

# 5 Exploiting big data to measure skills needs for the green transition

---

Big data can be a valuable source of information for timely and granular labour market analysis. In labour market research on skills, the most commonly used type of big data comes from online job vacancies that are analysed through Natural Language Processing. By means of examples from the public and private sector, this chapter discusses what big data is and how it is currently used in skills analysis linked to the green transition.

---

## Introduction

Big data are increasingly making their way into policy debates as a promising source of information for a more timely and granular analysis of economic phenomena. As mentioned in Chapter 4, a growing number of skills assessment and anticipation exercises exploit big data to get a timely picture of how green skills needs are evolving. This chapter of the report aims to shed light on the main features of big data analyses for the green transition, their drawbacks and success factors. Four case studies are examined in particular: (1) Jobs and Skills Australia’s big data work; (2) the study on “clean growth jobs” undertaken by the Department of Business, Energy and Industrial Strategy (BEIS) of the United Kingdom; (3) the green skills classification by ESCO; and (4) LinkedIn’s green skills research.

### **What is big data?**

First of all, it is important to clarify key terms and concepts related to big data analysis. While the use of the term “big data” often lacks precision, it usually refers to the large dimension of the datasets, that is coupled with the need to use large-scale computing power, non-standard software, and advanced methods to extract value from such data in a reasonable amount of time (OECD, 2016<sup>[1]</sup>). For instance, Chen et al. (2012<sup>[2]</sup>) define big data on the basis of how they perform vis-à-vis the three major challenges in data management: volume, variety and velocity, which are all ‘big’ in the case of big data.

One of the most commonly used types of big data in labour market research comes from online job vacancies (OJVs) – also referred to as job postings or job ads (Börner et al., 2018<sup>[3]</sup>; Hershbein and Kahn, 2018<sup>[4]</sup>; Modestino, Shoag and Balance, 2020<sup>[5]</sup>). Datasets on OJVs are typically compiled by web scraping online job boards (e.g. Indeed or Monster) and company websites on a regular and frequent basis. Some OJV data sources manage to reflect the quasi-totality of online job postings in a given country or area. Some of the most popular datasets are the ones produced by Lightcast (formerly Burning Glass Technologies) and by LinkedIn Economic Graph. One of the main advantages of these data is that they are a rich source of timely and granular information on skills and job requirements, which are typically difficult to gather via traditional methods such as labour market statistics. Thanks to this novel information, OJV analysis can provide valuable insights on labour market trends and enable the early identification of new or emerging jobs and skills (ILO, 2020<sup>[6]</sup>).<sup>1</sup> For example, Cedefop has been collecting and refining OJV data from EU member states since 2017, to understand the ongoing evolution of occupations and skills demand (Cedefop, 2019<sup>[7]</sup>).

The ILO was one the first international organisations to exploit OJVs in the context of the green transition. In 2019, the ILO used real-time big data on job ads from Burning Glass Technologies (now Lightcast) to conduct country studies examining the skill needs and occupational transitions linked to the shift towards low-carbon economies (ILO, 2020<sup>[6]</sup>). In particular, the ILO used a proprietary multiregional input – output model (EXIOBASE v3) to forecast the demand for each occupation in a green transition scenario, and then applied OJV big data to integrate the skills component. After identifying those industries and jobs that are expanding or decreasing due to the green transition according to their EXIOBASE model, the ILO used skills information from the jobs postings to identify the potential job transitions for workers in declining industries (ILO, 2020<sup>[6]</sup>).

Despite the clear value-added of data from online job postings for the estimation of current skills needs, several limitations of the approach should be highlighted. First, occupations and sectors where recruitment rarely takes place online are underrepresented in OJV datasets. Construction, fishing, or agriculture, for instance, are likely to be only partly covered. Similarly, at the occupation level, while data on high-skill occupations such as “managers”, “professionals”, or “technicians and associate professionals” showed good representativeness (Cammeraat and Squicciarini, 2021<sup>[8]</sup>), vacancies for the most elementary occupations are often not posted online. Furthermore, micro, small and medium enterprises are less likely to post online their job openings, thereby skewing advertisements towards those vacancies in larger firms.

The skills information in job ads does not always reflect the full skill profile of an occupation, since employers tend to include only some of the skills needed in the position, leaving aside either more transversal competencies or those skills that are so inherently linked to a profession that there is no need to include them explicitly in the job advertisement. For example, an online job posting for plumbers might mention “communication skills” but not “pipe installation” or “estimation of material requirements for projects”, both activities that takes up most of plumbers’ time but are considered so essential that they are taken for granted (ILO, 2020<sup>[6]</sup>).

## Global insights

### Measuring mismatch for green jobs in Korea through online job postings

In 2021, Korean researchers in Yonsei University analysed the degree of mismatch between the supply and demand for green jobs, exploiting online job postings from ‘Ecojob’ website, a green job-related recruiting services platform in Korea. Using web scraping techniques, the researchers obtained information on both the company side (e.g. industry, location, occupation, hire type, salary, experience required for each job advertisement), and jobseeker side (e.g. desired working area, employment type, desired industry and occupation, educational background, qualification). Thanks to this innovative source of data, the authors were able to quantify the degree of mismatch between supply and demand for green jobs in Korea by region, industry, and salary level (Song et al., 2021<sup>[9]</sup>).

### Forecasting jobs for the green transition in Germany

In a study financed by the Federal Ministry for the Environment, Bauer et al. (2021<sup>[10]</sup>) exploit OJVs to define occupations and sectors that are particularly relevant for the green transition in Germany, and to identify possible labour shortages that could hinder the green transition. Their analysis is split in two parts.

