

Jan Büchel, Jan Engler and Armin Mertens

The Demand for Data Skills in German Companies: Evidence from Online Job Advertisements

KEY MESSAGES

- **Data skills are required in every fifth online job advertisement in Germany**
- **Demand for data skills is increasing over time and heterogeneous between sectors**
- **Companies from the ICT sector demand more advanced data skills in every second job advertisement**
- **Sectors that are advanced in digitalization have a higher demand for jobs with data skills**
- **Policymakers should intensify the recruitment of foreign professionals and align curricula with growing business demand for data skills**

Data can be a strategic resource for companies, which they can benefit from in many ways. Based on data, products or services can be developed, processes can be improved, and supply chains can become more transparent if data is shared with other companies (Andres and Niebel 2022; Brynjolfsson and McElheran 2016; Kache and Seuring 2017). In addition, companies can develop data-driven business models (Gierten et al. 2021). For example, usage data of products that are already sold, such as machines, can be analyzed and used to predictively maintain machines in the future and thus reduce production downtimes. To fully exploit the potential of data, companies first and foremost have to be prepared for the data economy. For this purpose, firms have to store, manage, and use data in an efficient way. For example, different types

of data – like production, financial, or personal data – should be stored digitally. Standardized quality checks and interfaces can optimize the data flow from a management perspective. Lastly, the real added value of data lies in diverse usage purposes. For example, data can be used for analysis, visualization, documentation, forecast of processes, or for sale to other firms.

Given the importance of data as an invaluable asset in modern economies, scientific research in economics and management has increasingly focused on the dynamics and repercussions of the data economy (Börner et al. 2018; Cong et al. 2021; Farboodi and Veldkamp 2021). One of the key drivers for the data economy are data-related skills, since they enable individuals and organizations to effectively collect, analyze, and utilize data to make informed decisions and create new products and services (Pappas et al. 2018). However, there is still a lack of a deeper understanding of the demand in skills and jobs that are necessary to process and strategically use the large amounts of data companies accumulate. Considering the fast pace of the advancements of the data economy and the jobs that come with it, existing research is rather sluggish in analyzing current demand in the labor market (Börner et al. 2018; European Commission 2020). In this regard, job advertisements can provide timely information on the need for specific knowledge of companies (Büchel and Mertens 2021; Pejic-Bach et al. 2020). To provide tangible and up-to-date insights into the demand for jobs related to the data economy, we leverage data from online job advertisements and use a state-of-the-art machine learning approach to accurately estimate the demand for data skills in the German labor market.



Jan Büchel

is Economist for Data Economy in the research unit Digitalisation and Climate Action at the German Economic Institute.



Jan Engler

is Data Scientist in the Big Data Analytics Cluster at the German Economic Institute.



Armin Mertens

is Data Scientist in the Big Data Analytics Cluster at the German Economic Institute.

DATA ECONOMY READINESS OF GERMAN COMPANIES

Companies can demand data skills to be able to participate in the data economy or, if they are already data economy ready, to further improve their handling of data. In 2022, only 31 percent of companies in Germany were data economy ready, as a survey of 1,051 companies from the industrial and industry-related services sectors revealed (Büchel and Engels 2022c). This share is slightly higher than the previous year (29 percent), as a comparable survey found (Büchel and Engels 2022a). Therefore, over two-thirds of the surveyed companies do not meet the requirements to participate in the data economy in an efficient way. In concrete terms, this means that on average they fulfill less than 50 percent of the relevant aspects with respect to data storage, data management, and the use of data. One reason for this could be that relevant decision-makers still have to be convinced of the potential of data or corresponding data skills are generally lacking in the firm. If employees do not have the necessary digital skills and do not know how to use corresponding applications, companies may be limited in their ability to make progress towards participating in the data economy.

To address this, there are several ways for companies to build up data skills. On the one hand, firms can offer on-the-job training to improve the data skills of their employees. Hence, employees who have received relevant on-the-job training can be enabled to conduct and support the data strategy pursued by the firm. On the other hand, firms can hire new employees who already have necessary data skills. While on-the-job training activities are internal and thus not publicly known, job advertisements are published and can therefore be analyzed.

