

Kubitza, Dennis Oliver; Weißling, Katarina

Preprint

Whole Lotta Training - Studying School-to-Training Transitions by Training Artificial Neural Networks

Suggested Citation: Kubitza, Dennis Oliver; Weißling, Katarina (2025) : Whole Lotta Training - Studying School-to-Training Transitions by Training Artificial Neural Networks, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/310974>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Whole Lotta Training - Studying School-to-Training Transitions by Training Artificial Neural Networks

Dennis Oliver Kubitza^{1,2} and Katarina Weßling^{1,2}

¹Federal Institute for Vocational Education and Training (BIBB), Germany

²Research Centre for Education and the Labour Market, School of Business and Economics,
Maastricht University, Netherlands

Corresponding Author: Dennis Oliver Kubitza

Email: d.kubitza@maastrichtuniversity.nl

PREPRINT VERSION: January 2025

Abstract

Transitions from school to further education, training, or work are among the most extensively researched topics in the social sciences. Success in such transitions is influenced by predictors operating at multiple levels, such as the individual, the institutional, or the regional level. These levels are intertwined, creating complex inter-dependencies in their influence on transitions. To unravel them, researchers typically apply (multilevel) regression techniques and focus on mediating and moderating relations between distinct predictors. Recent research demonstrates that machine learning techniques can uncover previously overlooked patterns among variables. To detect new patterns in transitions from school to vocational training, we apply artificial neural networks (ANNs) trained on survey data from the German National Educational Panel Study (NEPS) linked with regional data. For an accessible interpretation of complex patterns, we use explainable artificial intelligence (XAI) methods. We establish multiple non-linear interactions within and across levels, concluding that they have the potential to inspire new substantive research questions. We argue that adopting ANNs in the social sciences yields new insights into established relationships and makes complex patterns more accessible.

Keywords: school-to-work transitions, vocational education and training, machine learning, artificial neural networks, explainable artificial intelligence, SHAP values, rule extraction

This paper uses data from the National Educational Panel Study ([Blossfeld, Roßbach & Von Maurice, 2019](#)). The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide network.

Contents

1	Introduction	2
2	Transitions in the German Context	4
2.1	The German Education System	4
2.2	Panel Data, Data Linkage, and Variables	4
3	Machine Learning Tools	6
3.1	Artificial Neural Networks	6
3.2	Explainable Artificial Intelligence	8
4	Implementation	8
5	Results	11
5.1	Extracted Rules	11
5.2	SHAP Values	14
5.2.1	Urbanity	14
5.2.2	Unemployment	15
5.2.3	Share of STEM employees	15
5.2.4	Female labor force participation	16
6	Discussion and Conclusion	16
7	Supplements	20
	References	22

1 Introduction

Educational transitions after school are well-researched in the social sciences. These transitions are key events in the life course because they have lasting consequences on careers. To empirically measure the success factors of these transitions with quantitative data, researchers apply different regression techniques. They typically aim at singling out one (group of) influencing factor(s) and 'control away' the confounding and mediating variables (Rohrer, Hünermund, Arslan & Elson, 2022). As influencing factors exist at various analytical levels (e.g., the individual level, the institutional (of schools and classrooms) or the regional level), multilevel approaches can be applied (e.g. Snijders & Bosker, 2012). When analyzing effects of higher-order levels, like regional contexts, studies focus on specific characteristics of these contexts (e.g., industry sectors Flohr, Menze & Protsch, 2020), the unemployment rate, or the provision of training places (Hillmert, Hartung & Weßling, 2017; Hartung, Weßling & Hillmert, 2022) as proxies representing (perceived) training or employment opportunities. Thus, various variables often represent the same or a similar theoretically assumed mechanism. This makes it challenging to assess the importance of regional factors. Furthermore, higher-order contextual variables show less variation and greater intercorrelation, making it challenging to determine their relationship to individual-level outcomes. In educational transition research, efforts are made to show that specific predictors, whether on the same or different analytical levels, interact or are moderated by other predictors. Hence, the effect of one predictor is reinforced or diminished by another predictor. To provide a substantive example, research has shown that the chance to secure a training position is lower for young people with lower levels

of schooling and from lower social backgrounds. This effect is stronger in socioeconomically poor regions with fewer training opportunities (Hartung, Weßling & Hillmert, 2022; Hartung & Weßling, 2024).

Only a few studies follow conceptual arguments provided by Cook (2003) suggesting that also the contextual variables on different levels (e.g., schools, neighborhoods or regions) can have a mediating or moderating influence on individual-level outcomes (Rich & Owens, 2023). Weßling and Meng (2021) show for the Netherlands that missing regional labor-market opportunities can be attenuated for ethnic minority youths by living in same-ethnic neighborhoods. To analyze such interactions, researchers have to make selections of predictors based on theoretical assumptions. Conventional regression techniques limit the possibilities to explore multiple and complex interdependencies between predictors on different levels simultaneously.

A growing body of research shows how artificial intelligence¹ (AI) can be used in the social sciences. Machine learning² (ML) algorithms can uncover overlooked patterns between predictors and summarize complex relationships, by relaxing some statistical assumptions (for an overview, see Lundberg, Brand & Jeon, 2022). For example is correlation between predictors a less severe problem for some ML algorithms, as they attribute lower impact to correlated features (De Veaux & Ungar, 1994).

However, ML is not a wonder drug for social research as its high complexity often poses challenges for interpreting and presenting models. Thus, it is essential to combine self-learning tools with 'human attention' (Lundberg, Brand & Jeon, 2022).

In this paper, we train artificial neural networks³ (ANNs) on survey data from the German National Educational Panel Study (NEPS): starting cohort 4 – 9th Grade (NEPS-SC4, NEPS Network, 2021) to replicate transitions of young people from secondary education to vocational training or to further general education. We extract the learned 'knowledge' from the ANNs in a human-readable form through explainable AI (XAI): We calculate importance scores for each input feature (i.e., predictor) and for each observation (i.e., student) to identify patterns in students' transition chances, and we extract human-readable rules from the ANN to explain these observed patterns. By applying this explorative approach, our results are not strictly causal, and we cannot falsify any theoretical assumption, but lay the groundwork to develop new research questions.

Our scientific contribution to the literature that implements ML in the social sciences (Di Franco & Santurro, 2021; Shu & Ye, 2023) is threefold: First, ANNs and XAI techniques represent the introduction of a new set of tools in sociological research. Second, this paper reduces the 'black box' image of AI methods by employing XAI to make the learned patterns about transitions interpretable and meaningful. Third, our work situates these patterns within well-established research on school-to-work transitions, identifying research gaps regarding multi-level influences, but also supporting existing relations.

The paper is structured as follows: In the subsequent section, we discuss the German school and vocational education and training (VET) system and introduce our data. Next, we offer a description of the machine learning employed in this study. Following this, we provide the implementation details. The subsequent section presents the results of the XAI approaches. Finally, the paper discusses the findings in the context of previous research and concludes with an evaluation of the study's limitations and broader implications.

¹Mimicking intelligent behavior through algorithms, often facilitated through machine learning.

²Algorithms that adapt themselves to solve a task, based on data that is presented to them.

³ML models that comprise a complex network of simple processing units.

2 Transitions in the German Context

2.1 The German Education System

The German education system is federally organized and characterized by (mostly) strict tracking of secondary school types. School types and school tracks can differ between states. Some federal states provide three, others two types of secondary schools, which offer different secondary education diplomas. The lower-level secondary education diploma (Hauptschulabschluss, 9th grade) and the intermediate-level secondary education diploma (Realschulabschluss, 10th grade) are considered pre-vocational, after which students typically can enter VET or continue with general schooling.

In Germany, approximately 60 percent of an age cohort obtain a professional qualification in the VET system (BMBF, 2024). Education in the VET system leads after 2 to 3 years to a qualification in one of the 324 legally defined professions. Most VET students are trained dually in the apprenticeship system, where students are employed at a company while attending part-time VET school. In the apprenticeship system, employers decide which applicant to recruit, which is comparable to the labor market entrance. Besides the apprenticeship system, the German VET system offers full-time school-based vocational training, which provides the possibility of obtaining a professional qualification without being employed at a company. The choice between the type of VET depends on occupations and sector. While trade and commercial occupations dominate the apprenticeship system, school-based VET focuses on health and care professions.

