INSTITUTE FOR EMPLOYMENT RESEARCH

The Research Institute of the Federal Employment Agency

IAB-DISCUSSION PAPER

Articles on labour market issues

09|2024 Predicting Job Match Quality: A Machine Learning Approach

Sabrina Mühlbauer, Enzo Weber

ISSN 2195-2663

Predicting Job Match Quality: A Machine Learning Approach

Sabrina Mühlbauer (Institute for Employment Research),
Enzo Weber (Institute for Employment Research, University of Regensburg)

Mit der Reihe „IAB-Discussion Paper“ will das Forschungsinstitut der Bundesagentur für Arbeit den Dialog mit der externen Wissenschaft intensivieren. Durch die rasche Verbreitung von Forschungsergebnissen über das Internet soll noch vor Drucklegung Kritik angeregt und Qualität gesichert werden.

The “IAB-Discussion Paper” is published by the research institute of the German Federal Employment Agency in order to intensify the dialogue with the scientific community. The prompt publication of the latest research results via the internet intends to stimulate criticism and to ensure research quality at an early stage before printing.

Table of Contents

Abstract

This paper develops a large-scale algorithm-based application to improve the match quality in the labor market. We use comprehensive administrative data on employment biographies in Germany to predict job match quality in terms of job stability and wages. The models are estimated with both machine learning (ML) (i.e., XGBoost) and common statistical methods (i.e., OLS, logit). Compared to the latter approach, we find that XGBoost performs better for pattern recognition, analyzes large amounts of data in an efficient way and minimizes the prediction error in the application. Finally, we combine our results with algorithms that optimize matching probability to provide a ranked list of job recommendations based on individual characteristics for each job seeker. This application could support caseworkers and job seekers in expanding their job search strategy.

Zusammenfassung

Dieses Papier beschäftigt sich mit einer groß angelegten Datenanalyse um die Matching-Qualität auf dem Arbeitsmarkt zu untersuchen. Hierfür verwenden wir einen sehr umfangreichen administrativen Datensatz zu Arbeitsmarktbiographien in Deutschland. Die Schätzungen werden sowohl mit maschinellem Lernen (extreme gradient boosting), als auch mit traditionellen statistischen Methoden (OLS, logit) durchgeführt. Bei der Gegenüberstellung beider Methoden wird deutlich, dass maschinelles Lernen insbesondere in den Bereichen Mustererkennung, Analyse von sehr großen Datensätzen und Minimierung der Fehlerrate deutliche Vorteile gegenüber den herkömmlichen Methoden aufweist. Schließlich werden die Prognosen für Matching-Qualität (Stabilität und Löhne) mit Matching-Wahrscheinlichkeiten kombiniert. Anhand dieser Ergebnisse wird für jede arbeitssuchende Person eine Liste mit Berufsvorschlägen generiert. Damit können Arbeitsvermittlern und Arbeitssuchenden Alternativen aufgezeigt werden, wodurch sich ihr Suchverhalten auf dem Arbeitsmarkt erweitern könnte.

JEL

C14, C45, C55, J64

Keywords

Job Match Quality, Job Recommendation, Machine Learning, Matching, XGBoost

Acknowledgments

We thank Hermann Gartner, Anja Bauer, Michael Stops and Ben Lochner for their valuable comments and suggestions. Special thanks go to the DIM department of the Institute for Employment Research (IAB), especially Ulrich Thomsen and Andreas Schneider, for their efforts to provide the data. We are grateful to the participants of the conference of the European Association of Law and Economics 2023 and the conference on Econometrics and Statistics 2023.

1 Introduction

In the labor market, job seekers encounter many different job offers in different occupations. In addition to matching probability, job match quality plays a key role in decision making. This role is underlined in a strand of literature that analyzes the effect of search strategies on job quality based on survey data (see Simon/Warner (1992), Olsen/Kalleberg/Nesheim (2010)). Several studies have shown that job searches via newspapers, personal contacts and especially online job searches (e.g., Holzer (1988), Freeman (2002)) lead to better job match quality. Mang (2012) showed, based on individual-level survey data, that online job searches provide better matches than do searches via newspapers, job agencies or personal recommendations. Another study from Cappellari/Tatsiramos (2015) used survey data to analyze the effect of friendship ties on job finding probability and job match quality.
The main goal of this paper is to provide a ranked list of job recommendations for each job seeker. The job recommendations include different criteria, such as matching probabilities and job match quality. Currently, caseworkers make job suggestions based on their experience. We try to support them by providing additional job alternatives. To do so, we analyze historical patterns of employment biographies and make predictions for current job seekers.
This study contributes to the literature by predicting individual job match quality to support caseworkers in German employment agencies. We investigate how job match quality can be analyzed with statistical methods, especially machine learning methods. Job match quality is measured by predicting job stability and wages using large amounts of administrative data containing information on employment biographies in Germany. Another focus is on analyzing which method is best for solving the underlying research question. In particular, the differences between common statistical methods and machine learning algorithms in terms of error rates, application and estimation efficiency are interesting points. Additionally, we show how hyperparameter tuning for extreme gradient boosting (XGBoost) works for an extremely large dataset for classification and regression problems. Furthermore, we create a combined indicator for job match chances and quality. To be more precise, we combine the expected wages and the predictions for job stability with the individual matching probability (see Mühlbauer/Weber (2022)). The components are calculated based on individual characteristics of the German labor force. Thus, we provide a ranked list of job recommendations that includes both matching probability and matching quality information for each job seeker. Caseworkers could benefit from this information because they can provide job seekers with alternative occupations in addition to their own suggestions.
Having additional information on job match quality could support caseworkers in German employment agencies and help job seekers find a job in the job category that suits best or

