Understanding Youth Unemployment in Italy via Social Media Data

Andrea Bonanomi[1], Alessandro Rosina[1], Ciro Cattuto[2], and Kyriaki Kalimeri[2]

[1] Università Cattolica del Sacro Cuore,
andrea.bonanomi@unicatt.it, alessandro.rosina@unicatt.it
[2] ISI Foundation,
kalimeri@ieee.org, ciro.cattuto@isi.it

1 Introduction

Transition to adulthood evolved over the past decades. While in archaic societies most transitions took place by discrete leaps and were signed by rites of passage, modern societies present a different approach: transitions are represented increasingly as being individual, relatively undefined (with respect to both modalities and timing), negotiable, and as having wide margins of choice. Due to the fact that nowadays the transition to adulthood has became longer, the family of origin has assumed more value and power in influencing young people. Italians, in particular, leave home on average at a later age with respect to young people in other European countries. The peculiarities of the Italian situation can be explained from both cultural and structural standpoints. On one hand, the presence of strong inter-generational ties is coherent with longer stays in the family of origin. On the other hand, the unfavorable labor market and the welfare system that is not generous to young generations tend to discourage individual autonomy and an active job search. [1].

Italians present several peculiar characteristics in the transition to adulthood.

1. They remain for a longer time at home than other European countries young people.
2. They can heavily count on their families, characterized by good communication, high levels of support and relationship quality. The results of research into enmeshment and the acquisition of autonomy and of an adult identity status are more complex.
3. These characteristics may be connected to the problem of increased youth unemployment in Italy, whose rate was 37.7% on April 2016 according to the Organisation for Economic Co-operation and Development.

This study aims at improving the current understanding of the conditions in which young Italians find themselves in during the acquisition of autonomy and in the active job search via Social Media Data and Digital Behaviors.

We place the focal point on the *Not in Employment, Education, or Training* population, hereafter NEET and on the grade of Autonomy that Young Italians have (i.e in the decision of leaving home). NEETS are young people from 18 to

32 years old, at high risk for social marginalization [1]. The failure to tap into the economic aspirations limits not only their income and skill development, but also their likelihood of later employability, autonomy and life-planning. The percentage of young Italian in this peculiar condition is equal to 20.3%, versus a European average equal to 10.3%. The most alarming aspect is the qualitative nature of the problem, in addition to the important quantitative value. NEETs composition is very heterogeneous: in this condition, there is the newly graduated with high motivation and high potential that is actively seeking a job in line with his expectations (possibly eventually downward with what the market offers), there is the young man who came out early from the school, slipped into a spiral of marginality and demotivation. But also people who do not have a job for choice because they want to take time for different types of experiences or to dedicate themselves to the family. NEETs are also the ones most likely to be in the path of transition to adult life, risking aging without taking significant steps in the realization of their own projects, not just employment but also in the life. They do not study and work, therefore they tend to associate other "no" on the side of the choices of autonomy, formation of a family, civic participation, of full citizenship. It is not enough to describe the characteristics of NEETs in terms of conditions and attitudes - useful for knowing the phenomenon but also potentially providing guidance for interception and engagement measures - going beyond not just the data of public statistics, but experimenting even the combination with innovative sources. A general characteristic of this population is the use of new technologies (Internet and social networks in particular). Internet access is great for all young people and for all working conditions. Analyzing specifically the use of Facebook, it emerges that 50% of NEETs passes at least an hour a day on Facebook, and just under 20% even more than three hours. Social networks, in this context of criticism, can be used more as an escape route than a tool to find useful information and share experiences that help to get out of the corrosive inactivity state as it might be. In this preliminary study we are going to look into in the analysis presented in the following paragraphs, that is to try to exploit this powerful tool, by turning it away, escape and further isolation, resource and opportunity to re-emerge and become part of the productive society and the community .

