Proceedings of the 20th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 681–686 DOI: 10.15439/2025F4081 ISSN 2300-5963 ACSIS, Vol. 43

Towards the automated classification of German job titles according to KldB

Ralf Dorau*, Kristine Hein*
* Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany
Email: dorau@bibb.de, kristine.hein@bibb.de

Abstract—**The automated classification of occupational titles is a pivotal component of labor market analysis, survey research, and administrative data processing. This paper explores the viability of mapping German job titles to the German Classification of Occupations (KldB) by employing conventional machine learning methodologies to examine the challenges and limitations inherent in the data itself. To this end, the present study leverages two complementary datasets—manually annotated survey data and a dataset of occupational synonyms—to assess the performance of established classifiers under varying levels of taxonomic granularity. The methodological challenges inherent to this study include class imbalance, semantic ambiguity, and linguistic variability, which are all characteristics of German job title expressions. The findings of the study suggest that while coarse-level classifications can be addressed with relatively simple models and text representations, finer-grained distinctions remain challenging to resolve using title-based features alone. The findings indicate that more expressive models and richer contextual information may be necessary for high-resolution occupational coding.**

I. Introduction

THE automated classification of job titles is an important topic in research, but also for practitioners in the field of labor market research. For example, in questionars people may answer the question about their occupation. However, this needs to be matched to a certain occupation, for example the German Classification of Occupations (Kldb) or the International Standard Classification of Occupations (ISCO). Other tasks include the classification of online job advertisements (OJAs) or the matching of other occupations, e.g., from online platforms like Kununu [1]. Other researchers try to tackle the question of identifiying occupations in textual data, for example in parliamentary debates [2]. The German language has several challenges, as occupational titles are not only single nouns but may also be a combination of nouns and may also contain additional data like the examining institution (e.g., 'IHK'). Next, different forms for male, female, gendered or neutral titles may exist. For generic texts, another challenge is that surnames may originate from (partly historical) professions, for example 'Bäcker' (baker).

Linking a variety of different sources on labor market is usually considered a very challenging task [3], however, in this paper, we limit the question to the mapping and automated classification of job titles in German language. Here, we find dictionary-based approaches which are widely used, but also other ML-based approaches. Training data was widely collected from survey data or job-titles from KldB, ESCO or other synonym data.

The main contribution of our paper is a large collection of novel training data and a systematical analysis of challenges to train classificators for German job titles according to KldB and in particular different levels of occupations and their level of performance.

This paper is divided into five sections. The first section provides an introduction, the second section gives the the state of the art, and related work. The third section describes the data and methodological background. The fourth section is dedicated to experimental results and evaluation. Our conclusions are given in the last section.

II. Related Work

A multitude of classification categories are recognized for occupations. The International Standard Classification of Occupations (ISCO) was developed by the International Labor Organization (ILO) and published in 1958, 1968, 1988, and most recently in 2008)[1]. The ISCO 2008 has also been utilized within the European Union (EU), with certain German-speaking countries (Germany, Austria, and Switzerland) developing a customized version of the classification. The International Standard Classification of Occupations (ISCO) is structured at a skill level and linked to the "European Skills, Competences, Qualifications and Occupations" (ESCO) ontology, which adds another hierarchy level to the data. In Germany, the Classification of Occupations (KldB) serves as the reference classification for the Federal Employment Agency (BA) and its research institute (IAB)[2]. In this organization, occupations are structured at a task level. The most recent version is the 2020 revision of the KldB 2010, which has undergone a comprehensive redesign, thereby rendering the previous versions from 1988 and 1992 obsolete. The development of this system was undertaken with the objective of ensuring compatibility with the ISCO-08 standard. The study of job titles and taxonomies has a long history, extending even before the advent of computer technology [4].

A portion of the research has focused on the classification of OJAs according to the O*NET framework [5]. This has included the application of normalization approaches [6] and

[1] See https://www.ilo.org/public/english/bureau/stat/isco/isco08/.
[2] See https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/KldB2010-Fassung2020-Nav.html.

