Big Data for Labour Market Intelligence

Capacity development programme 2024

Workshop 1: Big Data for LMIS: Digital and Green Analysis
Session 2

Big Data for LMIS - for skills demand analysis. Main concepts and methodological framework. Database. Dashboards to navigate the data.

Speaker: Mauro Pelucchi, Anna Clara Gatti
14/02/2024
What a Billion Jobs Can Tell Us
vs. Traditional Labour Market Information

Because of that speed and detail, the data are more actionable.

Greater speed, granularity compared to survey-based instruments.

In addition, job postings provide insight into real-world skill demands.
New source of data

Web Data ingestion is the process of obtaining and importing data from web portals and storing in a database.
What are the impacts on the Labour Market?

Some of the 2024 Key Trends

Talent shortage
- Digitalization of professions
- Relevance of Soft skills
- New professions and skills emerging
- Green economy and sustainability
- Artificial intelligence
Using Online Job Postings for many reasons:

- Up to date
- Detailed
- Adherent to reality

How observe these impacts?
A Complement, not a Replacement

Traditional survey-based sources will continue to be the best macroeconomic tools

But will also be slower and more expensive

Big Data is best for understanding changes in detail
<table>
<thead>
<tr>
<th>Type of data</th>
<th>Years of data</th>
<th>Ease of time series analyses</th>
<th>Data representative- ness</th>
<th>Compatibililty across economies</th>
<th>Real-time data access</th>
<th>Regular taxonomy (classification) updates</th>
<th>Data granularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional labor market data</td>
<td>~50</td>
<td>High</td>
<td>Apply statistical sampling methods and weights</td>
<td>✔</td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Big data</td>
<td>~10</td>
<td>Medium</td>
<td>Captures digitized labor market; can benchmark against public data to gain insight</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>High</td>
</tr>
</tbody>
</table>
Topics

- Value of Big Data for LMI
- Main concepts and methodological framework
- Skills demand analysis
- Dashboards to navigate the data
**Motivations**

**Labour market changes**

<table>
<thead>
<tr>
<th>Constantly changing labour market:</th>
<th>Improving skills needs analysis: the foundation of skills development</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Skills Evolution</td>
<td>- Up-to-date information</td>
</tr>
<tr>
<td>- New Emerging Occupations</td>
<td>- Information aligned with market demands</td>
</tr>
<tr>
<td>- Job Automatisation</td>
<td>- Prediction can be done to anticipate trends</td>
</tr>
<tr>
<td>- Mobility</td>
<td></td>
</tr>
<tr>
<td>- Remote Working</td>
<td></td>
</tr>
<tr>
<td>- Artificial Intelligence</td>
<td></td>
</tr>
</tbody>
</table>
Example of applications: Skills Change Index

Index of the Top 25 European Occupations by Skills Change

- Software Developers
- Web & Multimedia Developers
- Systems Analysts
- Graphic & Multimedia Designers
- Advertising & Marketing Professionals
- Systems Administrators
- Computer Network & Systems Technicians
- Industrial & Production Engineers
- Database Designers & Administrators
- Electronics Engineers
- Sales & Marketing Managers
- Software & Applications Developers & Analysts
- Management & Organisation Analysts
- Accountants
- Building Architects
- Mechanical Engineering Technicians
- Business Services & Administration Managers
- Biologists, Botanists, Zoologists & Related Professionals
- Commercial Sales Representatives
- Mechanical Engineers
- Research & Development Managers
- ICT User Support Technicians
- ICT Services Managers
- Manufacturing Managers
- Database & Network Professionals
**JUNIOR DATA SCIENTIST & ANALYST PLACEMENT**

London • Hybrid remote

Internship

You must create an Indeed account before continuing to the company website to apply

Apply on company site

---

As a Data Scientist at [Company Name], you will join the rapidly developing data team, who are responsible for measurement solutions and modelling expertise. Help us provide our clients with the true value of their media investment, help them understand and drive growth, and turn these insights into the planning cycle through our brand and integrated tech stack.

We are looking for inquisitive, articulate, numerate and above-all, enthusiastic placement students to support the wider team in delivering these critical insights and building on the capabilities of our product.

You will be part of a close-knit and friendly team who share results and celebrate success together.