The project tries to investigate the factors that drive the active job-research on the basis of online behaviors and activities ("likes", social networks, posts, clicked banners, visited sites) and electronic and / or physical participation at "labor market". Attendance can be "passive" (understood as basic online activities such as searching for open positions) or "active" (participation in groups, interviews, access to real-world job platforms / applications). Understanding participatory behaviors and job search patterns - or non-research - allows the profiling of the unemployed youth population to design personalized and targeted interventions to improve the employment results.

In detail, the main aim of the study is twofold; firstly, (i) to automatically identify the population of interest (i.e. NEET population, individuals who leave home not longer young or do not have a sufficient autonomy from their family)

inferring from their on-line digital traces and secondly, (ii) to uncover digital behaviors of the community of interest easily accessible from on-line social platforms, which can then be used as indicators of the most privileged communication channels for unemployment or educational advertising campaigns.

To this extent, we created a new Facebook application, namely *LikeYouth*, whose major functionality is to pave the way for self-assessed psychometric data collection from participants in form of questionnaires. Having obtained the participants' informed consent, the application gathers information regarding their public Facebook profile and their "Likes" on Facebook Pages. *LikeYouth* is an innovative data-collection tool of rich nonverbal cues for behavioural understanding and profiling along with validated information from self-reported psychological assessments. The application is designed to be easily extendable for new questionnaire while it is provide both in English and Italian language.

This approach takes advantage of the popularity of Facebook's social platform communicating with a potential audience of interest previously inaccessible. Additionally to the ease of cohort expansion, this approach offers the possibility of longitudinal studies monitoring at the same time level digital behaviours of high semantic information. Furthermore, information regarding the social ties of the participants and the influence of their digital social relationships now become available for social studies, creating a digital observatory.

Data from digital maps are integrated with information on the same subjects obtained from the most traditional Rapporto Giovani Survey. In fact, in this first phase of the project, LikeYouth was only proposed to Rapporto Giovani's sample, in order to test the validity of prediction and classification, and to be able to propose the Facebook App in the future on a large scale. The aim is to analyze the validity and usability of large amounts of information from online behaviors, and eventual advantage in the comparison with the classical analysis based on *ad hoc* analysis on small sample size.

2 Data Collection and Methods

The information gathered for this project originates from two different sources; from the survey of Rapporto Giovani" and from the *LikeYouth* application. Rapporto Giovani is a national-wide Italian survey, launched in 2015 by the Toniolo Institute of Advanced Studies with the inclusion of the CARIPLO Foundation and IPSOS LTD as executive partners. The survey was carried out by a mixed methodology CATI, CAPI with in-depth CAWI. More details are available at http://www.istitutotoniolo.it. The sample consists of 9,358 individuals aged between 18 and 32 years ($M = 25.7$, $SD = 4.7$), taking under consideration the age bracket as constituting emerging adulthood. The individuals were chosen with a stratified sampling technique. The population of "Rapporto Giovani" is a representative sample of the Italian youth population. The representativeness is given by a significant set of different variables (gender, age, geographical origin, education, marital status, etc.). Every subject, at the end of the survey, was invited to access to *Likeyouth*. Major areas investigated in Rapporto Giovani are related

to demographic, family, educational, behavioral, employment status and psychometric questions. Areas investigated in Likeyouth are referred to personality and morality assessments, questions regarding the pro-activity, the confidence, the vision of future, the perceived happiness, the current employment status, the vision about world work, the "*Likes*" of the users at Facebook Pages. The population gathering from Likeyouth consists of 1,858 individuals (about 20% of the total). The comparison (Table 1) between subjects of Rapporto Giovani and subjects of Likeyouth does not show significant differences respect to the most important demographic variables (i.e: Total Sample: mean age 25.7 ± 4.7, 50.8% male, 42.0% resident in the North of Italy, 18.0% with an high education level, 80.8% single, 19.9% NEET, mean age. Participants of Likeyouth: mean age 25.8 ± 4.4 48.9% male, 39.7% resident in the North of Italy, 19.7% with an high education level, 80.4% single, 21.9% NEET).