Graduates of the secondary education have the possibility - based on their performance and institutional restrictions - to continue in the general school system. In doing so, they can obtain a university entrance qualification, which allows them to continue further in higher education (HE). The upper-level secondary education diploma (Abitur/Fachabitur, 12th/13th grade) allows students to enter higher education at universities (of applied sciences). This is the most common alternative for these students. However, in the German context, at least 21 percent of students who enroll for VET have Abitur/Fachabitur (BMBF, 2024).

2.2 Panel Data, Data Linkage, and Variables

We use the National Educational Panel Study (NEPS): *Starting Cohort 4–9th Grade* (NEPS-SC4: NEPS Network, 2021). NEPS-SC4 is a school-based survey panel that targeted 9th graders in 2010 and followed them during school graduation, VET and/or HE, up to their labor-market entry. The data contains information on the place of residence, enabling us to link the data with regional data on the district level ('Kreise', corresponding to the European NUTS3 level). The NEPS-SC4 comprises a wide range of students', school, and parental characteristics.

We predict the first transition of students after obtaining a lower-level or intermediate-level secondary school diploma. Our dependent variable measures whether students enter 'dual VET', 'school-based VET', or 'stay in (general) school'. One of the most important predictors encodes the type of school students attended. We distinguish between a lower-level secondary school diploma (diploma.type_0, usually obtained at a so-called 'Hauptschule'), and intermediate-level secondary school diploma obtained at a so-called 'Gymnasium' (diploma.type_2), and an intermediate-level secondary school diploma obtained at a different school form (diploma.type_1). Figure 1 shows the transitions we are considering in the analytical part of this paper in relation to the whole Education System.

Of the 9,960 initially targeted individuals, 6,812 remained in the dataset after removing observations with missing or censored variables⁴. We merge the NEPS data set with regional information

⁴The data preparation matches 57,225 self-reported educational episodes to individual data and

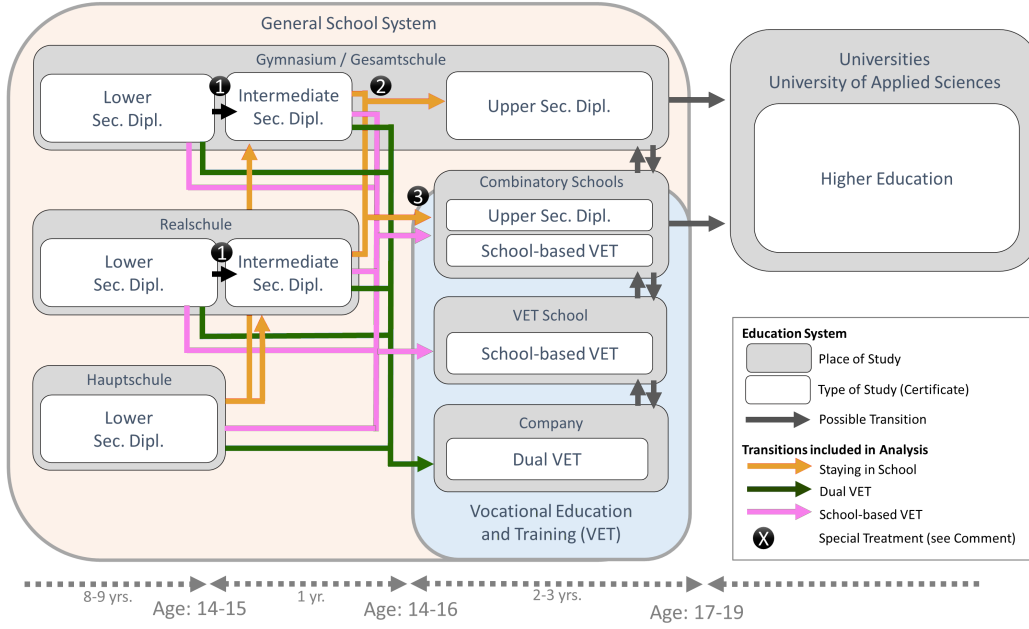


Figure 1: Overview of the education system and transitions in Germany. Depiction of the education system, showing schools, diplomas, transitions, and age context, distinguishing between the General School System and the VET System. Several types of secondary schools offer the same diplomas. **Notes:** (1) The transition from lower secondary education to intermediate secondary education is excluded from our analysis, since it usually occurs automatically by not dropping school. (2) Those who reported only finishing upper secondary school are classified as 'staying in school' during the age 14-16. (3) VET graduates that reported achieving a VET degree are counted as 'school-based VET', while others are classified as 'staying in school'.

from INKAR (BBSR, 2022) and IÖR data (Meinel & Schumacher, 2010) capturing local economic structure, demographics, land-use and electoral votes. INKAR is an interactive geospatially referenced platform provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR). The IÖR monitor developed by the Institute for Ecological Economy Research (IÖR) is a platform providing environmental indicators. All regional data is linked on basis of the NUTS3 identifier of the residential location of students.

The final dataset contains variables that allow for a comprehensive understanding of individual, school, and regional factors influencing transitional outcomes. Individual demographic factors, such as gender, migration background, and parental education, provide insights into social stratification and the role of family background in shaping transitions. Variables like last diploma score and school average score capture academic performance and educational quality, which are critical predictors of post-education transitions. Regional variables such as unemployment rates, shares of science, tech, engineering and mathematics (STEM) employees, shares of working females, industrial land use and urbanity reflect the economic and structural opportunities available in different areas and

school information. Educational episodes are self-reported by the participants and multiple may exist per student. 35,974 episodes remained after the merger and deletion of missing variables of which the relevant one was selected.

shares of conservative voters capture cultural differences between regions. A full list of all used variables with their descriptions and encoding is provided in the Appendix.

3 Machine Learning Tools

3.1 Artificial Neural Networks

ANNs are ML models that comprise a complex network of processing units, so-called neurons, resembling the structure of biological neural networks like brains. Each neuron implements a function projecting multiple inputs to a single output value, the so-called activation. Depending on its specification, a neuron can serve different purposes, like storing a single variable’s value, or performing a logit-transformation on the sum of its weighted inputs. By combining multiple neurons and implementing a weighted network between them, ANNs can extract knowledge from heterogeneous data sets, as their structural complexity allows them to learn non-linear relations. Since non-linear models are difficult to interpret, implementing ANNs in the social science has been tentative. However, ANNs provide also scientific benefits: The derived results are agnostic to researchers’ beliefs, existing models, and explanatory approaches, as long as the data captures such facts and the model is trained well enough. ANNs can solve tasks like image classification⁵, predictive maintenance, or speech recognition, just by appropriate selection of the internal organization of neurons (e.g., *feed-forward networks*, *recurrent networks*), the ANN training algorithms (e.g., *backpropagation of error*, *equilibrium propagation*) and the learning environment (*reinforced*, *supervised*, or *unsupervised*).

Especially for complex tasks with high-dimensional input data, deep neural networks⁶ (DNNs) are the preferred ML models. Since DNNs require huge amounts of ANN training data and computing capacities, they are only useful in specific scenarios where data amount and quality are sufficient. The above-mentioned applications to image classification or speech recognition fall under this category. Conventional ANNs are often more suited applications when predicting human decisions or behavior (Mason, Twomey, Wright & Whitman, 2018; Maszczyk, Golaś, Pietraszewski, Rocznik, Zajac & Stanula, 2014; Zeinalizadeh, Shojaie & Shariatmadari, 2015). This is particularly the case when data sampling is costly and data quantity is limited.

We use fully connected multi-layer perceptrons (MLP) without *thresholds*, which are one of the simplest forms of ANNs. MLPs for panel datasets are applied in epidemiology (Khan, Kaushik, Ji, Malik, Ali & Wei, 2019) but also in financial economics (Ding, Zhang, Liu & Duan, 2015) and engineering (Bojarski, Del Testa, Dworakowski, Firner, Flepp, Goyal, Jackel, Monfort, Muller, Zhang & others, 2016). They are *feed-forward* ANNs, in which information passes uni-directional from the input neurons through multiple layers of hidden neurons to the output neurons. Each artificial neuron applies an internal differentiable activation function to the weighted sum of outputs from the previous layer. We call the output of this function *neuronal activity*⁷. Each input layer neuron corresponds to a predictor (input variables like in our case the last diploma score, the regional unemployment etc.), and each output layer neuron corresponds to a class of the learned classification (output variable like in our case the class ‘staying in school’). Categorical variables are translated to multiple input and output neurons in the neural network, such that every predictor is represented by at least one neuron. Figure 2 shows a stylistic MLP, along with its terminology and relevant variables.