[Company Name] is a media agency that is made differently. We're purpose-led, data-driven and proudly independent. Our independence means we can focus 100% on doing the right thing to secure success for our clients and our brilliant people. We are trusted to deliver that success for some of the UK's most ambitious and complex organisations, including SunLife, Guide Dogs, RNLI, Laithwaites and National Trust.

**WHAT YOU'LL BE DOING**

- Work closely with business to identify issues that can be resolved using data solutions effectively for decision making
- Machine learning tools and statistical methods to solve complex problems
- Build algorithms and design experiments to merge, manage, interrogate, and extract data to supply tailored reports to colleagues, clients and wider areas in company
- Support the account management and planning teams across all facets of campaign measurement across media channels
- Develop automated data processes using Python/R
- Ability to organise a variety of large data sets
- Undertake regular analysis and reporting for retained clients
- Maintain clear and coherent communication, both verbal and written, to understand data needs and report results
- Working with the Datalab team and other senior business stakeholders to develop analytical propositions

**THE SKILLS YOU WILL BRING**

- Highly numerate undergraduate studying a relevant degree in mathematics, statistics, econometrics or computer science
- Pre-requisite skills: Strong Excel and MS Office usage
- Experience of coding in Python, R or SQL
- Experience of data visualisation tools like Tableau/Qlik/Power-Bl/Google Data Studio would be useful but not essential
- Theoretical understanding of statistical techniques such as regression and developing confidence measures.
- Strong data manipulation skills and a keen eye for detail.
JUNIOR DATA SCIENTIST & ANALYST PLACEMENT

London • Hybrid remote
Internship

You must create an Indeed account before continuing to the company website to apply

Apply on company site

As a Data Scientist, you will join the rapidly developing data team, who are responsible for measurement solutions and modelling expertise. Help a diverse client-set understand the true value of their media investment, create compelling data stories on how to drive growth, and automate the insights into the planning cycle through their advance and integrated tech stack.

We are looking for inquisitive, articulate, numerate and above-all, enthusiastic placement students to support the wider team in delivering these critical insights and building on the capabilities of our product.

You will be part of a close-knit and friendly team who share results and celebrate success together.

We are a media agency that’s made differently. We’re purpose-led, data-driven and proudly independent. Our independence means we can focus 100% on doing the right thing to secure success for our clients and our brilliant people. We are trusted to deliver that success for some of the UK’s most ambitious and complex organisations, including Sunlife, Guide Dogs, RNLI, Lakhaites and National Trust.

WHAT YOU’LL BE DOING

- Work closely with business to identify issues that can be resolved using data solutions effectively for decision making
- Machine learning tools and statistical methods to solve complex problems
- Build algorithms and design experiments to merge, manage, interrogate, and extract data to supply tailored reports to colleagues, clients and wider areas in company
- Support the account management and planning teams across all facets of campaign measurement across media channels
- Develop automated data processes using Python/R
- Ability to organise a variety of large data sets
- Undertake regular analysis and reporting for retained clients
- Maintain clear and coherent communication, both verbal and written, to understand data needs and report results
- Working with the Datalab team and other senior business stakeholders to develop analytical propositions

THE SKILLS YOU WILL BRING

- Highly numerate undergraduate studying a relevant degree in mathematics, statistics, econometrics or computer science
- Pre-requisite skills: Strong Excel and MS Office usage
- Experience of coding in Python, R or SQL
- Experience of data visualisation tools like Tableau/Qlik/Power-BI/Google Data Studio would be useful but not essential
- Theoretical understanding of statistical techniques such as regression and developing confidence measures.
- Strong data manipulation skills and a keen eye for detail
- Good communication skills.

DESIRED SKILLS

- Inquisitive analytical mind with a strong desire to find things out.
Challenges

High interest
• Observe micro-level labour demand (with some caveats)

Retrieval and analytics
• High volume
• High frequency velocity
• Many formats variety, noise
• Duplications
• No control over reference population veracity
How do you (a human) classify a job posting in an occupation?