Table 1: Comparison between subjects of Rapporto Giovani and subjects accessing Likeyouth

Attribute	Rapporto Giovani	LikeYouth
Population	9358	1858
% of higher degree school	18.0%	19.7%
% of medium degree school	49.5%	52.2%
% of lover degree school	32.5%	28.1%
Gender(% Females)	49.2%	51.3%
% Neet	19.9%	21.9%
% Student	37.7%	39.0%
% Worker	42.4%	39.1%
Age	25.7 ± 4.7	25.8 ± 4.4
Residence in Northern Italy	42.0%	39.7%
Marital Status (% of singles)	80.8%	80.4%

No statistical concordance index shows a significant difference also for some psychometric indicators (well-being, self esteem, BIG5, MFT, perceived happiness etc.). In particular, Figure 1 shows the Big5 personality traits for subjects in Rapporto Giovani ("Traditional") survey and Likeyouth App ("Facebook"), Figure 2 shows the comparison by the Moral Foundation Scale.

The Big Five (Big 5) personality traits, also known as the five factor model (FFM), is a model based that characterises personality based on five dimensions and has universal validity.

– **Openness to experience**: inventive/curious vs. consistent/cautious.

– **Conscientiousness**: efficient/organized vs. easy-going/careless.
– **Extraversion**: outgoing/energetic vs. solitary/reserved.
– **Agreeableness**: friendly/compassionate vs. analytical/detached.
– **Neuroticism**: sensitive/nervous vs. secure/confident.

The Moral Foundations Theory (MFT) focuses on the explanation of morality, its origins, development, and cultural variations [6,4]. The MFT may be considered to be at a higher level with respect to the dispositional traits of personality expressed in the Five-Factor model [3], providing insights on the characteristic adaptations of the individuals [5] as described by Dan McAdams's three-level account of personality (dispositional traits, characteristic adaptations, and life stories) [8,9]. MFT focuses on the psychological basis of morality, identifying the following five moral foundations (see [4,6]):

– **Care/Harm**: basic concerns for the suffering of others, including virtues of caring and compassion.
– **Fairness/Cheating**: concerns about unfair treatment, inequality, and more abstract notions of justice.
– **Loyalty/Betrayal**: concerns related to obligations of group membership, such as loyalty, self-sacrifice and vigilance against betrayal.
– **Authority/Subversion**: concerns related to social order and the obligations of hierarchical relationships such as obedience, respect, and proper role fulfillment.
– **Purity/Degradation**: concerns about physical and spiritual contagion, including virtues of chastity, wholesomeness and control of desires.

Fig. 1: Assessing *self-selection bias due to recruitment*. Mann-Whitney U test on the personality traits of those who participated in the survey but did not access the Facebook application (denoted as 'Traditional') against those who accessed the application (denoted as 'Facebook'). No statistically significant differences are observed. (a):Agreeableness, (b)=Openness, (c):Extraversion, (d):Conscientiousness, (e): Neuroticism