⁵Assigning predefined categories or labels to input data based on its characteristics or features.

⁶ANNs that have many layers and can therefore learn abstract knowledge and concepts.

⁷Response of a neuron in an ANN, representing either the continuous value of activation or the binary classification between active and inactive.

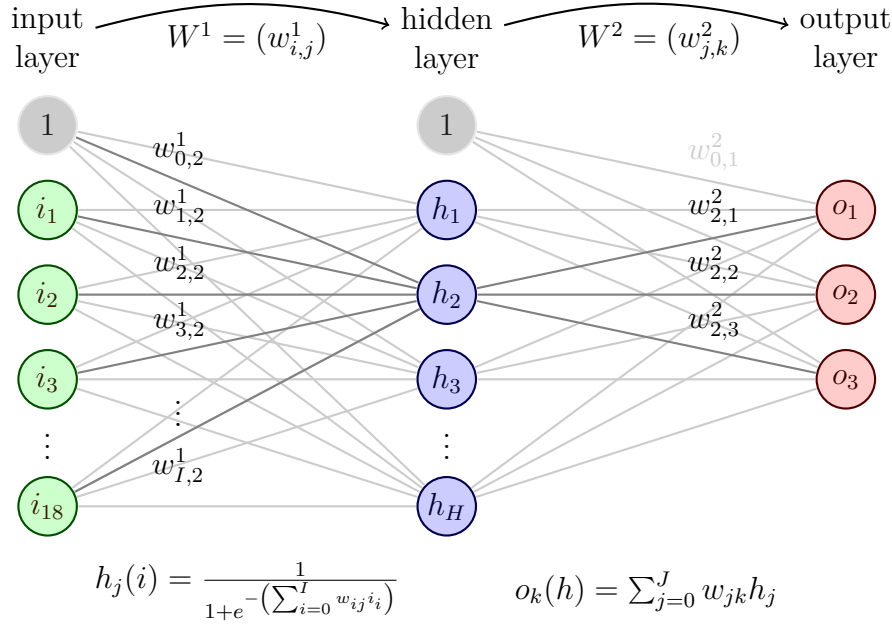


Figure 2: Specification for the used MLP

Depicted are the layers, neurons (including bias neurons labeled as (1)), weights between neurons, and activation function specifications. Input neurons correspond to parameters, output neurons to target classes. Design choices include the number H of hidden neurons and the activation function form (e.g., logistic, as depicted). Weights between layers are summarized as matrices W_1 and W_2 .

Before an ANN can be used in production, it needs to be trained to serve a specific task. ANNs can be trained using different approaches depending on the type of problem and data. *Supervised learning* involves training the network on labeled data, where the input-output pairs are known, and the goal is to minimize the difference between predicted and actual outputs. The ANN training process initializes all weights at random and updates the incoming weights of each neuron proportionally to the gradients of the overall error term. While the structure of the ANN remains unchanged, the relationships between neurons are adjusted during ANN training to assign more or less weight.

In the training of ANNs, multiple trade-offs need to be balanced. The bias-variance trade-off (Geman, Bienenstock & Doursat, 1992) involves balancing *underfitting* (high bias, where the model is too simplistic to capture patterns) and *overfitting* (high variance, where the model captures noise along with the signal). The trade-off between computational cost and performance (Carvalho, Ramos & Chaves, 2011) arises because larger models with more neurons or layers can improve accuracy but demand significantly more memory, processing power, and time, often exceeding available resources.

3.2 Explainable Artificial Intelligence

Despite being more flexible, ANNs can provide erroneous results, which are critical in realistic applications, such as autonomous driving. The detection of such errors is harder because to the black-box nature of ANNs. Explainable Artificial Intelligence (XAI) technology addresses this issue by equipping experts with tools to analyze and interpret ML models, making it possible to identify the sources of errors. As AI technologies are more frequently adopted across various industries and scientific disciplines, XAI are accordingly of increasing importance (see [Arrieta, Díaz-Rodríguez, Del Ser, Bennetot, Tabik, Barbado, García, Gil-López, Molina, Benjamins & others, 2020](#), for an extensive overview). The XAI methods for ANNs that we apply here belong to two groups of methods: feature relevance explanations and simplification-based explanations. Feature relevance explanation focuses on analyzing the ANN as a unified entity to understand the relationships between inputs and outputs. Various importance scores are estimated to quantify the contribution of each input feature to the network’s predictions and then visually analyzed. In the field of visual analytics, such methods are applied as a component of optimizing decision-making processes ([Keim, Andrienko, Fekete, Görg, Kohlhammer & Melançon, 2008](#)).

We perform such an analysis with *SHapley Additive exPlanations* (SHAP, [Lundberg & Lee, 2017](#)). SHAP values are based on the game-theoretic definition of Shapley values ([Shapley, 1951](#)) and analyze the differentials of classification probabilities for changes in input values. Since SHAP value calculations yield a separate importance score per predictor and observation, they are useful for identifying subgroups of observations with similar importance. For instance, SHAP values can indicate that a variable contributes to a 20% increase in the probability for a classification for one observation, while simultaneously contributing only 1% for a different observation.

In contrast, to ignoring the internal structure and analyzing the ANN as a unity, explanation by simplification reduces the ANN to smaller components that are easier to interpret. These methods can involve altering the structure of the ANNs, such as dropping irrelevant weights, replacing continuous with binary scales, or replacing certain neurons with alternative mechanisms. Algorithms for *rule extraction* (RE) transform ANNs into a set of easily understandable logical rules by recursively applying heuristics to the neurons. The form of the extracted rules depends on the purpose of the RE and the data types, whether variables are continuous, binary, categorical, or a mixture (see [Hailesilassie, 2016](#), for an overview of RE on DNNs).

Our method for RE is based on three techniques that are frequently used: *weight pruning* removes unimportant weights in the input of each neuron; *Discretization* regards the continuous space of values that a neuron can receive as inputs or send as outputs as a discrete set of values; *Linearization* replaces the non-linear activation function of the neurons with a linear equivalent.

4 Implementation

We use the R package *neuralnet* ([Wright & Stachelek, 2020](#)), which provides the implementation of the MLPs and the *resilient backpropagation* algorithm. The latter is applied for ANN training ([Riedmiller & Braun, 1994](#)), and ensures improving convergence rates compared to standard backpropagation of errors.

In the first step, we identify the best ANNs for replicating the dependent variable. The three transition alternatives - 'dual VET', 'school-based VET', and 'staying in school' — are defined as output neurons of the ANNs. Since the structure of the ANN and the specifics of the ANN training process heavily influence the performance of the ANNs, we evaluate multiple configurations of both structure and training methods. Each evaluation involves splitting the data randomly into ANN training (80%) and test (20%) data and averaging performance over 20 pseudo-random

