Computers and Education: Artificial Intelligence 3 (2022) 100043

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

On the radar: Predicting near-future surges in skills' hiring demand to provide early warning to educators

Ramtin Yazdanian [a,*], Richard Lee Davis [a], Xiangcen Guo [c,1], Fiona Lim [d], Pierre Dillenbourg [a], Min-Yen Kan [b]

[a] EPFL, Route Cantonale, CH-1015, Lausanne, Switzerland
[b] National University of Singapore, 21 Lower Kent Ridge Rd, 119077, Singapore
[c] Abu Dhabi Department of Economic Development, Al Falah Street, Fatima bint Mubarak St 6, United Arab Emirates
[d] SkillsFuture Singapore, 1 Paya Lebar Link, #08-08 Paya Lebar Quarter 2, 408533, Singapore

ARTICLE INFO

Keywords:
Lifelong learning
Skill needs
AI
Job ads
Predictive analysis

ABSTRACT

The AI-driven Fourth Industrial Revolution and the COVID-19 pandemic have one important thing in common: they both have caused significant and rapid changes to the skill set landscape of various industries. These disruptive forces mean that the early identification of the newly rising skills in a labour market — which we call its "emerging skills" — is crucial to its workforce. It is also crucial to the educators who, in order to provide lifelong training to the workforce, need to quickly adapt their curricula to the new skills.

We propose a classification methodology that uses the past job ad trends of skills to predict the emerging skills of a future period, defined as the skills that have experienced a surge in hiring demand in said period. This general definition allows for freedom in specifying the criteria for a skill being emerging (through thresholds on hiring demand and its growth), which could be important to educators. Applying our methodology to the Information and Communication Technologies (ICT) labour market in Singapore, we show that we are able to predict future emerging skills with good precision and recall and beat two baseline classifiers for multiple threshold sets. Our methodology also allows us to see where job ads fail to provide sufficient predictive signals, pointing to auxiliary data sources (such as Stack Overflow for ICT) and skill ontologies as potential remedies. The success of our method shows how AI can be used to empower learners and educators in the ICT domain (and potentially other domains) with useful and well-curated insights at a moment's notice, thus helping speed up the process of curricular change.

1. Introduction

Today, there is no shortage of disruptive forces acting on labor markets across the world. The AI and automation-centric Fourth Industrial Revolution is in full swing, bringing rapid change to the skills landscape of many domains (Brynjolfsson & McAfee, 2011, 2014; Maisiri et al., 2019). The COVID-19 pandemic has significantly disrupted most industries (Agrawal et al., 2020), making some skills outdated while raising others to prominence. What these disruptive forces have in common is how rapidly they change the skill sets required for many jobs and domains. Such changes require both educational institutions *and* firms to make changes to their course curricula and training programs, as employees may need retraining and students may need to be equipped with skills that have recently risen to prominence. The speed at which these changes happen complicates these processes, as organisational cycles might struggle to keep up with the pace (Brynjolfsson & McAfee, 2011; Ellis, 2003). This situation makes the early identification or prediction of these skill changes necessary (Ellis, 2003; ILO&OECD, 2018; Wilson, 2013), as such early identification can help training providers (both in educational institutions and in corporations) stay on top of the trends, thus speeding up curricular change. However, predicting the skills that are going to become important in the future is a challenging task, be it for the near future or the far future. The near-future prediction of skill needs becomes more challenging the earlier we wish to identify

* Corresponding author.
E-mail addresses: ramtin.yazdanian@epfl.ch (R. Yazdanian), richard.davis@epfl.ch (R. Lee Davis), xiangcen@gmail.com (X. Guo), Fiona_LIM@ssg.gov.sg (F. Lim), pierre.dillenbourg@epfl.ch (P. Dillenbourg), kanmy@comp.nus.edu.sg (M.-Y. Kan).
[1] Affiliated with SkillsFuture Singapore during course of the research project.

https://doi.org/10.1016/j.caeai.2021.100043
Received 3 August 2021; Received in revised form 9 November 2021; Accepted 6 December 2021
Available online 16 December 2021

rising skills (ILO&OECD, 2018).

In the past, different types of approaches have been used to tackle similar problems. Training Needs Analysis (TNA) has used questionnaires, interviews, and focus groups to identify worker skill gaps, often for one firm/organization or a group thereof (Gould et al., 2004). Surveys of university alumni and students have been used to assess the necessary skills that students do not acquire at university (Carnegie & Crane, 2019; Fowler et al., 2014). Analyses of job ads, enabled by the emergence of massive online job ad datasets, have investigated the historical trends of skill demand and projected their growth (Strack et al., 2020). However, none of the existing approaches tackle the problem of the early identification of fine-grained "emerging skills", i.e. skills that are rising to importance from relative obscurity. Traditional TNA approaches and other survey-based methods can be difficult to apply at large scales (due to their data collection methods) (ILO&OECD, 2018), while existing skill trend analysis methods often investigate coarse-grained skills rather than fine-grained ones (Boehm, 2006; CEDEFOP, 2018; Strack et al., 2020). In addition, many approaches focus on describing the present rather than predicting the future (Strack et al., 2020; Szabó & Neusch, 2015).

Defining "emerging skills" as **previously low-demand skills that have recently experienced a surge in hiring demand**, we design a classification pipeline with the aim of predicting the emerging skills of the near future. In other words, we aim to predict the surge in hiring demand before it occurs. Our hypothesis is that the job ad time series of each skill, indicating demand for the skill over time, contains signals that help predict whether or not it is going to emerge in the near future. Applying our methodology to data from the ICT sector in Singapore, we find that such a predictive task is feasible, confirming that job ads contain information that can be used to predict emerging skills. We also investigate the strengths and weaknesses of our classifier models, and examine the signals that distinguish the job ad trends of emerging skills from non-emerging ones, concluding that non-linear growth and spikes are the most important features of an emerging skill's job ad time series. The identification of training needs allows us to provide training providers, including university curriculum designers, Massive Open Online Course (MOOC) creators, and Human Resource managers with early warning on the skills that will soon rise to prominence, giving them time to prepare and/or procure their training material ahead of time.

We first discuss the existing literature on TNA and skill trend analysis with a focus on analyses of online datasets, and lay out the gaps in the literature that our work aims to fill. We then state our research questions and describe our methodology. Afterwards, we present the results of our classification pipeline, answer the research questions based on those results, and interpret the classification models and their predictions. In the end, we discuss the implications and limitations of our work and propose several directions for future work, aimed at rectifying the limitations of our methodology and improving our ability to predict the emergence of skills.

2. Related work

2.1. The need for curricular change

The disruptive effects that automation, AI, and other disruptors such as the COVID-19 pandemic have had on many industries cannot be overstated. For example, the COVID-19 pandemic brought about a sudden switch to teleworking, which in turn caused a surge in the need for basic digital skills, such as the use of teleconferencing software (Agrawal et al., 2020). The disruption brought about by AI and automation is even more fundamental, as many tasks that were previously only feasibly done by humans become doable by increasingly intelligent machines (Illanes et al., 2018). This covers a wide range of tasks, from driving a vehicle, delivering goods to customer care, even diagnosing disease (Forbes, 2019). Research based on recent economic trends shows that although these trends have led to increased productivity, they have spelled trouble for the median worker: as their skills (and at times even their *jobs*) are rendered obsolete through automation, these workers face worsening wages and employment prospects, leading to increasing economic inequality (Brynjolfsson & McAfee, 2011). At the same time, there is an explosion in the demand for skills relevant to the new industry, such as technological, programming, and data analysis skills (Goldfarb et al., 2021; Maisiri et al., 2019). All of these changes are happening in a short time frame, and research shows that many institutions, including educational institutions, have fallen behind (Brynjolfsson & McAfee, 2011). Therefore, educational institutions are in dire need of appropriately rapid methods for curricular change in order to keep up with these rapid developments and provide workers with the appropriate training.

2.2. Methodologies for curricular change

There is a significant body of literature dealing with methodologies for curricular change or for processes that are closely related to it.