The first part focuses on identifying current occupations and sectors in the “green economy”, and is based on a large-scale survey, containing information on the job advertisements posted on the portal [www.greenjobs.de](http://www.greenjobs.de) and on the job postings of the Federal Employment Agency. Based on desk research, the team manually creates a training set – i.e. a catalogue of keywords to detect occupations related to the “green economy”. They then run an automated text analysis (keyword filter) to detect these relevant keywords in the job postings data. This process is repeated and refined through manually testing random samples of occupations identified in the analysis and adding additional keywords or themes that were previously not included in the catalogue. Data on job postings by the Federal Employment Agency have been collected since 2012, which also allows a trend analysis of the development of jobs and sectors relevant to the “green economy”. To provide a forecast of the evolution of the identified occupations and sectors in the “green economy”, the study uses a macro-econometric input-output model called INFORGE. This approach allows estimating a range of future scenarios – from 2015 to 2025 and from 2025 to 2035 – on the basis of comprehensive economic and employment microdata capturing the complex interlinkages across sectors and activities, such as trade flows, household and government income generation and use, and investment dynamics.

The second part of the analysis provides simulations on the impact of selected green policies on occupations and sectors in the “green economy”, namely: (i) an increase in the investment rate in the refurbishment of buildings from 1% to 2%, (ii) a change in individual mobility behaviour towards public transport or non-motorised vehicles, and (iii) a change towards a more digitalised “Economy 4.0”.

Based on this two-step approach, the analysis by Bauer et al. (2021<sup>[10]</sup>) provides a remarkable level of granularity and allows forecasting by combining a big data approach with a macroeconomic model. However, this comes at the expense of high complexity and dependence on predetermined assumptions, which may strongly impact the projections.

### ***How does skills analysis with big data work?***

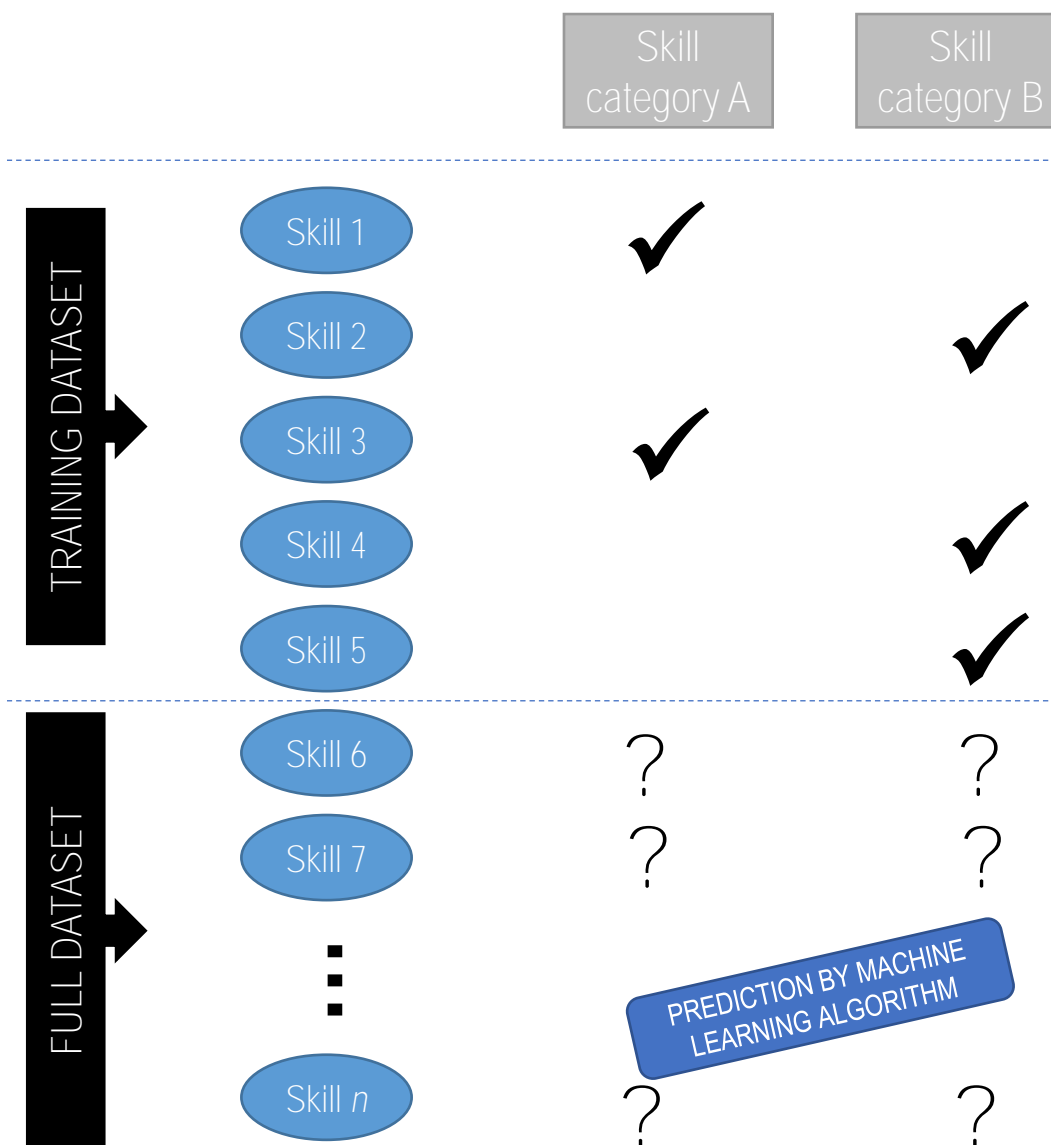
Big data has the potential of providing real-time, granular information compared to existing surveys and statistics sources. Since vast amounts of data are collected, however, selecting the correct methodology for proper data processing is imperative (ILO, 2020<sup>[6]</sup>). This is the case, for example, when identifying skills in big data, since information on skills typically appears as strings of text rather than pre-codified labels, and therefore require specific techniques to ensure that ambivalent text is mapped into meaningful skills categories.