motivates them to extend the search strategy. To do this, we weight the job match probabilities (see Mühlbauer/Weber (2022)) by including quality measures of wage and job stability.
We use administrative data from the German Employment Agency. The Institute for Employment Research provides these data in the Integrated Employment Biographies. This dataset contains employment biographies from each person in the German labor market. Naturally, this is a very large dataset that requires considerable computational power and the efficient application of estimation methods. We compare the estimations of different common statistical methods (i.e., multinomial logit or OLS) and machine learning methods (XGBoost or neural networks) regarding the error rates. Furthermore, we show how to address problems that occur due to the size of the underlying dataset.
Our study is related to several strands of literature. A broad strand of literature considers the effect of reforms on job match quality, especially on the effects of different schemes of unemployment benefits (e.g., Caliendo/Tatsiramos/Uhlendorff (2013)). For example, Van Ours/Vodopivec (2006), Giannelli/Jaenichen/Rothe (2013), van den Berg/Vikström (2014), Nekoei/Weber (2017) and Gartner/Rothe/Weber (2022) measure job match quality by wages or job stability.
We also connect to the literature that has emerged during recent years in which machine learning approaches have been applied to labor market research. In practice, machine learning (see Kianpisheh/Jalili/Charkari (2012) and Gulyas/Pytka et al. (2019)) is applied to process web texts to address big data problems or to perform policy evaluations. For example, Amato et al. (2015) and Boselli et al. (2018) analyzed vacancies via text classification.
The paper is structured as follows. Section two contains information on the data and the variables that measure job match quality (job stability and wages). Section three describes the empirical procedure and the results. Section four describes the construction of the indicator, and Section five concludes the paper.

2 Setting

2.1 Data

We use a ten percent sample of a large administrative dataset, namely, the Integrated Employment Biographies (IEB)[1] (see Antoni/Ganzer/vom Berge (2016)). The IEB contains the members of the German workforce who were covered by the social security system from

[1] Version 14.00.00-190927

1980 onward and provides information about the characteristics of each person, as well as the characteristics of their jobs. Using this dataset allows us to capture hard facts regarding employment history. The dataset combines different data sources. For the present research question, jobseeker histories (ASU/XASU) and employee histories (BeH) are needed.
It is crucial to have concrete information on occupations. In 2010, a new classification system (classification of occupations 2010) for occupations was introduced. From November 2011 onward, a new classification system was applied to the data. Due to the recoding, errors in the ”occupations” variable occurred. Thus, for our estimations, we use data from 2012 to 2018 (analogously to Mühlbauer/Weber (2022)).
The dataset allows us to include information about gender, nationality, education, age, the federal state a person lives in, the number of days of unemployment before finding a new job, the job category of vocational training, children (children under 15 years old), previous skill levels, previous occupations and tenure.

2.2 Defining Job Match Quality

In this paper, we use two indicators to define job match quality, namely, job stability and wages (see Nekoei/Weber (2017), Gartner/Rothe/Weber (2022)). Importantly, these indicators have the advantage that they can be accurately measured by building on rich process data.

2.2.1 Job Stability

We use the duration an individual is employed in an occupation as a measure of job stability. We distinguish between two different cases. One is the time of employment in the same occupation (i.e., job category), which is called ”occupation duration”. Here, we define the change in the job category or the transition to unemployment as an interruption. The other is the time of employment without interruption by a spell of unemployment, which is called ”employment duration”. We define the related binary outcome variable as follows:

$$\text{duration} = \begin{cases} \text{less than 6 months}, & \text{short-term employment} \\ \text{more than 6 months}, & \text{long-term employment}. \end{cases} \tag{1}$$

In Germany, the probation period ends after six months. This means that after this period, the Protection Against Dismissal Act (Kündigungsschutzgesetz) holds. Thus, six months is a practical threshold that is also used by Heisz (1999), for example. Furthermore, when choosing such a short period, only observations within six months of the end of the observation period must be dropped. To be more precise, we take observations from 2012

to 2016 to estimate the model and those from 2017 to evaluate the model. We find that for some people, there are multiple observations in the sample. When counting the number of employment days and the number of days in the same occupation, both ongoing and terminated employment are included. Note that the reported means do not show the true mean of being employed because we cannot determine how long one's ongoing employment will last.
Table 1 shows the share of people employed long-term or short-term for both the employment and occupation duration samples. Here, we see that the distributions of the observations are almost equal for the training and testing sets.

