Advanced Analytics in Statutory Health Insurance

AI-supported analysis of billing data for optimising care

Prof. Dr. rer. nat. Hendrik Annuth Mirko Oest BSc Umut Kurucay BSc Dr. rer. pol. Heiko Friedel Mika Santoro BSc Finn Mertens BSc Dr. rer. oec. Katja Gehrke

Billing data from statutory health insurance funds constitute an essential resource for healthcare research, as they are available in an anonymised, cross-sectoral and standardised format. Traditional statistical analysis methods offer advantages in data exploration as the complexity and volume of data increase. In Germany alone, around 3.17 billion diagnoses are documented annually (Barmer 2024), supplemented by further extensive health data such as medication, surgical and billing information. Artificial intelligence (AI) – through machine learning and established statistical methods – expands the scope of analysis by capturing non-linear relationships within large, diverse datasets. In practice, this is evident in ICD/ATC codes with a very large number of variants: compact, learned numerical representations (embeddings) condense the catalogues or codes into meaningful features and utilise similarities as well as historical information. Traditional alternatives exist, but require more explicit structural assumptions and implementation effort. AI-supported models can thus achieve higher predictive accuracy and open up additional perspectives, e.g. for targeted intervention programmes to avoid unnecessary healthcare expenditure. In doing so, AI by no means replaces medical expertise, but rather specifically enhances it with data-driven insights and hypothesis generation. The aim of this project was to use modern AI methods to systematically identify multimorbid insured individuals in whom the underlying condition is masked by further illnesses or complex combinations of diseases. By combining longitudinal cost data with data on healthcare utilisation, the aim is to identify outliers and create differentiated disease profiles.

>> For this study, an anonymised extract from longitudinal German insurance data, prepared for analytical purposes, was used. The study included insured individuals for whom a confirmed diagnosis of type 2 diabetes mellitus (ICD-10: E11) was documented on an outpatient or inpatient basis in the first quarter of 2019 (index quarter) . A further criterion was that they were subsequently either continuously insured for at least 15 quarters and could be followed up, or died during the observation period (16 quarters, 1 January 2019 to 31 December 2022).

Summary

Introduction: Multimorbidity poses challenges for healthcare provision, as complex disease trajectories are only partially captured by conventional methods. Using diabetes as a case study, this project investigates how AI methods can identify patient groups within anonymised routine statutory health insurance data. The aim is to

describe patterns of morbidity in order to identify outliers and complex care constellations.

Methods: Anonymised longitudinal statutory health insurance data were used, including ICD diagnoses, ATC drug prescriptions, cost data and socio-demographic characteristics. The data were converted into vectors using neural networks. Embeddings, HDBSCAN, UMAP, pattern mining and vector variance analysis were employed.

Results: The analysis shows that multimorbid patient groups can be described in a differentiated manner. Correlations between disease patterns, resource consumption, and stable and dynamic disease courses became apparent.

Discussion: The results highlight the potential of AI-supported analyses and open up new approaches for the further development of healthcare research and more targeted management.

Keywords

AI methods, statutory health insurance routine data, AI-based health services research

Crossref/doi

http://doi.org/10.24945/MVF.02.26.1866-0533.2813

For each insured person, quarterly sociodemographic characteristics (age, gender, date of death), cost and utilisation data, all coded diagnoses (ICD-10) and prescriptions (ATC codes) were extracted. The cost data included outpatient and inpatient services, medicines and other services. Service utilisation was recorded based on the number of doctor's visits, hospital admissions, days spent in hospital and prescribed medicines per quarter.

The billing data contained up to five-digit ICD-10 and ATC codes per quarter, with duplicates removed. Data preparation and reduction were carried out with the aim of making the analysis data and models compatible with the new Research Data Centre for Health (FDZ), thereby enabling future research projects involving sensitive health data.