Most studies attempting to classify skills based on OJVs use text classification, i.e. “the transformation of unstructured textual data (documents, books, reports, etc.) in a structured format” (Lassébie et al., 2021<sup>[11]</sup>). In recent years, a number of language models have been developed by researchers to correctly process textual data. These so-called Natural Language Processing (or NLP) models rely on machine learning, which is a type of artificial intelligence where a programme is able to learn rules from existing datasets without being explicitly programmed by humans (OECD, 2021<sup>[12]</sup>). In particular, the machine learning algorithms of NLP models often use the semantic context in which the text appears to transform strings into data (OECD, 2022<sup>[13]</sup>).

To simplify a very complex approach, many big data studies aimed at inferring skills from unstructured text (such as job postings or LinkedIn user profiles, as in the case studies below) follow a three-step methodology:

- First, the researchers create a training dataset to help the machine learning algorithm derive the logical rules to interpret text. The training dataset is composed of examples of skills and can take multiple forms – e.g. it can be a list of keywords or a collection of skills definitions.
- Second, learning from the training dataset, the NLP algorithms classify the text of, for example, the job postings or LinkedIn’s profiles into the relevant skill category (see an illustrative example in Figure 5.1).
- Finally, researchers often validate the machine learning output through manual checks, experts’ consultations, etc.

Figure 5.1. Example of skills analysis using machine learning algorithms



Source: Authors' elaboration.

### ***The intricacies of using big data***

All in all, two main features of big data analysis should be kept in mind to evaluate their usefulness for countries aiming to measure the skills needs stemming from the green transition. First, a key advantage of big data is their high frequency. These often-daily observations can clearly provide timely insights to understand better labour market trends. However, it is also important to note that these daily data cannot always be processed enough quickly by analysts, who therefore tend to publish their big data analysis results with a much lower frequency (e.g. quarterly or every six months). In a similar vein, users of these results – such as policy makers and journalists – risk having an information overload and cannot always digest high-frequency updates on employment and skills trends.

Another key advantage of big data is their granularity. Data such as online job vacancies, in fact, provide very detailed information on individuals' employment characteristics and skills, most of which are not available in standard labour market datasets. Yet, it needs to be reminded that this information only captures flows: i.e. it focuses on what are the new job postings and what are the skills required in today's openings. This is useful to nowcast current and short-term trends – i.e. to get a sense of what are the emerging occupations or competencies in the economy. By contrast, big data such as OJVs do not provide insights on what are the skills currently held by the overall population (i.e. the stocks), thereby making it difficult to use these sources of current skills assessments.

## Case study 1: Using big data to assess in-demand skills in Australia

In Australia, Jobs and Skills Australia (JSA), following on from work initiated by the National Skills Commission (NSC), is exploring data from Lightcast (formerly Burning Glass Technologies) to analyse real-time trends on job advertisements with the goal of identifying the skills that have grown in demand over the past five years. More specifically, the analysis matches skills information taken from the Lightcast job postings to the over 1 100 profiles outlining the required skills for each occupation contained in the Australian Skills Classification (ASC).

Each ASC skills profile for an individual occupation comprises three main elements – core competencies, specialist tasks, and technology tools (National Skills Commission, 2021<sup>[14]</sup>). *Core competencies* are universal skills that are required in all jobs although at different proficiency levels across sectors and professions – e.g. numeracy and literacy. *Specialist tasks* are the activities that workers undertake on a day-to-day basis within each occupation. Specialist tasks are only transferable to other occupations in the same *skill cluster*, i.e. jobs that have similar sets of specialist tasks.<sup>2</sup> Since specialist tasks are broadly transferable, it is likely that if a worker can perform one of the tasks in the cluster, (s)he can perform the other ones too. Finally, *technology tools* are the technologies required in each occupation. These can either be common technology tools that can be found across multiple jobs (such as email and search engines) or highly specialised, occupation-specific tools (like carbon monoxide analysing equipment).

### Definitions and methods

To assess which skills are most in demand, JSA distinguishes between emerging and trending skills:

- *Trending skills* are defined as already existing skills within an occupation that have grown in demand over the past five years. For example, the need for *social media skills* has grown more than ten times for hotel managers in the past five years (National Skills Commission, 2021<sup>[14]</sup>).
- *Emerging skills*, on the other hand, are trending skills which are also new to a certain occupation. For example, *infection control skills* are now required by 38 new occupations compared with five years ago.

Through a network analysis, JSA also identifies those trending and emerging skills that can be considered *gateway skills*, i.e. skills that are trending in multiple occupations and have many connections to other skills. Gateway skills provide a point of transferability between different occupations and uncover job transition pathways (National Skills Commission, 2021<sup>[14]</sup>).

In practice, skills reported in the Lightcast database have been matched with those in the Australian Skills Classification (either specialist tasks or technology tools). Qualitative analysis and desktop research was used to remove duplicate skills where two or more Lightcast skills mapped to one ASC skill within an occupation (for example, Lightcast's "Facebook" and "Twitter" skills were both mapped to one Australian Skills Classification skill, i.e. "Social media and web publishing software").

After the matching, a skill is considered trending for a given occupation if it has been mentioned in job advertisements for that specific occupation more frequently over the past five years. This is calculated by measuring the number of job ads that require a particular skill as a proportion of all jobs advertised for that occupation each year over five years. By contrast, a skill is considered emerging for an occupation if it has only emerged in job advertisements for that particular occupation in the last five years.

### **Application**

At the time of writing, JSA big data analysis based on Lightcast data and the Australian Skill Classification does not focus on assessing specifically green skills needs. This is partly because, albeit a useful source of timely information, Lightcast data also comes with a number of drawbacks. These include: (i) the sub-optimal representativeness of the data, as job advertisements are over-represented by professional jobs whereas many green jobs are in sectors such as forestry, agriculture and manufacturing; (ii) the lack of labelled data on what constitutes a green skill; (iii) the need to validate machine learning estimates through experts' work. Nevertheless, JSA uses Lightcast data to identify trending and emerging skills for occupations in the Australian Skills Classification. For example, Solar Installers are a specialisation of the occupation Electrician (General) which has a trending skill of Enterprise resource planning ERP software (used to provide a central data source for organisational information, and enable a variety of business functions including HR, other resource management, reporting, and financial management).