One such process is Training Needs Analysis (TNA), which is the process through which discrepancies are identified between the current workforce skills and the necessary workforce skills (Gould et al., 2004). This process can alternatively be referred to as the identification of skill needs (Wilson, 2013; ILO&OECD, 2018). These processes can be applied to individuals, departments, companies (or small groups thereof), or to a labor market as a whole (Gould et al., 2004). We are mainly interested in methods that can identify training needs in an entire labor market, as this is the scale that is most relevant to curricular change in educational institutions. The methodologies come in several varieties, as summarized by the International Labour Organization and the Organization for Economic Co-operation and Development in their 2018 report (ILO&OECD, 2018). Many of these methodologies are survey-based. **Focus groups, interviews, and surveys of domain experts** help gather their opinions on which skills are currently important or are rising in importance (Lee & Mirchandani, 2010). The closely-related **employer-employee surveys** are used to elicit the skill needs of the employees, both from the employers' perspective and from their own (Gould et al., 2004). **Graduate surveys**, in which graduates of an educational institution give their views on the necessary skills that their education had not given them (Carnegie & Crane, 2019; Stevens et al., 2011). **Job vacancy studies** are another type of methodology, and look at the jobs that employers have been unable to fill (Hosen & Alfina, 2016). Finally, **quantitative forecasts**, where the near-future or far-future demand for each skill is predicted based on past data, are quite important for the large-scale identification of skill needs. These methods often involve formal models of the underlying economic processes, such as E3ME (Economics C, 2019), but purely-predictive models also exist. The data used in these methods may involve both existing, continuously-generated data (e.g. online job ads) and collected data (e.g. population or economic censuses).

All the methodologies that involve collecting data (through surveys, interviews, or on the largest scale, censuses) have an important downside: the process of collecting the data is time-consuming and often difficult. For example, attaining an appropriate response rate can be a challenge for survey-based methods, especially when it comes to surveys aimed at experts and executives (Baruch & Holtom, 2008; Fan & Yan, 2010). Also, economic censuses and other such administratively curated data (which past labor market research has relied on) are only periodically collected due to the difficulty of their collection (Horton & Tambe, 2015). This is why big data from online labor market intermediaries such as hiring websites and Massively Open Online Courses (MOOCs) enable previously impossible research approaches: they are always-on and provide fine-grained data (Horton & Tambe, 2015; ILO&OECD, 2018).

Previous labor market research on skills using novel big data sources has often focused on higher-level skill or job trends (Gallivan et al., 2004; Gurcan & Cagiltay, 2019; Lee & Mirchandani, 2010; Matsuda

et al., 2019), and the potential of such data for curricular change remains mostly untapped. Most of the previous works on curricular change either use expert, graduate, or student surveys to effect it (Carnegie & Crane, 2019; Fowler et al., 2014; Stevens et al., 2011), or focus on personalizing education using student learning analytics (Cen et al., 2015; Williamson, 2017). The previous work that is of particular interest to us are analyses of more granular skill trends (BGT, 2019; Dawson et al., 2019; Strack et al., 2020). Some of these are conducted by the corporations that host or own the data, while others are academic research. For example, the whitepaper published by the Boston Consulting Group and Burning Glass Technologies in 2019 (Strack et al., 2020) groups skills into five categories, based on two factors: their overall hiring demand, and the growth of this demand. One of these categories, which they call "high-growth skills", is the main inspiration for our work. These are defined as skills with fewer than 10,000 ads in three years, whose growth over these years has been over 40%. These are the skills that are growing rapidly, but which are less likely to have already been identified as important due to their low previous popularity (compared to those skills that are growing fast *and* already enjoyed significant popularity to begin with). Our concept of "emerging" skills is essentially a generalization of this concept, without the specific thresholds, and combining this idea with the skill demand projections common in the literature is the basic idea behind our work. Another interesting work is (Dawson et al., 2019), where the authors use several hand-picked measures — including the growth in demand for a job title and its predictability — for detecting high-level skill shortages in Australian job ads. Their work particularly touches upon the difficulty of predicting hiring demand, although in their case, it is for job titles rather than skills.

2.3. Our contributions

Our work proposes a methodology for predicting emerging skills, making (Strack et al., 2020) and (Dawson et al., 2019) the closest existing literature to this study. What sets our work apart from these previous works is the fact that our study proposes a methodology for **prediction in the near future** for **granular skills** with **low previous demand**. Among the two previous works mentioned here, the former lacks a predictive focus, whereas the latter focuses on groups of skills, rather than individual, granular skills – a focus that may be necessary when dealing with large numbers of granular but fast-growing skills such as those in ICT.

3. Objectives and methodology

3.1. Data and definitions

Our data consists of all the job ads in the Singaporean Information and Communication Technology (ICT) sector between the beginning of 2017 to the beginning of Q2 2020,[2] although we only examine the data between the beginning of 2017 and the beginning of 2020, in order to exclude the disruptions caused by the COVID-19 pandemic in 2020 (since the effects of the pandemic are not a focal point of our study). Every job ad in our dataset contains the company posting the ad, the date the ad was posted, the textual description of the ad and skills extracted from it.[3] For the period we have chosen (2017–2020), the dataset contains a total of 31,350 job ads, spread across 2,264 companies and involving 987 skills that can in some way be called ICT skills, manually labelled as such by a doctoral student in computer science using multiple passes on the full set of hard skills. These skills range from programming-related skills to skills related to using specific computer software (such as Microsoft Office products, Adobe products, etc.) to a variety of subjects related to or using statistical analysis. Fig. 1 below shows the total number of job ads in the dataset on a monthly basis.

Our analysis relies on the job ad time series of each skill in order to predict the skills that will have a surge in hiring demand in the near future. However, looking at the *number of ads* that include the skill in a particular period of time (e.g. a month) is only one way to analyse the trends of that skill in job ads.

In order to formalize our point, we will define two concepts: The **hiring volume** of a skill is the number of job positions that have been announced for it in a particular time period. The **hiring spread** of a skill is the number of companies that have announced job positions for a skill in a particular period of time. Based on these two concepts, we will introduce three types of job ad time series for skills. These will serve two purposes: they will allow us to define emerging skills precisely, and will serve as competing data inputs to our classification pipeline.

1. Raw popularity (rawpop): The value of the skill's time series for each period of time t (whose length can be one month, one quarter, etc.) is simply the total number of job ads posted for it during that period:

$$rawpop_{s,t} = \sum_{c \in companies} ads_{c,t}$$

Where $ads_{c,t}$ is the number of ads posted during time period t by company c. This type of popularity (and by extension, time series) ignores hiring spread and only emphasises hiring volume.

2. Logarithmic popularity (logpop): For the value of the skill's time series for period t, instead of summing up the total number of ads

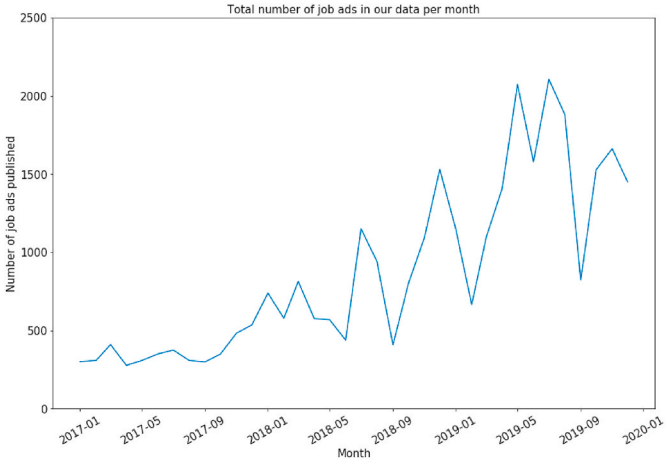


Fig. 1. Number of job ads per calendar month in the 2017–2020 period in our data. Note the rather drastic growth of the number of ads over time, which may be due to a growth in the popularity of JobTech itself. Considerable drops in job ad counts can be observed both in late summer and around the time of the Chinese new year in the later years.

