

Predicting Candidate Uptake on Individual Online Vacancies and Vacancy Portfolios

Corné de Ruijt^{†‡}, Sandjai Bhulai[†], Bram L. Gorissen[†], Han Rusman[‡] and Leon Willemsens[‡]

[†]Faculty of Sciences
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
Email: {s.bhulai, b.l.gorissen}@vu.nl

[‡]Endouble
The Netherlands
Email: {Corne, Han, Leon}@endouble.com

Abstract—The internet has undoubtedly had a substantial effect on how organizations and job seekers behave on the labor market, which has been beneficial for both job seekers and organizations. However, despite these benefits, it also comes with difficulties. Organizations might observe both applicant excess and applicant shortage on their vacancies. The problem of either applicant excess or shortage has been addressed by previous studies, which frequently conclude that this problem is inherent to the process of online recruitment. The usage of analytical techniques might reveal new insight into how organizations can account for this problem. This paper therefore studies how the number of job-applications on online vacancies in a particular week, which is referred to as the application rate, can be predicted and controlled. To answer this question, a dataset originating from a large Dutch organization was used. This dataset contains recruitment outcomes over a period of three years and just over 5,000 unique vacancies. This study trained multiple machine learning models on predicting the application rate. Furthermore, it analyses the predictability of the total number of weekly applications over the entire vacancy portfolio, and how both the application rate and the total number of applications is affected by the usage of online marketing campaigns.

Keywords—Recruitment analytics; HR analytics; Corporate career website; Predicting application rate

I. INTRODUCTION

This paper explores how online vacancies attract job seekers and how this process might be predicted and controlled. It thereby extends previous work [1], by discussing in more depth the predictability of the total number of weekly job applications over an organization's entire vacancy portfolio. Furthermore, this paper puts more emphasis on the effect of online marketing campaigns on the number of applications per vacancy per week, a metric that is further referred to as the application rate.

The internet has undoubtedly had a substantial effect on how organizations attract and select job seekers, and how job seekers search for new job opportunities. Already in 2003, 94% of the Global 500 companies reported having a corporate career website (CCW) [2]. Furthermore, there is a steady growth in the percentage of unemployed job seekers using the internet in different countries [3].

These results are not without reason: online recruitment has the potential for organizations to lower cost, shorten recruitment lead times, and attract a wider range of applicants [4], [5]. With employee turnover cost being easily over 150% of the departing employee's salary, lowering recruitment cost

could have a substantial positive financial impact [6][p. 88]. Furthermore, the usage of the internet also has a positive effect on job seekers: job seekers using the internet are less likely to become unemployed, and can expect a larger growth in terms of their wage [7].

The internet also comes with difficulties. One of these is the excess of applicants, many of which being unsuited or poorly suited for the job they apply to [5], [8]. The ease of using the internet to apply apparently removes some barriers for job seekers to also apply to jobs they might be less suited for. On the other hand, some organizations find more difficulties in attracting sufficient candidates to their corporate career website, in particular organizations with a less well-known employer brand [9].

For that reason, this paper studies the predictability and controlability of the application rate. A dataset originating from a large Dutch organization, describing recruitment outcomes over a period of three years and just over 5,000 unique vacancies is considered. We study both the predictability of the total number of applications over all vacancies in the vacancy portfolio in a particular week, as well as the predictability of the total number of applications per vacancy per week, i.e., the application rate.

This paper has the following structure: in Section II, previous research on the effectiveness of corporate career websites is discussed, as well as methods to overcome the problem of an excess in the number of applications. Section III discusses how data was obtained and prepared for this study. Section IV discusses results from preliminary data analysis, whereas Section V reviews the machine learning methods used to make accurate predictions over the application rate. Section VI shows the results that were obtained from running different machine learning models in an attempt to predict the application rate. Finally, Section VII provides short summary of this research combined with its conclusions, whereas Section VIII discusses ideas for further research.

II. RELATED WORK

Although to our knowledge no previous studies have considered predicting and controlling the application rate using data describing historic recruitment outcomes, previous research has studied how applicant shortage and excess can be controlled via other means. Firstly, by studying the effectiveness of corporate career websites as a whole, i.e., not per vacancy, using questionnaires. Secondly, by considering the

usage of machine learning in the applicant selection process in order to manage applicant excess, rather than using machine learning to attract or repulse job seekers to apply in the first place. This section will discuss results from both research perspectives.