## **Case study 2: Identifying clean growth jobs in the United Kingdom**

In the United Kingdom, the Department of Business, Energy and Industrial Strategy (BEIS) adopts a big data approach as one way to monitor the development of green jobs and skills. Exploiting data from Lightcast, they apply machine learning algorithms to produce granular, real-time insights about the increase of “clean growth jobs” in the United Kingdom. In practice, this approach consists of an analysis of web-scraped OJVs across all online job postings in the United Kingdom, based on a keyword filtering approach and a machine learning process. In this way, BEIS traces and monitors job growth in specific low-carbon or net-zero sectors, occupations, skills, job roles and geographic areas. This information is then fed back to the central government and used as an empirical basis to inform policy making.

The big data approach of BEIS is framed by the government's ‘Ten Point Plan for a Green Industrial Revolution’ (Government of the United Kingdom, 2020<sup>[15]</sup>). With the goal of putting the United Kingdom at the forefront of global markets for clean technology, this strategic document outlines the ten priority industries for the green transition, namely: (1) advancing offshore wind, (2) driving the growth of low carbon hydrogen, (3) delivering new and advanced nuclear power, (4) accelerating the shift to zero emission vehicles, (5) green public transport, cycling and walking, (6) jet zero and green ships, (7) greener buildings, (8) investing in carbon capture, usage and storage, (9) protecting our natural environment, and (10) green finance and innovation. The ‘Ten Point Plan for a Green Industrial Revolution’ expects that up to 250 000 jobs will be created and supported in these industries by 2030, and a big data analysis is undertaken to provide insights into the type and quantity of jobs in each of these industries.

### **Definitions and methods**

This approach adopts a definition of “clean growth job” based on the definitions set out in the Low Carbon and Renewable Energy Economy survey (LCREE) by the Office for National Statistics (Office for National Statistics, 2022<sup>[16]</sup>). A job is considered a clean growth job if: (i) a part of the role is related to “clean growth activities” (e.g. a technician who knows how to install a heat pump), or (ii) the role is located within a “clean growth company” (i.e. a company that operates in one of the sectors identified in the Ten Point Plan) and plays an active part in the company's activity (e.g. an accountant for a wind farm). The following table

provides concrete examples of how different online vacancies would be classified according to a keyword filter and a labelling method (see sub-section on methods and data below).

Table 5.1. Classification of job advertisements as “clean growth jobs”

Vacancy text	Keyword filter	Label	Notes
<i>‘... you will use your accounting skills to support development of our wind energy business’</i>	1	1	Counted based upon LCREE approach – in a clean growth sector business
<i>‘Wind turbine engineer required’</i>	1	1	Obvious clean growth sector job
<i>‘As a Blue Wind engineer you will support our advanced telecoms customers’</i>	1	0	These companies could supply furniture or telecommunication services to wind sector or other green companies, however, they are outside of the LCREE definition and not detectable as “clean growth jobs” from the vacancy text.
<i>‘you will work for this major supplier of office furniture’</i>	0	N/A	

Source: Information provided by the UK Department of Business, Energy and Industrial Strategy (BEIS).

As a result, the primary classification of “clean growth jobs” is done on an industry level, based on the prioritisation of the UK Ten Point Plan. BEIS also provides estimates according to occupations, skills, job roles or geographic areas, although this is not at the core of its analysis. Growth in demand for clean growth skills, for example, is extracted based on the skills taxonomy by Lightcast. This taxonomy currently consists of 17 000 individual labels which are organised into skill clusters (groupings of similar skills) and into skill types – specialised, software, and baseline skills (Burning Glass Technologies, 2019<sup>[17]</sup>).

Data on OJVs in the United Kingdom is available since 2013 and the database contains over 65 million job vacancies, covering the near universe of all OJVs. BEIS analyses the data in two main ways: (1) applying a keyword filter and (2) using a machine learning process. To identify potential “clean growth jobs”, BEIS produces a list of keywords associated with each priority industry covered by the UK Ten Point Plan. These keywords are selected by labour market experts. The text in Burning Glass job postings is then searched for mentions of these keywords and a data sample is obtained and mapped to the corresponding industry. It is worth stressing two aspects directly stemming from this approach. First, the same job posting may be assigned to more than one industry if the text contains multiple keywords. Second, some “clean growth jobs” can remain undetected as the filter may not fully cover all the relevant keywords for each industry.

In addition, Machine Learning techniques are used to analyse four industries that are identified as priority industries in the Ten Point Plan by the UK Government (Table 5.2).

Table 5.2. Methodology applied by industry

Industry	Keyword search	Machine learning
Wind	✓	✓
Hydrogen	✓	
New and Advanced Nuclear Power	✓	✓
Electric Vehicles	✓	✓
Green public transport and cycling	✓	
Jet zero and green ships	✓	
Heat and Buildings	✓	✓
Carbon Capture, Storage and Usage	✓	
Protecting the Environment	✓	
Green Finance	✓	

