International Journal of Scientific & Engineering Research, Volume 8, Issue 4, April-2017
ISSN 2229-5518

Social Recruitment: A novel technique to hire the potential knowledge workers

Kumar Nityan Suman[#1], Pragati[#2], Kanaad Pathak[#3], Animesh Giri[#4]

Department of Information Science and Engineering, PES Institute of Technology-Bangalore South Campus Bangalore, Karnataka – India

Abstract — In this model; we propose an innovative recruitment system based on job seeker's social networking profiles like Facebook, Twitter and LinkedIn. It is aimed to develop an advanced understanding to the recruiters regarding recruiting prospective candidates using various data mining and machine learning techniques. Using classic data mining techniques, user profiling and machine learning methods this will enable recruiters to effectively select appropriate job seeker for their organization in an automated manner.

Keywords — Facebook, LinkedIn, Twitter, Recruitment, Job Profile, Data Mining, Data Wrangling, Machine Learning, Logistic regression

1. INTRODUCTION

Recruiting is fundamental to the health and success of all businesses today. Finding, nurturing, and hiring the right talent is now a dynamic, diversified process, requiring recruiters to tap into social networks, analytics, mobile, and beyond. Traditionally the process involves grueling paper/pen based test followed by multiple other tests, group discussions, debates and interviews. There is a major flaw in this classic model of recruitment- a huge amount of time is devoted in the process, the amount of resources spent are very high plus the merit of a particular candidate is decided solely on his/her performance on the given day. Thus the quality of the performance measurement is on the lower side. Whereas social networks are free, practically everyone uses them already, and they're a good means for showcasing a brand and retrieving information. Not surprisingly, there are more than 2 billion people on social media sites like LinkedIn, Facebook & Twitter and its growing every second[4]. It is completely automated, saves time, cost effective and the candidate's potential is measured through his/her own contribution over a very long period of time which paints the digital picture of the candidate's potential.

Therefore, it is a continuous and a comprehensive assessment of the particular candidate. On an average every individual uploads 15% more data than his last year on these social networking websites every year[4]. Currently more than 65% recruiters use Facebook during recruiting process to assess candidate's brand[5]. More than 2 million jobs are posted on Twitter every month and 97.3% of the recruiters uses LinkedIn as a recruiting tool[5].

Then why not take it to the next level!

2. RELATED WORK

Social recruitment is recruiting candidates by using social platforms as talent database. Social recruiting is at the intersection of recruitment and social media. It uses social media profiles, blogs and online communities as a base to find the prospective candidates. Since 2009, there has been discussion in the recruitment and social media collaboration. According to Jobvite, an applicant tracking system (ATS) and social sourcing provider, released a report in May 2011 indicating that 80% of the employed surveyed said "Yes" to use of social media for recruitment. Companies like Ups, Twitter, Deloitte and npower moved their recruitment processes towards social media very aggressively last decade itself.

A) Facebook: Facebook is the largest of the social networking sites with over 1.5 billion monthly active users. [4] Facebook is a social utility that connects people with friends and others who work, study, and live around them. It is intended to connect friends, family, organizations, interests and business associates. 91% of millennial (15-34 years old) uses Facebook[4]. Daily active users in Asia itself are as high as

300 million[4].

B) Twitter: A popular micro blogging service enabling its users to send and read publicly visible messages called tweets. Tweets are text-based posts of up to 140 characters displayed on the users profile page. There are over 320 million users on Twitter, 34% of them login more than once in a day and with over 100 million monthly visitors[4].

International Journal of Scientific & Engineering Research, Volume 8, Issue 4, April-2017
ISSN 2229-5518

C) LinkedIn: LinkedIn is an interconnected network of experienced professionals from around the world, representing 170 industries and 200 countries territories (according to their website). You can find, be introduced to, and collaborate with qualified professionals that you need to work with to accomplish your goals." – LinkedIn. There are over 400 million users on LinkedIn with 100 million monthly unique visitors. LinkedIn is increasing at a speed of 2 users per second[4].

2.1 RECRUITMENT STATISTICS

In a competitive and complex job market, relationship reign supreme. Recruiters are facing an increasingly demanding and competitive talent market every year. From Survey conducted by Jobvite, completed by 1,404 recruiting and human resources professionals spanning several industries.