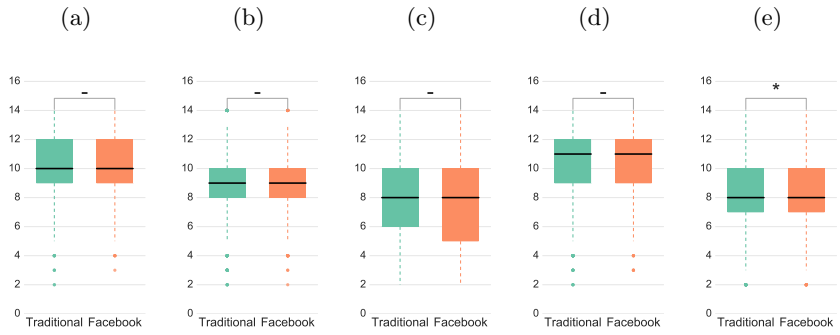
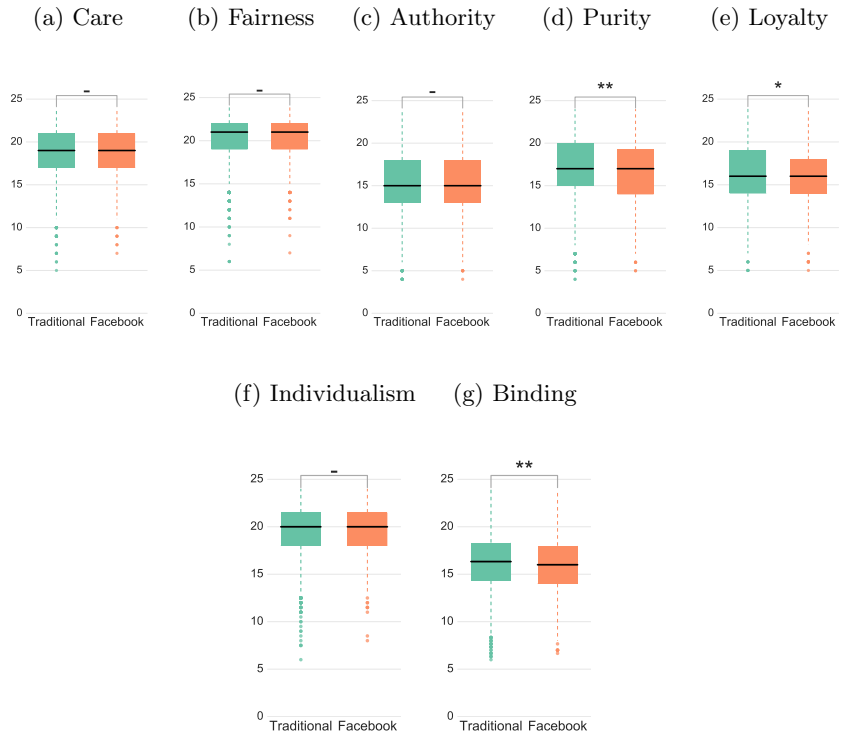


Fig. 2: Assessing *self-selection bias due to recruitment*. Mann-Whitney U test on the five basic moral foundations (MFT) of those who participated in the survey but did not access the Facebook application (denoted as 'Traditional') against those who accessed the application (denoted as 'Facebook'). The stars indicate the level of p-value. We note that in the 'Facebook' group tends to show a lower 'Purity' value, however, the differences are small. For all the attributes with statistically significant effects, the effect sizes are negligible with $d < 0.14$.



Participants of *Like Youth* may, therefore, be considered a representative sample of the Italian youth population. All statistical analyses have an inferential statistical significance. The goal is to assess how much of the information gleaned by questionnaires is behaviorally observable, which in turns relates to the scalability of the approach, since behavioral observation at scale is usually simpler and more cost-effective than large-scale survey campaigns.

Considering our limited size of the cohort (9358 people on the original cohort and 1858 people on the FB one) and its focus on a specific geographic location and age range, our findings suggest that surveys administered on Facebook do not exhibit major biases with respect to traditionally administered surveys, neither in terms of demographics, nor personality attribute assessment. This finding

indicates that engaging in a social media platform like Facebook slightly affects the individuals' behaviour reflecting also on their self-reporting. Overall, and considering the small size and homogeneity characteristics of our sample, our findings suggest that Facebook is an adequate tool for social and psychometric surveys, nonetheless, it is not an entirely neutral platform, and may have an impact depending on the nature of the administered survey.

Aim of this preliminary and exploratory study was to do prediction and classification. We tried to predict the working condition (dichotomizing it in two categories, NEET and NOT NEET) based on the digital behavior (especially on Facebook's *likes*) that act as predictors. We have tried to classify subjects on the basis to their tendencies and behaviors, and to investigate possible associations between the different conditions and digital behaviors. The aim is twofold: first, to better understand the NEETs condition through direct or indirect behavior, to reach them with targeted campaigns, marketing, social campaigns on pages and sites of interest, and secondly, to predict if a young man is potentially at risk for potential engagement and activation programs, estimating the probability of belonging to a particular category.