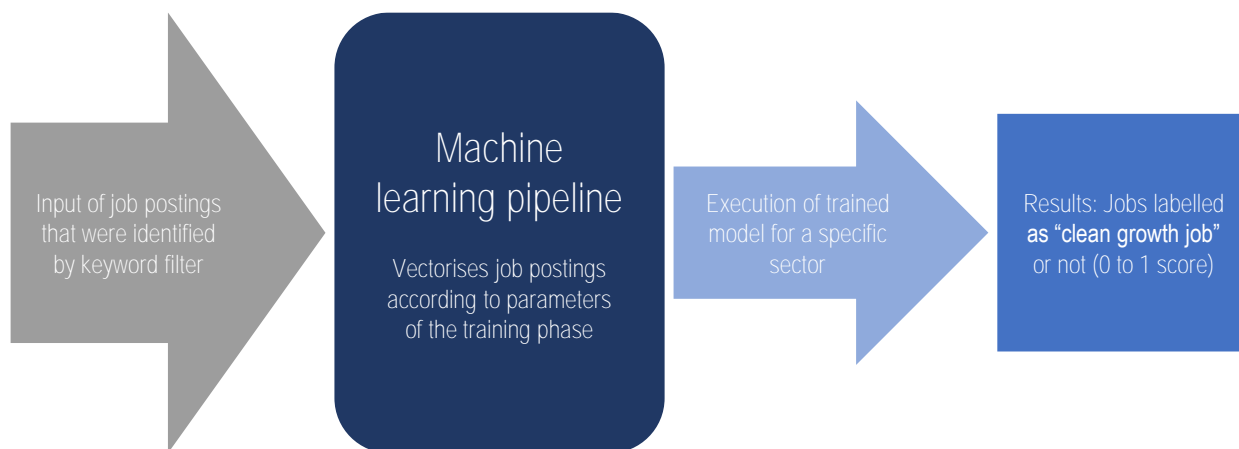
Source: Information obtained from the UK Department of Business, Energy and Industrial Strategy (BEIS).



The Machine Learning process is time-consuming and requires the development of a training dataset where a large sample of jobs from each industry are manually labelled to “train” and refine a machine-learning model. Experts provide labels for their sector, which encompasses a selection of relevant, “clean growth jobs”. A portion of these labels are then fed into a machine learning pipeline, to “train” an algorithm that produces a model which predicts whether the job is a clean growth job for that sector or not. This means that new text from job postings is vectorised by the algorithm (a process that allows the algorithm to operate on a large set of values at the same time), assigned a 0 to 1 score, and designated as a relevant “clean growth job” or not based on the parameters provided by experts. The remaining portion of jobs labelled by experts is then used to test each model’s performance and determine a threshold to increase accuracy.

The model (and corresponding threshold) is then applied to the full population of “clean growth jobs” to obtain a smaller but better-defined sample of relevant job postings. After this training phase, a similar process is followed to execute the trained machine learning algorithm (Figure 5.2).

Figure 5.2. Machine learning process applied by BEIS



Note: Authors' elaboration.

### **Application**

This big data approach by BEIS to identify green jobs is used to regularly provide evidence and insights to the central government. Findings on the geographical distribution of emerging “clean growth jobs”, for example, have high political relevance and may feed into policy making. Findings can also be shared with the Green Jobs Delivery Group, a forum between government and industry leaders (Government of the United Kingdom, 2022<sup>[18]</sup>).

The big data analysis by BEIS is currently not used to forecast the trajectory of “clean growth jobs” in the future. In addition, the approach focuses specifically on growth in “clean growth” occupations rather than green skills. Although the keyword filter is applied to tasks, activities and skills, these keywords are then used to label “clean growth jobs”. A skills summary is currently provided for each priority industry as one of a number of elements of the reporting.

### **Case study 3: Defining green skills and knowledge concepts in ESCO**

To support job mobility across Europe and a more integrated and efficient labour market, the European Classification of Occupations, Skills and Competences (ESCO) provides a dictionary of the skills and

occupations relevant to the European labour market and adult learning landscape that can be used by stakeholders to share a “common language”.<sup>3</sup> This taxonomy is very detailed, providing descriptions of 3 008 occupations and 13 890 skills linked to these occupations, and have therefore played a key role in shaping European-level policies on skills anticipation, job mobility and adult learning (European Commission, 2022<sub>[19]</sub>).

In 2021, the ESCO Secretariat conducted novel research to build a taxonomy of skills for the green transition, as part of the action plan of the European Skills Agenda, a five-year plan published by the Directorate-General for Employment, Social Affairs and Inclusion of the European Commission to strengthen sustainable competitiveness, ensure social fairness and build resilience. In particular, to support progress on the 2019 European Green Deal, the ESCO team set out to identify what skills and knowledge concepts are closely related to green activities, i.e. activities that reduce environmental degradation. To achieve this ambitious goal, big data analysis through machine learning (ML) techniques was conducted (European Commission, 2021<sub>[20]</sub>).

### **Definitions and methods**

The ESCO team created a dataset to train the machine learning algorithm to identify and classify skills as green, brown, and white. This training dataset was composed of about 4 800 strings of text (sentences and short definitions) extracted from over 30 European and international sources, describing activities as either environmentally sustainable, polluting, or none of the two. Text strings had been selected based on source reliability and their similarity to the structure and description of ESCO skills. The sources from which the text was collected include: 1) standard jobs and skills classifications, such as the EU taxonomy for sustainable activities, the O\*NET skills taxonomy and the classification by the French *Observatoire national des emplois et des métiers de l'économie verte* (ONEMEV); 2) online job vacancies, such as Indeed; 3) European or national legislation; and 4) related reports from international organisations, such as the OECD, ILO and UNIDO.

Strings of text are then distinguished as brown (400 elements), white (2 100 elements), and green (2 300 elements). For example, ‘production of electricity by coal’ was classified as a brown element, while ‘cogeneration of heat/cool and power from geothermal energy’ as a green element (Table 5.3). The ‘white skills’ label is added to account for skills that are difficult to clearly classify as green or brown.

Table 5.3. The classification of text strings in the ESCO training dataset

	ESCO definition	Example of string for training data (source)
Brown	knowledge and skills which increase the negative impact of human activity on the environment	Production of electricity by coal (ILO “Skills for a Greener Future”)
White	knowledge and skills which do not increase nor reduce the negative impact of human activity on the environment	Test computer or software performance (Australian Skills Classification)
Green	knowledge and skills which reduce the negative impact of human activity on the environment.	Cogeneration of heat/cool and power from geothermal energy (EU Taxonomy for Sustainable Activities)

Source: European Commission (2021<sub>[20]</sub>).

