NOTE

Seminars of the Big Data Knowledge Hub

AI-Machine Learning serving people and employment - profiling jobseekers

Notes of the online seminar promoted by the Big Data Knowledge Hub of the European Network on Regional Labour Market Monitoring (ENRLMM). June 30, 2022

The second of the Seminars of the Big Data Knowledge Hub took place on June 30, 2022. The aim of these series of seminars is to offer an opportunity to deepen the Network's knowledge on how to use Big Data for labour market research and consulting by presenting practical cases and demonstrations.

The project "AI-Machine Learning serving people and employment - profiling jobseekers" was presented by **Borja Pulido** (LANBIDE, Basque Country), and following comments were made by **Dorit Griga** (SECO, Switzerland). The open discussion counted with the participation of: **Christa Larsen** - IWAK Goethe University Frankfurt am Main, Germany), **Anne Lauringson** (OECD), **Ciprian Panzaru** (Universitatea de Vest din Timișoara, Romania), **Aleksandra Webb** (University of the West of Scotland, UK) and **Joanna Napierala** (Cedefop).

Introduction

Eugenia Atin (Speaker of the Big Data Working Group of the ENRLMM) after the initial greetings and thanks to the participants, contextualises this session in the work being done by the Big Data Working Group of the ENRLMM (European Network on Regional Labour Market Monitoring).

This is the second seminar of the series that we have named "Seminars of the Big Data Knowledge Hub". The Big Data Knowledge Hub https://bigdatahub.uvt.ro/ is a collaborative platform for mutual exchange and learning that was set up last year by the Big Data Working Group of the EN RLMM. The Knowledge Hub is the place where all the members of the EN RLMM can look for guidance when aiming to use big data in their labour market monitoring projects. It is an easy accessible source of information on the techniques used by other reference labour market observatories for a particular topic or challenge.

Christa Larsen (Founder and Coordinator of the ENRLMM), then presents the European Network on Regional Labour Market Monitoring (ENRLMM) http://regionallabourmarketmonitoring.net/, a network of around 400 members from 20 different countries. The network focusses on the process of monitoring regional and local labour markets and we are active in exploring what kind of data we can use at these levels. Since 2015 we are working on how to use big data from the internet for the purpose of labour market monitoring and the Big Data Working Group was established with the aim to collect examples on how Big Data can be used in local and regional labour market situations.

The annual topic of the Network is the relevance of Artificial Intelligence for regional and local labour market monitoring and in this sense, there has been an exchange among members and in September we will publish an Anthology and we will celebrate the annual meeting under this topic. So, today we have a very interesting presentation from a strategic point of view, that can be considered an appetizer for our Annual Meeting. The Conference will be held in Sardinia, at the EURISPES organization and will bring the state of the art on artificial intelligence as well as the perspective of the Public Employment Services, in relation to the activities they are carrying out related to the application of AI.

So, this second session is dedicated to a project that our partners from the Basque Country are conducting from Lanbide- the Public Employment Service. Borja Pulido will present today and describe the process used by the Lanbide-Basque Employment Service to profile jobseekers with the aim of making the public employment service more efficient in its mission to optimise the employability of users. He will explain how they have used AI and Machine Learning to do this profiling.

Presentation by Borja Pulido.

AI-Machine Learning serving people and employment - profiling jobseekers

The presentation of today is about a project that Lanbide (the Basque Employment Service) is currently developing. The tool has been deployed but it is not yet being used by the counsellors. It is a demonstration of how big data techniques can be used to profile jobseekers.

A Profiling process is, essentially, classifying or defining jobseekers according to their employability on the labour market. It could also be defined as a customised diagnostic tool to identify the risk levels of the individuals as regards their possibilities of returning to employment. The Profiling provides information on the employability of jobseekers. It allows the efficient management of active employment policies and

supports the professionals of the organisation, which results in a quality service provided to citizens.

The aim of the project is to develop a profiling methodology based on the following three types of profiling: expert profiling, quantitative profiling, and statistical profiling. The developed methodology can be applied in other organisations, as it has been shown that the current data analysis tools can be integrated in the business processes of an entity such as a PES or similar.

The main objective of the project is to model and classify the jobseekers to better know them. Once that segmentation is available, we can offer a proactive and personalised service, and make recommendations to improve the job prospects of our jobseekers. The implementation of this information system seeks to address four key challenges in the employment service:

- chronic long-term unemployment
- non-personalised service
- reactive service
- a low intermediation ratio

The first phase of the project was conducted with a very descriptive perspective, started from a fundamentally quantitative approach, where the emphasis was set on furthering knowledge of the segmentation of jobseekers in relation to their employability potential.