To this extend we designed a generic experimentation schema, aiming to assess and compare the predictive power of the digital behaviors. We postulated the study as a supervised classification process, automatically identifying the employment status and the potentiality that a specific person belongs to a category or not(i.e. NEET or NOT NEET), inferring only from their digital data. Additionally, we trained a predictive model for each of some demographic attributes. We employed a widely-used ensemble learning method for classification was employed, namely a Random Forest (RF) classifier [2].

We evaluated our performance with the *weighted area under of the receiver operating characteristic (AUROC)* statistic [7]. The weighted AUROC statistic was preferred over the commonly-used *accuracy* metric, since the former takes into account the effect of unbalanced labels, which holds true for most of our attributes. All the prediction scores reported through this paper are in terms of the weighted AUROC. In detail, the best model for each target variable emerged from a grid-search step and was employed in a k-fold cross-validation with $k = 5$. In each fold, the entire feature space, X, is randomly shuffled and split in two mutually exclusive sets, the training set Tr, (80% of X) and the testing set Ts (20% of X). Furthermore, for each target variable we estimated the relative rank (i.e. depth) of each feature, as emerged from the "Gini" impurity function, assessing in this way the relative importance of a specific feature to the predictability of the target variable [2]. We use as predictors the Facebook Pages the users visited and the respective Categories, as defined by the Facebook metadata. This information is represented as a sparse user-page and user-category matrix, the entries of which were set equal to the raw counts of visits and 0 otherwise.

3 Results

1858 young people from the Rapporto Giovani visited, during their "history" on Facebook, more than 330,000 different Facebook pages, grouped (from Facebook information) in 155 categories. The assignment of the category is determined, at the time of the making the page, by the administrator of the page itself. It is not, therefore, an objective assignment, but a subjective decision of the developer of the page. Thus, a new taxonomic categorization of the initial categories was made, creating 12 macro categories for a quantitative analysis of the results, and about thirty categories for a more qualitative interpretation of the same. On average, every subject has liked almost 400 Facebook pages, belonging to about 50 different categories.

The preliminary analysis focus on the classification and prediction of employment status. The employment state of the participants is predicted with accuracy 61%. For the first area and community of interest, the NEETs, we are able to predict whether an unknown person is potential NEET or not with accuracy 63%. The model is trained on the population originating from the "Rapporto Giovani"'s survey for which there has been a manual labeling of the NEET status characterisation by a field expert.

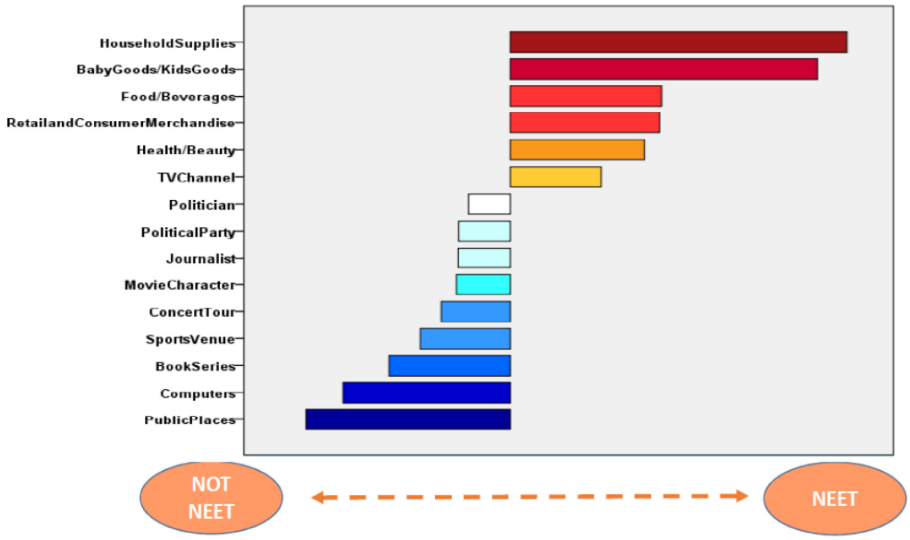
For a more qualitative and interpretative analysis of data, we chose to consider only a subset of the Facebook pages, that is those pages that satisfy two conditions: 1) have been visited by at least 20 young people in our sample; 2) had at least 100,000 likes, that is, they were very popular and well-known pages in Facebook. 2422 pages satisfied the two conditions. They were the predictors of the models in the second part of the analysis. Different data analysis techniques have been tested: Cohen's K, Phi Correlation Coefficient for dichotomous variables, Random Forest with Breiman's algorithm, CART Decision Tree, Logistic Regression. For each Facebook page, each technique produced a score of "importance". With a meta-analysis approach, the scores were weighted and normalized. At the end of this procedure, for each Facebook page, a normalized index was created in the closed subset [-1; +1]. Scores close to -1 are predictors of the condition of Not Neet, scores close to $+1$ are predictors of the Neet condition.