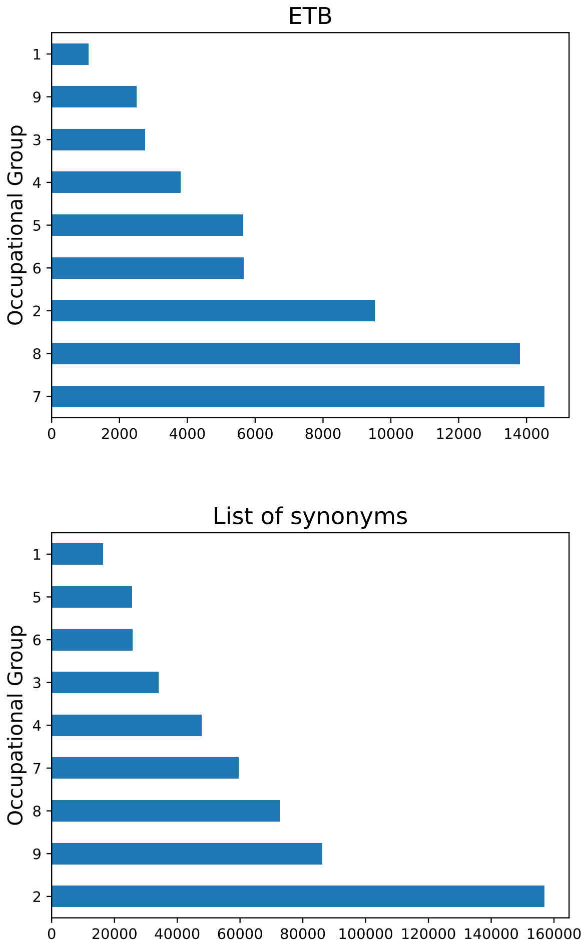

Fig. 1: Distribution of training data according to occupational groups

similarity-based methods [7]. The classification of job titles is also employed in the context of online job recruitment [8].

A limited number of publications have been published on the subject of German job titles, with a particular emphasis on the German KldB. For instance, a technical report based on OJAs [9] with challenges on level 4, but with promising results on level 1. Malte Schierholz's 2018 publication [10] introduced the concept of auxiliary classifications in the field of occupational coding. For further research on the subject of occupational coding in surveys, we refer to [11]. A master's thesis endeavors to predict KldB 5-digit job titles from survey data, thereby highlighting the persistent challenges associated with this endeavor, see [12]. In a similar vein, a scholarly article was published that compared the classification of survey data using BERT and GPT-3, see [13]. However, the absence of a standardized reporting methodology precludes the direct comparison of their results. Nevertheless, they evince analogous challenges to those observed in other studies. Consequently, both the classification of occupational areas and, in particular, the level of performance (5th digit) persist as arduous tasks.

III. Data and Methods

A. Data

The dataset aggregates 60,022 manually annotated occupational titles from the 2012, 2018, and forthcoming 2024 waves of the Erwerbstätigenbefragung (ETB; German Employment Survey)[3], each coded to KldB 2010. These representative surveys – conducted by BIBB and BAuA – target Germany's core workforce (employed individuals aged 15+ working $\geq$10 hours/week). While historical predecessor surveys (1979-1999 BIBB/IAB studies[4]) exist, their integration was precluded by fundamental incompatibilities between KldB versions, despite partial crosswalk feasibility.

[3]Main datasets: 10.7803/501.12.1.1.60 (2012), 10.7803/501.18.1.1.10 (2018). Supplementary variables: 2012: 10.7803/501.12.1.4.10 (full-text), 10.7803/501.12.1.3.20 (regional identifiers), 10.7803/501.12.1.5.30 (special variables); 2018: 10.7803/501.18.1.4.10 (full-text), 10.7803/501.18.1.3.10 (regional identifiers), 10.7803/501.18.1.5.10 (special variables).