Design and methodology

The methodological approach of this study combines modern machine learning techniques with classical statistical approaches to provide a nuanced representation of complex disease progression in diabetes mellitus. The individual steps of the analytical model are shown schematically in Figure 1:

First, the coded diagnoses (ICD codes) and prescriptions (ATC codes) are aggregated from the claims data for each insured person. These aggregated codes are then transformed into purely numerical values (embeddings), thereby making the originally categorical data usable for machine learning procedures. Furthermore, vector embeddings are created, i.e. a numerical representation of the data that captures the significance or characteristics of the insured individuals as points in a multidimensional space. This allows information such as cost, utilisation and sociodemographic data (age and gender) to be combined into a comprehensive multimodal vector (Pezeshkpour/Chen 2018).

Finally, various embedding methods, such as Doc2Vec and Transformer models, are trained and compared using quality metrics to identify the most effective variant for further analysis.

Figure 1 illustrates this methodological pipeline – from data preparation through the transformation and evaluation of embeddings, multimodal feature engineering, to clustering and analysis of disease progression.

First, the insured profiles are clustered at quarterly level; subsequently, temporal trends are mapped using a two-stage clustering process. Finally, the resulting groups of insured individuals are analysed to identify typical trajectories, risk profiles and notable patterns (Fig. 1).

Data preparation and feature engineering

The claims data used for this study consist predominantly of non-numeric, variable ICD and ATC codes. In order for this data to be used in machine learning procedures, a fixed-size numerical representation is required that preserves the clinical meaning and relationships between the codes (Pezeshkpour/Chen 2018). To achieve this, all ICD and ATC codes documented over 16 quarters are aggregated for each insured person. The resulting profile encompasses all diagnoses and medications from the last four years and, through repeated occurrences, reflects chronic or persistent conditions (Steiger/Kroll 2023).

By using the aforementioned embeddings, insured individuals with similar clinical presentations are also closely clustered within the numerical space. This allows both clinical similarities and interactions between diagnoses and medications to be represented (Fig. 2).

The proven Doc2Vec model, which has already been successfully applied to insured person data (Le/Mikolov 2014; Steiger/Kroll 2023), is used for embedding generation. The DBOW variant is employed, which focuses on the co-occurrence of codes – not their order.

Bayesian optimisation is used to fine-tune the model parameters (Frazier 2018). The Rank Match Score (RMS) is used to evaluate embedding quality; it measures how well the model correctly groups diagnostically similar patients in the numerical space (Taskesen 2025; Kendall 1938). The calculation is performed by comparing the spatial distance in the embedding space with the similarity, for example using the Jaccard coefficient. In addition, a pre-trained Pat2Vec model (Steiger/Kroll 2023) is adapted for the current dataset and validated against our own model.

The result is a multimodal vector per insured person that integrates medical, demographic and care-related characteristics and serves as the basis for all further analyses.

Evaluation of the embeddings

The quality of the generated ICD and ATC code embeddings was assessed using various complementary methods to evaluate both the internal structure and the practical utility of the embeddings. The evaluation was based on a large cohort of 341,887 patients, for whom all ICD and ATC codes were analysed over 16 quarters. The study examined both embeddings that represent ICD codes exclusively and those that represent

ICD and ATC codes jointly.

Overview of the methodological steps from data preparation to cluster analysis

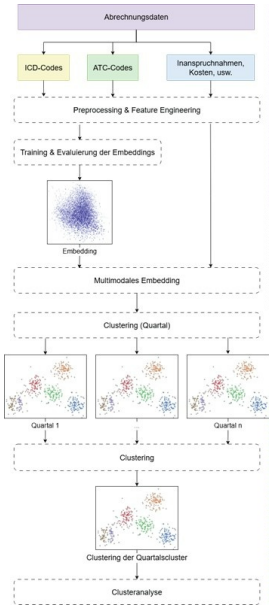


Fig. 1: **Overview of the methodological steps from data preparation to cluster analysis. Own illustration.**