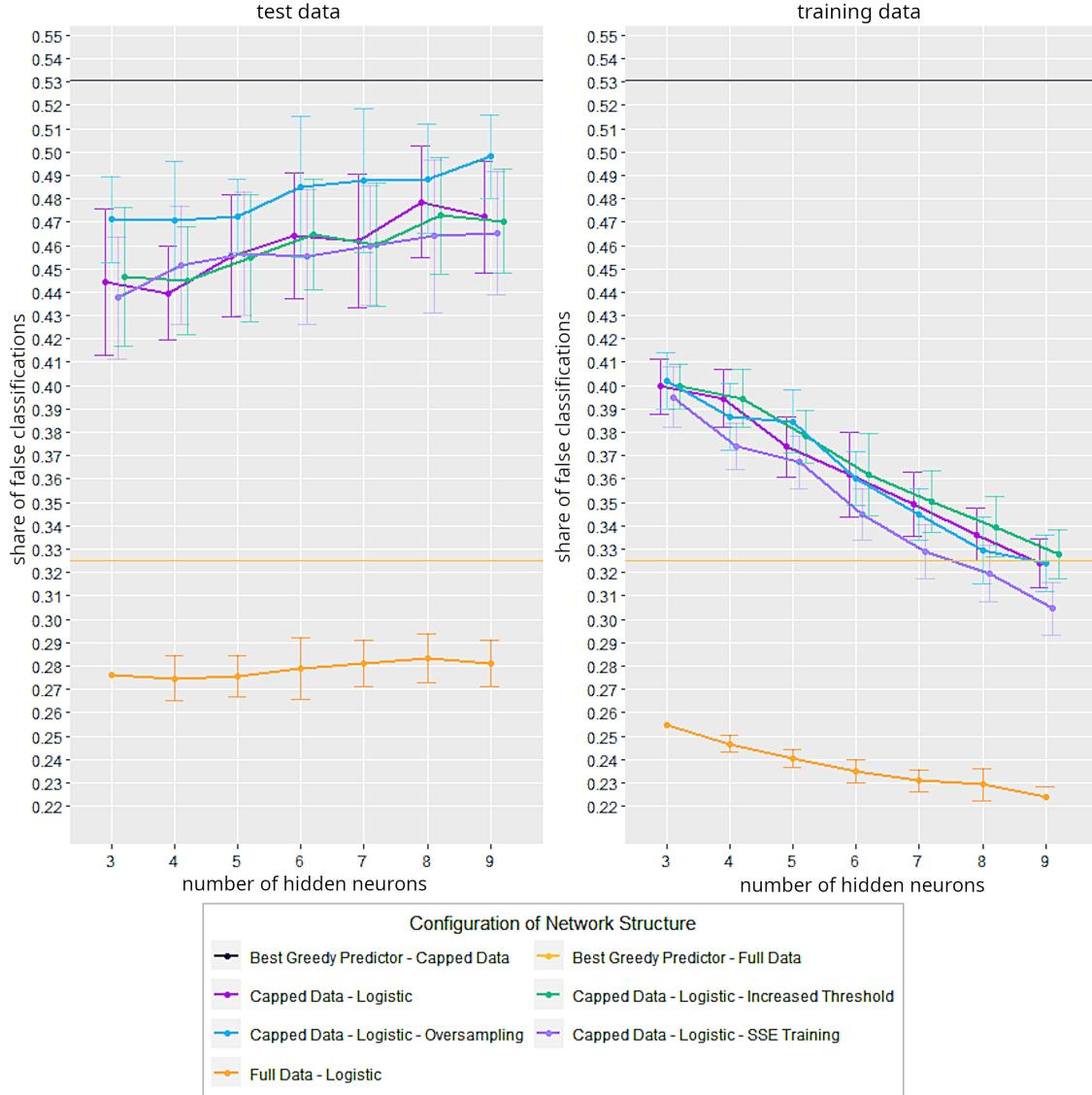


Figure 3: Performance Chart of Different ANN Models

Connected line plots of performance for different combinations of ANN Models, training and datasets. share of faulty classifications on y-axis. number of hidden neurons on x-axis. faceted by the test and training datasets. performance of greedy predictors are plotted for reference.

initializations for the weights. Performance statistics are calculated and evaluated on the test dataset. These measures help reduce *overfitting*, minimize random variance in performance, and ensure generalization.

The computation time increases when additional hidden layers are introduced, while the quality of the classification did not improve. Hence, we only consider variations in the number of hidden neurons within the first hidden layer, ranging from 3 to 10 neurons. Initial test runs reveal that

the distribution of target classes, particularly the underrepresentation of 'school-based VET', is unfavorable for ANN training. ANNs ignore such cases, leading to a high proportion of miss-classification for those underrepresented cases. Therefore, we allowed more variation in our 'full dataset' by *underfitting* the overrepresented classes or by *overfitting* the underrepresented 'school-based VET' class. To do so, we employ two strategies: *capping* and *over-sampling*.

Capping involves reducing the ANN training data to achieve a more balanced distribution between the different school diplomas and the type of ANN training. After removing overrepresented classes, 2,434 observations remain. Over-sampling involves duplicating underrepresented classes during ANN training to balance their representation. ANNs are trained either until the partial derivatives of the error function fall below 0.05 or until the ANN training reached 200,000 iterations. Corresponding with the trade-off between training time vs. model accuracy, we increased the target thresholds for the stopping criterion of the ANN training to reduce computing time. However, these thresholds often cannot converge with sufficient reliability for meaningful comparisons.

A *greedy predictor*, calculated on both the full and reduced dataset, served as a benchmark. This predictor classifies all observations into the most common class. Figure 3 presents the performance metrics for the test and ANN training datasets compared to these benchmarks. While the ANNs trained on the full dataset outperform the benchmarks, they show smaller relative improvements compared to models trained on capped or over-sampled data. This suggests that performance gains are disproportionately driven by reduced errors in overrepresented groups. A configuration with four hidden neurons, data capping, and a threshold of 0.05 emerged as the most effective. This setup shows the highest amount of learned information on the test data.

To further validate this configuration, we retrained the model ten times using different random seeds on the capped dataset. This approach ensured that the model can leverage all available relevant information. The best iteration achieved 68.62 percent correct classifications across all 6,812 observations, with an accuracy of at least 54.3 percent for the underrepresented classes of 'dual VET' and 'school-based VET'. Table 1 contains the multi-class precision-recall matrix. While the ANN experienced more difficulty distinguishing between the two types of VET, it performed better in replicating the choice between 'staying in school' and entering one of the two VET forms. Treating both types of VET as a single transition option results in a precision of 81.42 percent.

Table 1: Confusion matrix of selected ANN

		predicted class		
		staying in school	dual VET	school- based VET
true class	staying in school	$\begin{smallmatrix} 0.894 \\ 0.758 \end{smallmatrix}$ 3503	568	528
	dual VET	303	$\begin{smallmatrix} 0.531 \\ 0.543 \end{smallmatrix}$ 872	430
	school-based VET	108	200	$\begin{smallmatrix} 0.238 \\ 0.591 \end{smallmatrix}$ 300

Confusion matrix for the best-selected ANN structure (4 hidden neurons). Conditional shares of the predicted class in the exponent (multi-class precision) and conditional shares of the actual class in the index (multi-class recall).

We consider several XAI methods for application. The key limiting factors are the applicability to our ANN structure, the availability of implementations in CRAN-R (R Core Team, 2018), and compliance with GDPR⁸. The here used data is only available for on-site evaluation in the secure data room of the Leibniz Institute for Educational Trajectories. Advanced frameworks like LIME (Ribeiro, Singh & Guestrin, 2016) and G-REX (Konig, Johansson & Niklasson, 2008), which both provide rule-based explanations for network classifications, were not applicable due to their reliance on external non-R programs. The illustrations we present in the following use approximate SHAP values provided by the R package 'fastSHAP' (Greenwell, 2020). By illustrating these values as a beeswarm plot for a general overview and scatter plots for a detailed depiction, we visualize the input-output relations of the ANNs. Besides the visual interpretation of the SHAP values, we implement a RE for the hidden layers using simple heuristics from different RE algorithms⁹.

We begin with an analysis of the hidden neurons and apply a discretization of the outputs, where the distribution of activities shows a bi-modal pattern. Where discretization appears beneficial, we replace each suitable hidden neuron with if-then rules of the form: If the sum of the variables exceeds the bias, then the hidden neuron is active (=1); otherwise, it is inactive (=0). This approach directly implements both linearization and discretization. When discretization does not seem beneficial, we keep the neuron's internal structure. We apply weight pruning to remove input variables of hidden neurons with absolute weights below the mean of the absolute values. For mutually exclusive observations, like male and female genders, we add all but one weight to the bias and eliminate the corresponding variables.

The ANN training results in only four hidden neurons, making it feasible to analyze the interactions in the simplified ANN by hand. Let H1, H2, H3 and H4 denote the 4 hidden neurons. The weights of three of the four hidden neurons (H1-H3) (Table A1) exhibit exceptionally high values, indicating that their activations are also likely to be high in absolute terms. The distribution of the activity of the hidden neurons (Figure 4) across all observations supports the same conclusion. Consequently, the continuous neuronal activity of H4 is the primary determinant of classification. A value above a specific threshold triggers a change in classification. The activity or inactivity of the other hidden neurons defines the thresholds (Figure 5) and the target class of the classification.