[4]GESIS SUF: 10.4232/1.1243, 10.4232/1.12563, 10.4232/1.2565, 10.4232/1.12247

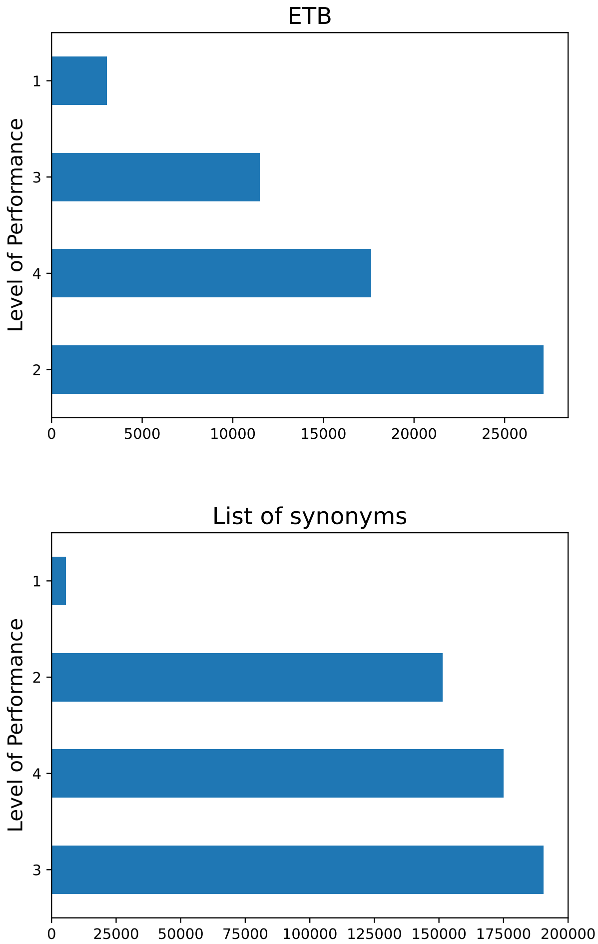

Fig. 2: Distribution of training data according to level of performance (5th-digit)

We use a second dataset comprising 526,535 synonyms and variants of male, female, and neutral job titles, provided by the German Federal Employment Agency (BA)[5]. Both datasets have their own biases, as can be seen in Figures 1 and 2: While in ETB dataset the majority of occupations are located in groups 7 and 8, they are in groups 2 and 9 for the list of synonyms. We find a similar bias with respect to the level of performance, coded in the 5th digit of KldB.

Data cleaning is an existentiell step. We remove occupational are 0 from the dataset as we omit military occupations. Second, we clean all data with errors, for example too large or small KldB numbers. In Dataset 1, 59,535 and in dataset 2, 522,197 entries remain.

B. Method

In this paper, we adopt the classification approach to labor market data presented in [14], [15]. The baseline model was trained using three standard classifiers with default values from the scikit-learn library [16] with logistic regression ($C = 10$), naive Bayes, and random forest. We split the data data so that 20% was used as test data and 80% as training data. As the main question is about specific challenges, we will apply traditional methods to identify the route for further research in this area.

However, another open question is the choice of text representation. As shown by this can significantly change the performance of classifiers. We can use the combination of term frequency and inverse document frequency as TF.IDF (term frequency–inverse document frequency). Another approach is the Word to Vector (W2V) which applies unsupervised learning to represent textual data and produces rather low-dimensional data. As we have seen, the later produces rather poor results and we will only showcase the results for Logistic Regression.

[5]Available at https://www.arbeitsagentur.de/institutionen/dkz-downloadportal.

TABLE I: Results for 1-digit on dataset 1

Approach	Metrics (Macro/Weighted) Precision	Recall	F_1 Score
Logistic Regression (Tf-idf)	0.88 / **0.86**	**0.79 / 0.84**	**0.83 / 0.84**
Logistic Regression (W2V)	0.45 / 0.47	0.26 / 0.46	0.26 / 0.38
Naive Bayes	**0.90** / 0.84	0.68 / 0.79	0.75 / 0.79
Random Forest	0.84 / **0.86**	0.76 / 0.79	0.78 / 0.81

TABLE II: Results for 1-digit on datasets 1+2

Approach	Metrics (Macro/Weighted) Precision	Recall	F_1 Score
Logistic Regression (Tf-idf)	**0.92 / 0.91**	**0.89 / 0.91**	**0.90 / 0.91**
Logistic Regression (W2V)	0.61 / 0.57	0.35 / 0.51	0.38 / 0.47
Naive Bayes	0.91 / 0.89	0.84 / 0.88	0.87 / 0.88
Random Forest	0.91 / **0.91**	0.88 / 0.90	0.89 / 0.90

IV. Experimental Results

A. Occupational areas (1-digit)

Tables I and I present the outcomes for occupational categories with a one-digit classification, contrasting the utilization of Dataset 1 exclusively or Datasets 1 and 2 collectively. The results of this study are largely consistent with those reported in the extant literature. As previously discussed, the W2V model demonstrated substandard performance. A comparative analysis reveals that, while all other models demonstrate comparable performance, Logistic Regression consistently exhibits superior performance in various applications.