The labelling process used in ESCO follows a 3-step methodology, which combines human labelling and validation, and the use of machine learning algorithms. Firstly, skills and knowledge concepts are manually labelled by ESCO experts based on the definition of green skills suggested by Cedefop.<sup>4</sup> This manual labelling was conducted by comparing green skills definitions and the description of each skill from ESCO v1.1.

In the second step, a machine learning classifier was applied to classify green, brown and white skills among all the ESCO skills. Using the training dataset, the classifier was built using a pre-trained ML

technique for natural language processing developed by Google called the Bidirectional Encoder Representations from Transformers (BERT) model, based on the Python programming language. The results of the classifier analysis yielded the likelihood of each skill and knowledge concept being green (European Commission, 2021<sup>[20]</sup>).

Lastly, the classification of skills developed through the ML classifier was reviewed and validated through a comparison with manually labelled results. This verification focused on minimising the possibility of 'false positive' and 'false negative'. The final round of validation follows the following rules:

- If a concept is labelled as green by the two methods (manual and ML classification), it is automatically accepted as green;
- If a concept is labelled as non-green by the two methods, it is automatically accepted as non-green;
- If a concept is labelled as green by only one of the two methods, it is revised.

As a result, a total of 571 ESCO skills and knowledge concepts were labelled as green, including 381 skills, 185 knowledge concepts, and 5 transversal skills. ESCO also provides additional information such as the essential and optional relationship between each skill and occupation.

### *Application*

ESCO's research on green skills is not technically an approach to SAA, as it focuses on identifying and categorising those ESCO skills and knowledge concepts that are relevant for the green transition, without undertaking a fully-fledged assessment of current skills needs or a forecasting of future skills for the green transition. Yet, ESCO's research identifies the skills that are currently the most prominent (therefore, expected to increase in demand) in relation to the green economy and green growth, and does so through the analysis of vast amount of data sources.

Policy makers in European countries could therefore build on the ESCO green skills taxonomy to construct their skills assessment and anticipation exercises and ensure the sufficient provision of training opportunities for these skills (ILO, 2015<sup>[21]</sup>). In this respect, ESCO encourages the public dissemination of their results and provides detailed information to guide those interested in using the ESCO green skills classification. Potential users include not only researchers, but also public employment services and education and training providers – as it offers a rich source of information to design training curricula around green skills (European Commission, 2021<sup>[20]</sup>).

## Global insights

Identifying the links between green jobs and training using a linguistic approach by O\*NET

In 2022, O\*NET, which provides a comprehensive taxonomy for occupations and skills in the American labour market, produced a novel big data analysis extracting green occupations and related training programmes using a linguistic approach algorithm similar to the efforts undertaken by ESCO. In particular, O\*NET extracted 72 keywords on green topics from previous green-related research (National Center for O\*NET Development, 2009<sup>[22]</sup>), and used these keywords as input in a machine learning algorithm to identify a range of 5 to 36 green-related occupations for each topic. Subsequently, a similar algorithm is used to identify which training courses included in the 2020 Classification of Instructional Programs produced by the Department of Education are related to each green topic. On average, 16 training programmes are connected to each topic. The full results are released on the O\*NET webpage so that they can inform individuals who wish to search for careers and training related to green jobs and skills (National Center for O\*NET Development, 2022<sup>[23]</sup>).

## Case study 4: Using LinkedIn’s Economic Graph to assess green skills

The popular social network LinkedIn houses one of the world’s largest database of professional profiles and vacancies. The platform has more than 850 million members worldwide that input their professional profiles in the database, including data on skill demand through vacancies and skill supply through member profiles. The LinkedIn Economic Graph team has been tasked with using the company’s data to carry out skills assessment in labour market areas that are particularly relevant to the new world of work.

In 2022, the team published the *Global Green Skills Report 2022*, a report that outlines the important role of human capital in greening the economy (LinkedIn, 2022<sup>[24]</sup>). The study features analysis on green jobs and green skills, how people’s skills profiles are changing, how demand for green skills is evolving, and assessments on whether the green transition is just and inclusive. The report provides analysis at the country and sector level and measures gaps in green skills for specific socio-economic groups.

### Definitions and methods

LinkedIn’s approach to classify green skills is based on two initial qualitative definitions: the definition of green projects and a preliminary definition of green skills.

- First, green projects are defined as those that involve a focus on one or more of 12 Green Activities.<sup>5</sup> The list is based both on internal analysis – such as the 2019 Green Economy analysis produced in-house by the LinkedIn Economic Graph research and Insights team – and external taxonomies – namely, the definitions of green jobs and green goods and services produced by the Bureau of Labour Statistics of the United States through their O\*NET Resource Center. Essentially, a green projects can be understood as a set of (economic) activities that are at least partially green.
- The second qualitative definition that feeds into the analysis is a preliminary list of green skills. This list is compiled using four sources of information: LinkedIn top skills (most cited skills by members in their profiles), inputs from the Economic Graph Team, interviews with industry experts, and ESCO.

This preliminary list on green skills is fed into a machine learning algorithm to identify which skills in the LinkedIn database can be labelled ‘green’. The database draws mainly on information about skills found in the individual profiles of users, located under the ‘Skills’ section of a member profile and in the free text areas of the profile. Through the process of identifying green skills in the data, the Economic Graph team adds additional filters to capture relevant skills. The new list of green skills is then classified into four categories outlined in Table 5.4.