[2] The job ads come from JobTech, a Singaporean online hiring platform who have provided the data to us through SkillsFuture Singapore (SSG) as an intermediary. Gratitude is due both to JobTech, who are the owners of the data and agreed to its sharing, and to SkillsFuture Singapore, who made it available for this study and provided indispensable input and feedback during the research process.

[3] The skills were extracted using JobTech's proprietary methodology and were given to us pre-extracted. For more information, please refer to https://jobtech.co/.

each company has posted for the skill, we first compute the logarithm of that number and then sum them up:

$$logpop_{s,t} = \sum_{c \in companies} log(1 + ads_{c,t})$$

What this type of popularity does is strike a balance between hiring volume and spread: one more ad by a company that has already posted an ad for the skill is worth less than an ad by a company that has not already posted an ad for it. However, it does not throw hiring volume out the window entirely, as more ads by the same company still matter, albeit less than they would in rawpop.

3. Binarized popularity (binpop): The value of the skill's time series for time period t is simply the number of companies that have posted an ad for it:

$$binpop_{s,t} = \sum_{c \in companies} I_{\{x>0\}}(x = ads_{c,t})$$

Where $I_{\{x>0\}}$ is the indicator function that is 1 for positive numbers. This type of popularity throws hiring volume out entirely and only focuses on spread.

An example demonstrating the differences between the three popularity types can be seen in Table 1.

3.1.1. Ground truth

The last preliminary to cover before describing our classification pipeline is to discuss the ground truth that we are going to predict. The definition we have given for emerging skills is a rather vague definition, and needs to be specified further for our prediction task. The two vague parts that need to be specified are "recency" and the "size of the surge".

Let's denote the rawpop of a skill *s* in the year *y* by $Pop_{s,y}$, and the n-th quantile of a set *T* as *Quantile*(*T*, *n*). Also, let's define the quantities $Prevpop_{s,y}$ and $Growth_{s,y}$ as follows:

$$Prevpop_{s,y} = Pop_{s,y-1}$$

$$Growth_{s,y} = Pop_{s,y} - Prevpops, y - 1$$

We then declare *s* an emerging skill in the year *y* if it satisfies:

$$Growth_{s,y} >= Quantile(\{Growths, y\}_{s \in Skills}, q_L)$$

$$Prevpop_{s,y} <= Quantile(\{Prevpop_{s,y}\}_{s \in Skills}, q_U)$$

where q_U and q_L are, respectively, quantile upper and lower bounds on previous year popularity and popularity growth. The first condition (with the quantile q_L) requires the skill to have grown considerably, and eliminates skills that have not experienced a surge in hiring demand from one year to the next. However, with q_L alone, what we have is growing skills, rather than emerging skills. This is why we have the second condition (with the threshold q_U): putting an upper bound on the previous rawpop of a skill enforces the recency part of the definition, as the skill must not have already been too popular in the year $y - 1$. Since q_U and q_L are quantiles, they allow the upper and lower bound values to be determined from the data itself, and for simplicity's sake, we use the same q_U and q_L pair for all years. These two quantiles are two degrees of freedom in our model, and they decide the general popularity level and growth of the skills we deem emerging. For example, lowering q_U will push some of the more popular skills into the non-emerging set, while increasing q_L will shrink the emerging set by making sure that only skills with larger growth values are deemed emerging. It is not a given that our model would work well for any choice of q_U and q_L (and we will see that it does not), and we will discuss how we can set their values. A caveat of our method for generating ground truth is that some skills can emerge in successive years: a skill *s* could be below the threshold q_U for both the year $y - 1$ and the year *y* and be above the threshold q_L for both years, thus putting it into the set of emerging skills twice in a row. We will discuss the implications of this situation in the results section.

Table 1
A comparison of the three popularity types for a hypothetical example. For example, Skill 2 has 2 ads posted by Company 1 and 2 each, and 10 ads by Company 4. Although the rawpop of Skill 2 is the highest, recruitment for it is mostly concentrated in Company 4, resulting in its binpop being lower than that of skills 1 and 3. In addition, the logpop of Skill 3 is again higher than Skill 2's due to greater spread, despite Skill 2's higher rawpop.

		Skill 1	Skill 2	Skill 3
# of ads by companies	**Company 1**	1	2	3
	Company 2	1	2	3
	Company 3	1	0	3
	Company 4	1	10	3
Popularity type	**Rawpop**	4	14	12
	Logpop	2.77	4.60	5.55
	Binpop	4	3	4

It is worth noting that the concept of emerging skills does not *have to* be defined through pure hiring volume (i.e. rawpop); it could also be defined based on hiring spread. Such a definition would focus on how much the skill has spread among companies, rather than how much hiring has happened for it. However, for our main objective of providing insights to training providers on which skills are more in need of training programs, we believe that the number of available positions for a skill is of much greater importance than its spread among companies. As such, we have decided to base our specific definition of emerging skills on hiring volume, rather than spread. However, the question of whether or not signals from hiring spread can help predict hiring demand is a different one; we will explore this question by pitting the three previously-defined popularity types (rawpop, logpop, and binpop) against each other as competing inputs to our predictive pipeline, and compare the performance of their respective models.

One final important subject remains to be discussed before we proceed further. The definition of emerging skills used in this study is a purely computational one, with no expert input. The reason for this design choice is that initially, we sought to use expert opinions to get ground truth on emerging skills. However, the expert-based approach failed due to two reasons:

1. The small number of experts (around 25 people) that ultimately answered the survey we had sent them (despite the fact that over 100 experts were contacted for this purpose).
2. The high level of disagreement that existed among them when they did respond. Since many different and granular areas of expertise exist in ICT, most experts are only intimately knowledgeable about a few of them, which makes disagreements even more likely and necessitates much larger numbers of respondents, which we were unable to attain.

As a result of this failure, we decided to use an approach that would be purely based on labor market demand and does not rely on expert input. This has the added advantage that labor market demand is what we really care about, since the goal of this study is to help training providers prepare workers for *the needs of the labor market they are in.*

3.1.2. Data points: Skill–Periods

Once we have the ground truth, for each type of job ad time series (i.e., each popularity type), we can create data points and create our training/test sets out of them. We call these data points **skill–periods**, with a skill-period for the year *y* consisting of the skill's job ad time series for the entirety of year $y - 1$, along with the ground truth label of the skill in the year *y* (1 if emerging, 0 if not).

Our full training set consists of all the skill–periods for the year 2018, whereas our full test set consists of all the skill–periods for the year 2019. This year-based split is necessary to avoid information leaking from the test set into the training set. The average *Rawpop* time series for

emerging and non-emerging skills for both years can be seen in Fig. 2. In order to compute confidence intervals for our performance measures, we also create multiple skill-based splits,[4] wherein each classifier is trained and evaluated on several random subsamples of the full training and test sets, respectively. In each such subsample, some skills are randomly selected to be in the test set, and are removed from the training set, thus making the training and test sets disjoint both in years and skills. This helps ensure that there is no information spillover from the training set into the test set.

3.2. Classification pipeline

3.2.1. Extracting features

The input to our classifier consists of features extracted from time series (where the time series come from the skill-periods). The features extracted include summary statistics (e.g., mean, various quantiles, variance), linear trends, measures of non-linearity and spikes, the coefficients of a Fast Fourier Transform (FFT) applied to the time series, and many more.[5] Many of these features are intuitively expected to be important (e.g., fast linear or non-linear growth or large spikes can be indicators of quick "emergence"), and the completeness of the set of features ensures that we do not miss out on signals in the data that could be useful to our prediction task.

Before feature extraction, we median-normalize each time point of each skill–period using the median of that time point. We then also apply moving average smoothing to reduce noise in our time series. Afterwards, feature extraction is performed on all the data points.