The findings of this study demonstrate that extending the dataset has a substantial impact on the quality of the results. For instance, the F_1 score increased from 0.84 to 0.91, a notable enhancement.

B. Occupational group (3-digit)

As illustrated in Tables III and III, the results for occupational group (3-digit) are presented. A direct comparison with occupational areas reveals that the performance is suboptimal. However, the employment of both datasets results in a substantial enhancement of the outcomes. However, a more thorough examination reveals that this approach is particularly ineffective for groups with limited data. A cursory examination of dataset 1 reveals that there is an absence of data points corresponding to area 114 (occupations in fishing), 213 (occupations in industrial glass-making and -processing), and 214 (occupations in industrial ceramic-making and -processing). For future research endeavors, there is a necessity for not only improved training data that is more balanced, but also for the evaluation of cross-validation approaches.

C. Level of Performance (5th digit)

The results of the classification of the level of performance (5th digit), as demonstrated in Tables V and VI, also exhibit a number of noteworthy outcomes. First, it is noteworthy that the extended dataset does not yield substantial improvements in the results. Secondly, we observe that different levels yield different F_1 scores. For Logistic Regression, a score of 0.86 is observed for both level 2 and level 4, while the level 1 (0.73) and level 3 (0.71) demonstrate inferior performance. A similar outcome is observed for all other methods, although level 1 and 3 demonstrate an even poorer performance. Therefore, it appears that the fifth digit is a significant factor in the classification of job titles. The available data from a job title alone is insufficient for effective classification.

V. Discussion and outlook

The classification of German job titles according to the KldB taxonomy remains a complex task, particularly at the more granular levels such as the 3-digit occupational group and the 5th-digit level of performance. The findings of this study underscore the strengths and weaknesses of conventional machine learning methodologies in this context. While the utilization of TF-IDF features in conjunction with Logistic Regression yielded robust outcomes, particularly for the 1-digit classification, performance exhibited a substantial decline for finer-grained levels. This phenomenon underscores the inherent intricacy of occupational data, wherein minor linguistic and contextual variations can wield substantial consequences for classification.

The present study demonstrates the significance of high-quality, balanced training data. The incorporation of the synonym dataset led to a substantial enhancement in performance at more extensive classification levels. However, this augmentation was accompanied by the introduction of biases stemming from the unequal distribution across occupational categories. This phenomenon is especially evident among underrepresented categories. Furthermore, our analysis indicates that classical models rapidly reach their limits when tasked with inferring the level of performance from job titles alone. In this context, the available information frequently lacks the requisite depth to accurately differentiate between skill levels. This finding underscores the necessity for more comprehensive or even external input data, such as job descriptions, qualifications, or contextual metadata, to capture the subtleties that may be obscured by simplistic job titles.

TABLE III: Results for 3-digit on dataset 1

Approach	Metrics (Macro/Weighted)		
	Precision	Recall	F_1 Score
Logistic Regression (Tf-idf)	**0.81** / 0.80	**0.60 / 0.74**	**0.67 / 0.75**
Logistic Regression (W2v)	0.19 / 0.36	0.09 / 0.28	0.10 / 0.24
Naive Bayes	0.49 / 0.74	0.28 / 0.60	0.33 / 0.61
Random Forest	0.79 / **0.81**	0.57 / 0.71	0.64 / 0.74

TABLE IV: Results for 3-digit on datasets 1+2

Approach	Metrics (Macro/Weighted)		
	Precision	Recall	F_1 Score
Logistic Regression (Tf-idf)	**0.87 / 0.85**	**0.80 / 0.83**	**0.83 / 0.84**
Logistic Regression (W2v)	0.49 / 0.44	0.25 / 0.35	0.29 / 0.33
Naive Bayes	0.85 / 0.80	0.55 / 0.70	0.63 / 0.71
Random Forest	0.86 / **0.85**	0.79 / 0.82	0.82 / 0.83

TABLE V: Results for Level of Performance (5th digit) on dataset 1

Approach	Metrics (Macro/Weighted)		
	Precision	Recall	F_1 Score
Logistic Regression (Tf-idf)	0.83 / **0.83**	**0.76 / 0.83**	**0.79 / 0.83**
Logistic Regression (W2v)	0.60 / 0.52	0.35 / 0.53	0.33 / 0.46
Naive Bayes	**0.87 / 0.83**	0.68 / 0.81	0.73 / 0.80
Random Forest	0.80 / 0.82	0.75 / 0.80	0.77 / 0.80

