

When a delay secretly brings joy

Any disruption to a train timetable is annoying. For Deutsche Bahn and for its passengers. With the up-to-the-minute traveler information from its new forecasting machine, the company is saving its customers unnecessary waiting times. It's a service that, from this year on, will go far beyond just long-distance travel.

COPY — Thomas van Zutphen

“ **W** *e talk for hours and hours about the sweet and the sour ...”* – Jenny’s favorite Ed Sheeran song suddenly pops into her head, but there’s no time for hours of chatting right now.

With a glance at the time – it’s just before 9 a.m. – she has just put down her mobile phone a second ago after hastily ending a call with her friend Kerstin with the words “You’re a sweetheart!” Because, from now on, every minute counts.

Who expects to be invited to a meeting with Ed Sheeran in Hamburg on a Saturday morning, in Jenny’s case right between her shower and breakfast? What a start to a weekend that had up until then she had expected to be pretty boring! Okay, “meeting” is perhaps a bit of an exaggeration. But a live concert by a star whose music has already given her and her friend many hours of joy, that’s quite something. Jenny had known for weeks that Kerstin’s sister had managed to get two tickets last summer. That alone was a stroke of luck, and you could only feel good for her. Almost every one of the British superstar’s few German concerts had sold out within minutes.

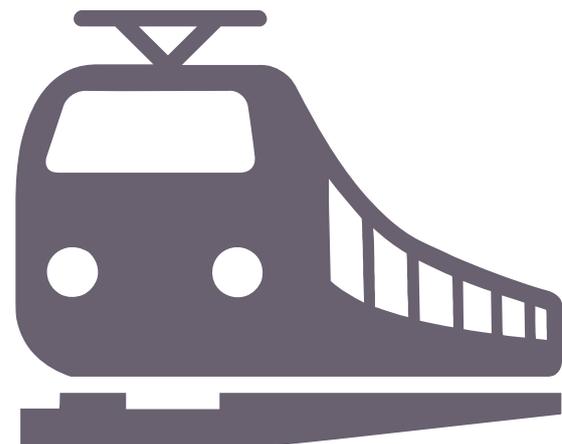
Now Kerstin was looking for a replacement for her sister of all people – “... that’s why I’m calling – she’s sick, with a 102 degree fever, so she can’t possibly...” – Jenny almost felt guilty. But with a quick look at the time – “The train won’t be quite on time, it only leaves at five past 10 from the main station” – she made a quick calculation: “I can make it.” Jenny is happy that the train is late today of all days. She’s actually a punctual person. But the DB Navigator app told the two women that she has a bit of extra time today. Jenny regularly travels by train and has been annoyed sometimes when a train has arrived too late. Especially when she was informed about the delay on the platform with only vague information like “in a few

Deutsche Bahn
operates up to

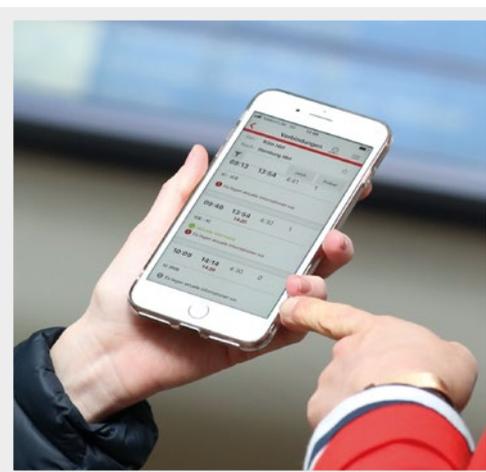
1,387

trains a day on its long-distance
network alone.

**Reliable forecasts in real time require
big data and intelligent algorithms.**



Rail passengers know: Not every train can be punctual. They therefore appreciate all the more receiving timely information about delays, in order to be able to make good use of the additional time.



minutes.” She gives brief thanks for the fact that her wish for a possible delay and timely information about it has been granted, thanks to Deutsche Bahn’s new forecasting machine. A closer look at the DB Navigator shows her that her friend was right. Instead of departing on schedule at 9:46 a.m., the Intercity train will leave Cologne 19 minutes later. If she hurries now, there might even be time for a quick breakfast at the station.