To estimate the demand for data skills in the German labor market, we utilize large amounts of data on job advertisements. These types of data are becoming increasingly popular in economic research on the labor market (Acemoglu et al. 2022; Azar et al. 2018; Deming und Noray 2020; Hershbein and Kahn 2018). Typically, job advertisements consist of a first part where the employer presents the hiring company, the working conditions, or special offers. This part is followed by a description of the underlying job profile and the required skills of the potential employee. These are most important for the following analysis. Indeed, this part of the job advertisement can be used to identify whether the advertisement requires data skills or not. For this reason, job advertisements can be an efficient instrument to analyze the extent to which companies in Germany require new employees to have relevant skills that help them to improve their storage, management, and use of data – the three components of data economy readiness. Thus, this can function as an indicator of the demand for data skills in Germany.

The following analysis is based on a data set provided by Textkernel that contains online job advertisements in Germany from over 60,000 different sources, including the major online job portals such as Indeed, Stepstone, or the German Federal Employment Agency. The data set encompasses job advertisements from the first quarters Q1 of the years 2019, 2020, 2021, and 2022. In total, 3.2 million unique job advertisements were posted in the first quarter of 2019, 3.0 million in 2020, 3.6 million in 2021, and 4.5 million in 2022. The aim of the following analysis is to classify each individual job advertisement depending on whether data skills are required for the job or not. This is possible by first identifying required data skills in the job advertisements. For this purpose, the basic concept and theoretical background focus on the three components of data economy readiness: data storage, data management, and usage of data (Büchel and Engels 2022a). The relevant aspects of these three components are compared to the job profiles and required skills in the posted job advertisements. If the new employee is intended to support the demanding firm in one of these aspects, the job advertisement is hence classified as requiring data skills. The following explains how this is implemented methodologically at the level of individual job advertisements.

MACHINE LEARNING APPROACH FOR IDENTIFYING JOB ADVERTISEMENTS THAT REQUIRE DATA SKILLS

It would be very time-consuming and costly to analyze each job advertisement manually. Instead, a machine learning (ML) approach is used to automatically classify the job advertisements as to whether they require data skills or not. For the model to make the correct decision, it first has to be trained to a sufficient extent with domain-specific information.

To train such an ML model, a set of job advertisements was annotated manually using the open-source text annotation tool doccano (Nakayama et al. 2018). The basic idea is that the model learns the classification decision from this template to independently make further predictions on new data. Therefore, the three authors annotated a set of job descriptions and determined whether these job advertisements require data skills or not based on the relevant aspects of data economy readiness.

Such a task can be rather subjective. To increase objectiveness, the authors have created annotation guidelines that provide detailed instructions for the widest possible range of different skills and phrases related to the data economy to correctly classify individual job postings. Initially, the authors used the terms listed in Büchel and Engels (2022a) as initial annotation guidelines and annotated a separate data set containing 100 job advertisements. During this process, the annotators discussed their disagreements with the aim of refining and updating the guidelines.

The authors decided that, for example, if the description of the job profile contains at least one keyword like “database maintenance,” “programming,” or the required skillset contains knowledge of cloud technologies, the job advertisement was classified as related to the data economy. If, on the other hand, the potential employee should only have Microsoft Office skills or it was unclear whether the activities were to be performed digitally or in an analog way (i.e., like storing customer data), this was not a decisive factor for classifying a job advertisement as data economy related.

Based on the annotation guidelines, an additional 1,000 job advertisements were annotated as training data for the ML model. Overall, a very high inter-annotator-agreement (~94 percent across all annotators) was reached. This is a high-quality signal of the guidelines and the corresponding training data. Remaining disagreements were solved by majority vote by the three authors.

