Deutsche Bahn Delay：
A Study of Official Statistical Bias and Empirical Modeling

Xiao SHEN, Yuanye XU

29.10.2025

1. Introduction

This study was motivated by a real travel incident that revealed the operational challenges of train travel in Germany. While taking trains operated by Deutsche Bahn (DB) in Germany, we were struck by the company's highly unpredictable punctuality. Sometimes trains were only ten minutes late—acceptable by most standards—but often delays exceeded an hour, accompanied by last-minute platform changes or even complete cancellations. Out of curiosity, we later examined DB's official punctuality statistics as published on its website and annual report, only to find a surprising discrepancy between the reported data and passenger experience.

Punctuality

Development in the year under review

PUNCTUALITY / %	2024	2023	2022
DB Group (rail) in Germany	89.4	90.1	90.9
DB rail passenger transport in Germany	89.5	90.3	91.0
DB Long-Distance	62.5	64.0	65.2
DB Regional [1)]	90.7	91.4	92.2
DB Cargo (Germany)	68.0	70.5	66.1
DB Regional (bus)	85.9	85.2	86.0
DB Cargo	68.2	69.7	66.3
Punctuality (whole journey) (DB Long-Distance)	67.4	68.9	69.3

Figure 1 DB Integrated 2024

According to DB Geschäftsbericht 2024, the overall punctuality rate for passenger rail transport in Germany was reported as **89.5%.** However, a closer look reveals that long-distance trains (DB Long-Distance) achieved only **62.5%**, while regional trains (DB Regional) reached 90.7%. A simple weighted calculation shows that long-distance services account for just 4.3% of total passenger traffic, while regional services make up 95.7%. As a result, the high punctuality rate of regional trains statistically "dilutes" the poor performance of long-distance services, masking the severity of long-distance delays. The gap between these official statistics and actual passenger experience motivated us to re-examine how DB defines and measures punctuality. Our goal is to reassess DB's reporting methodology and build a more realistic, data-driven delay prediction model that exposes structural biases in the official statistics.

2. Methodological Biases in DB’s Punctuality Statistics

2.1. Overly Lenient Definition of “Punctuality”

According to DB’s official documentation (Pünktlichkeitswerte - Zahlen & Fakten), “Züge gelten als pünktlich, wenn sie mit weniger als 6 Minuten Verspätung ankommen.” (“Trains are considered punctual if they arrive with less than six minutes of delay.”) In other words, any train arriving within 5 minutes and 59 seconds of schedule is officially classified as “on time”. This threshold is among the most lenient in Europe: France’s SNCF uses five minutes, while Japan’s JR applies a one-minute rule. By adopting the EU’s upper-limit tolerance of six minutes, DB’s official statistics tend to overstate its punctuality performance.

Punctuality

Share of on-time stops in relation to all stops along and at the end of lines in Germany. A stop is considered operationally punctual if the scheduled arrival time is exceeded by less than six minutes in passenger transport or by less than 16 minutes in freight transport. At DB Regional Rail, punctuality has been reported since 2024 on a third-weighted basis between S-Bahn (metro) alternating current, S-Bahn (metro) direct current and DB Regional Rail without S-Bahn (metro). At DB Regional Road, buses that leave more than one minute early have also been counted as unpunctual with retroactive effect since 2020. For the arrival time punctuality of passengers, see punctuality (whole journey).

Figure 2 Grossary of Punctuaclity

From a passenger’s perspective, however, even a one- or two-minute delay can disrupt tight transfer schedules, cause anxiety, or erode confidence in the system’s reliability -- especially in a densely interconnected rail network. The six-minute tolerance therefore fails to reflect the actual traveler experience and obscures the network’s underlying inefficiencies.

To address this methodological limitation, we adopt a stricter definition of punctuality, classifying a train as delayed if it arrives one minute or more after the scheduled time. We fully acknowledge that the operational complexity of the German rail network is not directly comparable to Japan’s relatively closed and highly standardized Shinkansen system. However, our purpose in adopting such a strict threshold is not to impose an unrealistic benchmark, but to enhance the sensitivity of the analysis, capturing even minor delays that meaningfully affect passenger experience and network performance.

This approach aims to uncover micro-level scheduling deviations and systemic inefficiencies often concealed under lenient tolerance rules, while also illustrating the tangible impact of minor delays on travelers’ perceptions. By adopting this passenger-centered and transparency-oriented perspective, we seek to provide insights that may help Deutsche Bahn enhance operational discipline and reaffirm its reputation for technical precision, ultimately contributing to a more reliable and efficient rail service.

2.2. Exclusion of Cancelled and Substitute Services

DB’s official punctuality statistics explicitly exclude Ersatzverkehr (substitute bus services) , whose delay rate, according to our data set, reachs 82%, and Ausfälle (cancelled trains), which exceed 6% of the proportion in our data set. This means an underestimation of true delay levels. From a passenger perspective, a sudden cancellation represents the most severe form of delay. In our analysis, **cancelled trains** are therefore treated as cases of extreme delay and included in the data set.