- 78% of recruiters find their best quality candidates through referrals.
- 69% of the recruiters have increased their initial salary offers in the last year.
- 88% of the recruiters report that culture fit is very important.
- 74% believes brand awareness is one of the biggest challenges to attract top talents and use social networks for it.
- 24% said that they expect to replace some jobs with technology in the next 2-3 years.
- Only mere 4% of the recruiters are not using social media in recruiting efforts. Mouth bobbling 72% of the recruiters believe data analytics is somewhat or very important in the hiring process.

2.2 TOOLS

There are a number of free and paid tools that can help you keep your ear's to the ground. LinkedIn Signal: A great tool to monitor and listen to what people are saying on LinkedIn outside of your network. Navigate from the "news" tab, add your key words and use the search fields to drill your results further. You can save searches and share them with people in your network too or use it for further analysis.

Topsy: Search the social web for conversations, photos, hash-tags, tweets etc. Search and compare up to 3 search queries at once and use the advanced search facility to narrow your results to specific web pages, times or languages.

Hashtagify: An open source tool to keep track of the candidates tweets on twitter. Recruiters can enter a hash-tag, and the tool will show related hash-tags with popularity trends, top influences using that hash-tag and usage patterns.

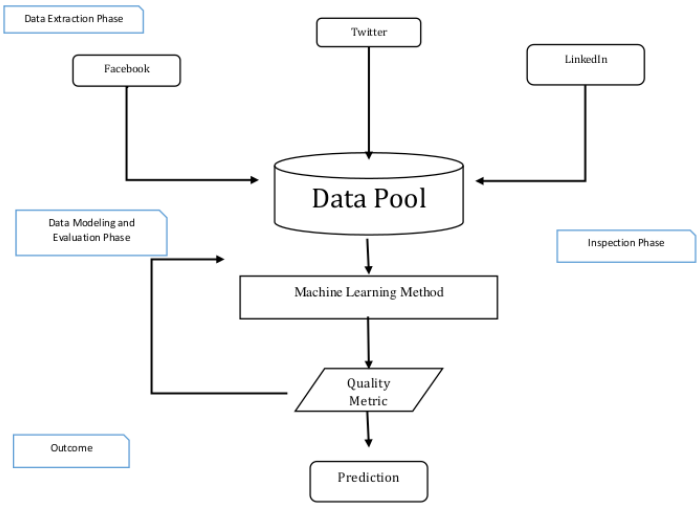
Entelo Search: It is a paid status analysis tool that predicts the trend of a candidate to change the job over 70 factors such as hiring trends, recent acquisitions, these candidates are six time more likely to leave their current job within 90 days than an average prospect.

Hootsuite: It is a paid (free- three profiles only) service that allows users to schedule posts on a variety of social media networks, and then get analytics on their success. Users can set up streams to do proper social listening of words, phrases and hashtags.

Workable: It is paid service for the recruiters looking for a user-friendly way to keep tabs on passive candidates. It saves the candidate's relevant profile information in the workable platform once you come across any candidate on the web – even on small sites like 'Behance'.

3. DESIGN AND METHODOLOGY

This section Talks about the design, methodology and process which makes up possible to automate the recruitment process which is based upon the social networking sites. The model clusters all relevant information about the prospective candidates from social networking sites Facebook, LinkedIn and Twitter.

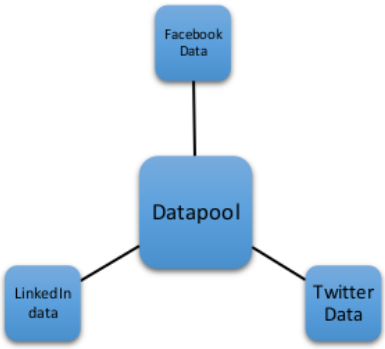


The data is analyzed using different data mining techniques and machine learning concepts to predict the status of the candidate—whether he/she is fit for the organization.

International Journal of Scientific & Engineering Research, Volume 8, Issue 4, April-2017
ISSN 2229-5518

3.1 DATA EXTRACTION AND CLEANING

This section talks about the data extraction using data mining techniques and different API's provided by the networking sites.



LinkedIn: The most important parameter that any employer would check for is the skills obtained by the applying candidate. Skills are derived from users profile mentioned by him/her over a period of time with particulars like date, time, place and durations. It also collects relevant information about the connections formed by the candidate with other professionals over time.