Table 1: Proportion of observations per job category [in percent]

	employment duration		occupation duration	
sample*	short-term	long-term	short-term	long-term
train set	31.15	68.85	41.47	58.53
test set	27.45	72.55	37.02	62.98

*Short-term and long-term employment are defined in Equation 1.
We can see that the proportions of long-term and short-term employment are similar for the test and train sets for both samples (employment and occupation duration).
Source: own calculations ©IAB

2.2.2 Wages

The second indicator for job match quality is the daily wage. Since the working time information allows one to differentiate only full- and part-time employment, we use only full-time employment. The wages are reported as daily wages in the data (i.e., the yearly wage is divided by the number of days in the corresponding year).
For data protection reasons, wages above the contribution limit for social security are right censored. Therefore, we apply the imputation procedure described in Gartner (2005) (see appendix). Table 2 shows some statistics for 2017 and 2018. For both years, the number of observations and the number of different persons are almost equal. The mean and the median for the daily wage are approximately 10 € higher for 2018.

Table 2: Descriptive statistics for the wage data

year	2017	2018
no. observations	1,050,210	1,092,315
no. different persons	856,636	896,990
mean wage	111.44 €	113.56 €
median wage	98.84 €	101.02 €
min wage	64.50 €	64.50 €
max wage	1,041.84 €	1,064.52 €

Source: own calculations ©IAB

2.3 Explanatory Variable

The IEB data contain information on both personal and job characteristics, for example, age, gender, marital status, children or residence at the federal state level. Furthermore, we observe the nationality variable. Naturally, the largest nationality group consists of Germans. For non-Germans, we distinguish between the nationalities of asylum seekers[2] and EU and non-EU nationalities. Table 3 provides an overview of the dependent variables. The data also contain variables concerning education or vocational training. Furthermore, we use information on the skill level required for certain jobs, preceding occupations or tenure.

Table 3: Explanatory variables

Variable	Description
Gender	Female, male
Federal state	Nordrhein-Westfalen), Bayern, Schleswig-Holstein, Sachsen-Anhalt, Hessen, Baden-Württemberg, Brandenburg, Mecklenburg-Vorpommern, Thüringen, Sachsen, Niedersachsen, Bremen, Berlin, Hamburg, Saarland, Rheinland-Pfalz
Nationality	German, EU, 8 main migration countries, Europe without EU, remaining countries
Marital status	Single/ lives alone, in relationship/ married
Children	minimum of one child under 15 years, no children/ unknown
Education	No school leaving certificate, primary/ lower school, intermediate school without vocational training, primary/ lower school, intermediate school with vocational training, upper secondary school leaving certificate without vocational training, upper secondary school leaving certificate with vocational training, university of applied sciences, university
Skill level	Skill level required for preceding employment(s)
Preceding occupation	Occupational group(s) employed in before starting a new job
Vocational training	Occupational group of vocational training
Age	Age at the start of employment
Days in unemployment	Number of days unemployed before starting a new job

Source: IEB; own calculations ©IAB

[2] Here, we refer to people whose nationality is equal to one of the eight countries of origin with the highest number of AS in July 2020 (source: https://de.statista.com/statistik/daten/studie/154287/umfrage/hauptherkunhslaender-von-asylbewerbern/). These eight countries are Syria, Iraq, Afghanistan, Turkey, Nigeria, Iran, Eritrea, Somalia and Georgia.

3 Empirical Strategy and Results

3.1 Prediction of Job Stability

3.1.1 Model

To predict job stability, a binary classification problem must be solved. The model is given by the following:

$$\text{P(duration } < 6 \text{ months)} = f(\mathbf{X}, \mathbf{Y}), \tag{2}$$

where **X** is an nxi-matrix covering i characteristics of every person and **Y** is an nxj-matrix covering j characteristics of the jobs of every person. Furthermore, n is the number of observations (i.e., spells). Finally, the vector P contains the probability of being employed in the short term.

3.1.2 Estimation

Since the aim of our research is to make predictions for future employment with historical data, we apply a test-train split by year. The training set contains observations from 2012 to 2016, and the test set contains observations from 2017 for evaluation purposes. When applying the model in practice, it would be feasible to work with continuously updated estimations based on the most recent data.
We use different methods to estimate the model. The classical application is to estimate a logit model, which is commonly used to solve binary classification problems. We make predictions for the test set and calculate the perfect cutoff value to classify the outcome. We find that for both samples, the cutoff value is approximately 50 percent.
In machine learning, tree-based methods are often recommended for solving a large range of different classification problems. Thus, we apply both random forest and extreme gradient boosting (XGBoost). We find that in our application, XGBoost performs better. The algorithm works for both regression and classification problems. In brief, XGBoost is a boosting method in which errors are minimized via an iterative optimization algorithm that minimizes a loss function. We choose the optimal specification by taking the model that minimizes the error rate[3] of the test set.

[3] There are different error rates that can be used. One must choose the error rate required for the underlying estimation problem. For predicting job stability, we choose the classification error as a measure of goodness for choosing the right model.

After the tuning process of the hyperparameters, the final model can be estimated. Due to the large number of observations, even a single tuning step takes a long time. Thus, in the first step, we test some modifications for one hyperparameter while the other ones are set to the default value. Then, we obtain the values that minimize the classification error and take the new values as default values and repeat the previous step. After this procedure, the error rates vary only around the third decimal. Since checking all possible hyperparameter combinations would be extremely time consuming for very small improvements, we apply this reduced version of the tuning process.

