthansa.”

RapidMiner enables Lufthansa to build powerful models using a large, diverse data set

The Lufthansa Industry Solutions project team started by assembling a large and diverse data set, any or all of which the team felt might have valuable predictive power. The data was comprised, naturally, off profiles of previous flights. But the team also supplemented it with other data that it felt might have a role to play - such as weather data, radar data, and other related factors. With RapidMiner, the team was easily able to prep this data for modelling, no matter how voluminous the data set became.

“Some data transformation was necessary and important to the success of the modelling effort,” said Dr. Fabian Werner, a data scientist on the Lufthansa team. “And it wasn't all automatic, as we needed to make some important judgement calls on how best to prep the data for modelling. For example, we decided to convert the timestamps of flight departures and arrivals from a linear scale to a measure of flight duration, which made more sense for our goals.”

RapidMiner enabled the team to create and refine model prototypes very quickly, which was critical to building the best models. “You never get it right the first time,” Dr. Werner said. Through this fast prototyping, the team not only found the best modelling approach, but also gained important insights into which data had the most predictive impact. For example, the start time of aircraft de-icing was not very helpful for predicting arrival time. Headwind speed, distance to destination, and time of year of the flight were important, unsurprisingly. But also important were some unexpected considerations, such as which runways planes took off from or landed on, and even how many passengers on the flight had connections to make (the team believes pilots may “try harder” to arrive on time if they know a lot passengers will miss connections if late).
In all of this, RapidMiner plays a pivotal role, but is not used in a vacuum. The Lufthansa Industry Team's technology ecosystem includes Hadoop, Oracle, and Teradata for data storage, and they still use R and Python for some of their work.

The data science team also built a simple front-end UI for the Lufthansa operations to use, for inputting flight numbers and quickly getting back a prediction of arrival time. To increase these business users’ confidence in the outputs of the modelling, they can also request historical views of past flights, to see what arrival times were predicted, and what the actual arrival times were.

“We need to build a tool that doesn’t just run well, but is actually used,” said Dr. Werner. “We involved the business users throughout the entire process, to make sure what we created in the end would serve their needs.”

RapidMiner saves Lufthansa money with better arrival time predictions

Lufthansa was already making arrival time predictions for each flight before the Industry Solution team's work began. But the team has determined that its new models built with RapidMiner are significantly more accurate at the time of aircraft take-off, that critical moment when the operations team want to be able to start making adjustments at the flight's destination if needed. This has resulted in cost benefits for Lufthansa in the tens of millions range.

“We have been able to show the business how much our work has reduced costs,” said Dr. Schmal. “Not only can we show it for each time period, but also cumulatively since the time our models went to work. It creates a compelling case for the value of applying data science to this problem.”

Future plans: models that update over the course of the flight based on real-time data

Arrival time predictions will soon get even better for Lufthansa. The Industry Solutions team is working on expanding its modelling approach so it can predict arrival time not just at the moment of aircraft take-off, but throughout the flight as well. This will involve processing real-time data that becomes available during the flight, updating the predicted arrival time continually.

“There’s even more we can do with RapidMiner to improve the service we provide our guests and decrease operating costs,” said Dr. Schmal.