Firstly, the analysis was done by using the original categories of Facebook, secondly by using the 12 Macro-Categories of Facebook pages recoded in the our categorization, and finally by using the individual Facebook pages. The analyses were conducted on the whole sample and on cohorts of particular interest (by gender and by age group).

Figure 3 shows the association between the most predictive Facebook categories and the status of NEET/NOT NEET. Figure 4 shows the association between the re-coded Facebook categories and the status of NEET/NOT NEET.

Both in the original categorization of Facebook and in our macro-categories classification, top indicators of NEET status are Facebook pages or categories related to consumer goods, in particular food, beverage, beauty and health (in particular baby and kids goods), TV Channel, Retail and Consumer Merchandise. The use of Facebook appears as a leisure tool and not much as utility

Fig. 3: Prediction of NEET or NOT NEET status from some Categories of Facebook pages



or service. Top indicators of NOT-NEET status are Facebook pages/categories related to travel, culture, humor and satire, performance art, news media and politics.

Further models are be developed for particular cohorts of the population of interest (male/female, under/over 24 years).

The analysis for each cohorts was conducted by considering the most predictive aggregated pages. Figure 5 and Figure 6 represent maps of the most predictive categories of Facebook page of the NEET or NOT NEET status, respectively by gender and age group (under and over 24 years).

In the prediction for gender, top indicators of NOT-NEET status are Facebook pages related to TV series (in particular for females), news media, culture, politics. Top indicators of NEET status are Facebook pages related to goods, food/beverage, kids and TV entertainment. In particular, for men, Facebook pages related to labor market or job-search pages. This aspect is very important, as it shows that, at least in part, Facebook activity for young men is, potentially, a virtual activity of research that needs to be considered to take action of engagement and active involvement.

In the prediction for age, top indicators of NEET status for over 24 years people are Facebook pages related to kids, TV channel and entertainment, while top indicators of NOT-NEET status are pages related to TV series, politics, culture and news media.

Fig. 4: Prediction of NEET or NOT NEET status from Macro-Categories of Facebook pages