3.1.3 Results

We use the classification error (CE) as a measure of goodness and compare the different methods. The CE is given by the following:

$$\text{error rate } = 1 - \text{accuracy } = 1 - \frac{\text{number of correct predictions}}{\text{total number of observations}}. \tag{3}$$

Table 4 shows the results for XGBoost and logit models. For both samples, we obtain clear

Table 4: Error rates for duration in percent

	Logit	XGBoost
Occupation duration	17.44	13.68
Employment duration	23.93	15.15

The results are calculated with R (required packages: xgboost, glm). The model contains all variables on personal and job characteristics available in the ASU/XASU and BeH. The models are estimated with the focus on minimizing the error rate for the test set (shown in the table).
Source: own calculations ©IAB

advantage of XGBoost in comparison to logit. To be more precise, the CE for logit is approximately 27.5 percent greater for the occupation duration sample and approximately 58.0 percent greater for the employment duration sample. Thus, for predicting the duration of employment, machine learning methods are preferred.
To obtain deeper insights into the results, we examine the importance of the variables. The gain is defined as the relative contribution of the corresponding variables to the model calculated by taking each feature's contribution for each tree in the model. The higher the value is, the more important a feature is for generating a prediction. In both XGBoost predictions, the most important variable is marital status, with a gain of 31.16 percent for the employment duration sample and a gain of 24.47 percent for the occupation duration sample. Furthermore, age and the number of days in unemployment before reemployment are very important for predicting job stability.
To control for the robustness of the results, we apply another threshold of one year. The aim is to check whether XGBoost is still preferred over logit. We find that XGBoost produces smaller error rates in this case. More details are shown in the appendix.

3.2 Prediction of Wages

3.2.1 Model and Estimation

Predicting wages is a classical regression problem. The model is given by the following:

$$\text{ln(wage)} = f(\mathbf{X}, \mathbf{Y}). \tag{4}$$

We solve the regression problem with OLS and XGBoost. The tuning procedure is equivalent to the XGBoost estimation of the classification problem described in Section 3.1. We also apply a test-train split by year. Since we are able to use the full sample, we take observations from 2012 to 2017 as the training set and those from 2018 as the test set.

3.2.2 Results

For the calculation of the , we check two different measures, namely, the mean squared error (MSE) and the mean absolute error (MAE), which are given by the following:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Z_i - \hat{Z}_i)^2 \tag{5}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|Z_i - \hat{Z}_i|, \tag{6}$$

where N is the number of observations, Z_i is the true value and $\hat{Z}_i$ is the predicted value. First, we apply an OLS to a linear regression model, which is given by

$$\text{ln(wage)} = \mathbf{X}\beta + \mathbf{Y}\gamma + \epsilon, \tag{7}$$

with β and γ containing the coefficients of **X** and **Y**, respectively. Furthermore, ϵ denotes the error term.

In the next step, we apply machine learning algorithms. We choose the best model by minimizing the MSE in the test set. Thus, the test MSE is calculated for each iteration step. Finally, the model with the lowest MSE is chosen. Table 5 shows the MSE and MAE for the log wage and Table 6 shows the results for the wage in euros.

Here, we see that the MSE for XGBoost is approximately 11.1 percent greater than that for OLS.

Table 5: Error rates of log daily wage

	OLS	XGBoost
MSE	0.0620	0.0558
MAE	0.1923	0.1820

The models are estimated with R (required package: xgboost). We include variables on personal and job characteristics included in the ASU/XASU and BeH. The models are optimized with respect to minimizing the MSE of the test set
Source: own calculations ©IAB

Table 6: Error rates of daily wage

	OLS	XGBoost
MSE	1246.61	1167.81
MAE	23.49	22.35

This table shows the transformation of the results for the log(wage) in euros to make the results more visible.
Source: own calculations ©IAB

Additionally, we find that XGBoost obtains better results with more training observations being included in the training set. The best results are produced by using the maximum period extending from 2012 onward in the training set. Concerning the MSE and MAE, in contrast to XGBoost, OLS hardly reacts to the variation in the number of observations. Since the number of observations from one year is already very large, an additional enlargement does not improve the accuracy. Thus, we suspect that XGBoost covers more complex patterns in the data, while OLS cannot find these structures. This result also highlights the potential to increase the accuracy even further when the algorithm is applied in practice with more long-term data that reaches to the current limit of what is available.

4 Creating a Matching Index

Jobs with high wages, high stability and high matching probabilities are naturally attractive for job seekers. Thus, job recommendations based on a combination of these variables could be crucial for placement decisions. For example, Allen/Van der Velden (2001), Green/Zhu (2010), Mavromaras et al. (2013) and Pecoraro (2014) show that mismatch could lead to negative labor market outcomes like wage penalties, absenteeism or high turnover. Thus, including variables that describe job match quality is crucial for making job recommendations.