5 Results

5.1 Extracted Rules

A classification based on the extracted rules instead of the ANN leads with a level of accuracy of 68.09 percent to similar performance like the ANN (68.62 percent), which is remarkable considering the reduced complexity. We analyze extracted rules in a bottom-up strategy. Thus, we first focus on the inputs of the hidden neurons. Here, we consistently observe a mixture between different levels of influences (i.e., individual, regional, and institutional). The exact values of each rule are provided in table 2.

The extracted binary rule for H1 contains diploma type and grades as variables on the individual level. Better grades lead to an activation together with either a lower secondary diploma or an intermediate secondary diploma at a Gymnasium. Contrary, the average last diploma scoring at the school has a negative impact on the activation and cancels out the individual score. This

⁸The General Data Protection Regulation (GDPR) is an EU regulation ensuring the lawful, transparent, and secure processing of personal data, while granting individuals' rights over their data. See regulation (EU) 2016/679 for details.

⁹As we found no implementation of a suitable RE algorithm in CRAN-R

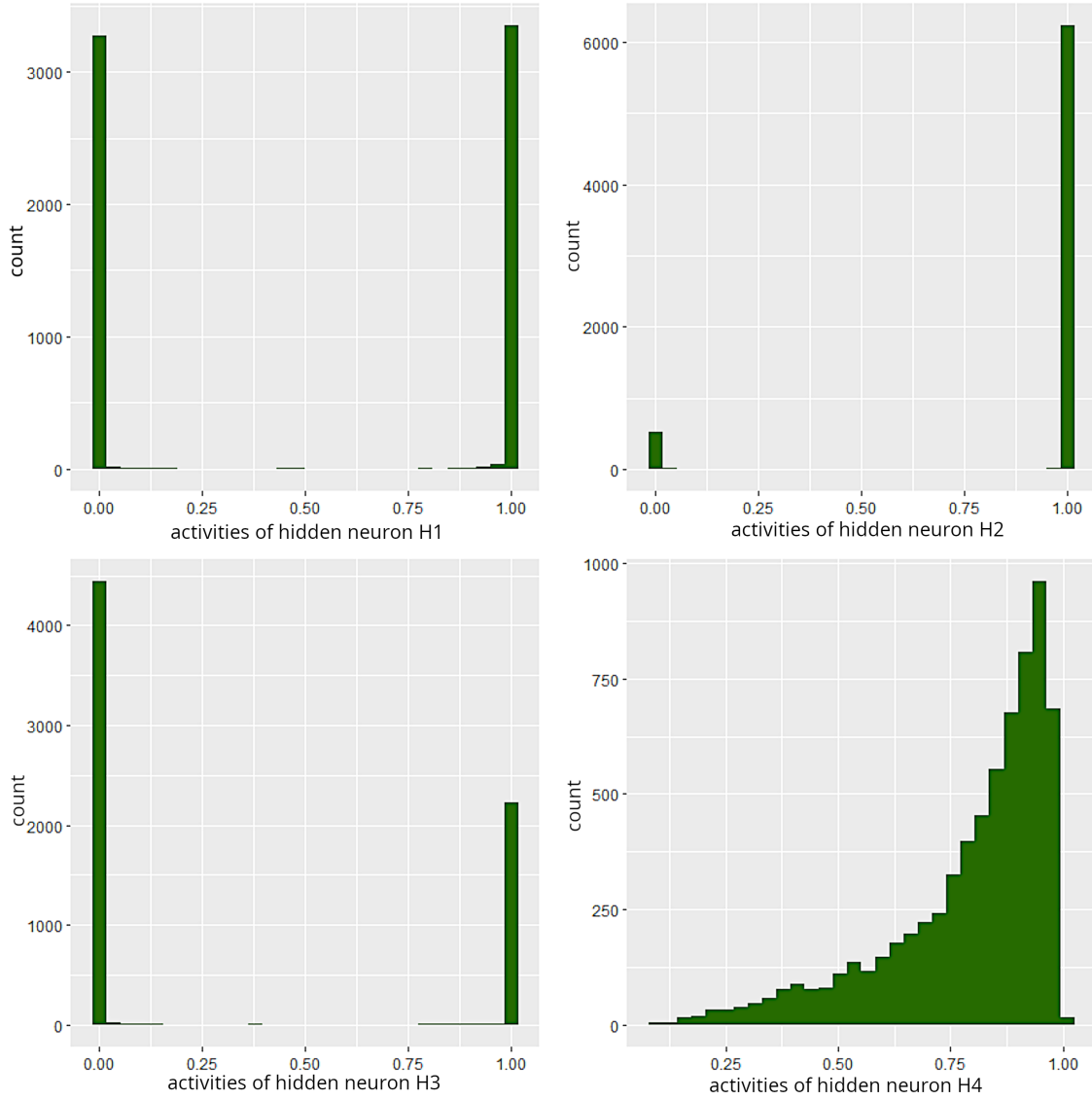


Figure 4: Bar plots of neuronal activities

The x-axis represents counts, while the y-axis shows neuronal activity. The plots are faceted by the four hidden neurons.

means H1 is more likely to be active for individuals significantly better than their school peers. High unemployment in the region, a high share of young inhabitants and urbanity accompanied by vacant VET positions are key drivers of the of activation. In turn, high shares of industry and commuters in the region are drivers of deactivation.

The extracted binary rule for H2 contains the diploma type as individual level information. Having a lower secondary diploma reduces the chance of activation of this neuron to that extend, that it can only become active for lower secondary diploma students if unemployment and urbanity

Table 2: Extracted rules for the activation of the hidden neurons

(a) Binary rule for H1		(b) Binary rule for H2	
Variable	Weights	Variable	Weights
bias	−222.68	reg_unemployed	−494.52
reg_industry_area	−197.89	reg_urban_permeation	−336.96
school_average_score	−158.89	last_diploma_type_0	−286.02
reg_commuters_150km	−124.47	reg_working_females	187.86
last_diploma_score	151.18	bias	374.10
reg_young_people	167.42	reg_industry_area	392.30
reg_unemployed	172.87	Active if SUM	> 0
last_diploma_type_2	187.27		
reg_vet_position	190.21		
last_diploma_type_0	195.09		
reg_urban_permeation	222.35		
active if SUM	> 0		

(c) Binary rule for H3		(d) Rule for H4		Depicted are
Variable	Weights	Variable	Weights	
bias	−363.36	last_diploma_score	−6.45	
reg_urban_permeation	−268.31	school_average_score	−1.98	
reg_unemployed	−194.56	reg_conservative_voters	−1.45	
male	195.2	reg_unemployed	1.85	
reg_working_females	248.31	reg_vet_positions	2.25	
reg_conservative_voters	373.58	bias	7.55	
reg_stem_employments	460.35	$1/(1 + e^{-SUM})$		
active if SUM	> 0			

the extracted rules for each hidden neuron. For H1 to H3, the rules calculate the weighted sum of the input variables, including the bias. If the result is above 0, the respective neuron becomes active (=1). For H4, the rules calculate the weighted sum and apply a logit transformation to determine the activation.

are both low. As the bias is high, H2 is likely to be active for all other diploma types. This is especially the case if the region is characterized by a wide-spreading industry sector and a high share of working women.

The extracted binary rule for H3 contains gender as individual level information. Being male increases the chance of activation. High shares of STEM employees, conservative voters, and a high shares of female employees in the region have positive impacts on the activation, while regional unemployment and urbanity have a negative impact.

The rule for H4 contains individual level information on students' last diploma scores and school level information on average scores. As the bias is the highest factor of influence, we can estimate that the neuronal activity of H4 is high by default, and even increases with high shares of unemployment in the region and open VET positions in the region. The most important factor in decreasing the neuronal activity is the score of the last diploma, followed by the school average score and the share of conservative voters in the region.