Jenny calculates that 20 minutes will be enough for a lightning quick check in the bathroom mirror, a quick change of clothes and packing the essentials for the night. Unlike her friend who lives in downtown Cologne, Jenny lives in the suburb of Brühl. It’s a good 12 miles from there to the city center, depending on the route. Taking the RB26 train to the central station, where Kerstin would wait for her, could work. But what if that train is also delayed, for example due to a disturbance on the track? Jenny decides against it.

“Leave and get in a taxi” – she thinks that following the advice from Ed Sheeran’s “Shape of you,” seems the better idea today. “9:20 a.m., Parkstrasse 17,” the taxi office confirms that a driver will pick Jenny up on time. And when she

steps out of the elevator, Jenny can already see the taxi waiting at the front door. She tells the driver her destination and takes a seat in the back – she looks at the app again. And indeed – should the taxi ride be delayed on the motorway ring to the A555 in the direction of the city center, this still shouldn’t be a cause for panic. But a quick check of the updated information from the forecasting machine shows: The IC 2310 on its way to the North Sea island of Sylt will now leave Cologne Central Station at 10:03 a.m. Nevertheless: That’s still 17 minutes later than scheduled, Jenny calculates and immediately informs her friend via WhatsApp. And she answers immediately: “Great! If it stays that way and you’re on time, there’ll even be enough time for an Iced Latte 😊😊😊”. At the thought, Jenny immediately thinks of Ed Sheeran’s “Cold Coffee.”

“Tell me if I’m wrong, tell me if I’m right”, it says in the song, but she hasn’t needed these kinds of hurried prayers for a while when travelling by train. That’s because for months, the forecasting machine has been making real time and down-to-the-minute forecasts of the arrival and departure times of all Deutsche Bahn long-distance trains. Jenny recently heard something about a “learning system”



Instead of spending unnecessary time waiting on the platform, users of DB Traveler Information can synchronize their arrival at the platform with that of their train down to the minute.

on the radio. She always wanted to read about it and plans to do so later (see box on p. 39). As expected, her taxi driver knows the way to Cologne Central Station even without a navigation device. But when his passenger asks him when they'll reach their destination, the man at the wheel just taps his dashboard display: "There's a little traffic jam at the turn-off into the city center. But we'll arrive at 9:45."

The driver is right. At a quarter to ten Jenny hands over the taxi fare and tip, gets out of the car and sees Kerstin waving to her in front of the station. Her friend approaches her with the words: "Do you know what's the best thing? There's another small correction. But the departure time of our IC is now one minute past 10. That's cool twice over!" "What do you mean?" "I just see it that way," she explains her view of things to Jenny: "A delay is a delay. But if you as a passenger are informed in time, you can use the time, make the best of it and don't have to rush to the platform to find out that the rush was for nothing." And Kerstin immediately calculated what that means for the two women right at that moment: "As a result, we have enough time not just for a coffee, but also for a croissant."

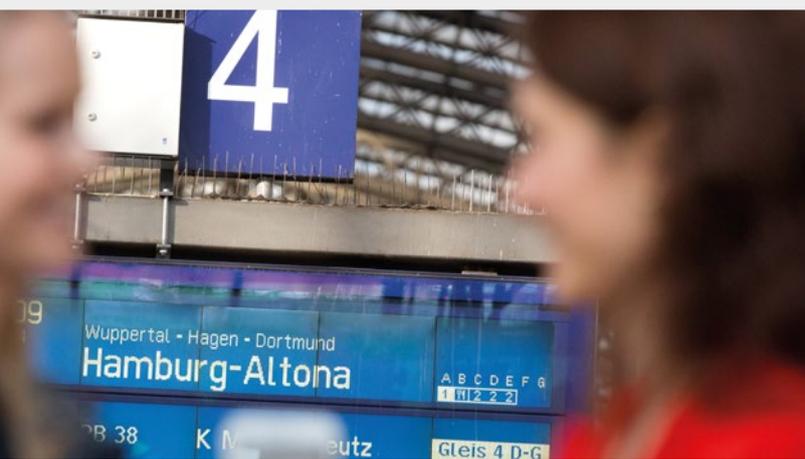
"My belly's sick to its stomach ...," Ed Sheeran sings in the ballad "Miss You." And that would have been exactly Jenny's feeling previously, so that she wouldn't have been able to eat a bite in this situation. The fear of missing the train – at least this particular one – would have hit her too hard in the stomach. But now, the two women are happy, everything is fine. At the café, Kerstin explains their itinerary to her friend in detail. The whole journey without changing trains takes just over four hours. But before the concert she wants to check in at the hotel. And anyone who really wants to "get closer" to Ed Sheeran at the concert at the race track in Hamburg's Bahrenfeld district, i.e. in the area 11 to 13 yards in front of the stage, should get there early.