Modern ML model have hundreds of million trainable parameters (Devlin et al. 2018). Training those models from scratch, using only these mere 1,000 annotated advertisements, would be impossible. Consequently, an already pre-trained model was used as a basis for the analysis. The manually annotated data was then applied to fine-tune the model to the specific task of identifying job advertisements including data skills. The model used here was the so-called BERT model by Gnehm et al. (2022). Their model was pre-trained on German job advertisements as well as general German text, for example using German Wikipedia. Fine-tuning such a model means that a new neuronal network classification layer was added to the BERT model and was then trained jointly with the other layers (Devlin et al. 2018).

After fine-tuning, the BERT model was applied to all 14.3 million job advertisements from the first quarters of the years 2019 to 2022, predicting the correct classification decisions with a very high accuracy. The prediction-making of the ML model was evaluated extensively. Initially, a separate dictionary-based model was created that classifies job advertisements by

searching for carefully selected domain-specific buzz-phrases, as “cloud” or “analyze data”. This second model was used as a baseline for the present task. To measure the prediction accuracy, an additional, so-called “test set” of 200 manually annotated job advertisements was created. Both models were then used to predict whether data skills are required or not in these job advertisements. The predictions were compared to the manual annotations. The ML model performed very well on the test set and achieved a higher accuracy compared to the dictionary-based model (94 percent vs. 92.5 percent), f1-Score (0.914 vs. 0.892), precision (0.928 vs. 0.909), and recall (0.902 vs. 0.877).

Considering the better performance of the ML model compared to the dictionary-based model, it was used in the subsequent analysis. The results of the application of the model to the entire data set are presented below.

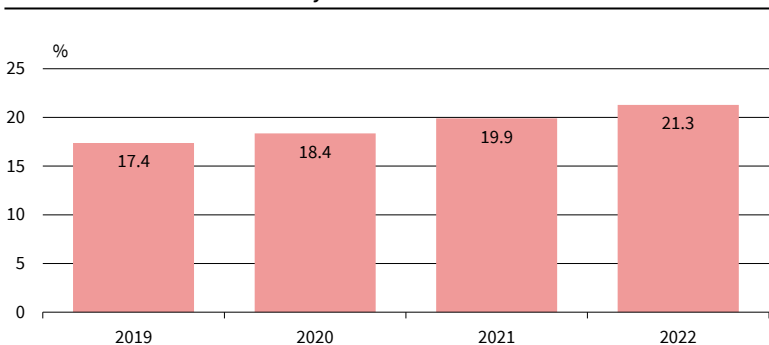
DEMAND FOR DATA SKILLS IN GERMAN COMPANIES

After applying the ML model, 21.3 percent of the 4.5 million online job advertisements in the first quarter of 2022 in Germany require data skills (Figure 1). The share has been rising steadily since 2019: In the first quarter of 2019, the share was at 17.4 percent, rising to 18.4 percent in the first quarter of 2020 and 19.9 percent in the first quarter of 2021, respectively.

The share of 21.3 percent in the first quarter of 2022 is plausible because just over 30 percent of companies in Germany currently are data economy ready (Büchel and Engels 2022c). Both aspects are closely linked since most companies in Germany are small and therefore unlikely to advertise a disproportionately high number of job advertisements on average. On the one hand, it is exactly the companies that use data efficiently that are looking for suitable employees, for example to further develop existing data-driven business models or processes. On the other hand, data skills are needed first and foremost in companies to become increasingly data economy ready. In comparison to the status quo of the data economy readiness, however, the analysis of job advertisements is slightly more dynamic and delayed because it shows which human capital investments companies are currently making to be able to manage data more efficiently in the future.

Figure 1 illustrates that data is taking on an increasingly important role for companies in Germany. However, it should be noted that we measure only the number of new job advertisements posted by companies. The analysis does not provide any information on the level of fluctuation. If fluctuation is high, it is more likely that some of the advertised jobs cannot be filled due to the shortage of skilled workers in digitalization professions in Germany, which has become even more acute in recent years (Burstedde 2021).