2.3. Oversimplified Classification System

DB divides passenger trains only into “long-distance” (ICE, IC, etc.) and “regional” (RE, RB, S-Bahn) categories. This simplification overlooks the operational and scheduling differences across service tiers. Although S-Bahn services account for the largest number of trains, they mainly operate within metropolitan areas and have limited relevance to intercity travel. In contrast, long-distance and regional trains (particularly RE and RB) determine the quality of medium- and long-haul passenger experience. Moreover, the high punctuality of S-Bahn services artificially inflates the average punctuality of the “regional” category.

To better reflect real operational structures, we introduced a three-tier classification: **long-distance, regional, and urban**. This refined taxonomy aligns with functional and scheduling characteristics and allows a more accurate understanding of how delays vary across the rail network.

Long-Distance	ICE, IC, EC, ECE, RJ, RJX, TGV, FLX, FEX, D, EN, NJ, UEX, ES, EST, WB
Regional	RE, RB, IRE, ALX RE, BRB RE, WFB RE, NWB, NWB RS, VIA, ME, MET, HLB RE, MEX, MEX a, MEX c, R, HBX, SVG
Urban	S, S X, SE

Figure 3 Three-tier Classification

3. Data Collection and Processing

Because DB does not publicly release detailed delay data, obtaining reliable samples was the first major challenge. The only source available is DB’s official real-time departure and arrival board, which refreshes every few seconds and displays only short-term operational data. This means that even if we conducted our own web scraping, the accessible data would be limited to a few days of recent records (e.g., several days in October), making it impossible to construct a long-term dataset.

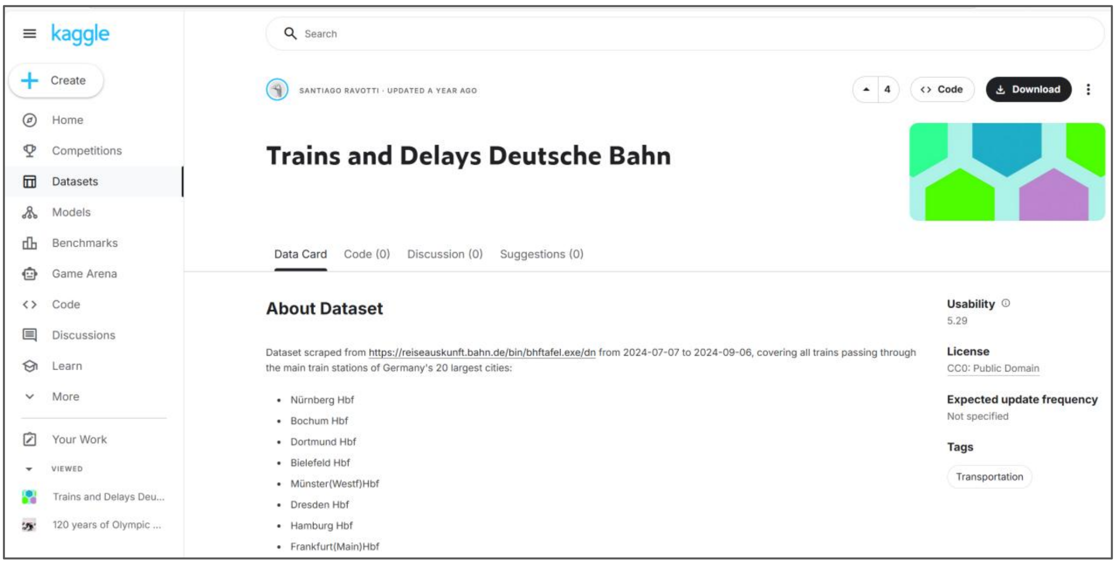

Figure 4 Datasets from *Santiago Ravotti* on *Kaggle*

To overcome this limitation, we therefore used a third-party dataset from Kaggle, compiled by independent developers through automated web scraping of DB's real-time arrival and departure boards. The dataset includes train records from July 20-25, 2024, and September 1-2, 2024, covering 20 major German cities. Although limited in time span, these samples are representative of typical operational conditions and are suitable for analyzing daily punctuality patterns.