In Mühlbauer/Weber (2022), individual matching probabilities were predicted. For this purpose, the authors used data on employment biographies, as described in the present paper. For each person, they predicted the probability of becoming employed in a certain occupation (i.e., 3-digit in classification of occupations 2010). They found that random forest (RF) is best for predicting matching probabilities (i.e., classification of 144 categories). In practice, it would be useful for caseworkers and job seekers to expand this information by including quality aspects. Thus, combining these aspects could lead to a more comprehensive approach than that obtained by solely looking at matching probabilities. We construct a matching index to provide a combined information. Both factors, namely, job match quality (i.e. stability and wages) and matching probability are equally weighted. Thus, the matching index Q_{rs} is given by the following:

$$Q_{rs} = \underbrace{0.25 \cdot P(\text{duration}_r > 6 \text{ months}|s) \cdot 0.25 \cdot E[\text{wage}_r|s]}_{50\% \text{ job match quality variables}} \cdot \underbrace{0.5 \cdot P(\text{M}_r = s)}_{50\% \text{ matching probability}} , \quad (8)$$

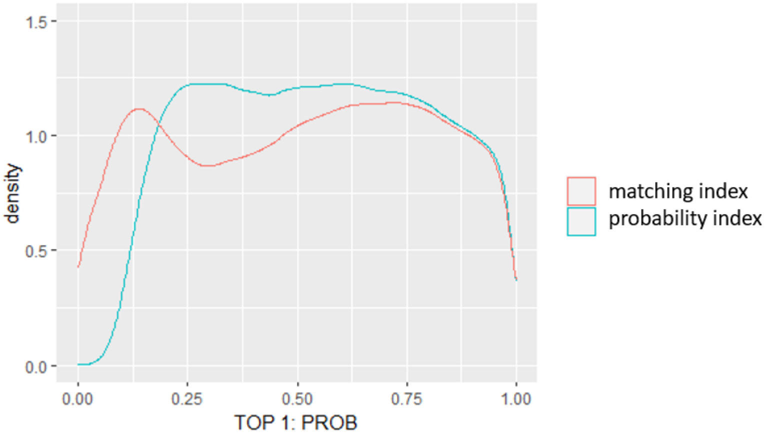
with r = 1, ..., N, N is the number of observations and s = 1, ... S, where S is the number of occupations. The matching quality indicator is calculated by the probability that person r is employed longer than six months in job category s, multiplied by the expected wage of person r in job category s and by the probability that person r is employed in job category s at all. Therefore, for each person, we obtain S index values. To be more concrete, the index is composed of the wage and duration conditional on each occupation and the matching probability for each occupation. This means that the expected value for a stable wage in a certain occupation is multiplied by the corresponding matching probability. For comparability reasons, we put all variables on the same scale before calculating the index. Afterwards, a ranking of the occupations for each person is possible.
To calculate the matching index (see formula 8), we need the predictions for each observation for 2017. For this purpose, wages are predicted based on a test set from 2016. Thus, we obtain 144 predicted wages for each observation. Additionally, we must re-estimate the matching probabilities (Mühlbauer/Weber (2022)) for 2017. Mühlbauer/Weber (2022) defined matching as a job seeker entering into employment by being matched to a specific occupation. They used the same data on employment biographies as those used in the current project to predict matching probabilities with statistical methods. The model is given by the following:

$$P(M_r = s) = f(\mathbf{X}_r, \mathbf{Y}_r), \quad (9)$$

with $s = 1, \ldots, S$ where S is the number of different occupations, $r = 1, \ldots, N$, N is the number of observations, and **M** is a Nx1-vector containing the occupational group in which someone is employed. Thus, M_r denotes the occupation under observation r. Furthermore, $\mathbf{X}_r$ is a 1xU-vector, where U is the number of variables denoting the characteristics of person r. $\mathbf{Y}_r$ is a 1xV-vector, where V is the number of variables denoting the characteristics of the jobs of person r. Thus, Equation (2) estimates the probability of a person being employed in a certain occupation and assumes that the process is a function of personal and job characteristics, both of which are included in the dataset. The authors found that random forest is best for making such predictions (i.e., solving a classification problem with 144 occupations).

Since each estimation is based on different persons or spells, the matching index can only be calculated for observations that are included in all samples. To be more precise, only full-time jobs subject to social security can be considered. Ultimately, the matching index can be calculated for 129,840 observations.

Figure 1: Cumulative density function for the matching index (equal weights) and probability index



Densities for the matching probabilities of the TOP1 job recommendations.
Source: own calculations ©IAB

Similar to the matching index, we create a matching probability index. This means that the matching probabilities for each person are scaled analogously to the scaling of the matching index and that both indexes can be compared directly. Figure 1 shows the density of the first-best recommendation for each index. As expected, considering the quality dimensions has some probability mass to the left. This difference is due to a tradeoff against higher job stability and wages.
Table 7 elucidates this tradeoff by showing the mean values of all three dimensions (i.e., probability, stability and wage) reached by both the probability index and the matching index for the job recommendations from first to third place (TOP 3). On average, the matching probabilities are greater for the probability index, while the wage and stability values are greater for the matching index. These differences are due to the influence of the job match quality variables and show that these variables have a significant influence on job recommendations.
Furthermore, we take a closer look at the TOP 3 job recommendations. We check how many coincide for both indexes without paying attention to the order. In 56 percent of the cases, two recommendations are different, while one recommendation is different in 31 percent of the cases. This shows that while matching probability plays an important role, there is a crucial part of the information provided by the quality variables.