ICD and ATC code embedding

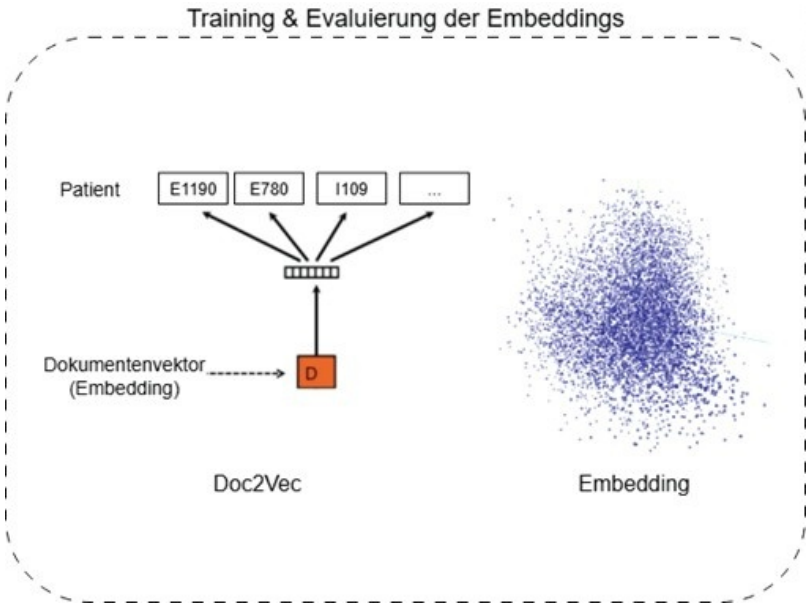


Fig. 2: **ICD and ATC code embedding. Own illustration.**

As part of the intrinsic evaluation, we examined how well the embeddings reflect the semantic proximity between medical codes and insured person profiles. The Neighbourhood Overlap Score was used for this purpose, which measures the extent to which neighbourhoods in the embedding space correspond to clinical similarities. The value ranges from 0 (no match) to 1 (perfect match); higher values indicate better semantic consistency (Chen/Buja 2009; Lin/Smith 2019; Flexer 2015).

For the extrinsic evaluation, we examine whether the representations of the data actually prove useful in practical application scenarios. The prototypical tasks selected here were predicting a patient's gender (classification) and aggregated healthcare costs (regression). These were then predicted using a classical random forest approach. The purpose of this analysis is to preserve the essential information within the data following embedding. Model quality was assessed for gender using accuracy and AUROC (Steiger/Kroll 2023), whilst the regression model used Mean Absolute Error (MAE) and R^2 as evaluation metrics (Steiger/Kroll 2023). A low MAE and high R^2 value indicate good model quality and high-quality embeddings.

By combining these intrinsic and extrinsic evaluation methods, both the internal structure and the practical applicability of the embeddings could be comprehensively assessed.

Two-stage clustering of the multimodal embeddings

The clustering of the multimodal policyholder embeddings was carried out using a two-stage procedure to account for both individual time points and the temporal development of patient trajectories. In the first stage, the individual embeddings of the policyholders were grouped separately for each quarter using HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise). This is a density-based clustering method that identifies groups (clusters) of data points without the number of clusters having to be specified in advance. To better structure the high-dimensional data, UMAP (Uniform Manifold Approximation and Projection) was previously used as a non-linear algorithm for dimensionality reduction (Steiger/Kroll 2023; Malzer/Baum 2021). In the second stage, the cluster assignments for a patient determined across all quarters were summarised into a trajectory vector (the temporal progression of an object or system through space), in which each dimension represents a quarter. These vectors were clustered again using HDBSCAN, this time without UMAP, employing the Hamming distance, as this optimally captures differences in categorical values (Allaoui/Kherfi 2020). This two-stage clustering makes it possible not only to analyse snapshots, but also to map the dynamics and development of disease and patient pathways in a differentiated manner over the observation period and to specifically identify groups of patients with similar progression patterns.

Cluster analysis

A comprehensive cluster analysis was carried out to identify the characteristic features for assigning an insured person to a specific cluster. The aim was to clearly distinguish the groups of insured persons from one another and to develop a distinct profile for each cluster in terms of disease burden, care and resource consumption.

The embedding of the insured data was central to this process, as it converts the multitude of non-numerical characteristics into a comparable numerical form whilst preserving the semantic relationships between diagnoses and medications. Only in this way was meaningful grouping based on complex medical patterns possible.