date	Hbf	scheduled_	expected_time	train_model	route	platform	real_time_du	has_delay
2024/9/1	Köln Hbf	13:00	13:00	RE 8(10819)	Rommerskirchen, 13:00	3	13:00	0
2024/9/1	Hannover Hbf	13:00	13:00	S 2	Nienburg(Weser), 13:00	1	13:00	0
2024/9/1	Münster(Westf)H	13:00	13:00	RB 89(90018)	Paderborn Hbf, 13:00	3	13:00	0
2024/9/1	Hamburg Hbf	13:01	13:01	S 5	Hamburg Elbgaustraße, 13:01	4Hamburg Hbf	13:01	0
2024/9/1	Frankfurt(Main)	13:01	13:01	S 4	Kronberg(Taunus), 13:01	101Frankfurt	13:01	0
2024/9/1	Hamburg Hbf	13:02	13:02	RE 60(4197)	Westerland(Sylt), 13:02	11D-F	13:02	0
2024/9/1	Wuppertal Hbf	13:02	13:02	S 8	Mönchengladbach Hbf, 13:02	5	13:02	0
2024/9/1	Berlin Hbf	13:03	13:03	FEX19832	Flughafen BER, 13:03	6	13:03	0
2024/9/1	Hamburg Hbf	13:03	13:03	ME RE3(82120)	Uelzen, 13:03	12A-B	13:03	0
2024/9/1	Hamburg Hbf	13:03	13:03	S 1	Wedel(Holst), 13:03	3Hamburg Hbf	13:03	0
2024/9/1	Hamburg Hbf	13:03	13:03	S 2	Hamburg-Altona(S), 13:03	4Hamburg Hbf	13:03	0
2024/9/1	Hannover Hbf	13:03	13:03	S 5	Paderborn Hbf, 13:03	2	13:03	0
2024/9/1	Hannover Hbf	13:03	13:03	WFB RE60(95778)	Braunschweig Hbf, 13:03	12	13:03	0
2024/9/1	Frankfurt(Main)	13:04	13:04	VIA RB10(25015)	Neuwied, 13:04	23	13:04	0
2024/9/1	Nürnberg Hbf	13:04	13:04	RE 32(3492)	Lichtenfels, 13:04	16	13:04	0
2024/9/1	Hamburg Hbf	13:05	13:05	S 3	Hamburg-Neugraben, 13:05	1Hamburg Hbf	13:05	0
2024/9/1	Hamburg Hbf	13:05	13:05	S 3	Pinneberg, 13:05	3Hamburg Hbf	13:05	0
2024/9/1	Münich Hbf	13:05	13:05	ICE 529	Dortmund Hbf, 13:05	19	13:05	0
2024/9/1	Köln Hbf	13:05	13:05	RB 48(17337)	Wuppertal-Oberbarmen, Köln Me	6	13:05	0

Figure 5 Raw data

departure_city(H	route_cleaned	train_model_cleaned	scheduled_time	arrival_hour_bucket	expected_time	expected_delay	real_tim	has_delc	real_delay_min	causes	T	U	V	W
Berlin	Baruth(Mark)	RE	13:11	13	13:14	3	13:14	1	0					
Duisburg	K?ln/Bonn Flughafen	RE	13:11	13	13:14	3	13:14	1	0					
Hannover	Norddeich Mole	IC	13:13	13	13:14	1	13:14	1	0					
Bonn	Walporzheim	RB	13:13	13	13:14	1	13:14	1	0					
Hamburg	Stuttgart Hbf	ICE	13:14	13	13:14	0	13:31	1	17	Grund: Versp?tung eines vorausfahrenden Zuges				
Hannover	Koblenz Hbf	ICE	13:14	13	13:14	0	13:34	1	20	Grund: Technische St?rung am Zug				
Duisburg	Dresden Hbf	FLX	13:14	13	13:14	0	13:35	1	21	Grund: Vorfahrt eines anderen Zuges				
Düsseldorf	Kaarster See	S	13:14	13	13:14	0	Fahrt f?l	1						
Hannover	Koblenz Hbf	ICE	13:14	13	13:14	0	Fahrt f?l	1						
Köln	Berlin Ostbahnhof	ICE	13:11	13	13:15	4	13:15	1	0					
Münich	Herrsching	S	13:13	13	13:15	2	13:15	1	0					
Leipzig	Hannover Hbf	IC	13:14	13	13:15	1	13:15	1	0	Grund: Versp?tung eines vorausfahrenden Zuges,Gle				
Nürnberg	Regensburg Hbf	RE	13:15	13	13:15	0	13:24	1	9	Grund: Warten auf einen entgegenkommenden Zug				
Duisburg	Arnhem Centraal	VIA RE	13:15	13	13:15	0	13:23	1	8					
Nürnberg	Neustadt(Waldnaab)	RE	13:15	13	13:15	0	13:24	1	9					
Köln	Paris Nord	EST	13:15	13	13:15	0	13:26	1	11					
Frankfurt(Main)	Darmstadt Hbf	S	13:13	13	13:16	3	13:16	1	0					
Bochum	Solingen Hbf	S	13:13	13	13:16	3	13:16	1	0					
Hamburg	Flensburg	RE	13:14	13	13:16	2	13:16	1	0					
Hamburg	Kiel Hbf	RE	13:14	13	13:16	2	13:16	1	0					
Frankfurt(Main)	Glauburg-Stockheim	RB	13:15	13	13:16	1	13:16	1	0					
Düsseldorf	Solingen Hbf	S	13:15	13	13:16	1	13:16	1	0					
Duisburg	Bocholt	VIA RE	13:15	13	13:16	1	13:16	1	0					
Berlin	Düsseldorf Hbf	ICE	13:16	13	13:16	0	13:35	1	19	Grund: Versp?tetes Personal aus vorheriger Fahrt				
Köln	Bergisch Gladbach	S	13:13	13	13:17	4	13:17	1	0	Grund: Versp?tung aus vorheriger Fahrt				
Münich	München Flughafen Termi	S	13:15	13	13:17	2	13:17	1	0	Grund: Beh?rdliche Ma?nahme				
Münich	Freising	S	13:15	13	13:17	2	13:17	1	0	Grund: Beh?rdliche Ma?nahme				
Stuttgart	Weil der Stadt	S	13:15	13	13:17	2	13:17	1	0					
Stuttgart	Osterburken	MEX	13:15	13	13:17	2	13:17	1	0					
Dortmund	Berlin Ostbahnhof	ICE	13:17	13	13:17	0	14:37	1	80	Grund: Technischer Defekt an einem anderen Zug				
Duisburg	Koblenz Hbf	RE	13:17	13	13:17	0	13:43	1	26	Grund: Polizeieinsatz				
Münster(Westf)	Paderborn Hbf	RB	13:17	13	13:17	0	13:22	1	5					
Wuppertal	Wuppertal-Oberbarmen	S	13:17	13	13:17	0	Fahrt f?l	1						
Dortmund	Berlin Ostbahnhof	ICE	13:17	13	13:17	0	14:44	1	87	Grund: Technischer Defekt an einem anderen Zug				
Münich	Deisenhofen	S	13:15	13	13:18	3	13:18	1	0					
Essen	Osnabrück Hbf	RE	13:18	13	13:18	0	13:23	1	5	Grund: Versp?tung eines vorausfahrenden Zuges				
Frankfurt(Main)	W?chtersbach	RB	13:15	13	13:19	4	13:19	1	0					
Dortmund	Brilon Stadt	RE	13:17	13	13:19	2	13:19	1	0					
Berlin	Flughafen BER	RE	13:18	13	13:19	1	13:19	1	0					