The second profiling had a more analytic component. The Lanbide Technical Office applied a second classification layer, based on the relationship between variables according to CHAID (Chi-squared Automatic Interaction Detection) decision trees, to the first pre-determined job-seeker segmentation. It was an initial attempt towards a more statistical profiling model.

We now turn to the third profiling phase. The aim is to build on the second phase, but automating the whole process. In this phase, the application of Artificial Intelligence is especially important, since it brings us different benefits:

- A dynamic model
- A model that is able to self-learn and adapt to the changing circumstances
- A model that is scalable and therefore allows the introduction of new information levels

This third phase seeks to exceed some of the shortcoming found in early attempts:

- Rigidity of models overly subject to the limits imposed on their design
- Excessive time between the start of the analysis and obtaining operating results.

Lanbide relied on big-data tools to overcome those constraints:

- they are more powerful in the processing of large amounts of data that grow exponentially when reused
- this type of techniques allows the system to be trained and learn while the data is analysed
- it provides information in easily interpretable formats that can be inserted in modules in real time to support Lanbide professionals in their work to design personalised employment guidance.

Methodology

The first step of the process is to calculate the Employability Indicator.

This is the variable on which the whole profiling model pivots. It provides an idea of the distance between the group or individual and the labour market, indicating a greater or lesser probability of being employed given the market's current characteristics and the variables associated to each group (or individual).

The historical data of the last 18 months are used to calculate this index and the following variables are taken into account:

- Number of days the jobseeker has been employed in the last 18 months (a)
- Number of days the jobseekers with the same occupation have been employed in the last 18 months (b)
- Number of days the jobseekers with the same qualification have been employed in the last 18 months (c)

A formula is applied (see the full presentation for more details) and the result is displayed in two different ways:

1. Numerical data: range between 0 and 100, where 0 is the worst figure and 100 is the best.
2. Group of indicators: the numerical indicators are grouped in quartiles (low/medium-low/medium-high/high employability index).

The following maxims help to better understand the concept:

- A jobseeker is more employable the more days they have been employed (stable job)
- A jobseeker is more employable if there is high employability of their peers in the target job sector (growth sector)

- A jobseeker is more employable if there is high employability of their peers in the completed education sector
- We consider that a jobseeker has full employability (employability index = 1, distance 0 to the market), if they have been employed as many days as possible during the period in question
- We consider the long-term unemployed to exhibit the greatest possible distance to the job market

Once the indicator has been prepared, the recommendations for improvement are based on the following groups:

1st level: 26 original groups are prepared and based on:

- Employment status of the jobseeker (employed or unemployed) and time in that employment status
- Age and academic background

2nd level: 6 sub-groups are prepared for each original group based on:

- Occupations requested by the jobseeker
- Proactivity in the offers, that is, how active the person is in the job search.

The preliminary analysis of the group and sub-group allow a set of personal routes to be defined to guide the jobseekers of low employability subgroups to their maximum employability subgroups, in other words, to improve their employability situation as effectively as possible. The optimum subgroup (highest employability index) of their group is sought for each jobseeker. The quality of an occupation is defined as its normalized average employability index, its range being from 0 (low quality) to 1 (high quality). For each applicant, the quality of all their occupations is calculated and the one with the highest value is taken.

The average employability indexes can be consulted for each subgroup disaggregated by province and sex, along with the optimum indicator for each subgroup, in other words, the highest index of each subgroup.

- The total number of jobseekers of each subgroup is also shown.
- This application allows us to consult the temporal analysis of the index (per months) and to apply various filters.
- However, the most important aspect of this tool is that it allows the information for each jobseeker to be displayed.
- The tool will show us the subgroup to which the jobseeker has belonged and belongs, and also the target optimum subgroup, along with their employability

index and the highest value of the current subgroup to which they belong. It will also show the transition rule, in other words, the recommendation that should be adopted by the user in order to move up to the optimum sub-group.