Table 5.4. Green skills classification

Green Skill Classification	Description
Green Skills	Clearly associated with “green” occupations
Ambivalent Skills	Utilised in both the Green Economy and elsewhere
Adjacent Skills	Tangentially associated with the Green Economy
Not Green Skills	Unassociated with Green Economy clearly, partially, or tangentially

Source: LinkedIn Economic Graph (2022<sup>[25]</sup>), <https://economicgraph.linkedin.com/data-for-impact>.

The green projects (i.e. projects that comprise of one or more of the 12 Green Activities) and green skills classification serve as inputs into the classification of what LinkedIn calls “green occupational titles” (Table 5.5). This terminology is used to identify which jobs are considered green, greening, or potential greening. LinkedIn essentially first identifies green skills and green projects, and then defines green occupational titles as those that contain green skills or green projects. This classification of green occupations can be used to assess the increase in green jobs in labour markets, but also shows how fast are green jobs growing compared to non-green jobs on a country and sector level.

Table 5.5. Green Occupational Classification

Green Titles	Occupations that usually work with Green Projects	or	Require Green Skills
Greening Titles	Occupations that sometimes work with Green Projects	and	Typically require some level of Green Skills
Greening Potential Titles	Occupations that occasionally work with Green Projects	and/or	May require some level of Green Skills

Source: LinkedIn Economic Graph (2022<sup>[25]</sup>), <https://economicgraph.linkedin.com/data-for-impact>.

The green skills classification is also used to quantify skill intensity – the extent to which different entities (e.g. countries, industries, and occupations) use these skills. This analysis relies on entities’ “most characteristic skills”, called a skills genome. The skills genome is an ordered list of the ~50 skills reported with most disproportionate frequency by members in any given entity. An algorithm called TF-IDF ranks each skill on the basis of i) how frequently it is reported by a member in the entity (TF), and ii) the logarithmic inverse entity frequency of the skill across a set of entities (IDF), which indicates how common or rare a skill is in the entire entity set. The more unique a skill is to a given entity, the more likely it is to be in the skills genome. Green skill intensity is then measured by assessing what share of the skills genome is comprised of green skills. Green skill intensity was used to help differentiate greening and green potential titles, and also track the flow of labour from greening, green potential and non-green jobs into green jobs.

### **Application**

The project was initiated prior to the 2021 United National Climate Change Conference, as the company saw a need by policy makers to have more information and analysis on the labour market changes due to the green transition. The report provides a global overview of skills trends along with some sector and country-specific analyses, and includes action plans for policy makers, business leaders, and the global workforce. Selected green data are available for multilateral institutions – such as the World Bank, the International Monetary Fund, Inter-American Development Bank and OECD – through LinkedIn’s “Data for Impact” programme (LinkedIn, 2022<sup>[26]</sup>). Currently, data-sharing with national governments is limited and takes place on request.

## References

- Bauer, S. et al. (2021), *Grüne Karrieren - Berufe und Branchen mit Green-Economy-Relevanz*, [10]  
[https://umweltbundesamt.de/sites/default/files/medien/479/publikationen/uiib\\_11-2021\\_gruene\\_karrieren.pdf](https://umweltbundesamt.de/sites/default/files/medien/479/publikationen/uiib_11-2021_gruene_karrieren.pdf).
- Börner, K. et al. (2018), “Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 115/50, pp. 12630-12637, [3]  
[https://doi.org/10.1073/PNAS.1804247115/SUPPL\\_FILE/PNAS.1804247115.SAPP.PDF](https://doi.org/10.1073/PNAS.1804247115/SUPPL_FILE/PNAS.1804247115.SAPP.PDF).
- Burning Glass Technologies (2019), “Mapping the Genome of Jobs. The Burning Glass skills taxonomy”, *Proceedings of the National Academy of Sciences*, [17]  
<http://hdl.voced.edu.au/10707/520491> (accessed on 7 October 2022).
- Cammeraat, E. and M. Squicciarini (2021), “Burning Glass Technologies’ data use in policy-relevant analysis: An occupation-level assessment”, *OECD Science, Technology and Industry Working Papers*, No. 2021/05, OECD Publishing, Paris, <https://doi.org/10.1787/cd75c3e7-en>. [8]
- Cedefop (2019), *Online job vacancies and skills analysis: A Cedefop pan-European approach*, [7]  
[https://cedefop.europa.eu/files/4172\\_en.pdf](https://cedefop.europa.eu/files/4172_en.pdf).
- Cedefop (2012), *Green skills and environmental awareness in vocational education and training*, [27]  
[https://cedefop.europa.eu/files/5524\\_en.pdf](https://cedefop.europa.eu/files/5524_en.pdf).
- Chen, H., R. Chiang and V. Storey (2012), “Business intelligence and analytics: From big data to big impact”, *MIS Quarterly: Management Information Systems*, Vol. 36/4, pp. 1165-1188, [2]  
<https://doi.org/10.2307/41703503>.
- European Commission (2022), *What is ESCO?*, <https://esco.ec.europa.eu/en/about-esco/what-esco> (accessed on 20 January 2023). [19]
- European Commission (2021), *Green Skills and Knowledge Concepts: Labelling the ESCO classification*, <https://esco.ec.europa.eu/en/publication/green-skills-and-knowledge-concepts-labelling-esco-classification> (accessed on 15 January 2023). [20]
- Government of the United Kingdom (2022), *Green jobs delivery steps up a gear*, [18]  
<https://gov.uk/government/news/green-jobs-delivery-steps-up-a-gear> (accessed on 7 October 2022).
- Government of the United Kingdom (2020), *The Ten Point Plan for a Green Industrial Revolution*, <https://gov.uk/government/publications/the-ten-point-plan-for-a-green-industrial-revolution> (accessed on 7 October 2022). [15]
- Hershbein, B. and L. Kahn (2018), “Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings”, *American Economic Review*, Vol. 108/7, pp. 1737-72, <https://doi.org/10.1257/AER.20161570>. [4]
- ILO (2020), *The feasibility of using big data in anticipating and matching skills needs*, [6]  
[https://www.ilo.org/wcmsp5/groups/public/---ed\\_emp/---emp\\_ent/documents/publication/wcms\\_759330.pdf](https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_ent/documents/publication/wcms_759330.pdf).