Figure 6 Cleaned Data for Analysis

The raw dataset originally contained records for more than 30 train categories, including both rail and substitute bus services. To ensure consistency with official railway operations, we removed all bus (Ersatzverkehr) records and reclassified the trains into three operational categories according to their service characteristics: long-distance, regional, and urban.

1	date	Hbf	arrive_station	train_category	depart_hour_bucket	has_delay
2	2024/7/23	Münich Hbf	Praha hl.n.	regional	21	1
3	2024/7/20	Münich Hbf	Praha hl.n.	regional	21	1
4	2024/7/21	Münich Hbf	Praha hl.n.	regional	21	1
5	2024/7/24	Münich Hbf	Praha hl.n.	regional	21	1
6	2024/7/24	Münich Hbf	Praha hl.n.	regional	21	1
7	2024/7/21	Münich Hbf	Praha hl.n.	regional	21	1
8	2024/7/22	Münich Hbf	Praha hl.n.	regional	21	1
9	2024/7/22	Münich Hbf	Praha hl.n.	regional	21	1
10	2024/7/24	Münich Hbf	Praha hl.n.	regional	19	1
11	2024/7/21	Münich Hbf	Praha hl.n.	regional	19	1
12	2024/7/23	Münich Hbf	Praha hl.n.	regional	19	1
13	2024/7/20	Münich Hbf	Praha hl.n.	regional	19	1
14	2024/7/22	Münich Hbf	Praha hl.n.	regional	19	1
15	2024/7/23	Münich Hbf	Praha hl.n.	regional	17	1
16	2024/7/24	Münich Hbf	Praha hl.n.	regional	17	1
17	2024/7/22	Münich Hbf	Praha hl.n.	regional	17	1
18	2024/7/20	Münich Hbf	Praha hl.n.	regional	17	1
19	2024/7/21	Münich Hbf	Praha hl.n.	regional	17	1
20	2024/7/21	Münich Hbf	Praha hl.n.	regional	17	1

Figure 7 Washed data for modelling

After extensive cleaning, each record contained the following attributes: **date, departure station, arrival station, train category, departure time bucket, and punctuality status** (1 = delayed, 0 = on-time). We defined a train as delayed if its arrival delay exceeded one minute, aligning with the revised punctuality standard adopted in this study. In addition, trains that were canceled for any reason (e.g., technical failure, construction, or substitution) were also treated as delayed and assigned a value of 1 in the binary variable “has_delay”.