**Data Scientists**
- **big data**: 83%
- **graph**: 66%
- **algorithms**: 87%
- **develop**: 33%
- **python**: 99%
- **database**: 51%

**Business Intelligence Analyst**
- **big data**: 12%
- **graph**: 3%
- **algorithms**: -20%
- **develop**: 20%
- **python**: 23%
- **database**: 77%

**Statistician**
- **big data**: 12%
- **graph**: 33%
- **algorithms**: 51%
- **develop**: -53%
- **python**: 53%
- **database**: 51%

---

**Junior Data Scientist**
- BIP Solutions
- Glasgow
- Hybrid remote
- £25,000 - £35,000 a year - Full-time

**Job details**
- **Salary**: £25,000 - £35,000 a year
- **Job type**: Full-time

**Full Job Description**
Are you passionate about product analytics and technology? The BIP product team is looking for enthusiastic analytics candidates that are passionate about data and want to make an impact.

The BIP Data Science team handle large volumes of text data, application data and business operations data. Our team is focused on developing data visualisation dashboards, text-based machine learning solutions, evaluating and optimising search applications, and implementing recommender systems.

Candidates will work with product analysts and engineers to translate data into meaningful insights to enable data driven decision making and new feature development.

The junior data scientist role will have a primary focus on the development and evolution of data visualisation dashboards with a clear growth path to develop your wider data science skillset.

You will also be encouraged to be innovative and put forward ideas that shape what data science is within BIP and ultimately drive the adoption of it within the business.
Big data and new considerations

Volume: Terabytes to exabytes of existing data to process

Velocity: Streaming data, milliseconds to seconds to respond

Variety: Structured, unstructured, text, multimedia

Veracity: Uncertainty due to data inconsistency & incompleteness

Value: Data philanthropy to high value monetization

Visibility: Data is generally open to anyone. Which raises issues of privacy. Security and provenance
Why and how to use Cognitive Computing and Machine Learning?

90% of data is unstructured. Cognitive computing and machine learning help us to analyze and leverage on 100% of data: this techniques provide the tools to solve problems that couldn’t be solved before.
Labour Market Information and Intelligence

LMI is any **quantitative** or **qualitative facts**, analysis or **interpretation** about the past, present or future structure and workings of the labour market and the factors that influence it.

- Economic and labour market conditions
- Education, qualifications, training and skills
- Current and future demand and supply of labour and jobs
- Vacancies and recruitment
Big Data Analytics in Action

Sensors receive information from the field after the actuations.

- Sensors: Raw data from the web
- Models: Algorithms analyzing raw data to extract predictive information needed to take better decisions
- Policy Makers: Automatic or manual action that interact with the context
Main concepts and methodological framework
What are the emerging skills and competencies in the labour market?

Source systems  
Collecting & transform  
Storage  
Provisioning  
Application and analytics

How draw meaningful insights

Turning big noisy data on job postings into clear and actionable data points

Sourcing and scraping
Machines trawl across websites looking for things that look like job postings, and then taking the data – we avoid aggregators and seek original sources.

Parsing
Broken down into what machines identify as likely important information elements – job title, salary, company, location, body text – to give a structure to each posting.

Quality filtering
Some of the things that look like postings just aren’t (e.g. they’re training courses), and so we apply filters to take them out of our source.

Deduplication
Websites repeat the same postings! We deduplicate daily across whole database – using job title, skills, period, employer name, location.

Final dataset
One unique posting for each opening, and key information such as job title, occupation, location, employer, skills, pay extracted.
Identifying reliable online job boards

A data-driven approach of more relevant job posting sources

A Landscaping activity is performed to produce a list of sources (web portals) that are relevant for the Web Labour Market in a given country.

International Labour Market Experts validate this list, that will become the initial step of the LMI System.

Conventional and New data: speaking the same language

Data classification using AI and Taxonomy

- **Occupations**: ISCO - International Standard Classification of Occupations
- **Skills**: ESCO - European Skills, Competences, Qualifications and Occupation
- **Location**: NUTS and ISO
- **Educational Level**: ISCED 2011 - International Standard Classification of Education
- **Sector**: NACE - Statistical classification of economic activities
The role of AI in Lightcast

Job title = Data Scientist
Data Science Analyst (Remote) : Yelp 3.4 ★
Remote

Full-time

- A knack for communicating quantitative results to a broad audience through writing and compelling data visualization.

Posted 30+ days ago · More...

Biostatistician - Global Biostatistics : IQVIA 3.8 ★
Remote

Full-time

- Experience with data analysis/statistical software and packages (pandas/statsmodels/scikit-learn within Python, R, etc.).

Posted 30+ days ago · More...