For quantitative characteristics such as costs and utilisation, classical descriptive statistics such as mean and standard deviation were calculated to analyse the distribution within the clusters and identify typical care profiles. Pattern mining was employed for the numerous diagnostic features (ICD and ATC codes), which often occur in complex combinations (Heaton 2017). This allowed both frequent individual codes and characteristic combinations and patterns within the clusters to be identified. For example, one cluster could be characterised by the frequent co-occurrence of diabetes (ICD E11), hypertension (ICD I10) and specific lipid-lowering drug prescriptions (ATC C10AA). Comparing such patterns across the clusters enabled the derivation of cluster-specific disease and/or medication profiles.

For further characterisation, the so-called vector variance was calculated. It measures the average dispersion of a patient's feature vectors across all quarters under consideration. A low variance indicates stable, consistent disease progression, whilst a high variance suggests dynamic or fluctuating progression and treatment approaches.

Results

The analyses enabled a differentiated characterisation of diabetes patients and revealed complex patterns in their disease progression, comorbidities and resource utilisation. The results pertain to the evaluation of the embedding models, the description of the identified patient groups through clustering, and the analysis of characteristic medical patterns.

Model development and evaluation of patient embeddings

To convert the complex, non-numerical patient data (ICD and ATC codes), various embedding approaches were developed and compared, including the reference model Pat-2Vec, a randomised null model, several Doc2Vec variants (including ICD-only and high-dimensional models), and a specially trained Transformer model (Custom BERT). Performance was evaluated using metrics such as overlap score, classification accuracy, AUROC, mean absolute error (MAE) and coefficient of determination (R^2).

The reference model showed the weakest results (overlap score 0.53; accuracy 54%; AUROC 0.50; MAE €22,626; R^2 0.00). The Pat2Vec model, with an overlap score of 0.61, accuracy of 77%, AUROC of 0.84, MAE of €18,831 and R^2 of 0.20, was already significantly better. The largest Doc2Vec variant (60 dimensions) achieved an accuracy of 82%, AUROC of 0.89, an overlap score of 0.62, an MAE of €19,502 and, with an R^2 of 0.27, the highest value of all models. The Custom BERT model outperformed all other models with an overlap score of 0.69, accuracy of 82%, AUROC of 0.90, MAE of €18,295 and R^2 of 0.26.

Overall, the Transformer model (Islam/Saidul 2023) demonstrated the best performance across all quality metrics, particularly in classifying patient gender and predicting costs. The overlap score, which measures the ability to group similar patients, was also highest here. The Pat2Vec model served as a reference for the basic validation of the approach, but the experiments with modern embedding methods highlight the potential of Transformer technology for even more precise analyses and improved clustering results (Table 1).

Übersicht der Leistungsmetriken aller getesteten Embeddings					
Metrik/Modell	**Overlap Score**	**Klassifikation Accuracy**	**Klassifikation AUC**	**Regression MAE**	**R^2**
Random/ Referenzmodell	0,53	0,54	0,50	22.626	0,00
Pat2Vec	0,61	0,77	0,84	18.831	0,20
Doc2Vec Dim 60	0,62	0,82	0,89	19.502	0,27
Custom BERT	0,69	0,82	0,90	18.295	0,26

Tab. 1: Übersicht der Leistungsmetriken aller getesteten Embeddings. Hinweis: MAE: niedriger ist besser; alle übrigen Metriken: höher ist besser. Eigene Darstellung.

Patient characterisation through clustering analysis

The two-stage HDBSCAN cluster analysis, based on multimodal embeddings, identified a total of 14 clusters within the cohort of over 340,000 people with diabetes. The seven most significant clusters can be divided into five main groups: clusters with a stable course of diabetes, high-risk clusters, clusters with high multimorbidity and mortality, clusters with high medication costs, and a large cluster with high variability.