Figure 1
Demand for Data Skills in Germany



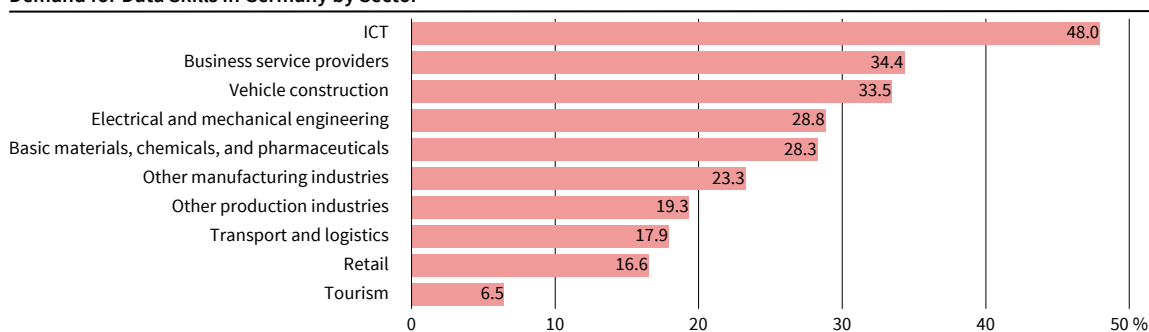
Note: Share of online job advertisements in Germany that require data skills; first quarter of each year; in % of all online job advertisements in Germany in the respective quarter.

Source: German Economic Institute based on Textkernel data.

© ifo Institute

Figure 2

Demand for Data Skills in Germany by Sector



Note: Share of online job advertisements in Germany that require data skills; first quarter of 2022; in % of all job advertisements in Germany; 389,748 job advertisements with data skills in total.

Source: German Economic Institute based on Textkernel data.

© ifo Institute

To gain a better understanding of the drivers of data skill demand in Germany, the share of such job advertisements in individual sectors is analyzed. For this purpose, we focus on ten sectors¹ from the areas of industry and industry-related service providers. In comparable studies, which have representatively surveyed the status quo of data economy readiness or digitalization, the same industry focus was applied. These studies can therefore be compared with the present analysis. Furthermore, the focus allows for more targeted statements on the sectors analyzed.

Information about the sector of each tendering company is included in the Textkernel data that contains an assignment of each job posting to the German Classification of Economic Activities (WZ 2008)². Based on these classifications, the companies can be assigned to the ten sectors and the job advertising activities of the respective sectors in the first quarter of 2022 can be analyzed (Figure 2). Generally, however, a large share of job advertisements is advertised through intermediaries. As a result, the searching company behind it and its sector cannot be identified. 38 percent of all job advertisements with data skills are advertised via intermediaries and thus cannot be assigned to any sector. The dataset also contains a few job advertisements that were not advertised through intermediaries but could not be assigned to any industry due to incomplete information, for example. Finally, there are job advertisements with data skills that do not correspond to the focus of the ten sector groups. These job advertisements are subsequently excluded from the analysis. In total, slightly less than 400,000 job advertisements with data skills can be assigned to the ten sector groups.

The ICT sector exhibits the highest demand: almost every second job advertisement in this sector requires data skills. This result is plausible, as companies in the sector offer, for example, data processing as a service to other companies or manufacture data processing equipment. Data is often a core component

of their business model. Other studies also illustrate that the German ICT sector is the leader in terms of both the level of digitalization and its data economy readiness among the sectors analyzed here (Büchel and Engels 2023 and 2022b).

Besides the ICT sector, the sectors business service providers and vehicle construction also exhibit a high demand for prospective employees with data skills: about one-third of job advertisements require data skills. This is particularly astonishing in the case of business service providers: in absolute terms, there are about 90,000 job advertisements with data skills, roughly the same number as in the ICT sector. However, since the total number of job advertisements is higher in the business service providers sector, the share of job advertisements with data skills is correspondingly lower. The demand in the two sectors of business service providers and vehicle construction also coincides with good results for the level of digitalization (Büchel and Engels 2023): employees with data skills are needed to initiate and implement digitalization projects in the company. However, these companies seem to aim to improve digitalization and data management even further. After all, they are still posting job advertisements for data skills on a large scale.