Naturally, one cannot define an optimal index construction. Weights cannot be generalized for each person due to differences in preferences. Nevertheless, there could be improvements made by applying different weights of job match quality variables and matching probabilities. Thus, we calculate the matching index for different combinations of

Table 7: Mean values for probability index (prob) and matching index (qual)

	probability [in %]		stability [in %]		wage [in Euro]	
	prob	qual	prob	qual	prob	qual
TOP 1	54.68	50.83	53.39	60.48	99.74	100.38
TOP 2	10.73	8.76	11.69	20.63	98.55	107.90
TOP 3	5.64	3.24	9.80	21.78	98.10	108.59

Source: own calculations ©IAB

weights and check the exchange ratio for the TOP 3, which is defined as follows:

$$\frac{PROB_{\text{matching index}} - PROB_{\text{probability index, } w}}{QUAL_{\text{matching index}} - QUAL_{\text{probability index, } w}}, \tag{10}$$

where w denotes the weight combination. Therefore, we calculate the mean wage, duration probability and matching probability separately for the TOP 3 job recommendations. For further analysis it is important to have the same scale for each variable. Therefore, we scale all variables and calculate the mean of the two quality variables. Thus, we have a mean probability (PROB) and a mean quality (QUAL) for each index. Now, the exchange ratio is calculated analogously to Equation 10. This calculation draws a similar picture for each of the TOP 3 recommendations. Table 8 shows that in any case, the exchange ratios considerably worsen beyond the quality weight of 50 percent also applied in Equation 8. This shows that for lower weights on matching probability the loss of probability becomes increasingly noticeable relative to the gain of quality information.

Table 8: Exchange ratios for different weights of quality variables and matching probabilities

weight	0.75 PROB	0.5 PROB	0.333 PROB	0.25 PROB	0.167 PROB
exchange ratio TOP1	-0.11	-0.57	-1.16	-1.30	-1.37
exchange ratio TOP2	-0.15	-0.33	-0.38	-0.72	-0.90
exchange ratio TOP3	-0.12	-0.24	-0.38	-0.42	-0.54

Source: own calculations ©IAB

5 Conclusion

In the present paper, we evaluate whether machine learning (ML) methods can play an important role in predicting job match quality. We measure job match quality by both job stability and wages. The underlying data are drawn from a large administrative sample containing employment biographies. Thus, we can observe hard facts about the characteristics of each person and the corresponding job. These data allow us to build statistical models that can be used to effectively map the current situation in the labor market.

The prediction of wages is performed by both OLS and extreme gradient boosting (XGBoost). We find that the MSE for XGBoost is 11.1 percent greater than that for OLS. Thus, XGBoost is clearly preferred over OLS. The better the ML results are, the larger the amount of data in the training sample is. In contrast, the OLS estimation almost does not react. Thus, the ML algorithm is able to find patterns in the data that cannot be obtained by OLS.

The prediction of job stability (i.e., job duration) is performed by logit and XGBoost. We distinguish between two dependent variables. On the one hand, we take the duration of being employed at all (i.e., employment duration), on the other hand, we take the duration of being employed in the same occupation (i.e., occupation duration). Using XGBoost produces a classification error that is 27.5 percent greater for the occupation duration and 58.0 percent greater for the employment duration. Thus, in this case, we clearly prefer the XGBoost approach.

In practice, having information on job match quality could be crucial for caseworkers. Furthermore, for each person, we predict the probability of being employed in a certain occupation, see Mühlbauer/Weber (2022). Combining the individual matching probability and job match quality aspects for creating a single list containing all the information is more practical. Therefore, we create a matching index containing matching probabilities, wages and job duration. By adding a weight of 50 percent to each index, matching probability and job match quality are found to comprise the best combination for constructing the index. Within the job match quality variables, we also assign equal weights for job stability and wages. The matching index could support the placement process in practice by providing information in addition to the caseworkers' impressions. Examining the variables more closely, we find that there is a significant difference between the matching index recommendations and those that consider only matching probabilities.

Certainly, there is a range of variables with the potential to cause discrimination. Mühlbauer/Weber (2022) showed a way to address this problem. Due to correlations, discrimination cannot be estimated by excluding critical variables. Thus, a solution could be to measure the influence of critical variables, finally, users (for example, job seekers or caseworkers) could decide which information they would like to receive or not.

Our results point to the potential of ML techniques in other matching applications. In future research, we plan to improve our results by adding information on competencies. Thus, the advantages of ML algorithms may increase even further due to potentially more complex patterns. We plan to add professional skills, as well as soft skills. Furthermore, the model could be evaluated in a field experiment. For this purpose, one would use the newest data to estimate the model. These additional job suggestions should support the common procedure of the placement process in employment agencies. To measure the effect of having this additional information based on statistical methods, we could randomly assign the treatment and control groups. Target variables could include both job findings and quality measures of the job. Of course, before we start such a field experiment, the model should be carefully checked for discrimination.

References

Allen, Jim; Van der Velden, Rolf (2001): Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. In: Oxford economic papers, Vol. 53, No. 3, p. 434–452.

Amato, Flora; Boselli, Roberto; Cesarini, Mirko; Mercorio, Fabio; Mezzanzanica, Mario; Moscato, Vincenzo; Persia, Fabio; Picariello, Antonio (2015): Challenge: Processing web texts for classifying job offers. In: Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), p. 460–463.

Antoni, Manfred; Ganzer, Andreas; vom Berge, Philipp (2016): Sample of Integrated Labour Market Biographies (SIAB) 1975-2014. In: Documentation on Labour Market Data 201604.

Boselli, Roberto; Cesarini, Mirko; Mercorio, Fabio; Mezzanzanica, Mario (2018): Classifying online job advertisements through machine learning. In: Future Generation Computer Systems, Vol. 86, p. 319–328.