After the feature extraction, feature reduction is necessary in order to avoid overfitting. This is because the number of data points is quite limited (around 1000 in each of the training and test sets), and the number of extracted features is relatively large (around 300). Our feature reduction pipeline has two steps. In the first, we perform feature selection on the training data to eliminate some of the less discriminating features. This is achieved by performing a one-way ANOVA for every feature and the output, choosing the top N_1 features in terms of F-value. In the second step, we apply Principal Component Analysis (PCA) to the training data and project both the training and test data into the new subspace spanned by the top N_2 principal components. The values of N_1 and N_2 (the number of features after feature selection and PCA, respectively) are, along with the model's other hyperparameters, determined using cross-validation, with the F1-score as the evaluation measure.

3.2.2. Classifier models

For our classifier, we design two competitor models: a **one-step binary classifier** model that predicts our binarized ground truth directly using logistic regression, and a **two-step regression model** that predicts $Growth_{s,y}$ itself using ridge regression. The output of the binary classifier model can be evaluated directly, whereas the q_L quantile will be used to binarize the output of the regression model for evaluation (making it a *two-step* classifier). The reason we have a regression model as our second model type is that the binary ground truth may be noisy near the *Growth* lower bound (i.e. the difference between a skill above the threshold and a skill below it may be quite small). Since the regression model predicts *Growth* itself, it avoids that noise entirely in its training (although the noise from the *Prevpop* upper bound will still be present).

The **one-step binary classifier** is a logistic regression model trained on the binarized ground truth with a post-filtering step, in which we only compute its predictions for skills with *Prevpop* below the upper bound, and predict the rest as negatives. The post-filtering step essentially means that our classifier only learns the indicators of growth that appear in the time series of emerging skills. In other words, among the skills below the *Prevpop* upper bound, it learns to discriminate between those that would grow considerably in the next year and those that would not. It does not, however, learn to discriminate between skills with *Prevpop* values above the threshold and those with values below the threshold. This makes sense, as this threshold is always a known value, even in a real future prediction scenario (since, for example, the upper bound for 2019 skill-periods is computed using the job ad time series in 2018, and uses no information from 2019). In line with this post-filtering step, we perform a pre-filtering step as well: we delete the skills with *Prevpop* above the upper bound from the classifier's training set. This means that the skills the classifier trains on are emerging and not-yet-emerging, while it sees none of the has-already-emerged skills.

The **two-step regression model** is a ridge regression model which, instead of training on the binarized ground truth, learns to predict *Growth* directly. If we denote the output of the model as $PredictedGrowth_{s,y}$, then we predict the skills where

$$PredictedGrowth_{s,y} >= Quantile(\{PredictedGrowth_{s,y}\}_{s \in Skills}, q_L)$$

$$Prevpop_{s,y} <= Quantile(\{Prevpop_{s,y}\}_{s \in Skills}, q_U)$$

as emerging skills, and all the rest as non-emerging. The same pre-filtering step is applied, involving the deletion of skills above the *Prevpop* upper bound from the training set. Much like the logistic regression model described above, this model learns the signals of growth in emerging skills. However, it has three advantages over the binary classifier model. First of all, as discussed before, it could potentially avoid the noise introduced by the binarized ground truth. Secondly, the direct forecasting of each skill's *Growth* can be useful *per se*. Finally, the forecasting of growth in demand means that the model could also be used to predict the skills that are expected to decline in popularity (although this paper is not concerned with such a prediction).

The two model types described above are among the many different types of models that we could use for our classification pipeline. However, more complex machine learning models (such as XGBoost) showed greater overfitting and lower robustness to changes in the problem's parameters such as the two quantile thresholds. Meanwhile, the two relatively simple models described above performed well without much overfitting, making them more suitable for the classification task given the (relatively small) amount of data involved in it.

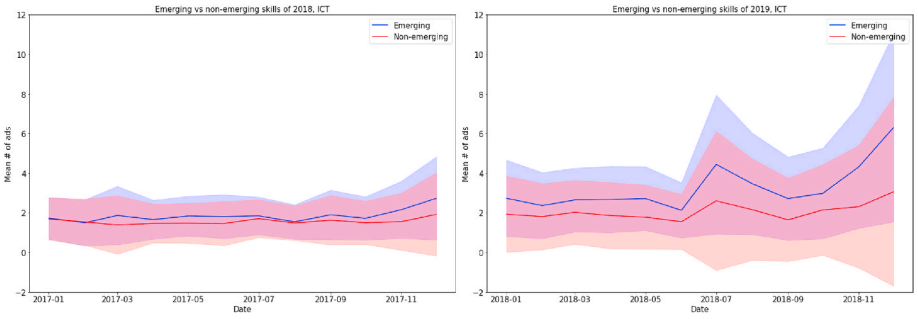
3.2.3. Baselines

In order to evaluate the performance of our model, we need to have some baselines that we can compare our models with. The structure of our problem lends itself to several types of baseline:

1. **Previous-year baseline**: This baseline reports the emerging skills of the previous year as positives and the rest as negatives. This corresponds to the idea that every skill that was emerging last year will be emerging again this year (which is made possible due to the fact that skills *can* be emerging two years in a row).
2. **Below Upper Bound baseline**: This improved baseline relies on the fact that many emerging skills already have some degree of popularity before they emerge. Using our own terminology, it relies on the fact that the most likely candidates for emergence in year y are the ones whose *Prevpop* is just below the upper bound. It reports the top K most previously popular skills as emerging (and the rest as non-emerging). This is a realistic baseline, since it only relies on our past knowledge of a skill. The Below Upper Bound baseline requires training, as we need to choose the value of K; i.e., how many of the below-threshold skills we want to report as emerging. This is done

[4] For the results presented here, the number of splits was chosen by taking one of the classifiers, starting from 10 splits, and continually incrementing the number of splits until the change in the size of the F1 confidence interval went below 0.000 1, which happened at 20 splits.

[5] The features have been extracted using the Python package tsfresh. Its documentation, including a full list of the extracted features, can be found at https://tsfresh.readthedocs.io/en/latest/.



(a) (b)

Fig. 2. The average time series for emerging versus non-emerging skills for **(a)** 2017–2018 and **(b)** 2018–2019. The line indicates the average value and the colored area indicates average ± standard deviation.

using a grid search for the best value of K on the training set, using the F1 score as the evaluation measure.

3.3. Research questions

With our definitions and methodology all laid out in detail, we can now specify our research questions as follows:

RQ1: How well can we predict the emerging skills of the near future?

RQ2: How much does performance degrade when predicting further into the future?

RQ3: To what degree does our ability to predict emerging skills depend on how we precisely define them, i.e. the upper and lower bounds used to specify the skills that are emerging? Are there areas where predictive performance is systematically worse?

RQ4: What are the features of a skill's job ad time series that indicate its near-future emergence?

4. Results

In order to answer our research questions, we have tested our models (each of which uses one particular popularity type and one of the two classifier types) against the two baselines (Previous-year and Below Upper Bound) for three different (q_U, q_L) pairs. To choose the three (q_U, q_L) pairs, we compiled a list of skills that came up as emerging/high-growth in the skill analysis white papers that we reviewed, and sorted them by popularity. We then tried to set the parameters in the three pairs such that the three corresponding emerging skill sets would cover different segments of this list. In this way, we can ensure that the ground truth skills are reasonable, and we are able to test the performance of our system for different (popularity-wise) specifications of emerging skills.

4.1. Predictive performance

Our first research question is concerned with the predictive performance of our model. Table 2 shows 90% confidence intervals for the performance of our best models for each of the three threshold sets, along with the performance of their respective baselines.[6] The three numbers in the parentheses are the 5th percentile, the median, and the 95th percentile, respectively. The first and most important takeaway from this table is that **our model outperforms both baselines for all three threshold sets where we employ the correct classifier**, with the gap between the median F1 scores of our best classifiers and the best baseline (which is the Below Upper Bound baseline) always being greater than 0.05.