Machine Learning Research Engineer : Evolution Artificial Intelligence
Remote

- You’ll be putting state of the art deep learning technology into production use.
- Designing, developing and rigorous testing of machine learning models.

Data Science Manager : Consortia
Remote

£120,000 a year · Permanent

- As Data Science Manager, you will lead and grow a group of world-class data scientists.

Senior Analyst, BI & Analytics : WorldRemit
Remote

Full-time

- Experience in an analytics or data science role.
- Strong SQL skills, able to handle large, complex data structures from multiple sources.

Posted 30+ days ago · More...
## Job Postings | Key dimensions

<table>
<thead>
<tr>
<th>Industry</th>
<th>Occupation</th>
<th>Type of employment</th>
<th>Geography</th>
<th>Time</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Job Title</th>
<th>Wage</th>
<th>Experience</th>
<th>Education required or preferred</th>
<th>Certifications</th>
<th>Employer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Skills demand analysis
# Identifying Future Skill Demands

**A Range of Lenses for Tracking Emerging Trends**

<table>
<thead>
<tr>
<th>#</th>
<th>Top IT Skills (Total postings)</th>
<th>Highest Paying IT Skills (Mean advertised salary)</th>
<th>Fastest Growing IT Skills (24 month projections)</th>
<th>Hardest to Fill IT Skills (Mean posting duration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>SQL</td>
<td>Zookeeper</td>
<td>TensorFlow</td>
<td>Public Cloud Security</td>
</tr>
<tr>
<td>2.</td>
<td>Java</td>
<td>TensorFlow</td>
<td>General Data Protection Regulation (GDPR)</td>
<td>Infrastructure as a Service (IaaS)</td>
</tr>
<tr>
<td>3.</td>
<td>JavaScript</td>
<td>Scala</td>
<td>Kubernetes</td>
<td>Cloud Technology Architecture</td>
</tr>
<tr>
<td>4.</td>
<td>Linux</td>
<td>AWS Redshift</td>
<td>Spring Boot</td>
<td>Cloud Infrastructure</td>
</tr>
<tr>
<td>5.</td>
<td>Python</td>
<td>AWS DynamoDB</td>
<td>Webpack</td>
<td>Ansible</td>
</tr>
<tr>
<td>6.</td>
<td>Data Analytics</td>
<td>Go Programming Language (Golang)</td>
<td>AWS Lambda</td>
<td>Apache Mesos</td>
</tr>
<tr>
<td>7.</td>
<td>Salesforce</td>
<td>Pig</td>
<td>Salesforce Lightning</td>
<td>Data Protection Planning</td>
</tr>
<tr>
<td>8.</td>
<td>C#</td>
<td>Apache Mesos</td>
<td>Redux</td>
<td>Work Breakdown Structure</td>
</tr>
<tr>
<td>9.</td>
<td>Scrum</td>
<td>AWS CloudFormation</td>
<td>Financial Microservices</td>
<td>Hadoop Cloudera</td>
</tr>
<tr>
<td>10.</td>
<td>C++</td>
<td>Deep Learning</td>
<td>Apache Kafka</td>
<td>OpenShift</td>
</tr>
<tr>
<td>esco_level_4_1</td>
<td>esco_level_4_2</td>
<td>Cosine</td>
<td>Motyka</td>
<td>Ruzicka</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Systems analysts</td>
<td>1.000000</td>
<td>0.500000</td>
<td>1.000000</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Software developers</td>
<td>0.221716</td>
<td>0.159765</td>
<td>0.190143</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Database and network professionals not elsewhere...</td>
<td>0.107499</td>
<td>0.135108</td>
<td>0.156214</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Engineering professionals not elsewhere classified</td>
<td>0.128942</td>
<td>0.127291</td>
<td>0.145858</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Information and communications technology user...</td>
<td>0.138873</td>
<td>0.120262</td>
<td>0.136702</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Web and multimedia developers</td>
<td>0.087979</td>
<td>0.108967</td>
<td>0.122293</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Research and development managers</td>
<td>0.090797</td>
<td>0.102387</td>
<td>0.114066</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Computer network and systems technicians</td>
<td>0.085416</td>
<td>0.084232</td>
<td>0.091980</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Sales and marketing managers</td>
<td>0.039574</td>
<td>0.073370</td>
<td>0.079180</td>
</tr>
<tr>
<td>Systems analysts</td>
<td>Information and communications technology service...</td>
<td>0.109083</td>
<td>0.071072</td>
<td>0.076509</td>
</tr>
</tbody>
</table>
LMI for Education