Caliendo, Marco; Tatsiramos, Konstantinos; Uhlendorff, Arne (2013): Benefit duration, unemployment duration and job match quality: a regression-discontinuity approach. In: Journal of applied econometrics, Vol. 28, No. 4, p. 604–627.

Cappellari, Lorenzo; Tatsiramos, Konstantinos (2015): With a little help from my friends? Quality of social networks, job finding and job match quality. In: European Economic Review, Vol. 78, p. 55–75.

Chen, Tianqi; Guestrin, Carlos (2016): Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, p. 785–794.

Freeman, R.B. (2002): The labour market in the new information economy. In: Oxford Review of Economic Policy, Vol. 18, p. 288–305.

Gartner, Hermann (2005): The imputation of wages above the contribution limit with the German IAB employment sample. In: FDZ-Methodenreport, , No. 02.

Gartner, Hermann; Rothe, Thomas; Weber, Enzo (2022): The quality-weighted matching function: Did the German labour market reforms trade off efficiency against job quality? In: Journal of Money, Credit and Banking.

Giannelli, Gianna Claudia; Jaenichen, Ursula; Rothe, Thomas (2013): Doing Well in Reforming the Labour Market? Recent Trends in Job Stability and Wages in Germany. Tech. Rep., IZA Discussion Papers.

Green, Francis; Zhu, Yu (2010): Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. In: Oxford economic papers, Vol. 62, No. 4, p. 740–763.

Gulyas, Andreas; Pytka, Krzysztof; et al. (2019): Understanding the sources of earnings losses after job displacement: A machine-learning approach. Tech. Rep., University of Bonn and University of Mannheim, Germany.

Heisz, Andrew (1999): Changes in job duration in Canada. In: Relations industrielles, Vol. 54, No. 2, p. 365–387.

Holzer, Harry J. (1988): Search Method Use by Unemployed Youth. In: Journal of Labour Economics, Vol. 6, p. 1–20.

Kianpisheh, Somayeh; Jalili, Saeed; Charkari, Nasrolah Moghadam (2012): Predicting job wait time in grid environment by applying machine learning methods on historical information. In: International Journal of Grid and Distributed Computing, Vol. 5, No. 3, p. 11–22.

Mang, Constantin (2012): Online Job Search and Matching Quality. In: ifo Working Paper, , No. 147.

Mavromaras, Kostas; McGuinness, Seamus; O'Leary, Nigel; Sloane, Peter; Wei, Zhang (2013): Job mismatches and labour market outcomes: Panel evidence on university graduates. In: Economic Record, Vol. 89, No. 286, p. 382–395.

Mühlbauer, Sabrina; Weber, Enzo (2022): Machine learning for labour market matching. Tech. Rep. 3, IAB-Discussion Paper.

Nekoei, Arash; Weber, Andrea (2017): Does extending unemployment benefits improve job quality? In: American Economic Review, Vol. 107, No. 2, p. 527–61.

Olsen, Karen M; Kalleberg, Arne L; Nesheim, Torstein (2010): Perceived job quality in the United States, Great Britain, Norway and West Germany, 1989-2005. In: European Journal of Industrial Relations, Vol. 16, No. 3, p. 221–240.

Pecoraro, Marco (2014): Is There Still a Wage Penalty for Being Overeducated But Well-matched in Skills? A Panel Data Analysis of a S wiss Graduate Cohort. In: Labour, Vol. 28, No. 3, p. 309–337.

Simon, Curtis J.; Warner, John T. (1992): Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure. In: Journal of Labor Economics, Vol. 10, p. 306–330.

van den Berg, Gerard J; Vikström, Johan (2014): Long-run effects of active labor market policy programs. In: the Fifth Workshop of the Centre for Research in Active Labor Market Policy Effects (CAFE), Citeseer.

Van Ours, Jan C; Vodopivec, Milan (2006): How shortening the potential duration of unemployment benefits affects the duration of unemployment: Evidence from a natural experiment. In: Journal of Labor economics, Vol. 24, No. 2, p. 351–378.

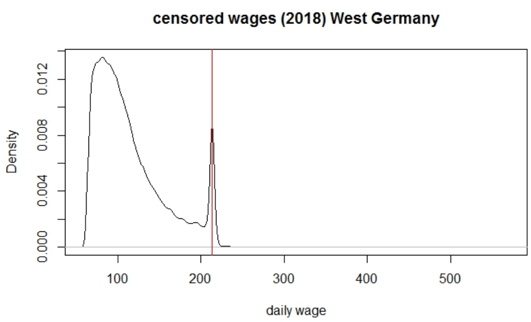
Appendix

Imputation of Wages above the Contribution Limit

Wages are right censored due to the contribution limit in social security. For reasons of data protection, it is only necessary to capture wages until this threshold is reached. Consequently, all values that lie above the contribution limit are automatically set to this value. Additionally, there are some outliers in the data with extremely high daily wages. This can occur, for example, if the wage for a long-term project of artists is reported only for one day. In contrast, there are also some implausibly low wages due to errors in the calculation of daily wages. To eliminate these values, we define a lower limit for daily wages. From 2012 to 2014, the limit is two times the so-called marginal earnings threshold (Geringfügigkeitsgrenze). Starting in 2015, the statutory minimum wage can serve as a threshold.