The two largest clusters (7 and 13) comprise approximately 131,000 patients with stable diabetes progression, a low average age (68 and 65 years respectively), relatively low vector variance, very low costs (600 euros per year) and a 100% survival rate. The high-risk clusters 9 and 10 comprise approximately 14,000 patients with significantly higher ages (79 and 81 years respectively), high mortality (only 18% and 4% survivors respectively), comparatively long hospital stays (10.5 and 12.1 days respectively) and the highest costs (€8,600 and €10,100 respectively). Cluster 4 (17,000 patients, age 81) is characterised by high multimorbidity and moderate mortality (54% survival rate). Cluster 1 (approx. 3,900 patients) is notable for particularly high expenditure on medicines. The subpopulation with high variability (–1), comprising 133,000 patients, exhibits relatively high vector variance (13.5), a median age (72 years), heterogeneous disease courses and a survival rate of 87%. Overall, there is considerable heterogeneity within the diabetic population, with relatively stable, highly complex and unstable disease courses, as well as specific ‘outlier groups’, being identified on the basis of the data.

A low vector variance indicates that the values within the vector space show only minor deviations from the mean. This means that the vector components are closely clustered around the mean and exhibit only a small dispersion, i.e. scatter.

Medical patterns and characteristic disease combinations:

The detailed cluster analysis enabled the identification of common and specific ICD and ATC code combinations. Particularly in typical diabetic patients (clusters 7 and 13), the well-known triad of type 2 diabetes mellitus (E11.9), essential hypertension (I10.9) and isolated hypercholesterolaemia (E78.0) was observed. Other combinations, such as diabetes with

hyperlipidaemia (E78.5) and hyperuricaemia (E79.0), indicated metabolic syndrome (Alberti et al. 2009). In addition, comorbidities such as knee osteoarthritis (M17.9), atherosclerotic heart disease (I25.1) (Zhang et al. 2016), presbyopia (H52.2) and astigmatism (H52.4) were typical (Klein/Klein 1997). In the high-risk clusters, rare but clinically particularly relevant combinations were found, such as diabetes in combination with hypokalaemia (E87.6) and acute respiratory failure (J96.0), with left heart failure (I50.1) and acute renal failure (N17.9), or acute haemorrhagic anaemia (D62) with diabetes and hypertension.

Beschreibung der aussagekräftigsten Cluster (Auszug 7 von 14)							
Gruppe	**Stabiler Diabetesverlauf**		**Hochrisiko-Cluster**		**Hohe Multimorbidität/ Mortalität**	**Hohe Arzneimittelkosten**	**Subpopulation mit hoher Variabilität**
Cluster	7	13	9	10	4	1	-1
Patientenanzahl (Tsd.)	60	71	6	8	17	4	133
Altersdurchschnitt	68	65	79	81	81	69	72
Anteil lebendig	1	1	0,18	0,04	0,54	1	0,87
Vektorvarianz	6,6	6,2	21	0,8	11,4	11,3	13,5
KH-Tage Ø	0,1	0,1	10,5	12,1	1,9	1,7	2,7
Gesamtkosten Ø (Tsd.)	**0,6**	**0,6**	**8,6**	**10,1**	**2,6**	**2,7**	**3,4**

Tab. 2: Beschreibung der aussagekräftigsten Cluster (Auszug 7 von 14). Eigene Darstellung.

These results demonstrate that AI-supported analysis of claims data can identify both typical and rare, clinically significant patterns based on data – a crucial foundation for targeted care and intervention strategies.

Discussion and clinical implications

The results of this study show that modern AI-based methods make it possible, for the first time, to evaluate large routine data sets such as statutory health insurance claims data in a data-driven and differentiated manner without the need to integrate clinical data, for example. By combining cost and utilisation data, including diagnoses and prescribed medicines, inmultimodal feature embeddings, together with a two-stage clustering approach, an innovative method for identifying insured person patterns has been developed. Particularly where additional clinical information is lacking, this approach opens up new possibilities for the evaluation and management of care. The method is suitable both for identifying specific features in insured individuals' profiles and complex patient trajectories, and for characterising typical disease trajectories over time.