4. Descriptive Analysis

After the initial data cleaning and preprocessing, we observed that across all train records (*Figure 6*), the average actual delay time was **4.6 minutes**. Among the trains that experienced delays, the average delay duration increased to **8.84 minutes,** with the maximum delay reaching **over 400 minutes**. Regarding the **causes** of delays, **track changes** (*Gleisweckseln*) accounted for 3.8%, delays caused by preceding trains represented 13.3%, **construction work** (*Bauarbeiten*) contributed 1.6%, and **technical issues** such as maintenance depot problems made up 7.31%. Among all delayed services, 1.3% were recorded as “Verkehrt” (wrong routing), and 5.9% were marked as “Fällt aus” (cancelled).

date	departure_city(	route_cleaned	train_model_cleaned	scheduled_time	arrival_hour_bucket	expected_time	expected_delay	real_tir	has_del	real_delay_min
2024/7/23	Berlin	Budapest-Nyugati	EC	18:43	18	18:43	0	1:35	1	412
2024/7/24	Berlin	Budapest-Nyugati	EC	18:43	18	18:43	0	1:37	1	414
2024/7/24	Berlin	Budapest-Nyugati	EC	18:43	18	18:43	0	1:40	1	417
2024/7/23	Dresden	Budapest-Nyugati	EC	16:50	16	16:50	0	23:48	1	418

Figure 8 Longest Delay Minutes

Furthermore, after basic filtering and statistical analysis, we found significant differences across train categories: the **S-Bahn** showed a delay rate of 36.9% with an average delay of 1.95 minutes, **ICE** trains had a delay rate of 77.3% and average delay of 16.91 minutes, **RE** services showed 57.1% and 6.69 minutes, **IC** services 65.3% and 13.60 minutes, **RB** trains 45.4% and 3.74 minutes, while **EC** trains exhibited the highest delay rate of 78.1% with an average delay of 41.57 minutes.

Train Type	Delay Rate (%)	Average Delay (min)
S-Bahn	36.9	1.95
ICE	77.3	16.91
RE	57.1	6.69
IC	65.3	13.60
RB	45.4	3.74
EC	78.1	41.57

Figure 9 Delay Information of Common Train Models

The analysis revealed clear structural patterns. Long-distance trains achieved an on-time rate of only 24.84%, meaning roughly one in four arrived as scheduled. Regional trains reached 48.74%, showing moderate stability. Urban trains (S-Bahn) performed best at 62.90%, reflecting the resilience of short-distance commuter operations.

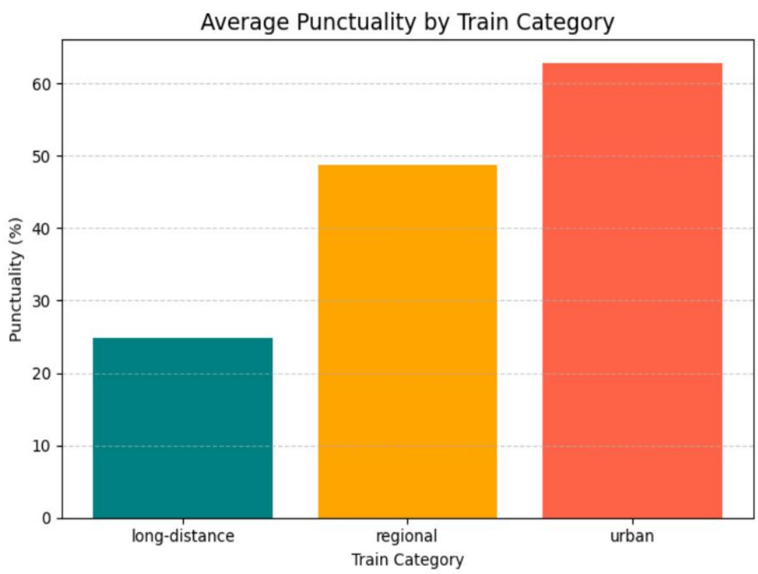

Figure 10 Punctuality by train category

Across the day, punctuality followed an inverted-U pattern: highest between 03:00–04:00 (67–74%), relatively low between 15:00–21:00 (46–50%), in which lowest at 17:00, and slightly recovers in the late evening. This pattern highlights how traffic density and infrastructure load strongly affect on-time performance, with daytime peak hours suffering the heaviest congestion and operational interference.

