

# Analyzing Social Media Data for Recruiting Purposes

Lucie Böhmová<sup>1</sup>, David Chudán<sup>2</sup>

## Abstract

Social media networks are tools that recruiters can utilize during a recruitment process. Most importantly, social media networks can be used in conjunction with applications capable of downloading information about their potential candidates. The aim of this article is to present a creation process of a model that could be helpful in recruiting area. A crucial part of this model is application software that downloads user's data, particularly from Facebook profiles. This model should propose appropriate analytical methods for data processing. The output of this article is employee recruitment model that can be used as a guide to utilize the potential of social media networks by HR professionals. Test run of this model on our population sample showed prediction accuracy of 68 % to 84 %.

**Keywords:** Cluster Analysis, Data Mining, Model, SNA, Social media networks.

## 1 Introduction

In the 21st century, social media became a phenomenon that is an integral part of our everyday life across all generations as well as companies. Social media are not used solely as a communication channel. Nowadays they are reaching many more areas, industries and denying a threshold amongst personal life and professional life (ČSÚ, 2015; Pavlíček, 2010). This potential has already been shown in HRM (Human Resources Management), particularly in the recruitment area.

The current Czech labor market situation is not very pleasant from an organization's point of view, mainly in the recruitment area. Companies are struggling to find suitable employees. The traditional methods do not work due to the low unemployment rate and high demand for employees (MPSV, 2016). Another reasons for this can be the decreasing number of economically active population (ČSÚ, 2013), the characteristics of the new generation—people from Generation Y and Z are independent, without sense for job commitment and leisure is a priority (Meister & Willyerd, 2010)—that is entering the labor market or the modern trend of the shared economy (PWC, 2015).

---

<sup>1</sup> Department of Systems Analysis, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic

✉ [lucie.bohmova@vse.cz](mailto:lucie.bohmova@vse.cz)

<sup>2</sup> Department of Information and Knowledge Engineering, Faculty of Informatics and Statistics, University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic

✉ [david.chudan@vse.cz](mailto:david.chudan@vse.cz)

Social media networks offer a solution that is innovative and potentially cost-effective. In practice, it is difficult for organizations to find out which social media networks should they use for the recruitment process and how to utilize their potential (Jobvite, 2014; HRnews, 2016). The combination of the above-mentioned facts raises current issues.

The general research problem of this article is the use of social media networks to support the recruitment process in modern HRM. The author's solution offers a few suggestions of suitable analytical methods for data extraction from social media networks. The output (and the goal of this article) is a model that supports recruitment.

The article structure is as follows. Literature review summarizes the current state of the social media recruitment area. Then there is a chapter devoted to the Social network analysis. The Data extraction from social media networks chapter describes how to download data from a custom-created application that is used to extract data and describe the most important data analysis outputs. The model development process is based on the chapter Model creation according to the methodology CRISP-DM, which includes 6 steps leading to a model creation. The final model is described in the Social media recruitment model chapter. After that this section is followed by a discussion, where are mentioned the benefits and limitations of the model. It also includes possible ideas for further research.

## 1.1 Literature review

Social media networks are a virtual space with a huge recruitment potential (Bartakova et al., 2017). People are voluntarily sharing so much personal information via social media networks, such as favorite movies, books, how, when and with whom they spend their time and sometimes also information and opinions about politics and religion (Böhmová & Malinová, 2013). It depends on privacy settings of every single user which information will be shared with the rest of the world and which one will be not (Pavlíček, 2016).

Research in cyberpsychology has examined how social media networks users engage in impression management (IM) to create specific impressions on friends or family members, and achieve a positive online identity. However, with organizations increasingly relying on cyber-vetting, job applicants are also likely to engage in IM tactics oriented towards employers in their social media networks profiles (Roulin & Levashina, 2016). There is already a new approach of personality prediction that is explored by merely evaluating the contents of a user's social media account (Ong et al., 2017; Annisette & Lafreniere, 2017; Park et al., 2015).

LinkedIn (2015) and server Ere Media are in an agreement on the topic of the Future forecast of world's trends for the year 2016. They had predicted that social media networks will play a key role in a company's HRM and also that social media networks will become a crucial source of talented candidates. On the other hand, on the social media networks is so much information, and these days it is not enough to just share job offers. Therefore organizations need effective hiring methods and tools (Sathya & Indradevi, 2017).

A challenge for the next years to come is to collect and analyze Big Data (McAbee, Landis & Burke, 2017). In the recruitment field this process comprises of users data gathering via social media networks. For these purposes, there exist recruitment models such as Proposed Practical Model for Media Driven Collaborator Recruitment (Khatri, 2015), Model COBRA (Muntinga, Moorman & Smit, 2011), Social Media Activity Model (Bender et al., 2017) etc. The weak point of current models is insufficient utilization of social media networks in terms of receiving candidate's references, completing candidates profile or acquiring the right candidates. Then there does not exist any model for user's behavior evaluation according to

the personality tests in terms of employees recruiting on the social media networks. Authors fill this gap by the suggested model for the employee recruitment.

## 2 Social network analysis

Social Network Analysis (SNA) is an interdisciplinary approach used to study a social structure. There are 2 types of data in this context (Toušek, 2015):

1. **Relational data:** Results from the relationships that participants have on a social media network, they are displaying a real social structure. In SNA terminology, relations are referred to as ties or edges, and units of analysis as nodes or vertices. These ties are properties of a set of factors that make up the structure of the social media network. The social media network can be defined in the most elementary way as a set of three or more actors, each of whom has at least one edge with any of the other actors. The SNA places a high level of importance on relational data, i.e. the relationships between the units of analysis within the social structure organized into sociograms, a diagram representing people as points and relationships between them as lines.
2. **Attribution data:** Are individual qualities of the actors (individuals or groups, e.g. socio-demographic characteristics such as age, gender, income, etc.) or attitudes and opinions (e.g. political preferences). These individual characteristics show possible contexts (e.g. the impact of income on political preferences) and social phenomena.

Every real social media network can be converted into a graph where the direction of relationships could be bidirectional as in the case of friendship on Facebook<sup>1</sup> (if the candidate is a friend of someone, he is also a friend of the candidate) or it may be one-way as in the case of Twitter (if someone is followed, you do not have to follow the candidate). Graphs where no direction is decisive are easier to interpret for some purposes, as is the case with LinkedIn links. (Newman, 2010)

Organizations can also use features of social media networks in order to recruit due to the fact they provide information about individuals such as their relationships and behavior. Density<sup>2</sup> says that any individual knows a lot of people, which can be very useful for business related positions. As the central role of the nodes suggests, depending on the centrality<sup>3</sup>, several types of personality can be observed. The organization can use it if it is looking for a specialist in the field, a company leader or the other way around, a human, who will bring new business opportunities to the company thanks to his friends.

There are numbers of software tools available for SNA to help with the measurement, layout and visualization of results (Molnár, 2011).

## 3 Data extraction from social media networks

The data about candidates from social media networks is significantly important for organizations. (Böhmová, Mcloughlin & Střížová, 2016) Therefore, the following section describes how data can be extracted. Most of the social media networks offer two different

---

<sup>1</sup> Facebook also offers a one-way connection if a person is followed, but this has to be enabled on their profile.