Algorithm3.1.1: Data from LinkedIn

1. Setup an app on LinkedIn's developer portal which requests for data.

2. Authorize the app using OAuth mechanism.

3. Profile API & JavaScript SDK will fetch data once user logs in and gives suitable permissions.

4. Clean the data, removing unwanted constraints and formatting accordingly.

Facebook: The data collected is focused on the cultural and personal front of the candidate which plays an important role to decide the fit of the candidate in the organization. It will also help to enhance the understanding of different social skills, communication skills, likes and dislikes.

Algorithm 3.1.2: Retrieving from Facebook

1. Setup a valid account on Facebook's developer portal which requests for data.

2. Using Facebook API's (Graph, Public Feed, and Keyword Insights) extract data.

3. Remove unwanted fields and format the data in a particular format.

4. Proficiency or dexterity required is acquired through training or experience.

Twitter: The data extracted from Twitter will enable the digital picture of the candidate in the world. The data will help achieve the goal towards a professionally fit candidate. It will help understand the ideology and beliefs of the candidate. The social and political aspects of his/her life. As every organization wants a workforce that believes in their values and follow it in the real life and not only on the paper.

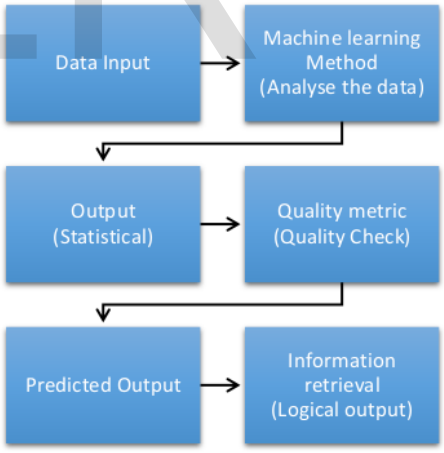
Algorithm 3.1.3: Extraction from Twitter

1. Setup an app on twitter's developer portal which requests for data in need.

2. Authorize the app using OAuth mechanism and log in through a valid official twitter account.

3. REST API's will fetch data once suitable permissions are given.

4. Proficiency, facility that is required is developed through training or experience.



3.2 ANALYZING DATA USING MACHINE LEARNING CONCEPTS

In this regression model we can consider more than one constraints and using several independent variables it is possible to come up with a relation between independent variable and the dependent variable (normally which is the output). An explanation of logistic regression can begin with an explanation of the standard logistic function.

International Journal of Scientific & Engineering Research, Volume 8, Issue 4, April-2017
ISSN 2229-5518

The logistic function is useful because it can take an input with any value from negative to positive infinity whereas the output always takes.

And the logistic function now is:

$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

We can now define the inverse of the logistic function, the logit (log odds):

$$g(F(x)) = \ln\left(\frac{F(x)}{1 - F(x)}\right) = \beta_0 + \beta_1 x,$$

And equivalently, after exponentiation both sides:

$$\frac{F(x)}{1 - F(x)} = e^{\beta_0 + \beta_1 x}.$$

Interpretation of these:

1. g() refers to the logit function. The equation for g(F(x)) illustrates that the logit (i.e., log-odds or natural logarithm of the odds) is equivalent to the linear regression expression.

2. ln denotes the natural logarithm.

3. F(x) is the probability that the dependent variable equals a case, given some linear combination of the predictors. The formula for F(x) illustrates that the probability of the dependent variable equaling a case is equal to the value of the logistic function of the linear regression expression. 4.

4. B0 is the intercept from the linear egression equation (the value of the Values between zero and one and hence is interpretable as a probability.

The logistic function is defined:

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

3.2.1 DEFINITION OF THE ODDS

The odds of the dependent variable equaling a case (given some linear combination x of the predictors) is equivalent to the exponential function of the linear regression expression. This illustrates how the logit serves as a link function between the probability and the linear regression expression.