- ILO (2015), *Anticipating skill needs for green jobs: A practical guide*, [21]  
[https://ilo.org/wcmsp5/groups/public/---ed\\_emp/---ifp\\_skills/documents/publication/wcms\\_564692.pdf](https://ilo.org/wcmsp5/groups/public/---ed_emp/---ifp_skills/documents/publication/wcms_564692.pdf).
- JobTech (2022), *Nytt dataset: JobSearch Trends | Jobtech*, [28]  
<https://jobtechdev.se/sv/nyheter/jobsearchtrends> (accessed on 14 October 2022).
- Lassébie, J. et al. (2021), “Speaking the same language: A machine learning approach to classify skills in Burning Glass Technologies data”, *OECD Social, Employment and Migration Working Papers*, No. 263, OECD Publishing, Paris, <https://doi.org/10.1787/adb03746-en>. [11]
- LinkedIn (2022), *Data for Impact*, <https://economicgraph.linkedin.com/data-for-impact> (accessed on 28 October 2022). [26]
- LinkedIn (2022), *Global Green Skills Report 2022*, [24]  
<https://economicgraph.linkedin.com/research/global-green-skills-report> (accessed on 11 August 2022).
- LinkedIn Economic Graph (2022), *LinkedIn Data Available through the Development Data Partnership: Data Explainer*, <https://economicgraph.linkedin.com/data-for-impact> (accessed on 10 February 2023). [25]
- Modestino, A., D. Shoag and J. Balance (2020), “Upskilling: Do Employers Demand Greater Skill When Workers Are Plentiful?”, *The Review of Economics and Statistics*, Vol. 102/4, pp. 793-805, [https://doi.org/10.1162/REST\\_A\\_00835](https://doi.org/10.1162/REST_A_00835). [5]
- National Center for O\*NET Development (2022), *Green Topics: Identifying Linkages to*, [23]  
[https://www.onetcenter.org/reports/Green\\_Topics.html](https://www.onetcenter.org/reports/Green_Topics.html) (accessed on 16 November 2022).
- National Center for O\*NET Development (2009), *Greening of the World of Work: Implications for O\*NET-SOC and New and Emerging Occupations*, [22]  
<https://www.onetcenter.org/reports/Green.html> (accessed on 12 October 2022).
- National Skills Commission (2021), *The state of Australia’s skills 2021*, [14]  
[https://www.nationalskillscommission.gov.au/sites/default/files/2022-03/2021%20State%20of%20Australia%27s%20Skills\\_0.pdf](https://www.nationalskillscommission.gov.au/sites/default/files/2022-03/2021%20State%20of%20Australia%27s%20Skills_0.pdf) (accessed on 23 September 2022).
- OECD (2022), *Skills for the Digital Transition: Assessing Recent Trends Using Big Data*, OECD Publishing, Paris, <https://doi.org/10.1787/38c36777-en>. [13]
- OECD (2021), *Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges and Implications for Policy Makers*, <https://www.oecd.org/finance/artificial-intelligence-machine-learning-big-data-in-finance.htm> (accessed on 27 October 2022). [12]
- OECD (2016), *Big Data: Bringing competition policy to the digital era*, OECD Publishing, Paris, [1]  
[https://one.oecd.org/document/DAF/COMP\(2016\)14/en/pdf](https://one.oecd.org/document/DAF/COMP(2016)14/en/pdf) (accessed on 27 October 2022).
- Office for National Statistics (2022), *Low Carbon and Renewable Energy Economy Survey - Office for National Statistics*, [16]  
<https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/lowcarbonandrenewableenergyeconomysurvey> (accessed on 7 October 2022).

Song, K. et al. (2021), "Matching and Mismatching of Green Jobs: A Big Data Analysis", *Sustainability*, Vol. 13/7, <https://doi.org/10.3390/su13074074>. [9]

## Notes

<sup>1</sup> OJVs are not only useful for SAA exercises. For example, in 2022, Sweden's Public Employment Service and JobTech Development released the *JobSearch Trends* dataset, which analyses job postings to understand what jobseekers look for when using the PES online platform (JobTech, 2022<sup>[28]</sup>).

<sup>2</sup> The motivation behind creating skill clusters is to deconstruct the traditional notion of occupational classifications and qualifications to get a more nuanced understanding of skill needs and their transferability (National Skills Commission, 2021<sup>[14]</sup>).

<sup>3</sup> ESCO is a European Commission project, run by the Directorate General for Employment, Social Affairs and Inclusion (DG EMPL).

<sup>4</sup> Cedefop (2012<sup>[27]</sup>) defines green skills as "the knowledge, abilities, values and attitudes needed to live in, develop and support a society which reduces the impact of human activity on the environment".

<sup>5</sup> The Green Activity Categories (and sub-categories) are as follows: pollution prevention, waste prevention, renewable energy generation, energy management, environmental remediation (including waste management, water quality management, environmental restoration), ecosystem management (including natural resource management, erosion control, biodiversity conservation, water resource management, climate change mitigation and climate change adaptation), sustainable education, sustainability research, environmental auditing (including environmental impact assessment and carbon accounting), environmental policy (including energy law and environmental law), sustainable procurement and environmental finance.





**From:**

## **Assessing and Anticipating Skills for the Green Transition**

Unlocking Talent for a Sustainable Future

**Access the complete publication at:**

<https://doi.org/10.1787/28fa0bb5-en>

### **Please cite this chapter as:**

OECD (2023), “Exploiting big data to measure skills needs for the green transition”, in *Assessing and Anticipating Skills for the Green Transition: Unlocking Talent for a Sustainable Future*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/521f19b6-en>

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.