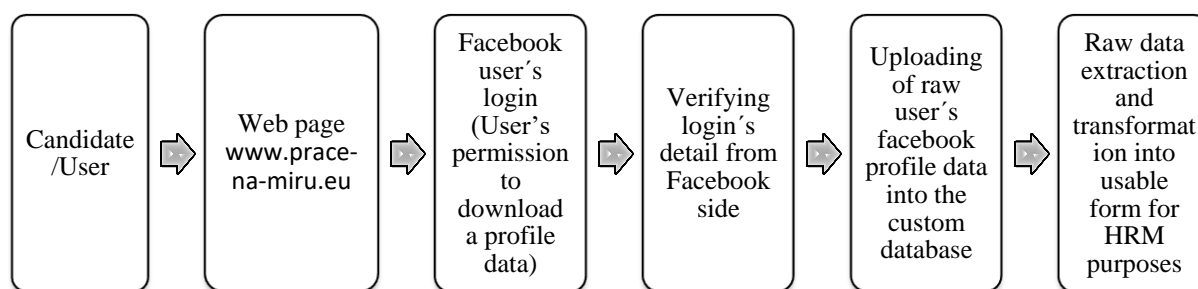
<sup>2</sup> Density is described as the ratio of the present network bonds to the maximum possible number of bindings. (Scott, 2000)

<sup>3</sup> Centrality is the value that tells how the top/peak of the network is significant. (Tore, Agneessens & Skvoretz, 2010)

options to integrate own applications. The first option is to place the application directly “inside” of the social media network where it is displayed in its determined space. For example, Facebook has a feature so-called canvas page, a home page of the application on Facebook with a unique URL that is chosen by a developer in the form [http://apps.facebook.com/\[selection\]/](http://apps.facebook.com/[selection]/). In order to get into the app, the user must access the Facebook URL via the apps.facebook.com domain.

The second option is to develop the app separately and implement it into an external web site that runs completely on its own URL. Connections can be made via Application Programming Interface (API).

For the purpose of this work, the authors used the second option and choose Facebook as a suitable network. The main goal of this application has been to gather information about users that are public and not publicly accessible (only information that user can see according to the privacy setting and can be seen by his/her friends, friends of friends) and analyze them afterward. The main purpose of this application is to serve in organizations as an addition to the traditional way (such as advertising on the job portals, companies’ websites etc.) of employee recruitment. Workflow of data extraction is shown below in Fig. 1. The authors have created an own application named “Práce na míru”, loosely translated “tailor-made work” which runs at web page [www.prace-na-miru.eu](http://www.prace-na-miru.eu).



**Fig. 1.** The data extraction workflow. Source Authors.

The candidate goes to the website “Práce na míru”, where he can find a login button to Facebook. After inputting his login credentials the initiation process begins. There appears a window where the user can find and check what will be downloaded. The candidate gives a permission to download data that will be stored in the database.

### 3.1 Data description

Information about “Práce na míru” application has been spread via the email newsletter to the target audience. This audience is students and fresh alumni<sup>4</sup> of the University of Economics in Prague. Also, the application has been promoted on social media networks in particular groups. 960 unique applicants have signed on to the application during the period of October 2016 to January 2017.

The data were transformed to a more appropriate form and also cleansed by using tool named Knime (Knime, 2017) together with MS Excel. The analysis of data that had been gathered from the “Práce na Míru” application shed some light on results, see in Table 1. This table shows a percentage representation of publicly accessible information about users according to the particular category. Users share information such as gender, device (used for log-in), the

<sup>4</sup> Absolutory amongst years 2011 to 2016.

list of friends, date of birth, etc. Some of these data are suitable for the model creation. For example, favorite music, favorite movie, etc. On the other hand, users do not share information about religion, politics opinion, inspiring people.

Category of publicly accessible data	Nr. of users in %	Category of publicly available data	Nr. of users in %	Category of publicly available data	Nr. of users in %
Gender	99 %	Tagged places	76 %	Favorite books	35 %
Login Device	92 %	Favorite music	73 %	Languages	30 %
Friend lists	91 %	Actual location	73 %	Relationship details	30 %
User's birthday	89 %	Hometown	67 %	Games activity	29 %
Profile Photo	87 %	Favorite TV series	66 %	Quote	16 %
Likes	84 %	Favorite films	56 %	Interest in a particular person	16 %
Education	83 %	Friend's posts on Timeline	56 %	Favorite sports	14 %
Email	82 %	Gallery of Photos	53 %	Bio	12 %
Events	81 %	Favorite Athletes	50 %	Religion	10 %
Own Timeline Posts	81 %	Favorite Athlete's Teams	45 %	Favorite inspirational person	8 %
Videos	77 %	Relationship	44 %	Politics	7 %
Photos	76 %	Work	36 %	Website	6 %

**Tab. 1.** A percentage of users that have publicly accessible information on their Facebook profiles.  
Source Authors.

For organizations, a very important source of information about candidates is the data from social media networks. Outcome of obtained data is that the 91% of the users have the number of friends as publicly accessible information. This information can HR managers use to see who the friends are and if they have a match. Afterward, they are able to acquire either good or bad references. 87% of users have the profile photography as publicly accessible information. It means that HR managers can use this information to verify who the applicant is and be more accurate when tracking their social media networks. On average users have 18 public photographs on their profile. Email address is publicly accessible information in 82% cases. HR managers can use this information to keep track of the user – Digital footprint.

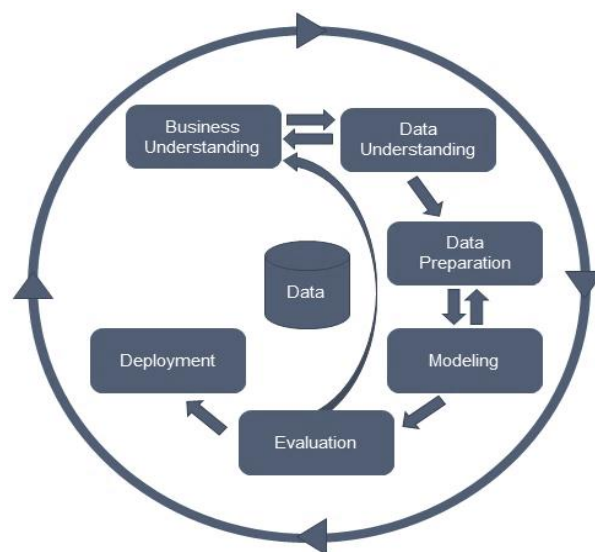
Posts on Facebook wall can be seen at 81% of users. This is very positive for HR managers due to the fact they can see a behavior of the candidate on the social media networks. They can see if the user's posts are polite and gather more behavioral information. For example, they can see if the person is emotionally unstable, etc. They can even see the construction of user's posts and find out if the user is thorough or the opposite. Also, the topics of the posts are very important.

76% of users are sharing on the Facebook information about visited place. This tells to HR managers how often the candidates travel. Pages and groups that people like and are members of or fans of give a picture of the user's hobbies and leisure activities. This is very important for the company's culture and further adaptation into the work-collective. This public

information about individuals can be very useful to create an objective image of the candidate in the recruitment process.

## 4 Model creation

The main goal of this work is to create a model for employee recruitment support, which will be based on data mining from social media networks. Therefore, for the purposes of this work, the authors were inspired by the **CRISP-DM methodology**<sup>5</sup>. This methodology serves as a unified framework that can solve various data mining tasks. The CRISP-DM methodology divides the whole modeling process into six basic stages, see Fig. 2. The outer circle in the figure symbolizes the cyclical nature of the process of knowledge acquisition from databases.