The method can improve epidemiological studies by accounting for biases caused by comorbidities. Furthermore, this type of segmentation could enable targeted adjustments to the Morbi-RSA. Alternatively, health insurance funds could use the algorithm internally to formulate more tailored offers for insured individuals. The latter, however, requires that statutory health insurance claims data be made available for analysis more quickly than has been the case to date. At the same time, it is evident that the data-driven identification of different policyholder profiles would benefit from greater integration of medical expert knowledge. The development of

reference datasets in which ‘outliers’ and typical patients are classified according to medical expertise could further enhance the validity of future analyses and classification models.

Conclusion

The project demonstrates how AI-supported methods can be used to identify complex multimorbid patient groups within routine data, whose underlying condition is masked by other diagnoses or combinations of diseases. The analysis was based on quarterly diagnoses (ICD codes) and medication prescriptions (ATC codes), which were transformed into numerical embeddings using neural networks. These vectors enable data-driven grouping of patients and the analysis of relationships between medical characteristics and resource utilisation.

References

1. BARMER Ersatzkasse. Arztreport 2024. Wuppertal: BARMER; 2024. p. 74. Available at: https://www.barmer.de/resource/blob/1255894/d58c1bd7c1a-99c78a54d410c490920c8/dl-arztreport-2024-data.pdf#page=74. Accessed: 19 June 2025.

2. National Association of Statutory Health Insurance Physicians. Uniform Assessment Scale (EBM), as of: 2nd quarter 2025 [Internet]. 2025. Available at: https://www.kbv.de/html/ebm. php. Accessed: 19 June 2025.

3. Alberti et al. (2009): Harmonising the Metabolic Syndrome. Circulation.

4. Klein R, Klein BE. "Diabetic eye disease." Lancet. 1997;350 Suppl 1:SII20-3.

5. Zhang et al. (2016): Association between diabetes mellitus and osteoarthritis: systematic review and meta-analysis. Osteoarthritis Cartilage.

6. Pezeshkpour et al. (2018): Embedding Multimodal Relational Data for Knowledge Base Completion. University of California, Irvine.

7. Le & Mikolov (2014): Distributed Representations of Sentences and Documents. Google Inc.

8. Frazier (2018): A Tutorial on Bayesian Optimisation.

9. Taskesen (2025): K-Nearest-Rank Scores: A Tool for Local Structure Analysis in Embeddings. Data Science Collective, Medium.

10. Kendall (1938): A New Measure of Rank Correlation. Biometrika, Volume 30, Issue 1-2, Pages 81–93.

11. Malzer & Baum (2021): A Hybrid Approach To Hierarchical Density-based Cluster Selection.

12. Steiger & Kroll (2023): Patient Embeddings From Diagnosis Codes for Health Care Prediction Tasks: Pat2Vec Machine Learning Framework.

13. Allaoui et al. (2020): Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study.

14. Heaton (2017): Comparing Dataset Characteristics that Favour the Apriori, Eclat or FP-Growth Frequent Itemset Mining Algorithms. College of Engineering and Computing.

15. Chen & Buja (2009): Local Multidimensional Scaling for Nonlinear Dimension Reduction, Graph Drawing, and Proximity Analysis . Journal of the American Statistical Association.

16. Lin & Smith (2019): Situating Sentence Embedders with Nearest Neighbour Overlap (N2O).

17. Flexer (2015): Improving Visualisation of High-Dimensional Music Similarity Spaces.

18. Islam et al. (2023): A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks.

The two-stage HDBSCAN clustering – supplemented by UMAP – captures not only snapshots but also temporal trends. The integration of pattern mining and the assessment of vector variance provide additional insights into the stability or dynamics of clinical presentations within the clusters. The results show that AI-supported analyses of anonymised billing data can effectively contribute to the identification of outliers and complex care constellations. The methodology provides a framework for the data-driven advancement of healthcare research and management and forms the basis for future work in which practical relevance can be further enhanced through the targeted integration of expert knowledge and additional contextual data. <<

Advanced analytics in the Statutory Health Insurance system: AI-supported analysis of billing data for healthcare optimisation

Introduction: Multimorbidity poses challenges for healthcare delivery, as traditional methods have limited ability to capture complex disease trajectories. Using diabetes as a case study, this project investigates how AI methods can identify patient groups within anonymised routine statutory health insurance data. The goal is to describe patterns of morbidity in order to identify outliers and complex care constellations.