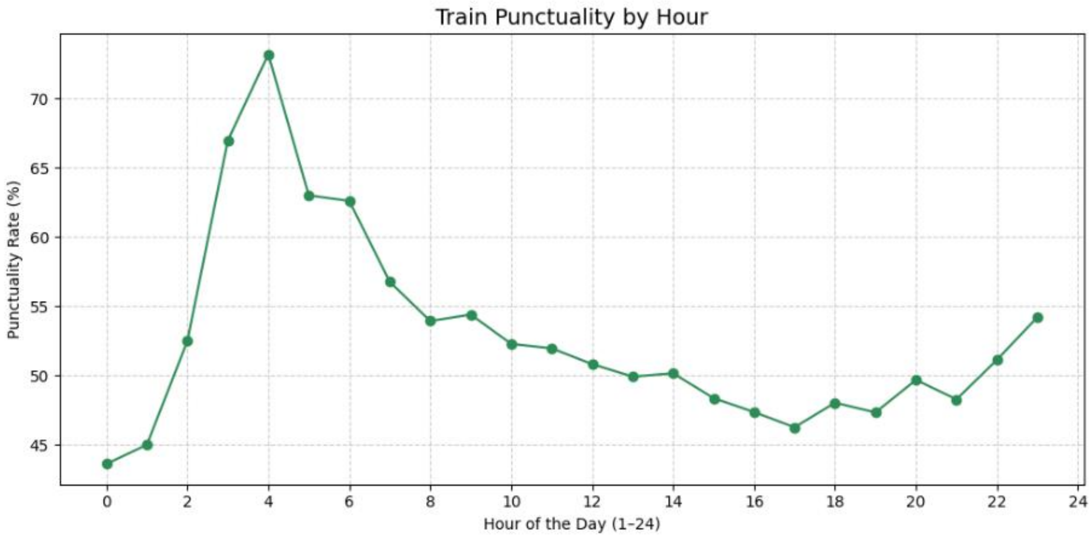

Figure 11 Punctuality by hour

Overall, our analysis indicates that under the stricter one-minute delay threshold, Deutsche Bahn's (DB) punctuality is substantially lower than the levels reported in official statistics. The overall on-time rate across all services falls below 50%, compared with DB's officially published value of 89.5% for 2024. While part of this gap can be attributed to differences in definition, the results nonetheless suggest that DB's reporting framework may systematically overestimate the network's operational reliability.

A closer look across service categories reveals pronounced structural disparities. Long-distance trains exhibit the weakest performance, with an actual on-time rate roughly half of the officially reported 62.5%. Significant variation is also observed between regional and urban services: in our sample, regional trains achieved only 48.7% on-time performance, whereas S-Bahn services reached about 62.9%. However, DB's official reports aggregate both under the single category of Regionalverkehr (regional transport), thereby masking the inferior punctuality of regional routes and artificially inflating the overall performance of the regional segment.

Taken together, these findings demonstrate that DB's punctuality issues are not random but structurally embedded -- delays concentrate along high-density, interregional corridors, where traffic congestion and network interdependence amplify disruptions. Moreover, the current reporting system's broad categorical aggregation further obscures these systemic differences. Adopting more granular classifications and stricter punctuality definitions can thus yield a more accurate understanding of Germany's rail network performance and provide a stronger empirical foundation for predictive modeling and operational improvement.

5. Modeling and Results

To quantify the influence of various factors on train punctuality, we applied machine learning methods. Because the available data covered only two time windows, we merged July and September samples and performed a random split (80% training, 20% testing) to ensure balanced distributions and robust generalization. Four models of increasing complexity were trained and compared.

5.1. Logistic Regression (Baseline Model)

As a linear model, logistic regression offers interpretability and efficiency. Using OneHotEncoder for all categorical variables (departure station, arrival station, train type, and departure time bucket) eliminated false ordinal relationships and allowed each category to be represented independently. The model estimates the probability of a train being on time (*p*) based on the logistic function:

$$\operatorname{logit}(p) = \ln \frac{p}{1-p} = \beta_0 + \beta_1(\text{departure station}) + \beta_2(\text{arrival station}) + \beta_3(\text{train category}) + \beta_4(\text{departure hour})$$

where β0 is the intercept and βi are the coefficients associated with each feature. The probability of punctuality is then given by: $p = \frac{1}{1+e^{-(\beta_0 + \sum_i \beta_i x_i)}}.$

In this formulation, a positive coefficient increases the log-odds of a train being on time, whereas a negative coefficient implies a higher likelihood of delay. This linear model provides a transparent baseline for comparison with more complex nonlinear models such as Random Forest and Gradient Boosting, balancing simplicity, interpretability, and predictive capability.

The model achieved 77.4% accuracy and an F1-score of 0.77, with recall values of 0.80 for on-time and 0.75 for delayed samples. Errors concentrated near borderline cases, such as minor delays close to the threshold. Overall, logistic regression provided a stable, interpretable baseline.

5.2. Random Forest

Building on the logistic regression baseline, a Random Forest model was developed to capture the nonlinear relationships that the linear model could not. The model achieved 78.2% accuracy and an F1-score of 0.78, showing a modest improvement over logistic regression while maintaining balanced precision and recall across both delayed and on-time classes.