As the analysis of each neuron on its own cannot quantify the decisions made by the ANN,

we need to analyze them jointly. A visualization of these joint interactions, conditioned on the neuronal activity of the four hidden neurons, is provided in Figure 5. As H2 has the most active cases, we start our analysis from there: If H2 is inactive, students' transition will either be 'staying in school' or 'dual VET' with different thresholds. If H1 is simultaneously active, it increases this threshold, such that with the same activation value of H4, more students would choose 'staying in school'. If H3 is also active, it lowers the threshold. If only H2 is active, the classification choice is between 'staying in school' and 'school-based VET'. If H3 is active in addition, the choice of classes changes to 'staying in school' and 'dual VET' with lower thresholds. If at least H1 and H2 are active, there is no choice anymore, and the classification is always 'staying in school', regardless of the values of H4.

In general, the cases where H1 and H2 are both inactive and both active seem to be irrelevant. We have no observations that share this condition (H1 and H2 inactive), or the outcome is that almost everyone stays in school (H1 and H2 active), regardless of the neural activation of H4. Thus we have not enough heterogeneity in our data that our ANNs can use.

For the other cases (where either H1 or H2 is active), a higher value of H4 always leads to the transition of 'staying in school', with the main determinants of H4 being individual school performance. Contextual factors such as peer effects on the school level and regional influences play a role - especially conservative voters, the unemployment rate, and the supply-demand ratio of VET positions in the region are important: If unemployment is high and relatively many vet positions are open students with comparatively high grades have a higher probability to enter VET, either 'dual vet' (H1 active) or 'school-based VET' (H2 active).

When only H2 is active, we observe the transitions to be almost equally likely with the same neuronal activity as H4. Such a grouping has two implications: First, this group is less affected by selection and sampling bias, as the ANN would have exploited such a hint. Second, there may exist an unobserved variable which would improve the ANN and also future studies on school-to-training transitions.

5.2 SHAP Values

Depending on the purpose of the analysis, researchers can apply different techniques for interpreting SHAP values. This subsequent section tries to identify subpopulations that 'behave' differently than the general population. For this purpose, the following list contains a description of visible deviations from the linear distributions of SHAP values to the continuous distribution of the original values within a variable. Figure 6 contains a beeswarm plot, for all variables and classifications, visualizing these differences. We limit the description to the most important variables for each of the three target classifications 'staying in school', 'dual VET', 'school-based VET'. The most common case of such deviations are groups of observations with a similar original value that show entirely different SHAP values. We append these observations with an explanation based on the mechanisms that work in the ANN, according to the extracted rules and interpret them. The interpretations of these patterns can be understood as exemplary and are limited to highly plausible and particularly interesting - based on the empirical knowledge of school-to-training transition - findings.

5.2.1 Urbanity

Higher urbanity increases the probability of 'staying in school' and decreases the probability of choosing VET. For some individuals that enter 'dual VET', this relation is reversed but with a smaller magnitude. Based on the extracted rules, we know that the urbanity has a negative effect on the activation of H2 and H3, a positive effect on the activation of H1, and no influence on

H4. An activation of H1 or deactivation of H3 always increases the chance of choosing 'staying in school'. Thus, urbanity can only have a reversed effect, like observed, if H2 is deactivated by the high urbanity. We interpret that, 'staying in school' is more attractive in urban areas because the schools might be more likely to be available that offer this alternative. In such contexts, the VET positions are more likely to be filled with students holding lower levels of schooling. In regions that have a more competitive job market and show higher shares of industrial land use, urbanization has a positive impact on transitioning to 'dual VET' instead of 'staying in school'. In that regard, our results align with already identified differences between rural and urban markets for students with lower and intermediate degrees (Schuster & Margarian, 2021).

5.2.2 Unemployment

Higher shares of regional unemployment increase the probability of choosing VET and decrease the probability of 'staying in school'. For all classifications, there are certain individuals where high and low shares of unemployment lead to the opposite decision, but with less magnitude. Based on the extracted rules, we know that: Regional unemployment has a positive effect on the activation of H1, a negative effect on the activation of H2 and H3, and a negative effect on the activity of H4. Higher activities in H4 always support a choice towards VET, as long as H1 and H2 are not active. An activation of H1 or an deactivation of H3 decrease the probabilities of entering VET. Most likely, H3 induces the reversed effect because unemployment has a higher effect on that neuron compared to other predictors of H3.

Thus, higher unemployment rates make it more attractive to start any type of VET, as VET provides earlier and more secure access to jobs. This is largely in line with the 'discouraged worker hypothesis' (Micklewright, Pearson & Smith, 1990; Hartung, Weßling & Hillmert, 2022; Brunello, 2009). In regions with a high share of STEM employees, a high share of conservative voters, and/or a high share of female labor force participation, the unemployment effect on entering VET versus staying on in school is reversed. In such regions, the status of VET occupations seems to differ. They have a higher value because they are more contested, which makes it difficult to enter them. This status effect is stronger for women, as we find that gender either moderates the previously mentioned predictors or is mediated by at least one of them.

5.2.3 Share of STEM employees

Higher shares of stem employees in the region increase the probability of choosing VET and decrease the chance of 'staying in school'. For 'dual VET', some of the highest STEM employee' shares also have the highest positive impact. For some individuals, the impact is larger than that of any other predictor. Based on the extracted rules, we know that: The share of STEM employees has a positive effect on the activation of H3 and no effect on the other hidden neurons. As the Bias of H3 is highly negative, the high values can only be attributed to cases where H3 will be inactive without the STEM predictor. Most likely, this is only achieved for regions in which the share of STEM employees is exceptionally large. Thus, STEM employees have, on average, no effect on the probabilities of transitioning. In rural areas with a high share of STEM, the effect is positive, increasing the transition to 'dual VET'. As stated in previous research, STEM is relevant for 'dual VET' in rural labor markets compared with any other options. The STEM effect is not only high in rural regions but also in regions with a high share of conservative voters. In such contexts, there seems to be a moderation effect on particular parts of the population.

5.2.4 Female labor force participation

Higher shares of female labor force participation increase the probability of transitioning to VET and decrease the chance of 'staying in school'. For 'dual VET', there are numerous individuals where high and low shares show a reversed behavior with the same magnitude. Based on the extracted rules, we know that: Female labor force participation has a small positive effect on the activation of H2 and H3 and no effect on H1 and H4. An activation of H2 reduces the chances for VET and an activation of H3 increases the chances of VET. This is as long H1 and H2 are not simultaneously active. The direction of the effect changes, depending on the other predictors that influence H2 and H3. Thus, the share of working women influences the transition to VET, especially for female students. The direction of the effect entirely depends on the regional labor market and whether it is a rural or an urban area. In rural areas with low unemployment, a high share of working women seems to motivate female students to enter VET. In urban areas with high unemployment, a high share of working women seems to increase young women's probability of 'staying in school'. Strong gender differences in the German VET system are known, but implicit drivers are hard to identify (Haasler, 2020). The here depicted rural-urban differences can affect policies addressing these disparities.

6 Discussion and Conclusion

By employing the capabilities of ML, our study conducts a comprehensive analysis and exemplary interpretation of school-to-training transitions in the German education system. We train multiple ANNs to emulate the transitions of adolescents during their secondary education. Through explainable artificial intelligence, we try to unveil the intricate ways in which these ANNs predict transitions. We use SHAP (SHapley Additive exPlanations) values for the identification of patterns and the extraction of human-readable rules for interpretation.

Our objectives with this study was: First to apply XAI to a well-studied application in sociological research - namely, post-school educational transitions -, to unravel the black-box of ANNs by making learned patterns interpretable, and finally to illustrate how these patterns could be interpreted and understood within the scope of empirical sociological research.

The interpretations we derive incorporate multiple predictors at different contextual levels simultaneously within the analytical framework, showcasing interactions of individual, regional, and institutional variables together. Notably and in frequently line with previous research (Micklewright, Pearson & Smith, 1990; Hartung, Weßling & Hillmert, 2022; Brunello, 2009; Haasler, 2020; Schuster & Margarian, 2021), regional unemployment rates, supply and demand for VET occupations, STEM employee shares in the region, regional female labor force participation, the highest school diploma, and, of course, individual grades and school performance proved to be the most viable determinants of transitions in our dataset. While some of the above interpretations have already been described in prior research, especially their interactions across multiple levels can be a starting point for future research.