The sector vehicle construction shows a development of its demand for data skills that differs from most other sectors: the latter exhibit a continuous increase of the relative share of job advertisements with data skills since 2019. In vehicle construction, however, the demand reached its peak of 37.3 percent in the first quarter of 2019, coming in second after the ICT sector (43.9 percent). In 2022, vehicle construction's demand for data skills had declined to 33.5 percent, ranking third. A comparable study on job advertisements requiring AI skills in Germany reached a similar result for vehicle construction (Büchel and Mertens 2022): since 2019, there was a similar decline in AI job advertisements in the regional clusters around Stuttgart and Wolfsburg, where large companies from the automotive and supplier sectors are located who advertise the majority of AI job advertisements in these regions. Since AI job advertisements make up a subset of job

¹ For the exact composition of the ten sectors, see Büchel and Engels (2023).

² The data provider Textkernel matches each company listed in each respective job advertisements to national company directories to gain information about the respective sector.

POLICY CONCLUSIONS AND FURTHER RESEARCH

The analysis shows that data skills are required in many job advertisements in Germany and are becoming more relevant. This is a positive sign for the data economy in Germany, as companies increasingly realize the potential of data and try to implement it in their own operations. However, a growing demand for employees with data skills poses challenges for companies in the future in view of the existing skills gap, especially in digitalization professions (Burstedde 2021). Policymakers should intervene by, for example, taking measures that make it easier for companies to recruit foreign workers. They could, for example, create incentives for immigration via higher education, as there is still a lot of potential in this field: currently, comparatively more people only immigrate after completing their university degree (Geis-Thöne 2022). Policymakers should start with the financial resources required for the granting of residence permits for the educational immigration of third-country nationals and set up a corresponding scholarship or guarantee program.

In addition, further research could focus on the specific tasks that the job advertisements with data skills are primarily advertised for. For example, new employees could primarily help the company to improve their data storage. Alternatively, the focus could be on data management or data use. However, the categories can also overlap, such that new employees could also help the company in several categories. A corresponding analysis can reveal in which data skill category companies have particularly high demand. This in turn can provide valuable implications for the vocational training, school, and higher education sector in Germany to address the growing skills gap. If, for example, skills are required on a large scale to technically create data interfaces in companies or to develop pragmatic data-based business models, this should be found accordingly in relevant curricula to be able to meet this demand in the future. This concerns, for example, the expansion of computer science as a school subject as well as the supply of appropriate teaching staff at schools in the STEM area and their continuous further training (Anger et al. 2022).

Beyond this, it could be analyzed in more detail whether there is indeed a sustainably growing demand for data skills in companies or whether a high fluctuation hinders sustainable growth. The challenge would be to determine the degree of fluctuation in companies and to separate the effect of fluctuation from the general growth of job advertisements. Nevertheless, a better understanding of the fluctuation could help to identify appropriate measures that companies or policymaker could implement to counteract the skills gap.

REFERENCES

- Anger, A. et al. (2022), *MINT-Herbstreport 2022: MINT sichert Zukunft*, Study for BDA, Gesamtmetall and "MINT Zukunft schaffen".