**Fig. 2.** CRISP-DM methodology. Source: (Chapman et al., 2000).

In the following subchapters, the individual phases of our recruitment model according to the CRISP-DM methodology are described in detail. Creation of the model is based on downloaded Facebook users data via the “Práce-na-míru” app created by the authors themselves.

### 4.1 Phase of business understanding

The phase of understanding the problem was carried out while defining the research problem and the main goal of this work.

The authors divided the data on the training and the testing part. Training data  $N = 960$  (see part 3.1 for more details) have been used in order to create a model PM<sup>6</sup> (see Fig. 11). Created model has been verified on testing data  $N = 198$  (see part 4.5).

### 4.2 Phase of understanding the data

<sup>5</sup> Shortcut CRISP-DM means CRoss-Industry Standard Proces for Data Mining.

<sup>6</sup> PM stands for Práce na Míru.

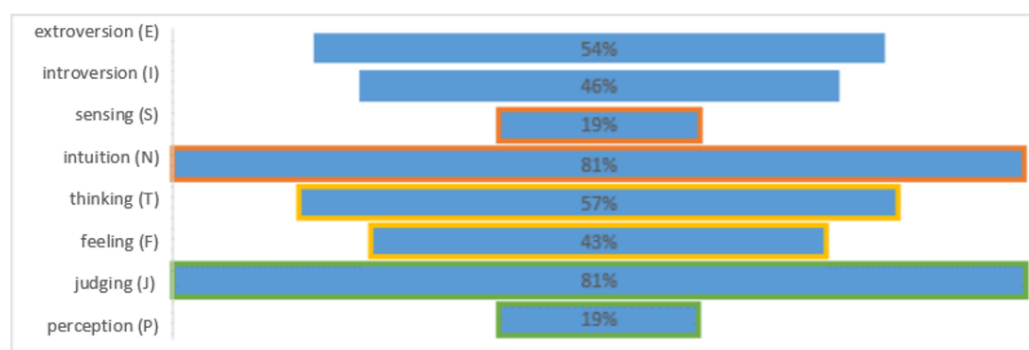
The phase of data understanding follows up the first phase. “Práce na míru” application has downloaded a lot of information about users from the Facebook, see Table 1. In order to evaluate their behavior from a recruitment point of view, it is necessary to determine the appropriate parameters. In terms of recruitment the best predictors are such that goes out directly from the personality test. Therefore, it was necessary to specify the requirements and choose suitable test of dependency of model purpose, which are:

- evaluation of personal characteristics,
- evaluation of interpersonal characteristics,
- evaluation of work characteristics,
- relevancy for recruitment,
- speed,
- transparency,
- option to fill the test online from everywhere,
- immediate evaluation without other expenses (e.g psychologist).

The requirements stated above are in an agreement with the MBTI personality test<sup>7</sup> (Mattare, 2015; Fretwell, Lewis & Hannay, 2013). In practice this test is usually used in Human Resources. It is used while creating job positions and recruiting. It is a part of psychological tests. The MBTI test determines personality type of potential candidates. Everything in this test is based on a combination of four basic characteristics groups (Myers, Mccauley & Most, 1985):

- perception of surrounding environment – extroversion (E) / introversion (I),
- way of obtaining information – sensing (S) / intuition (N),
- way of evaluating information – thinking (T) / feeling (F),
- life style – judging (J) / perception (P).

Target group that has registered into the “Práce na míru” app had been sent a MBTI test. Fig 3. shows the categorization results. Difference amongst the extrovert and introvert group of users seems to be balanced also in connection to thinking and feeling. The huge difference is amongst sensing and intuition in connection to judging and perception. These results are matching job offers that are relevant for this target group (Myers, Mccauley & Most, 1985; Böhmová & Vrnáková, 2015).



*Fig. 3. Division of criteria of particular groups. Source Authors.*

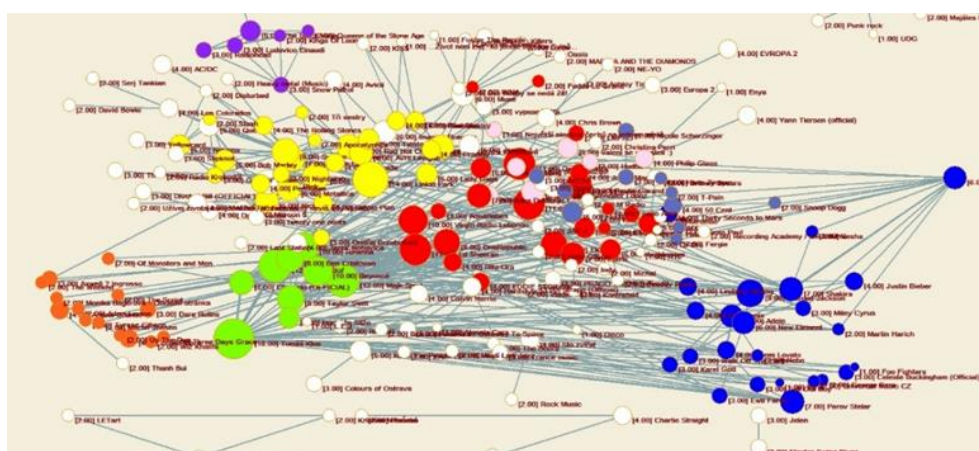
### 4.3 Phase of data preparation

<sup>7</sup> Myers-Briggs Type Indicator

The preparation of the data was based on selected analytical tool named Pajek, see (Pajek, 2017). For our cause this software serves as support of cluster analysis. This method has been chosen primarily due to the fact that there are too many unique values that are very similar for many attributes (see Table 1). Authors have used the hierarchical clustering method (Žambochová, 2008), called the Ward method (Mrvar & Batagelj, 2017).

After pre-processing, the authors performed segmentation of users into clusters that are used in the Social media recruitment model (see Fig. 11) as MBTI category predictors. Due to the fact there was a large amount of data, it was necessary to choose only clusters that had so-called “telling strength” as predictors<sup>8</sup>. From the 28 possible attributes (see Table 1), the authors identified 4 with the most prominent strength as predictors (specific interest categories: favorite music, favorite TV series, favorite movie and favorite athlete).

Graphical output is a graph<sup>9</sup>, which uses colors to highlight created clusters for an attribute such as favorite TV series, see Fig. 4. The more significant the cluster is, the bigger the point is. Colors indicate the cluster that the item belongs to.



**Fig. 4.** Clusters according to the connection for attribute favorite TV series. Source Authors.

The network graph is unreadable at the level of individual items, for example, the table below shows a list of clusters for the TV series' favorite attribute (see Table 2). Each cluster contains dozens of specific items, so for each cluster only three of the most common items are listed. The feature favorite TV series is represented by eight clusters that make up quite logical units, such as the F cluster is American popular sitcoms, cluster D represent Czech entertainment shows.

Cluster A	Cluster B	Cluster C	Cluster D	Cluster E	Cluster F	Cluster G	Cluster H
Prague Guide	A DOST	Červený trpaslík	StarDance	Game of Thrones	Gilmore Girls	The Big Bang Theory	Peklo na talíři
Man vs. Wild	Liga mistrů na Prima COOL	Knock knock: Penny?	Tvoje tvář má známý hlas	Partička	Dr. House	The Simpsons	Otázky Václava Moravce

<sup>8</sup> The authors used cluster analysis only for data of target group that was publicly accesible at least in 50 % of cases (see Table 1).