Artificial Intelligence Systems
(Tutte le nazioni)

Annulli per figure professionali associate al corso selezionato

Corso di Laurea: Artificial Intelligence Systems
Categoria del corso: Ingegneria Informatica
Classe di laurea: LM-32
Link alla scheda: URL

177.160
Skills Intelligence & Talent Manifesto

- New models for **skills projections** and **difficulty to fill occupations**

- The **future skill gaps** in the local market

- Talent Manifesto, a **public policy**, to attract talented individuals

- Enhance the region’s workforce by bridging the gap between demand and supply of skills

https://emiliaromagnainnodata.art-er.it/skills-intelligence-emilia-romagna/
The transition to a greener economy

INVESTMENTS MEANS INCREASED JOBS DEMAND AND CHANGING SKILLS NEEDS

Top 5 green skills in Germany in 2022

<table>
<thead>
<tr>
<th>Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renewable energy</td>
</tr>
<tr>
<td>Energy supply</td>
</tr>
<tr>
<td>Recycling</td>
</tr>
<tr>
<td>Waste management</td>
</tr>
<tr>
<td>Photovoltaics</td>
</tr>
</tbody>
</table>

Figure 1: Share of green postings by country, 2019 - 2022 (%)
AI disrupting the way we work?

AI RISING FAST

Source: Lightcast
Detailed datasets

8.5% of online

Digital economy and society (isoc) (Important notice)
- ICT usage in households and by individuals (isoc_i)
- ICT usage in enterprises (isoc_e)
- Digital skills (isoc_sk)
- ICT users (isoc_sku)
- ICT specialists (isoc_sks)
- ICT training (isoc_skt)
- Labour market demand for ICT specialists in online job advertisements – experimental statistics (isoc_sk_oja)
- ICT sector (isoc_se)
- Additional indicators on digitalisation (isoc_ad)

Labour market demand for ICT specialists in online job advertisements – experimental statistics (isoc_sk_oja)
- ICT sector (isoc_se)
- Additional indicators on digitalisation (isoc_ad)

Big data is the future of labor market information. Big data on the labor market can help augment traditional sources of data in providing real-time analyses, and could be especially useful in times of economic shocks. It allows for analyses such as calculating skill premia, understanding skill-adjacencies and creating career pathways, enabling reskilling and upskilling, and cataloguing emerging technology skills. This issues paper, which was made possible through a partnership with The Asia Foundation, conducts a mixed methods study of existing labor market data sources and compares them with big data available through digital platforms. It looks into integration of new data into labor market information systems, and assesses opportunities for bringing this data into public policymaking. This can open doors for policymakers to use big data to quickly respond to economic shocks (such as that caused by the COVID-19 pandemic), reskilling and upskilling workers, and matching job seekers with appropriate employers.
Skills for the Digital Transition

Assessing Recent Trends Using Big Data

This report presents the most recent trends in the labour market demand for digital professionals and skills, highlighting where bottlenecks are emerging and policy action is — and will be — needed to support individuals who aim to thrive in the digital transition. The report analyses a wide range of digital occupations and the associated skill and technology demands using a unique set of data collected from millions of job postings published online in Belgium, Canada, France, Germany, Italy, the Netherlands, the United Kingdom, the United States, Singapore and Spain. The evidence contained in this report is key for governments to design targeted retraining and upskilling policies, and for workers to fully benefit from the digital transition.
Dashboards to navigate the data
Inform decision makers and practitioners

Use billion job postings to inform on skills development

**Research**
- Understand how students (to look for a job) and employers (to hire) are using internet
- Analyse coverage of the country/region

**Develop ideas and plans**
- Identify dimensions, metrics and clusters
- Identify new emerging occupations and skills needs

**Test ideas and plans**
- Obtain feedback from employers, training providers, students
- Improve the data navigation paths

**Action**
- Apply metrics to decode billions of job postings in synthetic key indicators
Thanks!

Mauro Pelucchi - Head of Data Science -
mauro.pelucchi@lightcast.io

Anna Clara Gatti – Data Analyst –
anna.gatti@lightcast.io