This approach of using the temporal development also allows us to see whether a specific jobseeker is following our recommendations

- If the jobseeker has been given a recommendation to move up to the optimum subgroup after an interview with a counsellor, and the individual has followed the recommended steps, we should see a move from the subgroup to which the jobseeker belongs to one with higher employability values. If, on the contrary, we see that the individual remains in the same sub-group for several months, we can conclude that, indeed, the jobseeker had not followed our guidance.
- By introducing the "Employment Status" variable in the system, we can also determine whether that individual has been able to find a job after following our guidance.
- The recommendations that we are discussing are based on the two variables that help us to define the subgroups:
    - the occupations requested by the jobseeker (look for other occupations)
    - proactivity in the offers (to sign up to offers)

Conclusions

- The ultimate goal is for the counsellors to use this tool in their interviews with job seekers in order to be able to point them to the best opportunities of the labour market, in other words, that the profiling is used for the user to be able to reach the optimum subgroup.
- Given that the guidance is based on two variables that allow subgroups to be prepared, the recommendations will be at times to be proactive in the search for job offers and, at other times, to change the search for occupations to-wards others with better employment prospects (suggesting occupations to be requested).
- Lanbide is still at an experimental phase. Its use is still limited to consultations, in other words, it is merely used for observations.
- The employment service has over 200 variables in its databases. The aim is to incorporate them gradually, once their importance has been assessed when defining the employability index.
- Apart from the variables used by the employment services, the ESCO skills are also being incorporated, as it is known that they are increasingly more important for companies when selecting applicants.

- Macro-trends are also being added so that the occupations with the best future prospects and trends are rated best and those with poor future employment forecasts are poorly rated.
- The developed methodology can be applied in other organisations. As what has been proven is that the current data analysis tools can be integrated in the business processes of an entity such as Lanbide or similar.
- The optimum profiling design depends, fundamentally, on the appropriateness of the outcome variable to the goal of the programme and on the variety of the characteristics used to assess the employability level of the individuals. This second factor is crucial as the predictive power of the model clearly depends on the type of variables included in it.
- The profiling must be regularly reviewed and several designed according to the geographical location or the possible segments of the job market.

Comments by Dorit Griga

Dorit Griga *(SECO, Switzerland)*
The topic of the use of Artificial Intelligence for decision making is very interesting and very timely. We expect AI to benefit the quality of the services by standardising and reducing arbitrariness of judgments.

Another pilot experience was tested in a Swiss canton related to the use of AI in mediation of job seekers. The project "Job opportunities- Barometer" was developed to display a forecast of the expected duration of the unemployment. However, the forecasting quality of the Barometer was not satisfactory, and it was evaluated negatively by the personnel counsellors.

One of the lessons learned from this experience was that it is crucial to involve relevant stakeholders in the concept of the model: personnel counsellors and clients of the PES and social partners.

Dorit Griga (SECO, Switzerland) would like to spark the debate with the following questions:

- To what extent were the stakeholders involved during the design of the Basque experience?
- Has the tool been tested with the stakeholders (counsellors and jobseekers), if so, what were their reactions?
- Has the job seeking process improved thanks to the tool?
- Did you observe some resistance from behalf of the counsellors or jobseekers?
- Compared to other algorithms, it appears that the algorithm of this tool is transparent, and that jobseekers and counsellors can understand the variables of the basis. It seems that the variable age is the one that is taken into account but

other variables such as gender and social/ geographic variables are not considered.

- Is the aim of the tool, only to provide recommendations for search behaviour, or it is also to monitor and even to sanction?
- Have you already made concrete considerations with regard to future inclusion of variables? For example, the entire dataset of more than 200 variables?
- Can the counsellors also overrule the result of the algorithm if necessary?
- Concerning the target group, have you considered also perhaps employees registered with the PES looking to progress in their careers? Or only jobseekers (unemployed in particular)? It could be interesting for them to know their employability rate. Then we might have other variables such as, occupational status measured by income or working conditions. Then career options could be identified.
- The labour demand side is not taken yet into account in the model, would it be desirable to include an indicator on the number of vacancies on a certain occupation? The model would be better aligned with the LM.

OPEN DISCUSSION

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) stresses that this is a great example and that this is an approach, a concept, to drive how to automatically get real time information and have a good support in counselling. The questions of application that Dorit Griga (SECO, Switzerland) was raising are of course coming up. Perhaps we can discuss the conceptual approach and the ideas behind and then also separately, about the application.

Ciprian Panzaru (Universitatea de Vest din Timișoara, Romania) congratulates the Basque PES for the great job.

Anne Lauringson (OECD) was wondering about the user experience, often, the key on why some tools are uptaken or not is how the tool is integrated in the overall concept or system, how well it is integrated with other digital tools that the counsellors use. Perhaps it could be explained how it fits in the overall system. There are many good initiatives that are not taken up because the counsellor has so many websites and tools. So, we must face the challenge of these integration issues.

Aleksandra Webb (University of the West of Scotland, UK) agrees with Anne Lauringson, particularly in the era of client-centered methodologies and practices. But also the uptake and support from client facing employees would be important to consider and their particular role in helping integration of these solutions.