Given that the logit ranges between negative and positive infinity, it provides an adequate criterion upon which to conduct linear regression and the logit is easily converted back into the odds. So we define odds of the dependent variable equaling a case (given some linear combination x of the predictors) as follows:

$$\text{odds} = e^{\beta_0 + \beta_1 x}.$$

For a continuous independent variable the odds ratio can be defined as:

$$\text{OR} = \frac{\text{odds}(x+1)}{\text{odds}(x)} = \frac{\frac{F(x+1)}{1-F(x+1)}}{\frac{F(x)}{1-F(x)}} = \frac{e^{\beta_0 + \beta_1 (x+1)}}{e^{\beta_0 + \beta_1 x}} = e^{\beta_1}$$

3.2.2 MODEL FITTING

The regression coefficients are usually estimated using maximum likelihood estimation. In some instances the model may not reach convergence. Non convergence of a model indicates that the coefficients are not meaningful because the iterative process was unable to find appropriate solutions. A failure to converge may occur for a number of reasons: having a large ratio of predictors to cases, multicollinearity, sparseness, or complete separation.

- Having a large ratio of variables to cases results in an overly conservative Wald statistic (discussed below) and can lead to non-convergence.
- Multicollinearity refers to unacceptably high correlations between predictors. As multicollinearity increases, coefficients remain unbiased but standard errors increase and the likelihood of model convergence decreases. To detect multicollinearity amongst the predictors, one can conduct a linear regression analysis with the predictors of interest for the sole purpose of examining the tolerance statistic used to assess whether multicollinearity is unacceptably high.
- Sparseness in the data refers to having a large proportion of empty cells (cells with zero counts). Zero cell counts are particularly problematic with categorical predictors. With continuous predictors, the model can infer values for the zero cell counts, but this is not the case with categorical predictors. The model will not converge with zero cell counts for categorical predictors because the natural logarithm of zero is an undefined value, so that final solutions to the model cannot be reached. To remedy this problem, researchers may collapse categories in a

International Journal of Scientific & Engineering Research, Volume 8, Issue 4, April-2017
ISSN 2229-5518

theoretically meaningful way or add a constant to all cells.

- Another numerical problem that may lead to a lack of convergence is complete separation, which refers to the instance in which the predictors perfectly predict the criterion – all cases are accurately classified. In such instances, one should reexamine the data, as there is likely some kind of error.

As a rule of thumb, logistic regression models require a minimum of about 10 events per explaining variable (where event denotes the cases belonging to the less frequent category in the dependent variable).

In logistic regression analysis, deviance is used in lieu of sum of squares calculations. Deviance is analogous to the sum of squares calculations in linear regression and is a measure of the lack of fit to the data in a logistic regression model. When a "saturated" model is available (a model with a theoretically perfect fit), deviance is calculated by comparing a given model with the saturated model.

$$D = -2\ln\frac{\text{likelihood of the fitted model}}{\text{likelihood of the saturated model}}.$$

Two measures of deviance are particularly important in logistic regression: null deviance and model deviance. The null deviance represents the difference between a model with only the intercept (which means "no predictors") and the saturated model. The model deviance represents the difference between a model with at least one predictor and the saturated model. Given that deviance is a measure of the difference between a given model and the saturated model, smaller values indicate better fit. Thus, to assess the contribution of a predictor or set of predictors, one can subtract the model deviance from the null deviance and assess the difference on a chi-square distribution with degrees of freedom equal to the difference in the number of parameters estimated.

$$-2\ln\frac{\text{likelihood of the null model}}{\text{likelihood of fitted model}}.$$

In logistic regression, there are several different tests designed to assess the significance of an individual predictor, most notably –The Wald statistic.

Wald statistics: The Wald statistic, analogous to the t-test in linear regression, is used to assess the significance of coefficients. The Wald statistic is the ratio of the square of the regression coefficient to the square of the standard error of the coefficient and is asymptotically distributed as a chi-square distribution.

$$W_j = \frac{B_j^2}{SE_{B_j}^2}$$

As every coin has a head and a tail, Wald statistic also has its limitations as it tends to be biased when data are sparse.