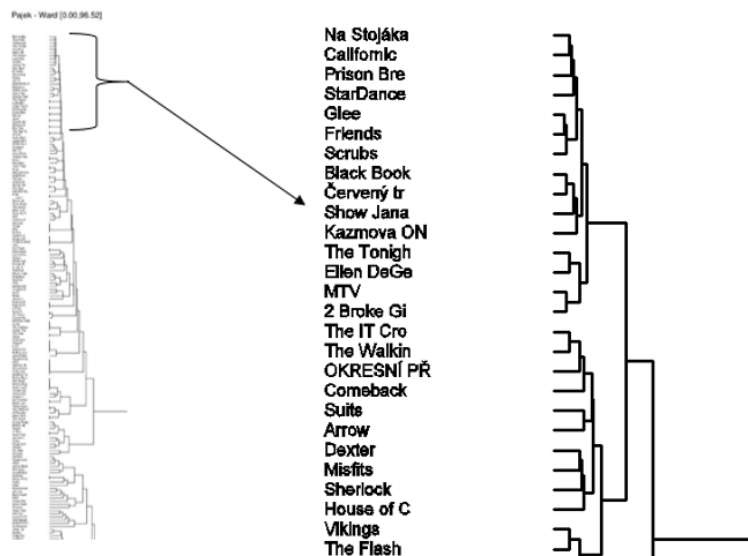
<sup>9</sup> Size of the points is determined by connectivity which is the number of nodes available from one node to one, two or three steps. (Mrvar & Batagelj, 2017)



Hlášky z Dva a půl chlapa	Vyšehrad	S Italem v Kuchyni	Show Jana Krause	FRIENDS	3v1	How I Met Your Mother	Hyde Park
---------------------------------	----------	-----------------------	---------------------	---------	-----	--------------------------	-----------

**Tab. 2.** Clusters overview for attribute favorite TV series. Source Authors.

Next possible visualization of clusters is with a help of dendrogram, see Figure 5.



**Fig. 5.** Dendrogram for attribute favorite TV series. Source Authors.

## 4.4 Modelling phase

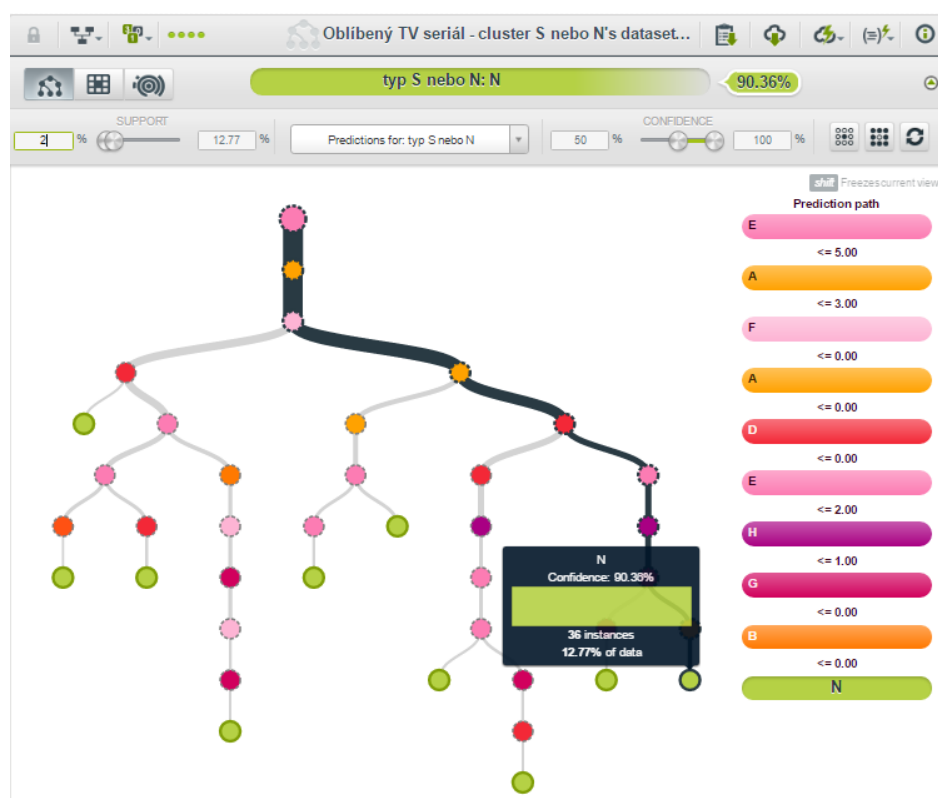
In this phase, we sculpture a decision trees with a help of the BigML tool (BigML, 2017b). “BigML is a consumable, programmable, and scalable Machine Learning platform that makes it easy to solve and automate Classification, Regression, Anomaly Detection, Association Discovery, and Topic Modeling tasks.” (BigML, 2017a) The reason why we used this tool is that it is very intuitive, user friendly and can create attractive graphical output of models.

The decision trees are chosen by the authors because they are a machine learning tool designed for classification and prediction tasks. Machine learning provides a number of more complex algorithms for classifying and predicting variables. The authors chose decision trees for several reasons. First, they process both categorical and numerical variables. Furthermore, it is relatively easy to find nonlinear relationships between input attributes. Another reason is that the result of the decision tree can be graphically represented and interpreted.

From possible algorithms<sup>10</sup> for a creation of decision trees the most suitable solution for the purpose of this work is the CART (Classification and Regression Trees) algorithm that generates the binary tree, a decision tree where each parent node has two child nodes (Žambochová, 2008). This algorithm is used in case that we have one or more independent variables (continuous or categorical). Next, we need to have one dependent variable, which can also be continuous or categorical. At each step, the algorithm goes through all possible divisions using all the values of all the independent variables and searches for the best of these divisions.

<sup>10</sup> Many algorithms exist for making decision trees. The most used are ID3, C4.5, AID, CHAID, QUEST, and CART. (Žambochová, 2008)

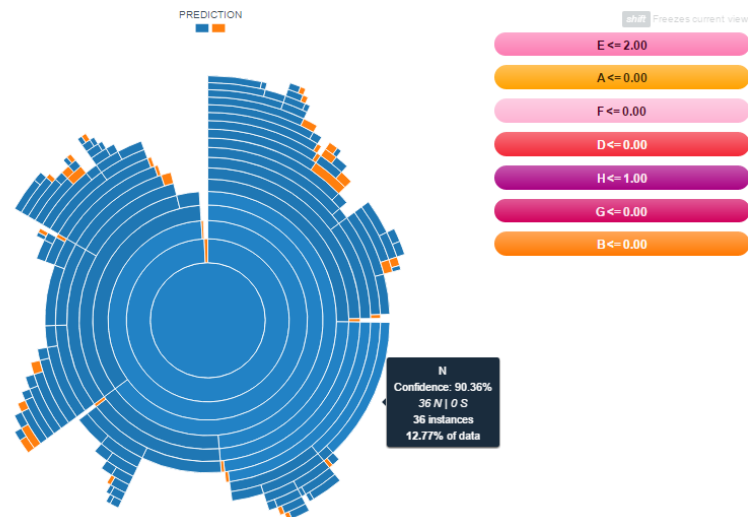
For each target area (attributes: favorite book, favorite music, favorite TV series, favorite movie), the authors created decision trees that determine one of the personality categories MBTI tests, in total 16 decision trees. Trees represent the absolute frequency of occurrence in a given cluster for each user. A tree trained in historical data can be used to predict who most likely fall into a category of personality type. Fig. 6 displays the decision tree is composed from clusters as the favorite TV series attribute. A key transformation here is the individual MBTI group criteria. Fig. 6 specifically illustrates the category - obtaining information. The beginning of the tree shows that the key factor is cluster E. On the figure below is the description of the branch that is bold and grayed out: If a user on his Facebook's profile has marked his favorite TV series falling within the E and H clusters and he did not mark one TV series that falls into A, F, D, G, B, then he fits with 90.36 % confidence<sup>11</sup> into characteristics (N – intuition) from MBTI personality categories. In this way, we can easily read the rest of the branches of the tree.