On their way to platform 4, Kerstin wants to look at the app again, just to be certain. When booking the train tickets for herself and her sister, she had "ticked" via the app that she

wanted to be informed automatically about changes in the departure time of "IC 2310." Her smartphone therefore notifies her of any changes to the timetable both acoustically and visually. But better safe than sorry.

"Give a little time to me or burn..." – well, it's clear that sums up the way train passengers perceive every minute they wait on the platform as wasted time. The forecasting machine continues to predict "10:01 a.m." as the final departure time. "We can actually have a quick look in the station pharmacy." Minutes later the two women climb the last few steps up to the platform and see their Intercity already arriving into the station from the south. Exactly a quarter of an hour late, "... but also somehow on time, down to the last minute," to the delight of the two women.

 Joachim.Betz@t-systems.com



Why algorithms train every day

T-Systems has implemented a learning system for Deutsche Bahn to forecast the departure and arrival times of trains. Both the hypothesis and the driving force is the assumption that patterns are hidden in the historical delays of train connections, whose identification will immediately contribute to a better prediction of delay situations (see interview p. 40).

For this purpose, the T-Systems experts divided the functionality of the forecasting machine into two components. A learning component performs non-real time-critical pattern recognition and provides the process with complex mathematical models. These so-called support vector machines (SVMs) are essentially permanently teaching the system how to more precisely predict incremental delays on the route segments of each individual train. As soon as a result is available, it is read by the machine in the real time component, processed and used to create a forecast for the running time of a train within seconds.

In this ML process (ML = Machine Learning) different algorithms have to work together reciprocally. An SVM algorithm, for exam-

ple, makes it possible to make predictive statements about the "timetable accuracy" of trains based on the characteristics of the train journeys and the accompanying delays of the past. Another one produces static predictions for train connections whose routes and journey numbers are fixed. It analyses statistically which static delay value would have best affected the respective arrivals and departures of the trains in the past. However, since journey characteristics and delay patterns change almost daily, every 24 hours the algorithms undergo a nightly retraining of their modeling procedures.

During the forecasting process, the machine reacts in real time to new incoming timetable information, current train positions and manual entries from the control center. It is conceivable, for example, that the system could also react in the future if a star who is playing a concert – such as the above mentioned Ed Sheeran – spontaneously decides at the end to give not one or two encores, but five or six.



Joachim Betz

Joachim Betz, Principal Transport Solution at T-Systems, in conversation with Peter Schütz, Head of Traveler Information at Deutsche Bahn AG.

Learning from data

Mr. Schütz, what exactly is the “forecasting machine”?

The forecasting machine is used to provide customers with highly precise information on train arrival and departure times along their planned itineraries and to show alternatives in the event of deviations from the plan. This information is “played” centrally by the forecasting machine from a “Single Point of Truth” to all of our output media. These are, for example, the display boards in the train stations, various apps, websites and travel portals in the network.

What’s the status quo?