Most of our models soundly beat the Previous-year baseline, proving that they learn more than to simply predict all the emerging skills of the previous year as the emerging skills of the next. When it comes to the Below Upper Bound baseline, they almost always have greater precision and lower recall. However, for the purpose of our work, the results of our models are much more useful than those of the Below Upper Bound baseline, as we will demonstrate with an example. Let us take the Logpop + two-step classifier for the (0.8, 0.65) threshold set, whose confidence intervals versus those of the baselines can be seen in the boxplots of Fig. 3. We train it on the full training set, and call the resulting model the **reference classifier**. This model, which beats the ensemble Below Upper Bound baseline (F1 of 0.645 vs 0.600), predicts a total of 243 skills as emerging, whereas said baseline predicts 310 as emerging (some examples of these skills can be seen in Table 3). The baseline has 11 more true positives (and thus 13 fewer false negatives) than our classifier, at the cost of 56 more false positives. Since the emerging skills predicted by our classification pipeline are to be reviewed by experts, it is desirable to keep the number of false positives low, as they make experts' job harder. The reference classifier achieves better performance than the baseline while predicting over 20% fewer skills as emerging, and is therefore much more suitable for our goal of providing experts with insights.

When it comes to a comparison of the different model types, the best-performing model across the board is Logpop + two-step, while Binpop + two-step also generally shows good performance. One for one (i.e. keeping every other factor constant), Rawpop classifiers fails to outperform any Logpop or Binpop classifiers. This has a very interesting implication: hiring spread is a very important component in predicting hiring volume. Comparing one-step and two-step classifiers shows that the former fail badly for the threshold set (0.7, 0.65), showing the greater robustness of the two-step classifier for different threshold sets.

4.1.1. Predicting the further future

We now move on to the second research question, which is the question of whether our performance drops when trying to predict further into the future. It is a rather difficult question to answer with the data that we have, since its length is limited to 3 years. To answer it, we define **first-half emerging skills** as follows: skills that are emerging if we only consider hiring demand in the first half of the year and delete the second half of the year from our calculations. In a similar way, we can define **second-half emerging skills**. We then define first-half-only emerging skills as those that are first-half emerging but not second-half emerging, and define second-half-only as the inverse.[7] The first-half-

[6] Other baselines, such as simplified versions of our classifier models where we use a handful of simple features as our input, were also possible. However, we generally found such classifiers to be outperformed by the Below Upper Bound baseline, and thus excluded them from the results.

[7] Bear in mind that, although unlikely, it is possible for first-half-only or second-half-only emerging skills to *not* be emerging when considering the whole year. In our analysis, we only consider those that *are* emerging.

Table 2
Confidence intervals for the test-set performance of our models versus the baselines for three sets of ground truth thresholds. Each parenthesis is in the format (5th percentile, median, 95th percentile). The classifiers (including the Below Upper Bound baseline) are trained on ground truth from 2018 and tested on ground truth from 2019. For each threshold set, those of our models that significantly beat all baselines in terms of F1 (based on a Kruskal-Wallis test, significance level of 0.05) are in bold font. The thresholds q_U and q_L are the percentiles (between 0 and 1) used for getting the upper and lower bounds, respectively. For example, 0.8 means the 80th percentile.

Ground Truth Threshold Set	Popularity Type	Classifier Type	Precision	Recall	F1
$q_U = 0.8$ $q_L = 0.65$ (166 positives in 2019)	Binpop	Two-step	(0.565, 0.648, 0.715)	(0.5, 0.58, 0.681)	(0.555, 0.605, 0.681)
		One-step	**(0.531, 0.603, 0.656)**	**(0.658, 0.72, 0.801)**	**(0.6, 0.658, 0.713)**
	Logpop	Two-step	**(0.486, 0.534, 0.58)**	**(0.739, 0.79, 0.842)**	**(0.587, 0.637, 0.683)**
		One-step	(0.577, 0.631, 0.701)	(0.516, 0.59, 0.72)	(0.546, 0.616, 0.693)
	Rawpop	Two-step	(0.549, 0.616, 0.736)	(0.499, 0.55, 0.64)	(0.533, 0.594, 0.66)
		One-step	(0.444, 0.493, 0.588)	(0.619, 0.69, 0.761)	(0.528, 0.573, 0.655)
	–	Baseline Previous year	(0.375, 0.434, 0.481)	(0.399, 0.44, 0.481)	(0.392, 0.436, 0.481)
	–	Baseline Below Upper Bound	(0.404, 0.448, 0.495)	(0.8, 0.86, 0.921)	(0.537, 0.591, 0.639)
$q_U = 0.8$ $q_L = 0.7$ (116 positives in 2019)	Binpop	Two-step	(0.489, 0.619, 0.68)	(0.371, 0.543, 0.714)	(0.468, 0.579, 0.677)
		One-step	(0.446, 0.5, 0.607)	(0.51, 0.643, 0.689)	(0.485, 0.554, 0.643)
	Logpop	Two-step	**(0.4, 0.463, 0.502)**	**(0.684, 0.771, 0.859)**	**(0.511, 0.592, 0.633)**
		One-step	**(0.489, 0.559, 0.609)**	**(0.513, 0.586, 0.686)**	**(0.507, 0.570, 0.644)**
	Rawpop	Two-step	(0.457, 0.523, 0.636)	(0.314, 0.486, 0.6)	(0.399, 0.506, 0.578)
		One-step	(0.405, 0.462, 0.556)	(0.429, 0.514, 0.629)	(0.434, 0.493, 0.578)
	–	Baseline Previous year	(0.225, 0.288, 0.346)	(0.227, 0.329, 0.373)	(0.23, 0.307, 0.346)
	–	Baseline Below Upper Bound	(0.377, 0.427, 0.474)	(0.599, 0.714, 0.773)	(0.48, 0.533, 0.579)
$q_U = 0.7$ $q_L = 0.65$ (99 positives in 2019)	Binpop	Two-step	**(0.472, 0.54, 0.741)**	**(0.2, 0.367, 0.502)**	**(0.307, 0.431, 0.572)**
		One-step	(0.056, 0.2, 0.45)	(0.032, 0.067, 0.167)	(0.04, 0.101, 0.229)
	Logpop	Two-step	**(0.36, 0.424, 0.503)**	**(0.565, 0.667, 0.738)**	**(0.449, 0.519, 0.598)**
		One-step	(0.165, 0.317, 0.573)	(0.065, 0.1, 0.168)	(0.087, 0.165, 0.24)
	Rawpop	Two-step	(0.393, 0.517, 0.701)	(0.167, 0.3, 0.402)	(0.263, 0.367, 0.474)
		One-step	(0.091, 0.258, 0.503)	(0.033, 0.1, 0.167)	(0.049, 0.145, 0.234)
	–	Baseline Previous year	(0.166, 0.215, 0.301)	(0.197, 0.25, 0.4)	(0.188, 0.235, 0.339)
	–	Baseline Below Upper Bound	(0.225, 0.256, 0.296)	(0.598, 0.667, 0.768)	(0.328, 0.373, 0.427)

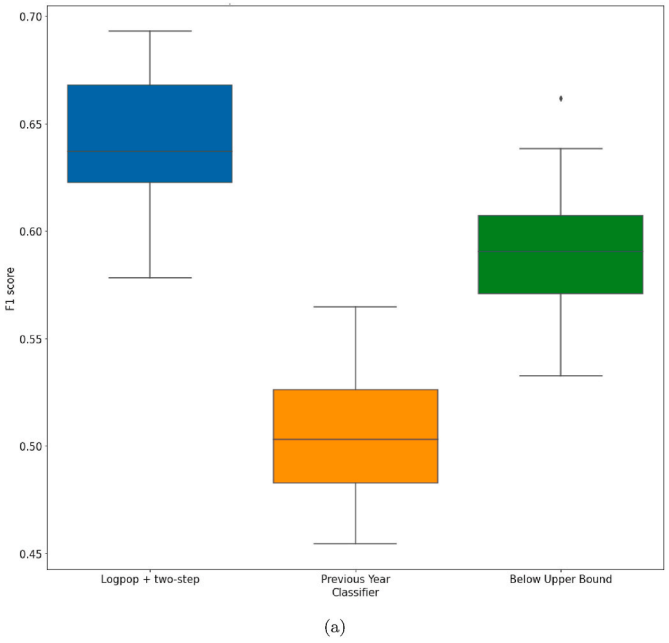


Fig. 3. Boxplots comparing the F1 score of the 20 Logpop + two-step classifiers versus the respective Previous Year and Below Upper Bound baselines for $q_U = 0.8$ and $q_L = 0.65$.

only emerging skills of 2019 should, intuitively, be easier for our models to predict than the second-half-only emerging skills of 2019, as the latter are essentially being predicted 6 months further into the future. Unsurprisingly, by reviewing the predictions of our reference classifier, we find this intuition to be true. Our true positives correctly predict 38 out of 47 of the first-half-only emerging skills, while only predicting 23 out of the 39 s-half-only emerging skills. A chi-squared test to see if the recall on second-half emerging skills is significantly different from recall on the rest rejects the null hypothesis (i.e. the recall being the same) at a significance level of 0.01, whereas the same test for the first-half emerging skills fails to reject the null hypothesis. Therefore, we can conclude that **performance does deteriorate significantly when trying to predict further into the future**, making this an important direction for future improvement.