Although our application points the way to new approaches and possibilities of data analysis in the social sciences, it bears notable limitations. The plausibility of the derived interpretations depends on different factors. Mainly the used algorithms and the data itself. The overall quality of the classification in this study is better than expected, with ANNs and extracted rules both showing an accuracy of over 68 percent. Thus, while XAI methods could foster the implementation and application of ML in the social sciences, the possibilities in open-source implementations (especially in R) are still fairly limited. Furthermore, survey data does not prove to be an optimal data source for the application of ML, as it is impossible to generate new data. The only way to account for

underrepresented data is to adapt the training process of the ANNs, which is inferior to adding better and new data. Better-suited data sources are, for example, social media or job advertisement web-pages, where new observations can be extracted and mined on demand.

Beyond that, including posterior predictors is time consuming and might lead to different optimal ANNs. To ensure reproducibility of results, software that enables weight initialization independent of the amount of input variables is necessary. Missing software and the computational constraints limited the computational scope of our ANN models.

Notwithstanding these limitations, the application of ANNs and XAI to survey data effectively replicated well-established relationships concerning school-to-training transitions. Moreover, it generates valuable new insights, particularly regarding non-linear cross-level interaction effects, serving as a promising foundation for future research.

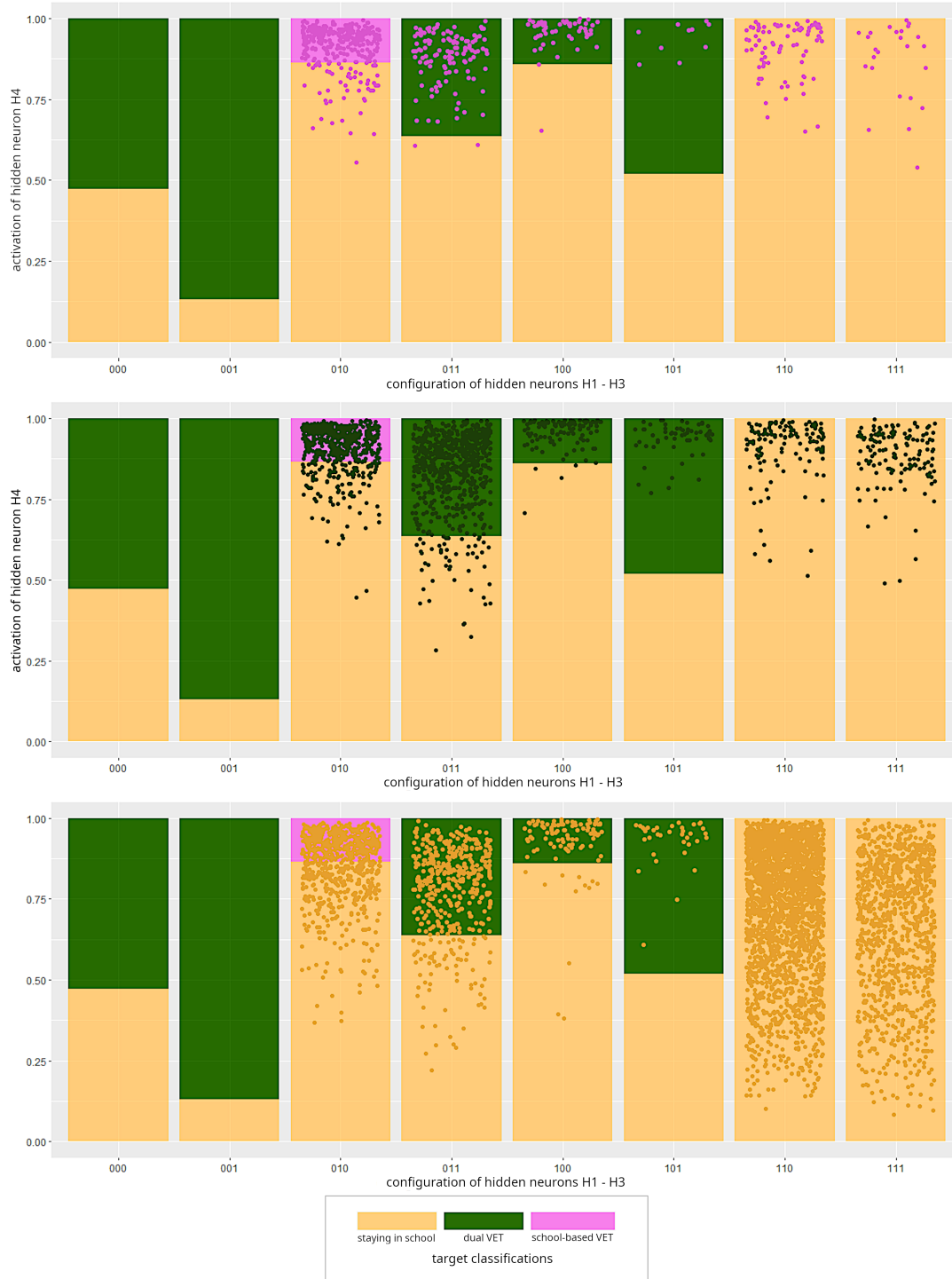


Figure 5: Measured activities for observations depending on the binary activation of H1-H3 and the continuous activation of H4

The activation value of H4 is shown on the y-axis, while combinations of active and inactive neurons H1-H3 are plotted on the x-axis. The plots are faceted by the three target classes and observations belonging to these true classes. The combination '000' corresponds to all three neurons (H1, H2, and H3) being inactive, whereas '111' indicates all three neurons are active.



Figure 6: Beeswarm Plots

The plots illustrate SHAP values on the x-axis and predictors on the y-axis. Colors represent the original value of each predictor, scaled from 0 to 1. The plots depict all correctly classified samples and are faceted by the three target classes.

7 Supplements

Names and description of variables

Individual information

type_train	'0' for continuation in school-based secondary education. '1' for company based VET. '2' for 'school-based VET'. <i>(categorical, derived from ts15201)</i>
last_diploma_score	Grades of school diploma scaled from 1 (worst) to 5 (best). <i>(interval, derived from tf11218)</i>
mig_background_0	No migration background. <i>(binary, derived from t400500-g3R)</i>
mig_background_1	Migration background. <i>(binary, derived from t400500-g3R)</i>
parent_academic_edu_0	No parent studied at a university or comparable institution. <i>(binary, derived from t731422, t731472)</i>
parent_academic_edu_1	At least one parent studied at a university or comparable institution. <i>(binary, derived from t731422, t731472)</i>
parent_school_edu_0	Highest educational degree of both parents is at most at lower secondary education diploma. <i>(binary, derived from t731320, t731370)</i>
parent_school_edu_1	Highest educational degree of both parents is intermediate secondary education diploma. <i>(binary, derived from t731320, t731370)</i>
parent_school_edu_2	Highest educational degree of both parents is higher secondary education diploma. <i>(binary, derived from t731320, t731370)</i>
female	Participant categorizes herself as 'female'. <i>(binary, derived from t700031)</i>
male	Participant categorizes himself as 'male'. <i>(binary, derived from t700031)</i>
last_diploma_type_0	Participant finished lower secondary education <i>(binary, derived from ts11209)</i> .
last_diploma_type_1	Participant finished intermediate secondary education at a different school than Gymnasium. <i>(binary, derived from ts11209)</i>
last_diploma_type_2	Participant finished intermediate secondary education at a Gymnasium. <i>(binary, derived from ts11209)</i>

School Contexts

avg_school_score	Average final scores of all participants of this school. Grades of scaled from 1 (worst) to 5 (best). (<i>continuous, derived from tf11218</i>)
------------------	---

Regional Contexts

reg_unemployment	Share of People without employment
reg_vet_positions	Ratio of Supply of VET Contracts to Demand of VET contracts as of 2013
reg_stem_employees	Share of workers working in STEM oriented services as of 2013
reg_conservative_voters	Share of Votes for the Christ-Democratic Union 2017
reg_big_companies	Share of Businesses with more than 250 employees as of 2013
reg_commuters_150km	Share of worker with more than 150km distance commuting as of 2017
reg_female_employees	Share of female employees out of all employees as of 2013
reg_young_people	Share of inhabitants between 18 and 25 years as of 2011
reg_urbanity	Urban permeation as of 2010
reg_industry_area	Share of Land Area used for Industry and Trade as of 2011

Table A1: First layer weight matrix for neural network with 4 hidden neurons (base model)