Borja Pulido (Lanbide, Spain) tries to answer some of the questions that were raised. First, about the variables used in the algorithm. To accelerate the process, they chose 3 variables to start analising the potential of the tool. They used the CHAID decision tree to identify the 3 most critical variables that are: the proactivity in the offers, the occupations requested and the employability status. The variables where Lanbide cannot act (such as the demographic variables) had to be dismissed. The aim is to gradually incorporate new layers of variables to improve the system.

About the testing, Lanbide is at an early stage and it has not been tested yet by counsellors in interviews (it will in the near future). It has been used only to consult the jobseekers situation.

The labour demand side information such as the number of vacancies for a given occupation is expected to be included soon in the Macrotrends. The occupations with better prospective will be then better rated.

8-10 people from Lanbide are involved in this project working jointly with a private consultancy. Lanbide (as a PES) has a Commission formed by different stakeholders such as Trade Unions, Business associations and social partners. All projects must be approved by this Commission, which meets regularly, so this means that the project is supported by this type of stakeholders.

It is not possible now to overwrite the result of the algorithm, but perhaps during the testing with counsellors, they face this problem, but not now. The aim is to complement this system with the skills, so maybe the skills help to better rate if one person was over or under- rated.

Nowadays there are sanctions by the Basque PES when a jobseeker is receiving benefits and declines 3 job offers. Who knows, maybe in the future, if the tool says that you have to follow some recommendations and you do not follow them, maybe we can think about some kind of sanctions.

The system is only for unemployed jobseekers, it does not allow the use of the tool with employed people. At the moment, if a person is employed, they have a 100% employability index. To implement it for employed people, we would need another algorithm, the methodology and maxims would need to be changed.

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) raises a question about the methodology. Big data tools allow to find correlations between variables that you would not have expected. How was the selection process of the tool? Borja Pulido (Lanbide, Spain) explains that the statistical programme indicates the variables that need to be used. They started with the basic variables, age, gender, academic level, if the jobseekers has a subsidy or not, etc. and they used the CHAID

decision tree, to see what happened when they introduce a new variable into the system. For example, they introduce the variable gender and the tree says, yes, men have a higher employability index than women. So, this is a significant variable. Having a subsidy or not was not a significant variable. So we build the system with the significant variables only, even if it differs from expectations. The level of introduction of the variable is also important and you get different results depending on this.

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) points out that this allows to explore on the administrative data that we have in the PES. It is very worthwhile to have these discussions that brings us forward. Apart from the user involvement, the acceptance, she wants to ask Dorit Griga (SECO, Switzerland) what would be the recommendations for the beginning of the process, if there are technical considerations. Dorit Griga (SECO, Switzerland) answers that stakeholder involvement must happen the sooner the better. And about the sociodemographic variables, we need to be careful, an Austrian algorithm was judged to be discriminating women and had to be readapted. Perhaps we need to be more flexible, the Basque experience seems rather strict, maybe women get higher employability indexes in “female- typical” occupations. There is a danger if we use the gender CHAID and not the whole population. But there are academic discussions and methods to prevent this unfairness.

Anne Lauringson (OECD) thinks that this is a very relevant point - then an indication for a female jobseeker that on this occupation is the highest chance to be employed, but there is another opportunity for you, but you might need to break the glass ceiling...

Borja Pulido (Lanbide, Spain) explains that the tool needs to be used with the supervision of a career advisor to avoid this kind of situations. It is a tool for the counsellors, not for the jobseekers. It is always an expert the one who provides the results to the jobseekers and can over-rule the results if necessary.

Dorit Griga (SECO, Switzerland) will share a paper (find attached) that a colleague from SECO has developed as part of a further training course on the topic of "Fairness in statistical jobseeker profiling". The paper does not represent a general view of SECO, however, SECO is currently working on a framework paper on various guidelines that are regarded as a basis for the use of AI in public employment services.

Joanna Napierala (Cedefop) thanks Borja Pulido (Lanbide, Spain) for the interesting presentation.

The next of the series will take please in Autumn, at the end of October, the Network will send the invitation.

Bilbao, June 2022

References

Speaker: BORJA PULIDO
Lanbide- Servicio de Empleo Vasco, Basque Country, Spain
Email: b-pulido@lanbide.eus
EUROPEAN NETWORK ENRLMM
Email: jenny@jennykipper.de
Email: c.larsen@em.uni-frankfurt.de
Website: www.regionallabourmarketmonitoring.net
BIG DATA WORKING GROUP
Eugenia Atin
E-mail: e.atin@prospektiker.es
Website: https://bigdatahub.uvt.ro/