TABLE VI: Results for Level of Performance (5th digit) on datasets 1+1

Approach	Metrics (Macro/Weighted)		
	Precision	Recall	F_1 Score
Logistic Regression (Tf-idf)	**0.85 / 0.85**	**0.78 / 0.85**	**0.81 / 0.85**
Logistic Regression (W2v)	0.71 / 0.65	0.50 / 0.63	0.50 / 0.62
Naive Bayes	**0.85** / 0.82	0.67 / 0.81	0.71 / 0.81
Random Forest	0.84 / 0.84	0.77 / **0.85**	**0.81 / 0.85**

In terms of future research, several avenues emerge as potentially fruitful directions. Initially, the integration of extensive and heterogeneous datasets, encompassing historical KldB mappings or annotated survey data, could assist in mitigating class imbalance and enhancing generalizability. Secondly, the implementation of advanced language models, such as BERT, domain-specific transformers, or large language models (LLMs), has the potential to substantially improve classification accuracy, especially when fine-tuned on labor market data. Finally, methodological improvements, such as stratified cross-validation or ensemble learning approaches, have the potential to offer more robust evaluation and enhanced performance. While the present approach is both effective and stable, these insights serve as a foundation for more sophisticated future work in the automatic classification of job titles.

References

[1] K. Hein and J. Dörpinghaus, “What is said about vet on social media in germany? trends, demands, and opinions.” in *NORDYRK BOOK OF ABSTRACTS*, 2024, p. 109.

[2] J. Binnewitt, “Recognising occupational titles in german parliamentary debates,” in *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*, 2024, pp. 221–230.

[3] A. Fischer and J. Dörpinghaus, “Web mining of online resources for german labor market research and education: Finding the ground truth?” *Knowledge*, vol. 4, no. 1, pp. 51–67, 2024.

[4] N. R. Council, D. of Behavioral, S. Sciences, C. on Occupational Classification, and Analysis, “Work, jobs, and occupations: A critical review of the dictionary of occupational titles,” 1980.

[5] F. Javed, M. McNair, F. Jacob, and M. Zhao, “Towards a job title classification system,” *arXiv preprint arXiv:1606.00917*, 2016.

[6] Y. Zhu, F. Javed, and O. Ozturk, “Document embedding strategies for job title classification.” in *FLAIRS*, 2017, pp. 221–226.

[7] I. Rahhal, K. M. Carley, I. Kassou, and M. Ghogho, “Two stage job title identification system for online job advertisements,” *IEEE Access*, vol. 11, pp. 19 073–19 092, 2023.

[8] F. Javed, Q. Luo, M. McNair, F. Jacob, M. Zhao, and T. S. Kang, “Carotene: A job title classification system for the online recruitment domain,” in *2015 IEEE First International Conference on Big Data Computing Service and Applications*. IEEE, 2015, pp. 286–293.

[9] R. Baskaran and J. Müller, “Classification of german job titles in online job postings using the kldb-2010 taxonomy,” 2023.

[10] M. Schierholz, “An auxiliary classification with work activity descriptions for occupation coding,” *AStA Wirtschafts-und Sozialstatistisches Archiv*, vol. 12, pp. 285–298, 2018.

[11] A. Müller, “The implementation of the german classification of occupations 2010 in the iab job vacancy survey: documentation of the implementation process,” IAB-Forschungsbericht, Tech. Rep., 2014.

[12] V. P. V. Karanam, “Occupation coding using a pretrained language model by integrating domain knowledge,” 2022.

[13] P. Safikhani, H. Avetisyan, D. Föste-Eggers, and D. Broneske, “Automated occupation coding with hierarchical features: A data-centric

approach to classification with pre-trained language models," *Discover Artificial Intelligence*, vol. 3, no. 1, p. 6, 2023.

[14] R. Fechner, J. Dörpinghaus, and A. Firll, "Classifying industrial sectors from german textual data with a domain adapted transformer," in *2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS)*. IEEE, 2023, pp. 463–470.

[15] R. Fechner and J. Dörpinghaus, "No train, no pain? assessing the ability of llms for text classification with no finetuning," in *Proceedings of the Position Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS), Belgrade, Serbia*, 2024, pp. 8–11.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.