**Fig. 6.** Decision tree (favorite TV series) for MBTI category – obtaining information. Source Authors.

Another display form is the beam graph, see Fig. 7, which shows the decision tree in total, with a proportional representation of the number of people (according to the circle segment) and the level of confidence for a given personality category. Individual circles represent the path of the tree through the number of occurrences in each cluster. The color represents the level of confidence in the decision tree according to the personality type (blue color means N – intuition, orange color means S – sensing).

<sup>11</sup> The quality of the decision trees indicates the percentage of confidence. The higher the percentage, the more accurate the decision tree is.



**Fig. 7.** Beam graph (favorite TV series) for MBTI category – obtaining information. Source Authors.

While using BigML, you can detect data types of individual columns and divide data into separate instances. In the next step, it is possible to use the selected number of instances to create a model above in which predictions can be made. Fig. 8 thus shows the form of the predictive model for determining the MBTI type according to the clusters of particular data categories (here specifically for the popular TV series), exported to MS Excel. Into yellow cells is possible to input values, specifically how many times and in which clusters the particular candidate fits. Afterward, the tab counts a probability of personality type (based on the MBTI test).

Obilbeny TV serial - cluster S nebo N's dataset's model						Powered By BigML			
Use the cells in yellow below to input data to generate a new prediction									
A	B	C	D	E	F	G	H	Prediction: typ S nebo N	Confidence
								N	75,48%
Predicate	Node Output	Node Instan	Node Support	Node Confide	Prediction	Prediction	Confidence		
if 'E' is missing	N	282,00	1,00	0,75	N		75,48%		
if 'E' > 5	S	2,00	0,01	0,34	N		76,07%		
if 'E' <= 5		280,00	0,99	0,76	N		76,07%		
if 'A' is missing	N	280,00	0,99	0,76	N		76,07%		
if 'A' > 3	S	1,00	0,00	0,21	N		76,38%		
if 'A' <= 3		279,00	0,99	0,76	N		76,38%		
if 'F' is missing	N	279,00	0,99	0,76	N		76,38%		
if 'F' > 0		183,00	0,63	0,67	N		78,17%		
if 'D' is missing	N	96,00	0,34	0,67	N		66,61%		
if 'D' > 3	N	8,00	0,03	0,68	N		63,82%		
if 'D' <= 3		88,00	0,31	0,64	N		63,82%		
if 'E' is missing	N	88,00	0,31	0,64	N		63,82%		
if 'E' > 0		36,00	0,13	0,68	N		47,36%		
if 'E' is missing	N	52,00	0,18	0,68	N		68,10%		
if 'E' > 1		25,00	0,09	0,52	N		75,03%		
if 'C' is missing	N	27,00	0,10	0,52	N		51,52%		
if 'C' > 0		6,00	0,02	0,41	N		60,97%		
if 'B' is missing	N	21,00	0,07	0,41	N		40,88%		
if 'B' > 0	N	3,00	0,01	0,44	N		33,72%		

**Fig. 8.** Form of predictive model for MBTI. Source Authors.

## 4.5 Evaluation phase

### Formal verification of the model to support recruitment

On the training data (N = 960), the model PM learned how to make the decision about personality types. Model PM validation was run on new data collected from users who had

signed up for the “Práce na míru” application between February and March 2017. There were 198<sup>12</sup> people in the target group.

Verification of the model on test data confirmed the accuracy of the prediction. The MBTI personality category is placed in range of 68 % to 84 % in individual cases with confidence levels of 43 % to 81 %. These numbers show high reliability of PM model’s outcomes. This model is used in the next chapter.

## 4.6 Deployment phase

The final phase is the deployment of the model PM in the real usage for yet mentioned recruiting purposes at RPC VŠE and xPORT VŠE Business Accelerator. “Práce na míru” application will still be active for students and alumni of VSE University. Students taking part in this project will receive relevant job offers. These offers should be tailored for them related to the results of their character type according to the model PM.

The environment for which the model PM has been created is constantly changing. The model needs to be continuously checked, expanded and updated to maintain reliability and accuracy.

## 5 Social media recruitment model

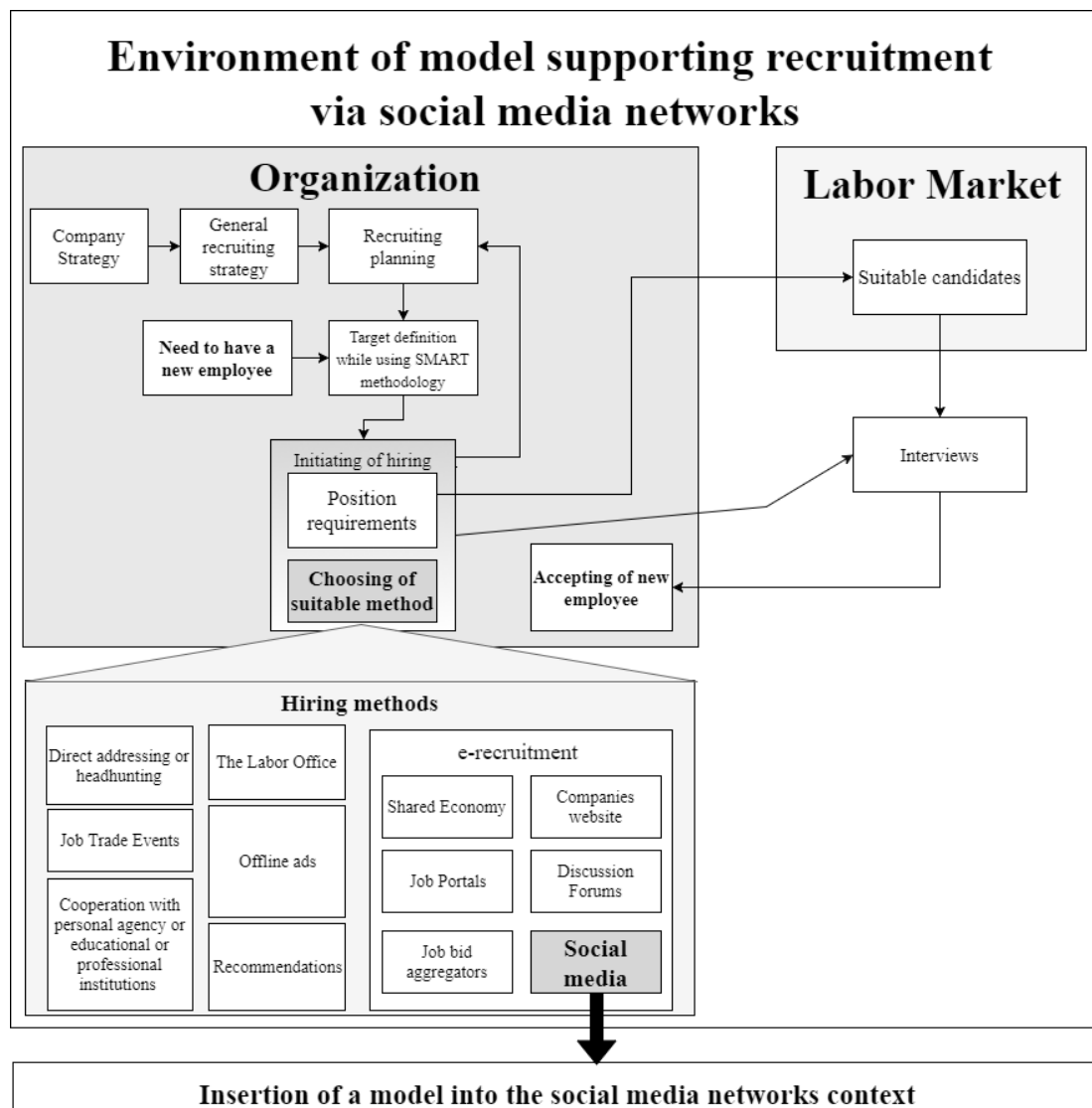
The purpose of this chapter is to meet the goal of this article, which is to create a model for employee recruitment. The graphical form of the general model was based on the previous chapter 4. Model creation. For a better understanding of how the model is embedded into an organization and its surroundings, everything is illustrated graphically below in Fig. 9. Furthermore, the deployment of the model into the context of social media networks is shown visually, see Fig. 10.