On long-distance services, the best networked system in the country with up to 1,000 trains a day, our passengers and those picking them up at stations have been benefiting from forecasting machines since May of last year. An improved travel experience in this form was also the aim of the project when we launched it in 2015. The initial question was: How do we get the best information on connections or delays for the arrival and departure times of our trains? Our hypothesis at the time was: Huge amounts of data and intelligent algorithms should make it possible to make and deliver a more reliable forecast in real time than our old forecasting system could achieve. The algorithm always calculates the actual data in relation to the current operating status of the customer’s itinerary. And that’s refreshed every minute. Particularly on long-distance routes, the operating status can change every minute. When a customer travels directly from A to B, one minute is usually meaningless. However, a minute can be decisive when changing trains. That’s why it’s important that we get a good forecast for changing trains at an early stage and can offer the customer a consistently reliable travel chain forecast.

What kind of data does the forecasting machine use?

Train-running information, scheduling decisions, but also secondary information that the system can obtain from delays of other trains and that the algorithm has learned to deal with. In today’s live operation of the forecasting machine, these are all inventory data that we guide through the algorithms. In the future, this will be supplemented by circulation information, weather data, GPS data, etc. in order to achieve a shorter and at the same time even more valid information cycle. On the technical side, there are only two levers of improvement for this service: on the one hand, the data – the diversity of its sources and their provision frequency – and, on the other hand, the algorithm.

Speaking of the “future” – what are you planning next?

The aim is to further improve the forecasting quality for long-distance services and, in parallel, to start integrating regional and metropolitan trains in 2019. It’s always a matter of learning from the data and deducing from it which improvements can be rolled out. That means nothing more than reading the right things from the data. If you like, it’s a standard process in the Big Data world. This also includes fast failing, sorting out data in order to say in a portfolio tunnel: These are the drivers with which we can once again create the largest hub for improving forecasts. The next step is to adopt these elements and introduce them into production. To this end, we want to successively expand our data source landscape and our long-term goal is to also include local public transport, i.e. to integrate buses, subways and suburban trains.



Peter Schütz

And when does a cooperation partner on the IT side come into play?

First of all: we use state-of-the-art technologies for the many millions of forecasts that are made every day. Microservices, for example, which consolidate the data by the minute, are automated and scalable, on Big Data platforms in the cloud. This gives us the scalability that allows us to be very flexible. Today, modern IT is a cloud operation. But the idea behind our cooperation with T-Systems is a different one: We bring the railway knowledge, while T-Systems brings the algorithmic knowledge – the knowledge about how to find the right algorithm to solve a certain problem. They provide the competent team that knows everything about analytics and can handle both AI and Big Data. If both groups, and not just the respective data scientists, work together in a highly integrative manner, then there is real power behind it. Analytical power, implementation power and a great deal of competence. We have made a really great start, but we are sure that we can continue to improve here as well. And with every improvement, we will raise our bar. That is the so-called “forecast quality.” In long-distance services today, 30 minutes before a train arrives, it is 87.5 percent. However, we want to aim for a forecast quality of 95 percent, ideally even 99 percent.

How will you achieve this?

A travel guide that is as comprehensive as possible needs information from all the modes of transport involved. And we’re talking about over 800,000 trips a day on public transport in Germany. That’s why we are already integrating a large number of external companies into the travel information platform so that we can supply the information channels consistently and well. We have deliberately said that we are

making the basic information available not only to DB companies, but also to all external transport companies, so that the entire industry can benefit. In the long term, this also includes air traffic, i.e. the real time supply of take-off and landing information to airlines and airports. That is the future. In principle, the ability to forecast is not a USP in itself. This is more of an industry standard. But to make forecasts of this quality, as the forecasting machine can already do for long-distance train services today, is something special compared to other countries. That’s why we are investing so that we can expand this service even further and make it faster.

You can’t get much faster than “real time”.

It’s not about systemic speed, but about the speed of project implementation. The quality of the algorithm accounts for 30 percent. The quality of the data accounts for 30 percent. But 40 percent is down to the way we work. At some points we still work redundantly as a team. We have to get to the market even faster through the way we work. That’s the next step. And that’s where I have confidence in the partnership. It’s working.



Joachim.Betz@t-systems.com