Methods: We used anonymised longitudinal statutory health insurance data including ICD diagnoses, ATC drug prescriptions, cost data, and sociodemographic characteristics. The data were converted into vectors using neural networks. We employed embeddings, HDBSCAN, UMAP, pattern mining, and vector variance analysis.

Results: The analysis shows that multimorbid patient groups can be described in a differentiated manner. Correlations became apparent between disease patterns, resource utilisation, and stable and dynamic disease courses.

Discussion: The results highlight the potential of AI-supported analyses and open up new approaches for the further development of health services research and more targeted management.

Keywords

AI methods, SHI billing data, AI-based healthcare research

Authors' declaration

The authors declare that there are no conflicts of interest.

Citation

Annuth et al.: "Advanced Analytics in Statutory Health Insurance: AI-supported analysis of billing data for care optimisation", in: "Monitor Versorgungsforschung" (02/26), pp. 57–63.
http://doi.org/10.24945/MVF.02.26.1866-0533.2813



Prof. Dr. rer. nat. Hendrik Annuth ORCID: 0009-0007-1797-8063 took up the Chair of Data Science & Artificial Intelligence at Wedel University of Applied Sciences in 2020 and teaches courses in Data Literacy, Exploratory Data Analysis, Machine Learning and Deep Learning. He is the programme director for both the Bachelor's and Master's programmes and supervises numerous practice-oriented student projects as part of his teaching duties. Contact: Hendrik.Annuth@fh-wedel.de



Mirko Oest BSc ORCID: 0009-0000-3532-2737 holds a Bachelor's degree in Industrial Engineering from Wedel University of Applied Sciences. He is continuing his studies in the Master's programme in 'Data Science & Artificial Intelligence' at the same university. His academic career combines engineering expertise with modern data science methods. Contact: MirkoOest@googlemail.com



Umut Kurucay BSc ORCID: 0009-0004-3543-0227 is a dual Master's student at Airbus, specialising in data-driven applications in aviation. Prior to this, he worked as a .NET developer at best4bps and completed a dual Bachelor's degree at Deutsche Telekom with a focus on process automation. He studied Business Informatics at FHDW Bielefeld and is currently pursuing a Master's degree in "Artificial Intelligence & Data Science" at FH Wedel. Contact: hello@umutkurucay.com



Dr. rer. pol. Heiko Friedel works in the field of healthcare research and health economics. Among other things, he is responsible for conducting

scientific studies and projects, as well as for expert reports assessing the cost-effectiveness of various services provided by statutory health insurance funds. Contact: Heiko.Friedel@teamgesundheit.de



Mika Santoro BSc ORCID: 0009-0003-5631-5423 is a dual Master's student at Airbus. Prior to this, he completed a dual Bachelor's degree at Hewlett Packard Enterprise (HPE) specialising in 'Computational Data Science'. He studied Computer Science at the DHBW Stuttgart and is currently enrolled in the Master's programme 'Data Science & Artificial Intelligence' at the University of Applied Sciences Wedel. Contact: Mika.Santoro@gmail.com



Finn Mertens BSc ORCID: 0009-0003-8124-7856 is a Master's student in Computer Science at Wedel University of Applied Sciences. He also works as a student trainee software developer at adesso SE. Previously, he worked as a student trainee in Software Quality Assurance at Diebold Nixdorf and as an IT Security Consultant at KPMG. Contact:

inf103461@stud.fh-wedel.de



Dr. rer. oec. Katja Gehrke ORCID: 0000-0001-5516-457X is Executive Researcher at the Institute for Descriptive Health Data Analysis (DGDA). Prior to this, she held a senior position at the German Pharmaceutical Industry Association, focusing on health economics and healthcare research, as well as strategy and market access. She studied economics and worked at the Chair of Public Finance and Health Economics at TU Berlin and at the Université libre de Bruxelles (ULB). Contact: KGehrke@dgda-institut.de