### Model embedding to support recruitment in an organization and its surroundings

Close environment of the recruitment model is the relevant labor market where all the candidates are located. The goal of any organization is to invite them for an interview. In order to find suitable candidates, an organization must initiate a recruitment process, which includes, among other things, the selection of a suitable recruitment method. Fig. 9. indicates that there are many recruitment methods. One method is using social media networks. The model supporting recruitment takes advantage of that potential.

---

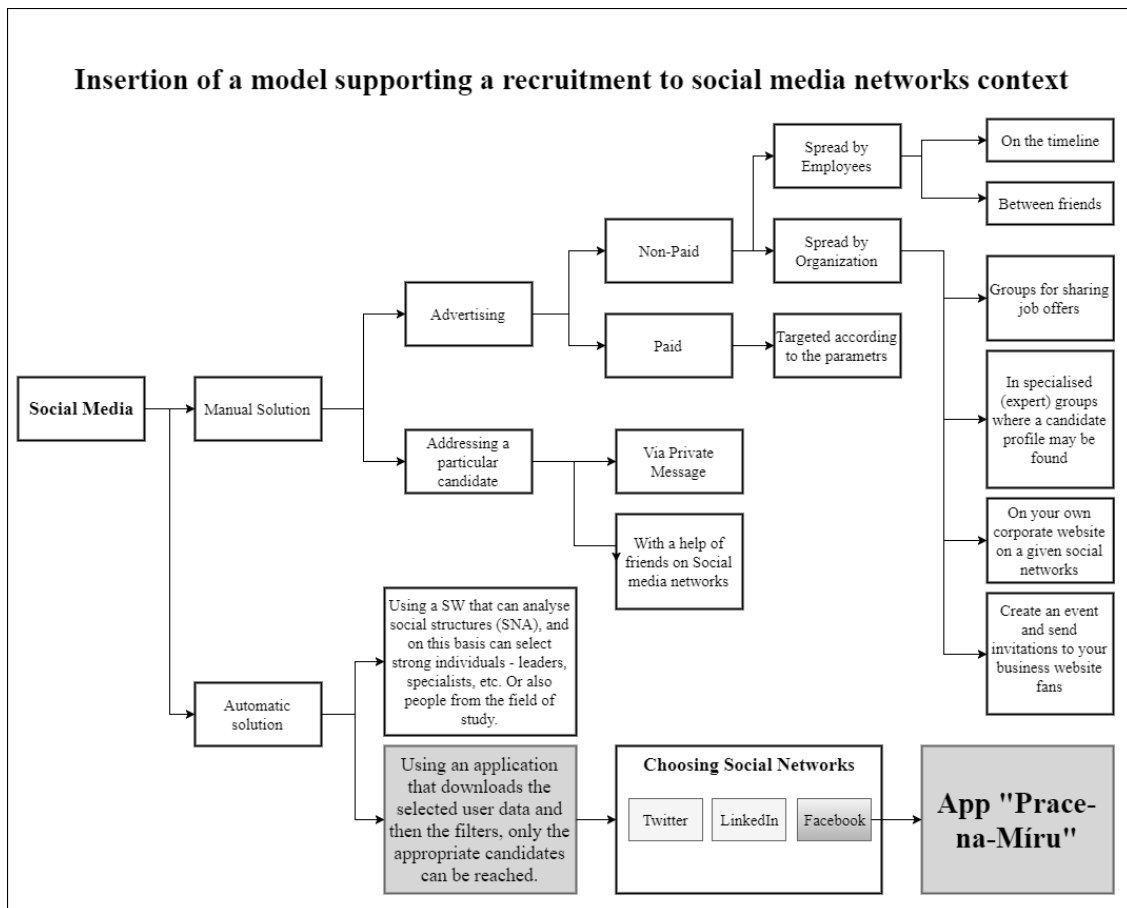
<sup>12</sup> Testing data.



**Fig. 9.** Environment of model supporting recruitment via social media networks. Source: Authors.

## Model in the social media context

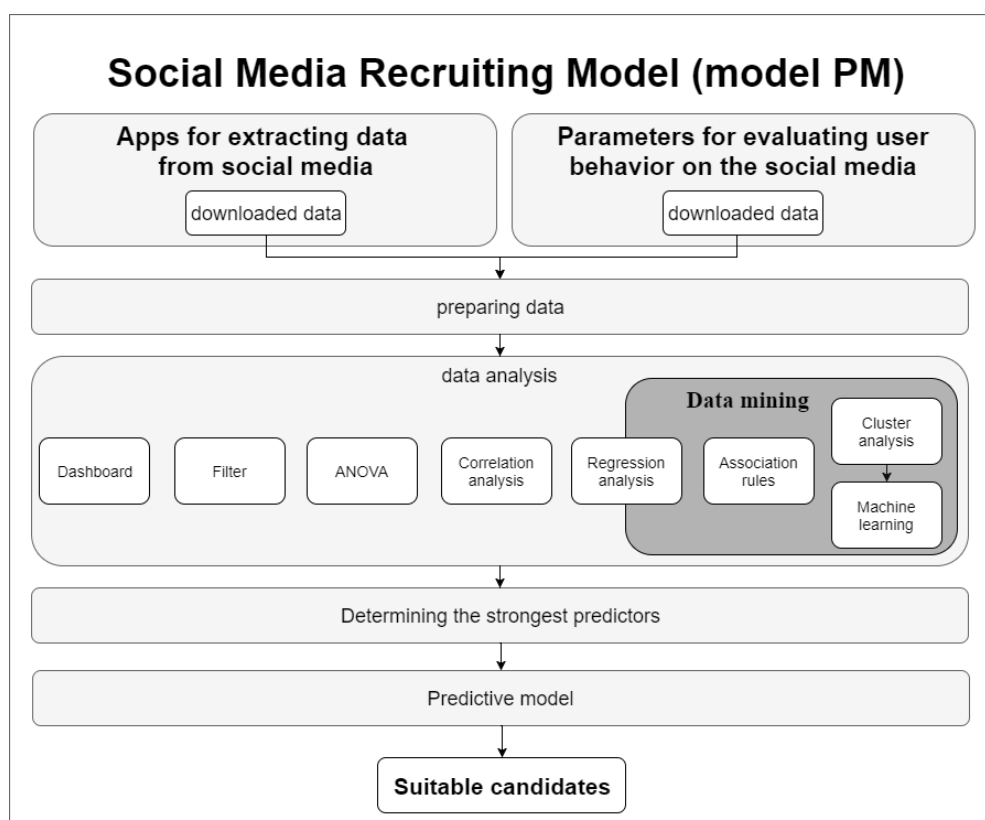
Fig. 10 illustrates how an organization use social media networks for recruitment through a process map. The organization can do it in two ways. One is a manual solution, which is described in more details below. The second option is an automated recruitment solution, which includes, among other things, the possibility of creating a custom recruitment application tailored to the needs of an organization, so the authors named it “Práce na míru”.