Linear predictor function: The basic idea of logistic regression is to use the mechanism already developed for linear regression by modeling the probability p i using a linear predictor function, i.e. a linear combination of the explanatory variables and a set of regression coefficients that are specific to the model at hand but the same for all trials. The linear predictor function f(i) point i is written as:

$$f(i) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_m x_{m,i},$$

Where are B0, B1, …, regression coefficients indicating the relative effect of a particular explanatory variable on the outcome. The regression coefficients β 0 , β 1 ... β m are grouped into a single vector β of size m + 1.The resulting explanatory variables x 0, i , x 1, i , x m, i are then grouped into a single vector X i of size m + 1.This makes it possible to write the linear predictor function as follows:

$$f(i) = \boldsymbol{\beta} \cdot \mathbf{X}_i,$$

4. CASE STUDY

Case A: Sodexo - leading provider of integrated food and facilities management services in the US, Canada and Mexico and the 21 st largest employer worldwide. Sodexo has built a big community online: Facebook – 7355 fans and Twitter – 7206 followers. All recruitment properties online are connected, and the company has built a talent community of 300,000 people.

Hiring manager satisfaction has increased from 4 to 4.63 on a scale of 5. The time-to-hire has decreased by 8 whole days from 2007. Recruitment advertising costs have been reduced by $300,000 annually. The success of the social recruiting strategy in the US has prompted Sodexo to roll out the same in other markets, including France which is Sodexo's homeland.

Case B: Deloitte - In 2009/2010 had about 4600 employees in Netherlands- one of the most difficult markets to hire. Only 10% of the workforce is looking for a new job. They built a

International Journal of Scientific & Engineering Research, Volume 8, Issue 4, April-2017
ISSN 2229-5518

separate career website for recruiting new workforce. There was an increase in number of visitors by 177%. Traffic to their career site came mainly from social networks- 33,266 and mere 9,971 through other job sites.

There was an increase of 234% candidates coming from social media than other sources. Direct and Indirect benefits of social media are much larger than anyone can imagine.

Case C: UPS - United Parcel Service, switched from 90% of their budget in print media in 2005 to 97% of their budget in social media in 2010. UPS decided to change their recruitment strategy and started to move aggressively to online recruitment. From the very beginning, UPS even used mobile technology to reach out to their target audience.

The number of hires of all different channels rose dramatically. They hired 12 in 2009 to 226 on Facebook following year. The cost of hire reduced from $600/700 to $60/70 per hire. Time to hire was reduced especially for seasonal staff and during on peak times. UPS is today one of the leading companies in using social media in their recruitment strategy.

FUTURE SCOPE

The next logical step would be to implement a working prototype of the described model. Data sets need to be generated for training as well as testing from the various websites mentioned. Our focus will be on making the model intelligent enough to differentiate the logical differences while analyzing the profile of a prospective candidate for accurate predictions. The quantity and quality of data would play a vital role in the process of prediction. Our foremost challenge would to acquire good quality data.

CONCLUSION

Our model aims to become a viable alternative to traditional methods of recruitment processes. It has several advantages over existing conventional methods of recruitment as described. Using openly available data, it proposes a model for effective recruitment by employers. The personal as well as professional background of a candidate is accounted for, thereby giving a wholesome view of the potential employee. The implications of this model can be completely understood only after the deployment in a real life scenario. The Data mining techniques and Machine learning algorithms plays a very important and centralized role which makes it possible to construct and implement the entire model.

ACKNOWLEDGMENT

We would like to thank developers of Facebook, LinkedIn, and Twitter for making their API's openly available to us. We would like to acknowledge the support of our teaching faculty, friends for their thoughts and support. We are grateful to the data science community all over the world for inspiring research in the areas of data mining and machine learning.

REFERENCES

1. LinkedIn Developer API's [Online]. Available: https://developer.linkedin.com/

2. Facebook Developer API's [Online]. Available: https://developers.facebook.com/docs/

3. Twitter Developer API's Available: https://dev.twitter.com/rest/ [Online].

4. Statistical Data (Facebook, Twitter, LinkedIn). Available: http://expandedramblings.com/

5. The Jobvite Recruiter National Survey 2015. [Online]. Available: http://jobvite.com/

6. Case studies: Social media in recruitment. [Online]. Available: http://linkhumans.com/

7. Recruitment Technology Trends Report 2014. [Online]. Available: http://recruiter.com/

8. ARCHANA L; NIVYA V G; THANKAM S M. "Recruitment through Social Media Area - Human Resource".(IOSR-JBM) e-ISSN: 2278-487X, p-ISS

9. Ajay Sharma, Swaraj Prince, -- Placement Prediction System using Logistic Regression‖, IEEE International Conference on MOOC, Innovation and Technology in Educationp.337-341