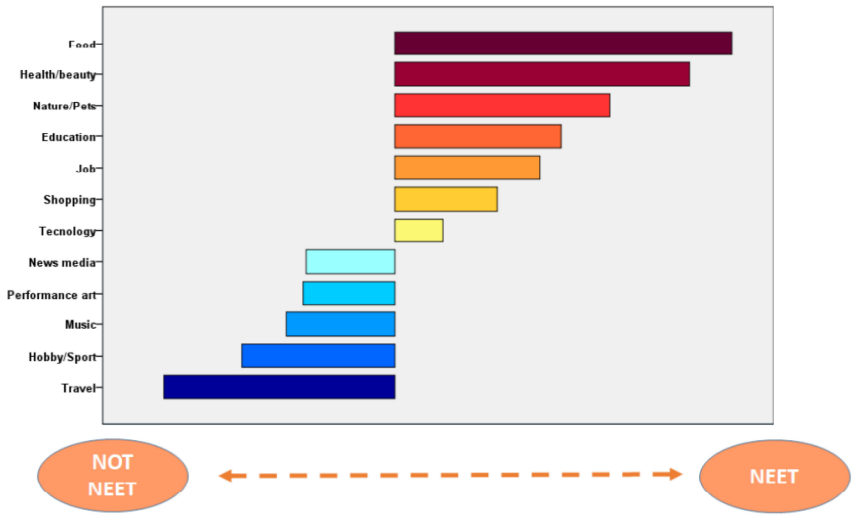


Fig. 5: Prediction by Gender

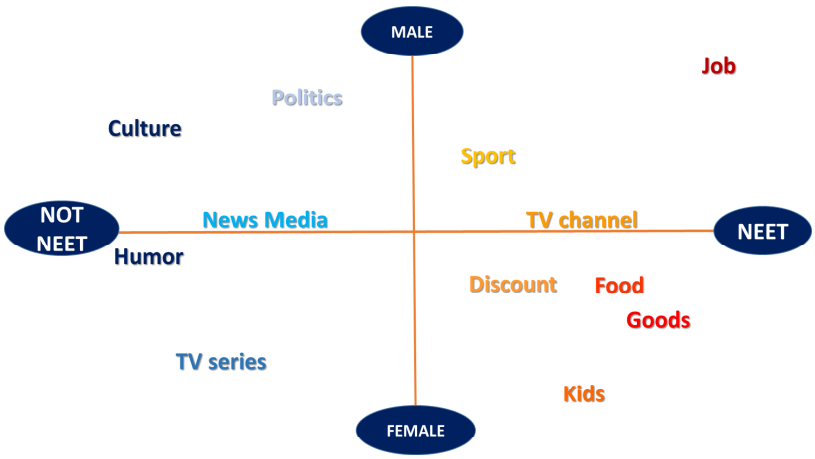
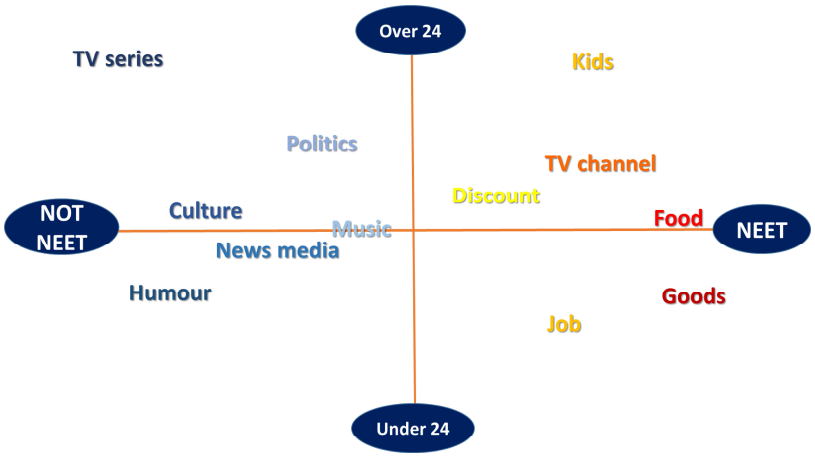


Fig. 6: Prediction by Age



For example, for each cohort, Table 2 shows some particularly predictive pages of the two different status.