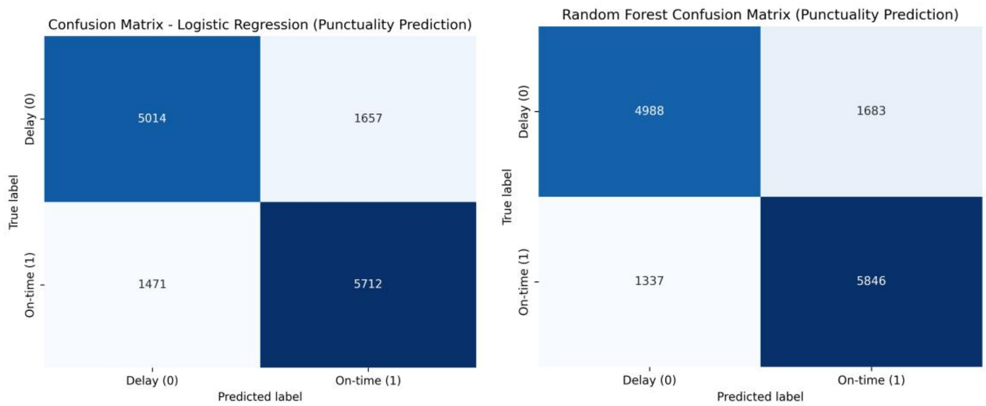

Figure 12 Comparison of Two Confusion Matrices

Comparing the confusion matrices of Logistic Regression and Random Forest reveals that the latter provides a modest but meaningful improvement in classification performance. The Random Forest correctly identifies more on-time trains (5846 vs. 5712) and reduces false negatives (1337 vs. 1471), resulting in higher recall for the on-time *class*. This improvement stems from its ability to capture nonlinear interactions between spatial and temporal variables, such as station combinations and departure times. While Logistic Regression remains more conservative and balanced, Random Forest offers stronger generalization and better recognition of borderline cases, reflecting the clustered and systemic nature of DB’s delays.

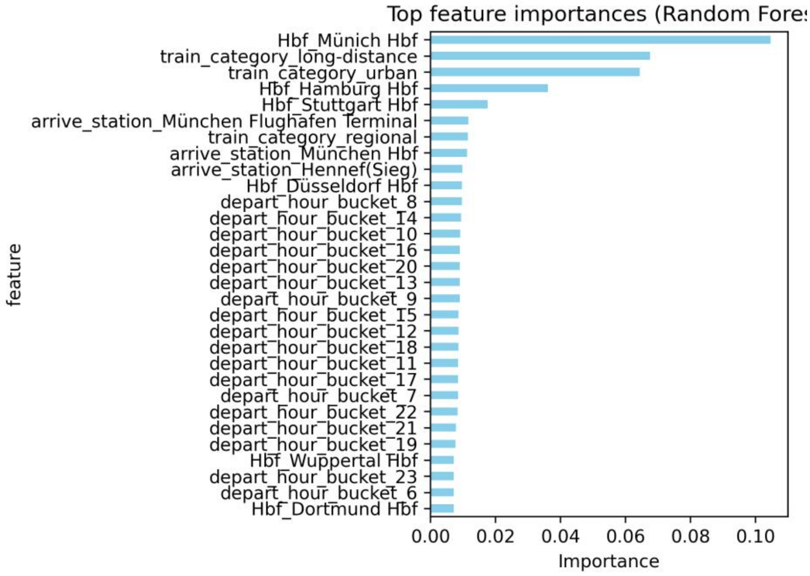

Figure 13 Top feature importances

Feature importance analysis further revealed that departure and arrival stations were the most influential predictors of punctuality, accounting for about 63% of total importance, followed by departure time bucket (25%) and train category (12%). Major hubs such as Munich, Hamburg, and Stuttgart Hbf dominated the top ranks, confirming the spatial concentration of delays across Germany's rail network. Temporal factors also played a secondary role, with peak-hour operations showing higher delay risks. These findings suggest that train delays are not random but are strongly associated with spatial and temporal factors, particularly along heavily congested routes such as the Frankfurt - Köln - München corridor, highlighting the clustered and systemic nature of DB's delay patterns.

5.3. Gradient Boosting Models: LightGBM and XGBoost

For further performance gains, we implemented two gradient boosting decision-tree frameworks: LightGBM and XGBoost. Both are well suited to high-dimensional, sparse data and can efficiently capture feature interactions. LightGBM achieved the best results: accuracy 0.7925, F1 0.8019, AUC 0.8727. XGBoost performed similarly: accuracy 0.7864, F1 0.7969, AUC 0.8689.

LightGBM's advantage stems from its histogram-based split and leaf-wise growth strategy, which enhances efficiency in sparse environments. XGBoost's level-wise growth produces slightly lower accuracy but greater stability on smaller feature spaces.