- Acemoglu, D., D. Autor, J. Hazell and P. Restrepo (2022), "Artificial Intelligence and Jobs: Evidence from Online Vacancies", *Journal of Labor Economics* 40, 293-340.
- Andres, R. and T. Niebel (2022), Big Data Analysen und Produktivität. Ein Kurzbericht zur Nutzung von Big Data Analysen in Unternehmen auf Basis administrativer Daten aus den Niederlanden, ZEW, Mannheim.
- Azar, J., I. Marinescu, M. Steinbaum and B. Taska (2020), "Concentration in US Labor Markets: Evidence from Online Vacancy Data", *Labour Economics* 66, 101886.
- Brynjolfsson, E. and K. McElheran (2016), "Data in Action: Data-driven Decision Making in US Manufacturing", *US Census Bureau Center for Economic Studies Paper*, CES-WP-16-06.
- Börner, K., O. Scrivner, M. Gallant, S. Ma, X. Liu, K. Chewing, L. Wu and J. A. Evans (2018), "Skill Discrepancies between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy", *Proceedings of the National Academy of Sciences* 115, 12630-12637.
- Büchel, J. and B. Engels (2022a), "Datenbewirtschaftung von Unternehmen in Deutschland", *IW-Trends* 49(1), 73-90.
- Büchel, J. and B. Engels (2022b), "Branchentrends beim Data Sharing. Status Quo und Use Cases in Deutschland", *IW-Report* 53.
- Büchel, J. and B. Engels (2022c), "Viele Unternehmen sind nicht bereit für die Datenwirtschaft", *IW-Kurzbericht* 96.
- Büchel, J. and B. Engels (2023), *Digitalisierungsindex 2022. Digitalisierung der Wirtschaft in Deutschland*, Expertise for the German Federal Ministry for Economic Affairs and Climate Action, Forthcoming.
- Büchel, J. and A. Mertens (2021), *KI-Bedarfe der Wirtschaft am Standort Deutschland. Eine Analyse von Stellenanzeigen für KI-Berufe*, Study as Part of the Project "Development and Measurement of the Digitalisation of the Economy in Germany" for the German Federal Ministry for Economic Affairs and Energy.
- Büchel, J. and A. Mertens (2022), *KI-Bedarfe in Deutschland. Regionale Analyse und Entwicklung der Anforderungsprofile in KI-Stellenanzeigen*, Expertise for the German Federal Ministry for Economic Affairs and Climate Action.
- Burstedde, A. (2021), *Fachkräftesituation in Digitalisierungsberufen – Beschäftigungsaufbau und Fachkräftemangel*, Study as Part of the Project "Development and Measurement of the Digitalisation of the Economy in Germany" for the German Federal Ministry for Economic Affairs and Climate Action.
- Deming, D. J. and K. Noray (2020), "Earnings Dynamics, Changing Job Skills, and STEM Careers", *Quarterly Journal of Economics* 134, 1965-2005.
- Devlin, J. et al. (2018), "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding", *arXiv*, arXiv:1810.04805.
- European Commission (2020), *The European Data Market Monitoring Tool. Key Facts, First Policy Conclusions, Data Landscape and Quantified Stories*, D2.9 Final Study Report, Brussels.
- Farboodi, M. and L. Veldkamp (2021), "A Model of the Data Economy", *NBER Working Paper* 28427.
- Geis-Thöne, W. (2022), "Fachkräftesicherung durch Zuwanderung über die Hochschule. Aktueller Stand und Handlungsansätze für die Politik", *IW-Trends* 49(3), 67-88.
- Gierten, D., S. Viète, R. Andres and T. Niebel (2021), "Firms Going Digital: Tapping into the Potential of Data for Innovation", *OECD Digital Economy Papers* 320, Paris.
- Gnehm, A. S., E. Bühlmann and S. Clematide (2022), "Evaluation of Transfer Learning and Domain Adaptation for Analyzing German-speaking Job Advertisements", *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*, 3892-3901.
- Hershbein, B. and L. B. Kahn (2018), "Do Recessions Accelerate Routine-biased Technological Change? Evidence from Vacancy Postings", *American Economic Review* 108, 1737-1772.
- Kache, F. and S. Seuring (2017), "Challenges and Opportunities of Digital Information at the Intersection of Big Data Analytics and Supply Chain Management", *International Journal of Operations & Production Management* 37, 10-36.
- Nakayama, H., T. Kubo, J. Kamra and Y. Taniguchi (2018), *doccano: Text Annotation Tool for Human*, <https://github.com/doccano/doccano>.
- Pappas, I. O., P. Mikalef, M. N. Giannakos, J. Krogstie and G. Lekakos (2018), "Big Data and Business Analytics Ecosystems: Paving the Way towards Digital Transformation and Sustainable Societies", *Information Systems and e-Business Management* 16, 479-491.
- Pejic-Bach, M., T. Bertoncel, M. Meško and Z. Krstić (2020), "Text Mining of Industry 4.0 Job Advertisements", *International Journal of Information Management* 50, 416-431.