Table 3
Examples of skills predicted as positive (+) or negative (−) by the reference classifier (Logpop+two-step; $q_U = 0.8$, $q_L = 0.65$) versus the corresponding Below Upper Bound baseline. The skills have been selected to be recognizable and are not randomly sampled. The skills in green are ground truth positives; i.e., emerging (per the definition of emerging skills, and for the thresholds $q_U = 0.8$ and $q_L = 0.65$), while those in red are ground truth negatives.

	Reference classifier (243 predicted positives)	Below Upper Bound baseline (310 predicted positives)
Kubernetes	+	+
Kotlin	+	−
AR/VR	+	+
Tensorflow	+	+
Keras	−	+
Logstash	−	+
Apache Cordova	+	−
Bigquery	−	−
Numpy	−	+
D3.js	−	+
Cryptocurrency	+	+

4.2. The limits of predictive performance

Our third research questions concerns the relationship between the specific definition of emerging skills (or in other words, the values of q_U and q_L) and the performance of our models. As we saw in the results shown in Table 2, reducing q_U led to a considerable worsening of performance across the board. The fact that the threshold set (0.7, 0.65) differs from the threshold set (0.8*and*0.65) only in terms of q_U, which sets the upper bound on *Prevpop*, suggests that our models are generally worse at predicting the less popular emerging skills.

To see whether or not this is true, we investigate the true positives, false positives, and false negatives of our reference classifier by examining their *Prevpop* and *Growth* values. Fig. 4 shows the violin plots of these distributions. According to Fig. 4a, the *Prevpop* values of the reference classifier's false negatives (with a median of 10) are generally much lower than the *Prevpop* values of the true positives (with a median of 22) and false positives (with a median of 36), and the *Prevpop* distribution of false negatives is much more different from the *Prevpop* distribution of true positives, compared to the *Prevpop* distribution of false positives. Meanwhile, Fig. 4b shows that when it comes to *Growth* values, the false negatives (with a median of 28) are more similar in distribution to the true positives (with a median of 46), whereas false positives have much lower values (with a median of 6). For both *Prevpop* and *Growth*, the distribution for the false negatives is significantly different from that of the true positives (Kruskal-Wallis test, significance level of 0.01), but the effect size is much smaller in the case of *Growth* distributions, as we have seen. All this evidence means that the set of false negatives is comprised of skills whose annual growth was relatively comparable to the true positives, but whose previous-year popularity was much lower. This resulted in much weaker signals from their past job ad time series, which caused our model to incorrectly classify them as negatives. In other words, their surge in popularity emerged *too rapidly* for our model to appropriately detect. Therefore, the answer to our third research question is that **classification performance deteriorates for skills whose past popularity is too low**, and this is another area for future improvement, which we shall discuss further.

4.3. Important features

Our last research question is concerned with feature importance. For this analysis, we take the reference classifier, and we investigate the coefficients of the model's features (made possible by the fact that both of our classifier types are linear models), which are themselves linear combinations of the original time series features (since we have used PCA). To compute an ad-hoc importance score for each original feature, we multiply its coefficient in each of the model's features by the weight of that feature in the model, and sum these values up. We then rank the original features using this ad-hoc score. The ranked features and their scores can be seen in Appendix A (Table 4). Based on the values seen in the table, the most two important families of features that contribute positively to the "emergence" of a skill are as follows:

- Features pertaining to non-linearity, sudden growth, and spikes, such as **the number of data points below the mean**, **the value of the time reversal asymmetry statistic**, **skewness**, and **the longest strikes below and above the mean.**
- Features related to the amount of variation in the time series, such as **variance**, **mean absolute sum of changes**, and **whether the variance is larger 1.**

The features that contribute negatively to emergence are murkier in general. The most important family is **the number of recurring data points**, which would penalise time series where many of the values are the same number. Some nonlinearity features show up as negative contributors, but the positive contributors of that family outweigh the negatives.

Looking at these positive and negative contributors, it seems that the most important signals of skill emergence are sudden growth, spikes, and generally larger variation, which is something we would intuitively expect, given the definition of emerging skills. This is also consistent with the false negative problem that we had previously discussed: when a skill's job ad counts are low, almost all of the positive features will have reduced values, making it much more likely for the skill to be classified as non-emerging.

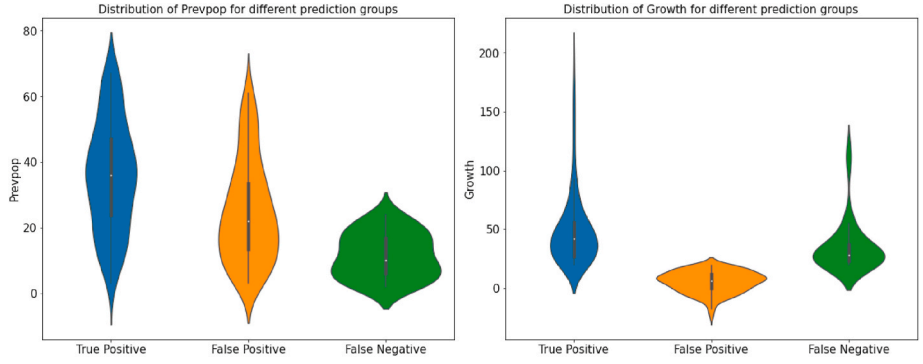


Fig. 4. The distributions of **(a)** *Prevpop* and **(b)** *Growth* values for true positives, false positives, and false negatives of the reference classifier (on the test set, meaning that these are the rawpop values of these skills in 2018). Note that *Prevpop* values are non-negative.

5. Discussion

5.1. Implications for educational institutions

Our results showcase the feasibility of forecasting emerging skills in the ICT domain, although work remains to be done on its generalizability to other ICT labor markets and other professional domains. This success shows that AI can help enable educational institutions to keep up with rapid changes in the labor market, especially since the ICT domain is among the most rapidly evolving professional domains. Although our methods do not have perfect precision or recall, they often only fail to predict the emergence of much less popular skills, which are usually (but not always) related to larger skills that our methods do classify as emerging (e.g. even though Keras is not predicted as emerging, TensorFlow is). This, plus the fact that our methods predict a manageable number of skills as emerging, means that they are able to provide insightful information about the evolution of the skills landscape to training providers and decision makers, be they in institutions providing formal education (e.g. universities), lifelong learning platforms (e.g. MOOC websites), or in the training departments of corporations.

Our methodology pairs especially well with MOOCs. The early warning provided by our methods allows MOOC creators (e.g. on Udemy, which is a popular MOOC platform where anyone can create a course) to create short, skill-based online courses in anticipation of the emergence of particular skills, thus speeding up curricular change. Here, the unique focus of our methodology on less popular and more granular skills (which are the ones more likely not to have already been identified as important skills) provides an advantage: finer-grained skills require shorter courses, which would in turn be quicker to make.

This does not mean, however, that the results of our methodology are not useful for universities. The purpose of analyzing emerging skills is not only to upskill the existing workforce, but also to appropriately train the upcoming workforce, or in other words, students. Therefore, the insights generated by our predictive pipeline are just as important for curriculum and course designers in universities as they are for MOOC designers.