Table 2: Most predictive Facebook pages of NEET or NOT-NEET status

Cohort	NEET	NOT-NEET
Males	Worky.biz	Expo2015
	Verissimo	Repubblica.it
Females	Prenatal	MatteoRenziCheFaCose
	Lidl Italia	Grey's Anatomy
Under 24 years	Just Cavalli	Dr.House
	Girella	Barack Obama
Over 24 years	Humana Italia	Il Milanese Imbruttito
	ScontieBuoniAcquisto.it	Report

In general NEETs visit discount pages, promotions, prize competitions, offers, which can be explained for various reasons. Certainly, NEETs have more free time than workers and students, so they can spend their time searching this kind of pages on Facebook. Moreover, they are pushed by their weaker economic situation, to ensure themselves an acceptable quality of life. NOT NEETs have a more active and conscious digital behavior, and typically they like pages of culture, information, satire, so not closely linked to advertising and particular

media exposure, while NEETs visit more pages linked to the Consumer Merchandise world, and therefore linked to massive advertising campaigns on Facebook. NEET's digital behavior appears to be more passive. As in life choices they appear to be more passive and less enterprising, even in on-line behavior they transfer this negative trait of personality. This negative feature can, however, be overturned in a positive way: when we decide to undertake advertising campaigns and actions strictly directed at the NEET population, the probability of a success of such campaigns would have a greater yield compared to the population average, since this population seems to be more receptive and "catchable". In addition, by knowing the most predictive pages, targeted campaigns may be implemented directly within these pages, so to reach the target set with high probability of success.

4 Conclusions

In conclusion, we present an innovative, flexible way to perform social experiments, by developing *LikeYouth*, a Facebook application easily extendable to new questionnaires, without the limitations of the traditional survey methods on the cohort. Since taking advantage of the Facebook's platform popularity and its advertising tools, it can reach out to the population in need with relatively limited economic and temporal requirements. Having obtained a basic understanding of the interests of a particular community and the most effective communication channels, targeted advertising campaigns can be tailored to these interests and their effectiveness can be measured by traditional A/B testing techniques. The challenges emerging from this approach are related to the engagement strategies, since Facebook is not adequate for extensive questionnaires common to the traditional sociological and psychological research practices, which for this kind of studies should be redesigned in a more creative and interactive approach. At the same time, engagement and rewarding strategies become important to creating an active self-sustainable community of participants.

The analysis conducted by LikeYouth on 1858 young people of Rapporto Giovani panel gave comforting results, both on the sample's representativeness and on the predictive and classifying performance of modeling used on Social Media Data. This analysis also showed the passivity of the young NEETs, which, compared to other peers, appear to be more passive, less entrepreneurial and less oriented to cultural, educational and information interests. However, especially in the males cohort, a part of young people visits pages related to labor market and job search portals. Therefore, they begin to implement possible strategies to get out of their status of inactivity. The aims of the study were: 1) Better understanding the NEET condition; 2) easily reaching them for targeted social campaigns and marketing, to engage them, to activate them and to involve them in the digital world; 3) overturning the vision that society and politics often have on the digital social networking tool, thus using them as social activator. The extreme flexibility of social networks has to be exploited in its full potential.

References

1. Sara Alfieri, Emiliano Sironi, Elena Marta, Alessandro Rosina, and Daniela Marzana. Young italian neets (not in employment, education, or training) and the influence of their family background. *Europe's Journal of Psychology*, 11(2):311, 2015.
2. Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
3. Paul T Costa Jr and Robert R McCrae. The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, 6(4):343, 1992.
4. Jonathan Haidt and Jesse Graham. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1):98–116, 2007.
5. Jonathan Haidt, Jesse Graham, and Craig Joseph. Above and below leftright: Ideological narratives and moral foundations. *Psychological Inquiry*, 20(2-3):110–119, 2009.
6. Jonathan Haidt and Craig Joseph. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4):55–66, 2004.
7. Jialiang Li and Jason P Fine. Weighted area under the receiver operating characteristic curve and its application to gene selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(4):673–692, 2010.
8. Dan P. McAdams. What do we know when we know a person? *Journal of Personality*, 63(3):365–396, 1995.
9. Dan P. McAdams and Jennifer L. Pals. A new big five: Fundamental principles for an integrative science of personality. *American Psychologist*, 61(3):204–217, 4 2006.