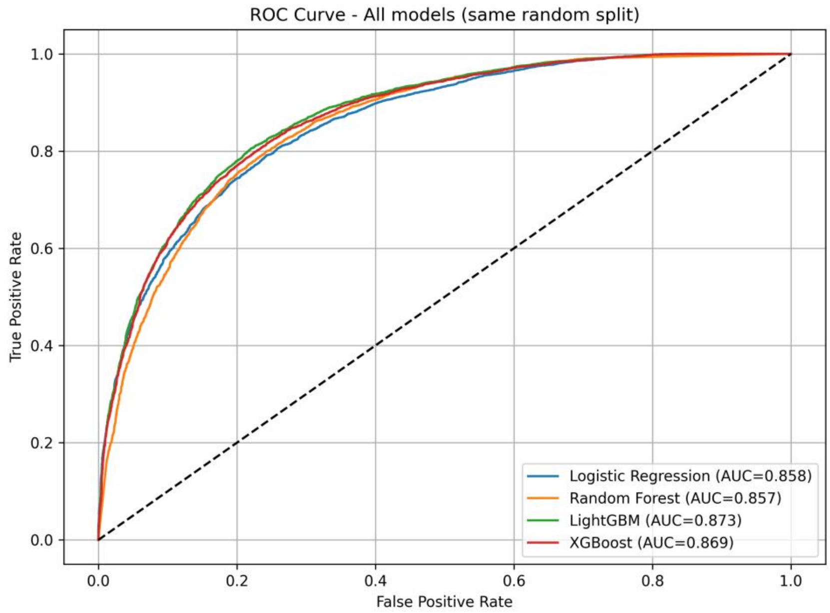

Figure 14 4 Models ROC Curve

6. Key Findings

The modeling results demonstrate that DB's punctuality prediction problem is not linearly separable but influenced by multiple nonlinear interactions. Train type, departure time, and departure station interact strongly—verified by the feature importance of LightGBM. The combination of departure and arrival stations remains the most significant factor (over 60% of total importance), underscoring the tight link between train reliability, geography, and network load.

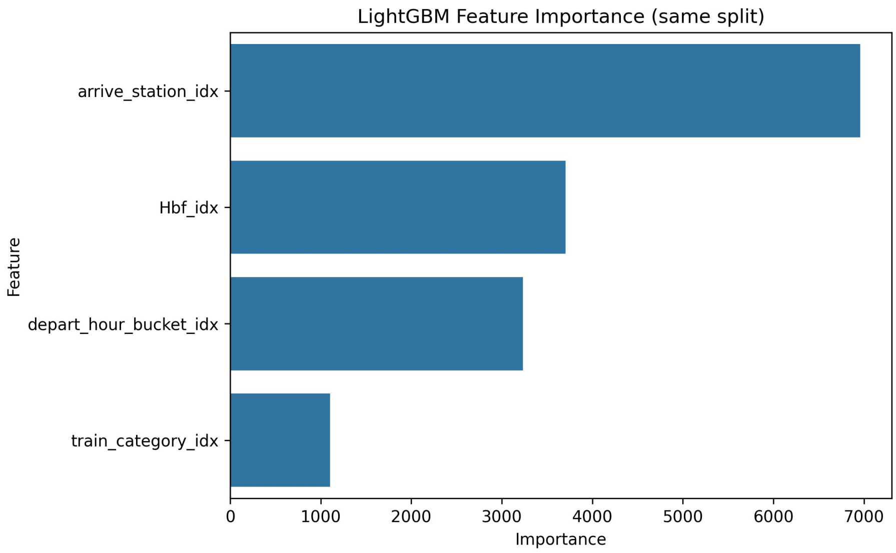

Figure 15 LightGBM Feature Importance

From a traveler's perspective, this finding is intuitive. Journeys between major hubs such as Frankfurt, Munich, and Cologne carry a significantly higher risk of delay. These corridors handle the densest traffic and most complex scheduling in the German rail network. In practical terms, travelers should plan for possible delays when moving between major cities—by allowing buffer time for transfers or avoiding peak hours and holidays. For DB itself, these high-risk routes represent critical areas for capacity management and scheduling optimization to improve overall punctuality and passenger satisfaction.

7. Limitations and Future Work

Although the four models differ in accuracy, their overall performance is similar (77–79%), indicating that the available features nearly exhaust the dataset's explanatory power. Since the input variables are mostly categorical (stations, train types, time buckets), their relationships with punctuality are largely linear. Nonlinear models such as LightGBM and XGBoost offer stronger flexibility but limited improvement without additional dynamic features like weather, congestion, or maintenance data.

Nevertheless, complex models remain valuable. LightGBM and XGBoost exhibit better robustness on borderline samples and provide richer feature-importance insights, revealing hidden interactions that can guide future feature engineering. Future research should therefore expand the data scope—integrating weather, construction schedules, holidays, and historical delay patterns—to enable models to capture broader temporal dynamics and achieve higher predictive precision.

8. References

1.https://www.kaggle.com/datasets/santiagoravotti/trains-and-delays-deutsche-bahn?resource=download

2.https://ibir.deutschebahn.com/2024/en/combined-management-report/product-quality-and-digitalization/the-customer-is-at-the-center-of-our-actions/punctuality/?utm_source=chatgpt.com

3.https://ibir.deutschebahn.com/2024/en/glossary/

4.https://github.com/FabreXUYY/Deutsche-Bahn-Delay