Neuron	H1	H2	H3	H4
bias	-1.121	14.856	-18.531	2.010
last_diploma_score	151.177	-108.015	2.476	-6.451
reg_unemployed	172.868	-494.516	-194.556	1.850
reg_mint_employments	55.825	64.034	460.345	0.389
reg_conservative_voters	-75.422	5.810	373.581	-1.454
reg_vet_positions	190.214	107.901	145.152	2.250
reg_big_companies	-81.698	8.302	129.199	-0.081
reg_commuters_150km	-124.473	58.628	134.255	0.343
reg_working_females	-7.871	187.858	248.308	-0.293
reg_young_people	167.420	-13.386	-115.686	0.825
reg_industry_area	-197.889	392.302	51.112	-1.333
reg_urban_permeation	222.349	-336.960	-268.306	-0.082
school_average_score	-158.886	-8.021	-150.826	-1.982
last_diploma_type_0	3.783	-2.827	-101.900	2.307
last_diploma_type_1	-191.306	283.195	-27.169	1.845
last_diploma_type_2	-4.037	295.054	0.895	1.213
female	-20.204	35.344	-194.090	0.274
male	2.407	1.738	1.113	-0.448
mig_background_0	-6.393	2.558	-0.706	1.251
mig_background_1	-0.934	30.548	-88.940	1.177
parents_school_edu_0	-60.080	78.037	-6.478	0.846
parents_school_edu_1	-3.234	37.352	-122.375	1.144
parents_school_edu_2	8.197	0.875	-157.792	0.772
parents_academic_edu_0	-0.421	0.790	-0.483	1.027
parents_academic_edu_1	-24.854	15.788	-36.086	0.366

Table A2: Second layer weight matrix for neural network with 4 hidden neurons (base model)

Neuron	dual VET	school-based VET	staying in school
bias	-1.602	-4.608	2.375
H1	-1.058	-2.185	2.210
H2	-2.136	-0.700	2.130
H3	1.694	-0.774	-1.185
H4	2.711	5.631	-5.717

References

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. & others (2020). Explainable Artificial Intelligence

- (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, 58, 82–115.
- BBSR (2022). Laufende Raumbeobachtung des BBSR - INKAR. *Bundesinstitut für Bau- Stadt- und Raumforschung (BBSR) im Bundesamt für Bauwesen und Raumordnung (BBR)*.
- Blossfeld, H.-P., Roßbach, H.-G. & Von Maurice, J. (2019). *Education as a lifelong process: The German National Educational Panel Study (NEPS)* (2 Ed.), Volume 3 of *Edition ZfE*. Springer VS.
- BMBF (2024). Berufsbildungsbericht 2024. *Bundesministerium für Bildung und Forschung (BMBF)*.
- Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J. & others (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.
- Brunello, G. (2009). The effect of economic downturns on apprenticeships and initial workplace training. A review of the evidence. *Empirical research in vocational education and training*, 1(2), 145–171.
- Carvalho, A. R., Ramos, F. M. & Chaves, A. A. (2011). Metaheuristics for the feedforward artificial neural network (ANN) architecture optimization problem. *Neural Computing and Applications*, 20(8), 1273–1284.
- Cook, T. D. (2003). The case for studying multiple contexts simultaneously. *Addiction (Abingdon, England)*, 98 Suppl 1, 151–155.
- De Veaux, R. D. & Ungar, L. H. (1994). Multicollinearity: A tale of two nonparametric regressions. *Selecting models from data: artificial intelligence and statistics IV* (pp. 293–302).
- Di Franco, G. & Santurro, M. (2021). Machine learning, artificial neural networks and social research. *Quality & quantity*, 55(3), 1007–1025.
- Ding, X., Zhang, Y., Liu, T. & Duan, J. (2015). Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*. Buenos Aires, Argentina: AAAI Press.
- Flohr, M., Menze, L. & Protsch, P. (2020). Berufliche Aspirationen im Kontext regionaler Berufsstrukturen. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 72(S1), 79–104.
- Geman, S., Bienenstock, E. & Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 4(1), 1–58.
- Greenwell, B. (2020). Package ‘fastshap’.
- Haasler, S. R. (2020). The German system of vocational education and training: challenges of gender, academisation and the integration of low-achieving youth. *Transfer: European Review of Labour and Research*, 26(1), 57–71.
- Hailesilassie, T. (2016). Rule extraction algorithm for deep neural networks: A review. *arXiv preprint arXiv:1610.05267*.
- Hartung, A. & Weßling, K. (2024). Discouraged and hedged – why students enter VET after obtaining university eligibility. *Journal of Vocational Education & Training*, 0(0), 1–24.

- Hartung, A., Weßling, K. & Hillmert, S. (2022). Interplay between family background and labour-market conditions in shaping students’ occupational status expectations. *Journal of Education and Work*, 35(4), 405–421.
- Hillmert, S., Hartung, A. & Weßling, K. (2017). A decomposition of local labour-market conditions and their relevance for inequalities in transitions to vocational training. *European Sociological Review*, 33(4), 534–550.
- Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J. & Melançon, G. (2008). *Visual analytics: Definition, process, and challenges*. Springer.
- Khan, M. T., Kaushik, A. C., Ji, L., Malik, S. I., Ali, S. & Wei, D.-Q. (2019). Artificial neural networks for prediction of tuberculosis disease. *Frontiers in microbiology*, 10, 395.
- König, R., Johansson, U. & Niklasson, L. (2008). G-REX: A versatile framework for evolutionary data mining. In *2008 IEEE International Conference on Data Mining Workshops* (pp. 971–974). IEEE.
- Lundberg, I., Brand, J. E. & Jeon, N. (2022). Researcher reasoning meets computational capacity: Machine learning for social science. *Social Science Research*, 108, 102807.
- Lundberg, S. M. & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Mason, C., Twomey, J., Wright, D. & Whitman, L. (2018). Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a backpropagation neural network and logistic regression. *Research in Higher Education*, 59, 382–400.
- Maszczyk, A., Gołaś, A., Pietraszewski, P., Roczniok, R., Zajac, A. & Stanula, A. (2014). Application of neural and regression models in sports results prediction. *Procedia-Social and Behavioral Sciences*, 117, 482–487.
- Meinel, G. & Schumacher, U. (2010). Konzept, funktionalität und erste exemplarische Ergebnisse des Monitors der Siedlungs- und Freiraumentwicklung (IOER-Monitor). *Flächennutzungsmonitoring II. Konzepte – Indikatoren – Statistik*.
- Micklewright, J., Pearson, M. & Smith, S. (1990). Unemployment and Early School Leaving. *The Economic Journal*, 100(400), 163.
- NEPS Network (2021). National Educational Panel Study, Scientific Use File of Starting Cohort Grade 9. *Leibniz Institute for Educational Trajectories (LIfBi) Bamberg*.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ribeiro, M. T., Singh, S. & Guestrin, C. (2016). ”Why should i trust you?” Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144). Association for Computing Machinery.
- Rich, P. & Owens, A. (2023). Neighborhood–School Structures: A New Approach to the Joint Study of Social Contexts. *Annual Review of Sociology*, 49(1), 297–317.
- Riedmiller, M. & Braun, H. (1994). Rprop-description and implementation details. Technical report, Technical Report. Univer.
- Rohrer, J. M., Hünermund, P., Arslan, R. C. & Elson, M. (2022). That’s a Lot to Process! Pitfalls of Popular Path Models. *Advances in Methods and Practices in Psychological Science*, 5(2).

- Schuster, K. & Margarian, A. (2021). Vocational training choice from a regional perspective. *Empirical Research in Vocational Education and Training*, 13(1), 3.
- Shapley, L. S. (1951). *Notes on the N-person Game*. Rand Corporation.
- Shu, X. & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817.
- Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd Ed.). London: Sage Publishers.
- Wessling, K. & Meng, C. (2021). The mediating and moderating role of neighbourhoods and regions on second-generation immigrant youths' school-to-work transitions in the Netherlands. *Population, Space and Place*, 27(2), e2384.
- Wright, M. & Stachelek, J. (2020). R neuralnet.
- Zeinalizadeh, N., Shojaie, A. A. & Shariatmadari, M. (2015). Modeling and analysis of bank customer satisfaction using neural networks approach. *International Journal of Bank Marketing*.