**Fig. 10.** Insertion of a model into the social media networks context. Source Authors.

### Social media recruiting model (model PM)

The Model to support recruitment via social media networks is illustrated graphically below, see Fig. 11. The model consists of an application for automated download of user's data from social media networks and also parameters and predictors for evaluating user's behavior. It also includes a predictive model that evaluates predictors. This application must have an open API, relevant user's information and must be useful from the recruiting point of view. To determine useful predictors, it is necessary to perform data analysis that is using appropriate analytical tools (e.g. cluster analysis, regression analysis, ANOVA etc.). The social media recruitment model should make it easier to find suitable candidates for organizations while using the predictive model.



**Fig. 11.** Social media recruiting model (model PM). Source Authors.

## 6 Discussion

It is clear from the modeling process that the application of custom created model requires deeper technical and analytical knowledge. That is why the persons in charge of recruitment (or HR department) need to already obtain a functional instance of the artifact that will be based on their needs for the given segment or specific job positions and also the target group of candidates. The usage of an artifact instance must be very intuitive and fast, with no additional cost.

Organizations can choose any social media network that has an open API to create a recruiting application with data extraction ability. Additionally, they can choose a personality test or other typological-evaluating test form to evaluate user's behavior in order to determine parameters to evaluate predictors. This is also related to the selection of an appropriate analytical method.

One of the possible failures in recruiting via social media networks could be a false identity of a user who can purposefully create or modify his profile according to the requirements for the position. This is typical, for example, for LinkedIn, which serves primarily for recruitment purposes. That is why the human factor is always important in the form of a physical interview (personally or remotely) with HR or other authorized person. Organizations may also experience mistrust of candidates in their recruitment application and a dissuasive attitude towards providing their data. Other possible limitations that organizations will have to deal with in terms of social media network recruitment arise from GDPR rules across the European Union, which implies more rights for candidates, more responsibility for data controllers. (OJEU, 2016)

### Benefits of the model:

- Satisfying of informational needs of an organization while recruiting.
- Filling a gap in existing models for recruiting.
- Prediction of personality type based on candidate's behavior on social media networks.
- Analysis of existing data on social media networks, categorization and description how they can be obtained (automatically or manually).

**The limitations** of this model arise from several areas. The basic limitations are the scope of work, focusing only on the Czech labor market and the sustainability of the outputs as it is a rapidly changing and constantly evolving interdisciplinary topic.

Social media recruiting model (PM) is not suitable for finding and evaluating all people on the labor market, but only for those who have an account on covered social media networks. The model also does not ensure finding suitable candidates, but it only selects from people who are registered in an application that is extracting user's data on a given social media network. At the same time, the model is affected by the segment of users who log in to the application.

A necessary condition for selecting right social media network that can be used for the model is the openness of the social media network in terms of development environment (API). Only if this condition is matched then the proposed application for extracting and mining the user's data can be used.

Legislation is a major limitation, which makes it impossible to use all available social information in the practical application of the model. The authors are aware of possible model distortions, despite testing the model on real data. Model distortions may be a false correlation type, development sequence, or missing intermediate member (Molnár et al., 2012).

### Possible ideas for further research are:

- Create an automated solution for other social media networks like LinkedIn and Twitter.
- Create a comprehensive methodology to support recruitment through social media networks.
- Adding a dictionary for emotional colored words into the model.

## 7 Conclusion

Data from social media networks is an important addition to information about candidates for organizations. The results of the research on publicly accessible information on Facebook have shown that the target group of users has on their profiles much useful information for recruitment purposes.

The authors used the MBTI personality test to help diagnose the characteristics of the candidates, such as the perception of the environment or the way information is collected. The model includes an automated solution, application named "Práce na míru", for user's data mining, specifically from the Facebook. Also, it includes a proposal for analytical data processing, specifically the model describes a cluster analysis, decision trees and a predictive model to determine a personality type (based on the MBTI test).