5.2. Limitations

Our pipeline and its results have a few limitations. Firstly, we effectively only had access to three full years of data, meaning that we could only compute ground truth emerging skills for two years (2018 and 2019). As a result, the only test of future prediction we could perform was to train models that predict one year into the future for one specific year. The fact that we are able to predict the emerging skills of 2019 with good performance means that after our preprocessing steps, the skill time series from 2017 to 2018 and those from 2018 to 2019 look reasonably similar. In other words, the changes in the market from the year 2017 to the year 2018 are not so big as to make the signals learned based on 2017 data useless for 2018 data, and our model is able to pick up signals that are relevant for both years. However, we do not know if this phenomena would hold for other periods, since we have only tested the forecasting ability of our methodology for a single pair of years. Additionally, as we saw, our model's predictive ability deteriorates when we try to predict further into the future. However, if training programs can be created rapidly enough; e.g. in 2–3 months, we believe this would not be a big issue. Such a time frame is not unreasonable for skill-based MOOCs, due to their smaller size than full-fledged training programs, and previous research has shown decentralized MOOC platforms such as Udemy to be quite agile (Yazdanian et al., 2020).

The second important limitation of our work is that it ignores all trends that are larger in scope than individual skill trends. This means that it ignores the relationships between skills and is oblivious to larger trends, such as the collective rise of a group of skills (e.g. the simultaneous rise of several deep learning-related skills) or the rise of one group accompanied by the fall of another (e.g. a new wave of JavaScript-based technologies phasing out older web technologies).

Finally, as we have discussed before, our methodology struggles to correctly predict emerging skills whose previous popularity is too low, due to the limited signals available in their job ad time series. In other words, our methodology does not work well when a skill emerges very quickly and unexpectedly. This can be a particular limitation when it comes to new skills, which have no ads before their creation date. However, even very fast-growing new skills such as TensorFlow often experience a short period of low demand in which they are not yet sufficiently well-known to gain widespread adoption, and the data from that period can be used to predict their emergence. As a result, this is not much of a problem, unless very early detection is desired. Therefore, the impact of this limitation on the value of our results depends on the experts that wish to use our results: if they only consider lower-popularity skills to be truly emerging or if they are looking for very early identification of emerging skills, then the impact of this issue will be larger. Therefore, addressing this problem is a high priority for future work on our methodology.

5.3. Future directions

Our work opens up multiple avenues for future work, both in the form of generalizations and improvements to the existing system.

First is the generalization of our methodology to other domains. Since our methodology is self-contained, with the emerging skill ground truth and the predictive signals all coming from job ads, it is generalizable to any professional domain and any labour market where formal job ads exist. The main question, when it comes to generalizability, is whether emerging skills are a viable practical concept in the professional domain in question. The most important factor in answering this question is the rate at which the labour market evolves, both in terms of the appearance of new skills and the growth of existing skills. This is something we essentially took for granted in the ICT domain, as it is probably the fastest-evolving professional domain at the moment. Our proposed methodology provides a framework for verifying the viability of emerging skill prediction through the question of whether or not we can beat baseline emerging skill predictors, and a study of the professional domain's rate of evolution would further strengthen our method's ability to verify the applicability of the concept of emerging skills to said domain.

Secondly, since our results imply the importance of hiring spread in the prediction of hiring volume, we can ask the following question: Is there a particular set of companies that anticipate skill trends well? In other words, is a skill's spread among certain companies more important than its spread among others? The idea that such a set of companies exists makes intuitive sense in the ICT domain, where Big Tech are often the creators of new technologies, and following these corporations alone can yield valuable insights into the direction of the market in the near future. This idea could provide an improvement to our pipeline: An approach where the spread of a skill among companies is weighted by the "influence" of each company (as opposed to the current approach, where every company has the same weight), with more "trend-anticipating" companies having larger weights. The "influence" concept would have to be defined based on the company's past ability to predict emerging skills.

Another direction for future work is analyzing emerging skills by finer-grained geographical regions (e.g. analyzing by state instead of analyzing the entire country at once). A more complex, hierarchical model can incorporate this geographical information and identify the emerging skills of each region while also identifying and incorporating higher-level trends (national in this case). Such an analysis was not possible in the present study due to the fact that Singapore is a city-state and no meaningful geographical subdivisions were possible.

There are several research directions directed towards rectifying the limitations of our current pipeline. One of the limitations of our models is that they predict the emergence or non-emergence of each skill only

using its own job ad trends, essentially ignoring the relationships between skills. In reality, many skills are related, and their trends are part of larger trends. For example, the simultaneous rise of "Tensorflow", "Keras", and "PyTorch" is not accidental, but rather due to the rise of "Deep Learning" in general. This points towards an approach incorporating a skills ontology: if the relationship between the four skills mentioned above is made explicit (e.g. through "Deep Learning" being a parent of the other three), it could be incorporated into a model that looks not only at the job ad trends of the skill it's predicting, but also at those of its parents and children. Such a classifier model would need to be more complex than the linear models we have used in this study.

Another limitation to address is the fact that our current pipeline only uses one source of data. On one hand, this is a strength, since it makes our method self-contained and applicable to any domain. On the flip side, however, auxiliary data sources that are domain-specific could provide additional signals and improve our predictive ability. For the ICT domain, an interesting auxiliary data source is Stack Overflow, a massively popular Q&A platform for software developers. Incorporating signals from Stack Overflow could improve our ability to forecast skill trends, since previous work has shown that it tends to be faster than job ads at manifesting new skills. This could be a potential solution to another limitation of our method, which is the problem of low-popularity skills becoming false negatives: Stack Overflow could show earlier and stronger signals of these skills' rise in popularity.

6. Conclusion

We have presented a generalizable methodology to predict emerging skills, and have showed its feasibility in the ICT domain, which is one of the fastest-changing domains. Our methodology's early identification of rising skills provides training providers and domain experts with insights that help speed up the process of curricular change. This allows educational institutions to keep up with the trends and to equip workers with the right skills for a changing labor market. Our work shows that AI is a double-edged sword, disrupting labor markets but also able to help institutions and people adjust to the new markets, thus addressing some of the problems it causes. We believe that future work incorporating skill ontologies and auxiliary signals can help address the limitations of our method and push the boundaries of emerging skill prediction even further, providing more accurate and more comprehensive insights to experts.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Importance scores for original features

Table 4
A list of features used in the reference classifier, along with their ad-hoc importance scores. Explanations of the features and their names can be found at https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html.

Original Feature	Ad-hoc Importance
cid_ce__normalize_True	3.260 370 174
last_location_of_minimum	3.041 300 917
longest_strike_above_mean	1.819 125 048
time_reversal_asymmetry_statistic__lag_2	1.771 156 6
Skewness	1.746 095 463
Kurtosis	1.721 839 525
longest_strike_below_mean	1.704 475 687
Minimum	1.490 589 562
variance_larger_than_standard_deviation	1.470 279 171
has_duplicate_min	1.470 279 171
benford_correlation	1.470 279 171
mean_abs_change	1.453 978 262
standard_deviation	1.300 465 484
Variance	1.254 512 743
mean_second_derivative_central	1.218 262 811
first_location_of_maximum	1.210 340 323
last_location_of_maximum	1.174 841 021
Median	1.107 915 982
variation_coefficient	1.050 612 425
Maximum	1.000 516 271
absolute_sum_of_changes	0.985 774 444
has_duplicate_max	0.944 529 448
sum_of_reoccurring_data_points	0.944 529 448
abs_energy	0.944 529 448
c3__lag_3	0.747 451 9
ratio_value_number_to_time_series_length	0.559 784 466
c3__lag_1	0.449 628 367
sum_of_reoccurring_values	0.391 383 83
first_location_of_minimum	0.258 098 861
count_above_mean	0.204 591 848
time_reversal_asymmetry_statistic__lag_3	0.030 081 356
mean_change	−0.218 127 186
sum_values	−0.233 810 196
percentage_of_reoccurring_values_to_all_values	−0.353 926 206
count_below_mean	−0.529 535 18
percentage_of_reoccurring_datapoints_to_all_datapoints	−0.571 704 456
has_duplicate	−0.655 769 556
Mean	−0.850 531 591

(continued on next page)

Table 4 (*continued*)

Original Feature	Ad-hoc Importance
time_reversal_asymmetry_statistic__lag_1	−1.468 221 572
c3__lag_2	−1.631 146 116

References

Agrawal, S., Smet, A. D., Lacroix, S., & Reich, A. (2020). *To emerge stronger from the COVID-19 crisis, companies should start reskilling their workforces now*. Technical Report McKinsey Insights.

Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations, 61*, 1139–1160. https://doi.org/10.1177/0018726708094863. Publisher: SAGE Publications Ltd).

BGT. (2019). Mapping the genome of jobs: The burning Glass skills taxonomy. Technical Report Burning Glass Technologies. URL: https://www.voced.edu.au/content/ngv:84406.

Boehm, B. (2006). Some future trends and implications for systems and software engineering processes. Systems Engineering, 9, 1–19. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sys.20044. doi:https://doi.org/10.1002/sys.20044. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sys.20044.

Brynjolfsson, E., & McAfee, A. (2011). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Brynjolfsson and McAfee.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Carnegie, T. A. M., & Crane, K. (2019). Responsive curriculum change: Going beyond occupation demands. *Communication Design Quarterly, 6*, 25–31. https://doi.org/10.1145/3309578.3309581

CEDEFOP. (2018). Skills forecast. URL: https://www.cedefop.europa.eu/en/publications-and-resources/data-visualisations/skills-forecast.

Cen, L., Ruta, D., & Ng, J. (2015). Big education: Opportunities for big data analytics. In *2015 IEEE international conference on digital signal processing (DSP)* (pp. 502–506). https://doi.org/10.1109/ICDSP.2015.7251923. iSSN: 2165-3577.

Dawson, N., Rizoiu, M.-A., Johnston, B., & Williams, M.-A. (2019). Adaptively selecting occupations to detect skill shortages from online job ads. In *2019 IEEE international conference on big data (big data)* (pp. 1637–1643). https://doi.org/10.1109/BigData47090.2019.9005967

Economics, C. (2019). E3ME - our global macro-econometric model (version 6). URL: https://www.e3me.com/what/e3me/.

Ellis, S. P. (2003). Anticipating employers' skills needs: The case for intervention. *International Journal of Manpower, 24*, 83–96. https://doi.org/10.1108/01437720310464981. Publisher: MCB UP Ltd).

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior, 26*, 132–139. https://doi.org/10.1016/j.chb.2009.10.015. URL: https://www.sciencedirect.com/science/article/pii/S0747563209001708.

Forbes. (2019). *Tech experts predict 13 jobs that will Be automated by 2030*. URL: https://www.forbes.com/sites/forbestechcouncil/2019/03/01/tech-experts-predict-13-jobs-that-will-be-automated-by-2030/.

Fowler, D., Poling, N., Anthony, W., Morgan, J., & Brumbelow, K. (2014). Data-driven curriculum redesign in civil engineering. In *In 2014 IEEE frontiers in education conference (FIE) proceedings* (pp. 1–9). https://doi.org/10.1109/FIE.2014.7044242. iSSN: 2377-634X.

Gallivan, M. J., Truex, D. P., & Kvasny, L. (2004). Changing patterns in IT skill sets 1988-2003: A content analysis of classified advertising. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 35*, 64–87. https://doi.org/10.1145/1017114.1017121

Goldfarb, A., Taska, B., & Teodoridis, F. (2021). *Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings*. URL. https://doi.org/10.2139/ssrn.3468822. SSRN Scholarly Paper ID 3468822 Social Science Research Network Rochester, NY https://papers.ssrn.com/abstract=3468822.

Gould, D., Kelly, D., White, I., & Chidgey, J. (2004). Training needs analysis. A literature review and reappraisal. *International Journal of Nursing Studies, 41*, 471–486. https://doi.org/10.1016/j.ijnurstu.2003.12.003. URL: https://www.sciencedirect.com/science/article/pii/S0020748903001913.

Gurcan, F., & Cagiltay, N. E. (2019). Big data software engineering: Analysis of knowledge domains and skill sets using LDA-based topic modeling. *IEEE Access, 7*, 82541–82552. https://doi.org/10.1109/ACCESS.2019.2924075 (Conference Name: IEEE Access).

Horton, J. J., & Tambe, P. (2015). Labor economists get their microscope: Big data and labor market analysis. URL: *Big Data, 3*, 130–137. Publisher: Mary Ann Liebert, Inc., publishers https://www.liebertpub.com/doi/abs/10.1089/big.2015.0017.

Hosen, A., & Alfina, I. (2016). Aggregation of open data information using linked data: Case study education and job vacancy data in Jakarta. In *2016 international conference on advanced computer science and information systems (ICACSIS)* (pp. 579–584). https://doi.org/10.1109/ICACSIS.2016.7872725

Illanes, P., Lund, S., Mourshed, M., Rutherford, S., & Tyreman, M. (2018). *Retraining and reskilling workers in the age of automation*. Technical Report McKinsey & Company.

ILO, & OECD. (2018). *Approaches to anticipating skills for the future of work*. URL: Report International Labour Organization & Organization for Economic Co-operation and Development, 2nd Meeting of the Employment Working Group, Geneva, Switzerland https://www.ilo.org/global/about-the-ilo/how-the-ilo-works/multilateral-system/g20/reports/WCMS_646143/lang–en/index.htm.

Lee, K., & Mirchandani, D. (2010). Dynamics of the importance of IS/IT skills. *Journal of Computer Information Systems, 50*, 67–78. URL: https://www.tandfonline.com/doi/abs/10.1080/08874417.2010.11645432. doi:10.1080/08874417.2010.11645432. Publisher: Taylor & Francis _eprint https://www.tandfonline.com/doi/pdf/10.1080/08874417.2010.11645432.

Maisiri, W., Darwish, H., & Dyk, L. (2019). An investigation of industry 4.0 skills requirements. *South African Journal of Industrial Engineering, 30*. https://doi.org/10.7166/30-3-2230

Matsuda, N., Ahmed, T., & Nomura, S. (2019). *Labor market analysis using big data: The case of a Pakistani online job portal*. SSRN Scholarly Paper ID 3491253 Social Science Research Network Rochester, NY. URL: https://papers.ssrn.com/abstract=3491253.

Stevens, D., Totaro, M., & Zhu, Z. (2011). Assessing it critical skills and revising the mis curriculum. URL: *Journal of Computer Information Systems, 51*, 85–95. https://doi.org/10.1080/08874417.2011.11645489. Publisher: Taylor & Francis _eprint https://www.tandfonline.com/doi/pdf/10.1080/08874417.2011.11645489 https://www.tandfonline.com/doi/abs/10.1080/08874417.2011.11645489.

Strack, R., Kaufman, E., Ádám, K., Sigelman, M., Restuccia, D., & Taska, B. (2020). *What's trending in jobs and skills*. Technical Report Boston Consulting Group & Burning Glass Technologies. URL: https://www.bcg.com/publications/2019/what-is-trending-jobs-skills.

Szabó, I., & Neusch, G. (2015). Dynamic skill gap analysis using ontology matching. In A. Kő, & E. Francesconi (Eds.), *Electronic Government and the information systems perspective lecture notes in computer science* (pp. 231–242). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-22389-6_17.

Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. SAGE. Google-Books-ID: yFAsDwAAQBAJ.

Wilson, R. (2013). Skills anticipation—the future of work and education. *International Journal of Educational Research, 61*, 101–110. https://doi.org/10.1016/j.ijer.2013.03.013. URL: https://www.sciencedirect.com/science/article/pii/S0883035513000335.

Yazdanian, R., West, R., & Dillenbourg, P. (2020). Keeping up with the trends: Analyzing the dynamics of online learning and hiring platforms in the software programming domain. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-020-00231-1. URL:.