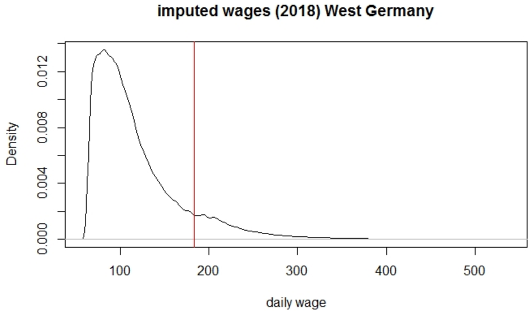
In the next step, we have to estimate the wages that lie above the contribution limit. Since we have right-censored wages, we use a tobit model. To cover small inaccuracies in the calculation of daily wages, we use 95 percent of the contribution limit as a threshold for the tobit estimation. Furthermore, we estimate separately for year, gender and educational group. Since the contribution limit differs for East and West Germany, the estimations are conducted separately for each region. Therefore, the daily wage is predicted for 112 subsamples. To calculate the expected value, we apply the imputation of wages described in Gartner (2005). After this procedure, we are able to replace daily wages that lie above the contribution limit with imputed values. Figure A2 and figure A4 show the density function of daily wages for 2018 before and after the imputation for West and East Germany. The graphs for the remaining years look similar.

Figure A1: Density function of non-imputed daily wages for West Germany



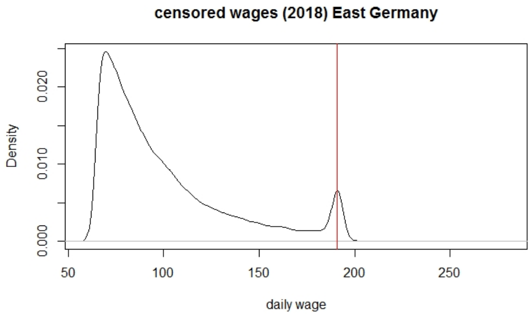
Source: own calculations ©IAB

Figure A2: Density function of imputed daily wages for West Germany



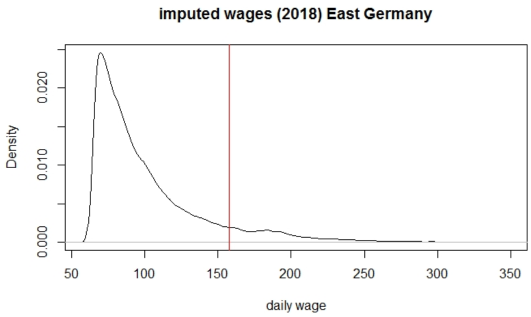
Source: own calculations ©IAB

Figure A3: Density function of non-imputed daily wages for East Germany



Source: own calculations ©IAB

Figure A4: Density function of imputed daily wages for East Germany



Source: own calculations ©IAB

A brief Overview of Extreme Gradient Boosting

Extreme gradient boosting (XGBoost) is a parallel tree boosting algorithm that is used to solve many different research questions in an efficient way. Chen/Guestrin (2016) provided an introduction to and the mathematical background of this approach. In brief, the algorithm can be used for solving different kinds of research questions, such as classification, regression or ranking. Another advantage is the handling of missing values because they are handled automatically, thus, preprocessing is redundant.
XGBoost is based on gradient boosted decision trees. The decision trees are created in

sequential form. Since single decision trees have a high variance, multiple decision trees hare combined. They are built based on weighted independent variables. Boosting means that after having built a weak model from the training data, a second model is built that attempts to correct the errors present in the first model. This procedure is repeated until a certain maximum number of models or complete training data are correctly predicted.
In practice, XGBoost works well for large datasets and is highly customizable due to the wide range of hyperparameters. The output also contains feature importance; thus, one can better determine which variables are most important in a certain model. Naturally, the algorithm has several limitations. As described in Section 3.1, finding the optimal set of hyperparameters is very time consuming because the estimation of large datasets is computationally intensive. In this context, sufficient memory resources for estimating large models are crucial.

Job Stability: Choose a Different Threshold

Additionally, we check if taking another threshold would change the results dramatically. We take one year instead of half a year. Table A1 shows the results which are similar to the results with one year as a threshold.

Table A1: Error rates of daily wage

	XGBoost	Logit
employment duration	18.72 %	21.83 %
occupation duration	18.38 %	23.97 %

The models are estimated with R (required packages: xgboost, glm). For logit, we calculate the optimal threshold for classifying the results. For the employment duration sample, the value is 0.47. For the occupation duration sample, the value is 0.26.
Source: own calculations ©IAB

List of Figures

List of Tables

Imprint

IAB-Discussion Paper 09|2024

Publication Date
12. July 2024

Publisher
Institute for Employment Research
of the Federal Employment Agency
Regensburger Straße 104
90478 Nürnberg
Germany

Download
https://doku.iab.de/discussionpapers/2024/dp0924.pdf

All publications in the series “IAB-Discusssion Paper” can be downloaded from
https://iab.de/en/publications/iab-publications/iab-discussion-paper-en/

Website
https://www.iab.de/en

ISSN
2195-2663

DOI
10.48720/IAB.DP.2409

Corresponding author
Sabrina Mühlbauer
Telefon 0911 1799743
E-Mail sabrina.muehlbauer@iab.de
Enzo Weber
Telefon 0911 1797643
E-Mail enzo.weber@iab.de