## References

- Annisette, L. E., & Lafreniere, K. D. (2017). Social media, texting, and personality: A test of the shallowing hypothesis. *Personality and Individual Differences*, 115, 154-158. doi: [10.1016/j.paid.2016.02.043](https://doi.org/10.1016/j.paid.2016.02.043)
- Bartakova, G. P., Brtkova, J., Gubiniová, K., & Hitka, M. (2017). Actual trends in the recruitment process at small and medium-sized enterprises with the use of social networking. *Economic Annals-XXI*, 164(3-4), 80-84. doi: [10.21003/ea.V164-18](https://doi.org/10.21003/ea.V164-18)
- Bender, J. L., Cyr, A. B., Arbuckle, L., & Ferris, L. E. (2017). Ethics and privacy implications of using the internet and social media to recruit participants for health research: A privacy-by-design framework for online recruitment. *Journal of Medical Internet Research*, 19(4), e104. doi: [10.2196/jmir.7029](https://doi.org/10.2196/jmir.7029)
- BigML. (2017a). About BigML. Retrieved September 6, 2017, from: <https://bigml.com/about>
- BigML. (2017b). BigML. Retrieved September 6, 2017, from: <https://bigml.com/>
- Böhmová, L., & Malinová, L. (2013). Facebook user's privacy in recruitment process. In *Proceedings of the IDIMT 2013 – Information Technology Human Values, Innovation and Economy* (pp. 159-166). Linz: Trauner Verlag.
- Böhmová, L., & Vrnáková, I. (2015). Uplatnění absolventů VŠE a hodnocení absolvovaného studia. Na základě průzkumů absolventů z let 2005/06, 2010/11, 2011/12, 2012/13, 2013/14 a 2014/15. Retrieved September 6, 2017, from: [https://absolventi.vse.cz/wp-content/uploads/2013/03/160118\\_Dotazn%C3%ADk.pdf](https://absolventi.vse.cz/wp-content/uploads/2013/03/160118_Dotazn%C3%ADk.pdf)
- Böhmová, L., Mcloughlin, S., & Střížová, V. (2016). How time changes human resources's view on social networking sites as recruiting tool in the Czech Republic. In *Proceedings of the System approaches'16 – Responsible development of systems* (pp. 80-86). Prague: Oeconomica.
- Chapman, P., Clinton, J. C., Kerber, R., Khabaza, T., Reinartz, T., Shearer C., & Wirth, R. (2000). CRISP-DM 1.0, Step-by-step data mining guide. Retrieved September 6, 2017, from: <ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf>
- ČSÚ. (2013). Projekce obyvatelstva České republiky. Retrieved September 6, 2017, from: <https://www.czso.cz/csu/czso/projekce-obyvatelstva-ceske-republiky-do-roku-2100-n-fu4s64b8h4>
- ČSÚ. (2015). Šetření o využívání ICT v domácnostech a mezi jednotlivci. Retrieved September 6, 2017, from: [https://www.czso.cz/documents/10180/32631125/061004-15\\_C.pdf/eccbfabd-8824-45e6-acf5-b7a4c710211a?version=1.1](https://www.czso.cz/documents/10180/32631125/061004-15_C.pdf/eccbfabd-8824-45e6-acf5-b7a4c710211a?version=1.1)
- ČSÚ. (2016). Věkové složení obyvatelstva – 2015. Retrieved September 6, 2017, from: <https://www.czso.cz/csu/czso/vekove-slozeni-obyvatelstva>
- Fretwell, CH. E., Lewis, C. C., & Hannay, M. (2013). Myers-Briggs Type Indicator, A/B Personality Types, and Locus of Control: Where Do They Intersect?. *American Journal of Management*, 13(3), 57-66.
- HRNews. (2016). Twitter pomáhá hledat práci. Retrieved September 6, 2017, from: <http://www.hrnews.cz/lidske-zdroje/trendy-id-148711/twitter-pomaha-hledat-praci-id-738597>
- Jobvite. (2014). Social Recruiting Survey. Retrieved September 6, 2017, from: [https://www.jobvite.com/wp-content/uploads/2014/10/Jobvite\\_SocialRecruiting\\_Survey2014.pdf](https://www.jobvite.com/wp-content/uploads/2014/10/Jobvite_SocialRecruiting_Survey2014.pdf)
- Khatri, C., Chapman, S. J., Glasbey, J., Kelly, M., Nepogodiev, D., Bhangu, A., & Fitzgerald, J. E. (2015). Social media and internet driven study recruitment: Evaluating a new model for promoting collaborator engagement and participation. *PLoS ONE*, 10(3), e0118899. doi: [10.1371/journal.pone.0118899](https://doi.org/10.1371/journal.pone.0118899)
- Knime. (2017). Knime. Retrieved September 6, 2017, from: <https://www.knime.com/>
- Mattare, M. (2015). Revisiting Understanding Entrepreneurs Using the Myers-Briggs Type Indicator. *Journal of Marketing Development and Competitiveness*, 9(2), 114-119.
- McAbee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review*, 27(2), 277-290. doi: [10.1016/j.hrmr.2016.08.005](https://doi.org/10.1016/j.hrmr.2016.08.005)
- Meister, J. C., & Willyerd, K. (2010). *The 2020 workplace: how innovative companies attract, develop, and keep tomorrow's employees today*. New York: Harper Business.
- Molnár, Z. (2011). Jak využít sociální sítě v podnikání. *Systémová integrace*, 18(1), 134-154.
- Molnár, Z., Mildeová, S., Řezanková, H., Brixí, R., & Kalina, J. (2012). *Pokročilé metody vědecké práce*. Zeleneč: Profess Consulting.
- MPSV. (2016). Hledání zájemců o práci. Retrieved September 6, 2017, from: <http://portal.mpsv.cz/sz/zamest/hledejinz>

- Mrvar, A., & Batagelj, V.** (2017). Programs for Analysis and Visualization of Very Large Networks. Pajek. Retrieved September 6, 2017, from: <http://mrvar.fdv.uni-lj.si/pajek/pajekman.pdf>
- Muntinga, D., Moorman, M., & Smit, E.** (2011). Introducing COBRAs. *International Journal of Advertising*, 30(1), 13-46. doi: [10.2501/IJA-30-1-013-046](https://doi.org/10.2501/IJA-30-1-013-046)
- Myers, I. B., Mccauley, M. H., & Most, R.** (1985). *Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*. Palo Alto: Consulting Psychologists Press.
- Newman, M. E. J.** (2016). *Networks: An introduction*. New York: Oxford University Press.
- OJEU.** (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council. Retrieved September 6, 2017, from: [http://ec.europa.eu/justice/data-protection/reform/files/regulation\\_oj\\_en.pdf](http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf)
- Ong, V., Rahmanto, A. D. S., William, & Suhartono, D.** (2017). Exploring personality prediction from text on social media: A literature review. *Internetworking Indonesia Journal*, 9(1), 65-70.
- Pajek.** (2017). Pajek: analysis and visualization of large networks. Retrieved September 6, 2017, from: <http://mrvar.fdv.uni-lj.si/pajek/>
- Park, G., et al.** (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934-952. doi: [10.1037/pspp0000020](https://doi.org/10.1037/pspp0000020)
- Pavlíček, A.** (2010). *Nová média a sociální sítě*. Praha: Oeconomica.
- Pavlíček, A.** (2016). Facebook – privacy settings and personal information disclosure. In *Proceedings of the IDIMT 2016 – Information Technology, Society and Economy Strategic Cross-Influences* (pp. 133-144). Linz: Trauner Verlag.
- PWC.** (2015). The Sharing Economy. Retrieved September 6, 2017, from: <https://www.pwc.com/us/en/technology/publications/assets/pwc-consumer-intelligence-series-the-sharing-economy.pdf>
- Roulin, N., & Levashina, J.** (2016). Impression management and social media profiles. In *Social media in employee selection and recruitment: Theory, practice, and current challenges* (pp. 223-248). Cham: Springer. doi: [10.1007/978-3-319-29989-1\\_11](https://doi.org/10.1007/978-3-319-29989-1_11)
- Sathya, R., & Indradevi, R.** (2017). Influence of social media in recruiting talents. *Man in India*, 97(4), 219-231.
- Scott, J.** (2000). *Social network analysis: a handbook*. Thousands Oaks: SAGE Publications.
- Tore, O., Agneessens, F., & Skvoretz, J.** (2010). Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths Social Networks. *Social Networks*, 32(3), 245-251. doi: [10.1016/j.socnet.2010.03.006](https://doi.org/10.1016/j.socnet.2010.03.006)
- Toušek, L.** (2015). Analýza sociálních sítí. In L. Toušek et al. (Eds.), *Kapitoly z kvalitativního výzkumu* (pp. 75-100). Plzeň: Západočeská univerzita v Plzni.
- Žambochová, M.** (2008). Data mining methods with trees. *E+M: Ekonomie a Management*, 11(1), 126-131.



Copyright © 2018 by the author(s). Licensee University of Economics, Prague, Czech Republic. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY), which permits use, distribution and reproduction in any medium, provided the original publication is properly cited, see <http://creativecommons.org/licenses/by/4.0/>. No use, distribution or reproduction is permitted which does not comply with these terms.

The article has been reviewed. | Received: 13 July 2017 | Accepted: 26 September 2017

Academic Editor: Stanislava Mildeova