A. Effectiveness of corporate career websites

Maurer and Liu [9] introduced a model that describes how corporate career websites influence job seekers in their decision to apply. In their model, Maurer and Liu define three inputs that affect job seekers in their usage of the recruitment website. The first input are job seekers' characteristics, such as his/her search query, motivation, current job, and knowledge. The second input, influence emphasized routes, defines how the job seeker navigates through the website. The third input are source features: features and content that are actually on the website. To analyze the influence emphasized routes (input 2), the elaboration likelihood model of persuasive communication [10] is used. The elaboration likelihood model defines two types of cues that can attract job seekers' attention. Central cues are cues that trigger the job seeker's careful thought and consideration. Vacancies that offer more salary or offer improved working conditions might persuade the job seeker to apply. Peripheral cues are cues that attract job seekers by raising an emotional response, such as employee testimonials.

Using this model, decisions can be made in the website design to attract a certain audience. Maurer and Liu argue that job seekers with a low level of search motivation or experience are more influenced by peripheral cues than central cues, the opposite holds for high levels of search motivation and experience. Furthermore, to attract job seekers, web designers should consider that the effectiveness of an application as a function of the amount of information is U-shaped. Hence, more information is not always better. Information richness should also be considered from different dimensions including active control (the ability to maintain control over the environment and information received), reciprocal communication (speed and direction of communication), vividness degree (auditory or visual channels used), and vividness depth (quality of information). Maurer and Liu stress that although online recruitment is more unbounded in the information that can be presented, potential excesses should be well managed.

Parry and Tyson [8] performed a longitudinal study to the usage of internet recruitment methods and the perception towards internet recruitment by employers. The study conducted seven surveys in the period from 1999 until 2006, where 16,000 employers were contacted per e-mail. Also, the study conducted fifteen semi-structured interviews with individuals from different industries, organization sizes, and geographical locations. Five interviews were held with providers of online recruitment technology. The study revealed four best practices that companies could exercise to increase the effectiveness of their internet recruitment practices. First, while prominent brands automatically attract job seekers to the organization's recruitment website, this is more difficult for organizations that are less known within the job seekers' audience. Hence, these organizations should use tools such as job boards and advertising channels to drive traffic towards the organization's recruitment website. Second, for prominent brands the large number of applicants is rather a burden than a blessing. Processing all these applicants can take a significant amount

of time and resources. Therefore, these companies should consider using automated initial screening to already filter out those applicants who are unsuited for the job. Third, data from candidates can be saved in a talent pool. Finally, companies should attempt to make the job description and company description realistic. Unrealistic descriptions might generate traffic to the recruitment website, but also contain job seekers that in reality have a bad person-job or person-organization fit.

In another study, Braddy et al. [11] instructed 48 undergraduate students to explore a pair of corporate career websites. Afterwards, the participants were asked which of the two was most strongly associated with each of the following nine cultural dimensions: innovative, emphasis on rewards, supportiveness, outcome orientation, attention to detail, team orientation, aggressiveness, decisiveness and diversity. Based on this information, the researchers were able to identify which website content was associated with which cultural dimension. For example: explicitly mentioning that risk taking is encouraged was found to associate with an innovative culture, where stating the awards won in the past and the plans for company growth is associated with aggressiveness. The researchers also found that explicitly stating the companies culture was the most cited reason for associating a culture with an organization. Furthermore, website design features were found highly important for conveying perceptions of innovation, attention to detail, team orientation and diversity. Stating relevant organizational policies seemed instrumental for conveying perceptions of rewards, supportiveness and diversity.

Ton et al. [12] asked 100 Hong Kong students to complete four assignments on two Chinese job boards, namely creating an account, creating a resume, conduct a job search, and complete a job application. The job boards had an a priori high service level and low service level respectively. The service quality of the two websites was measured over the dimensions: overall service quality, time spent on the website, mental workload, general service quality, accuracy and efficiency, interface, maneuvering speed, and additional support. The latter five dimensions were measured via a questionnaire. The study found that the total time spent on the website was significantly less for the job board having the a priori high service level. The mental workload was significantly higher for the a priori low service level website, which could be caused by the job seeker having to put more effort in searching and interpreting information on the low quality website. The research concludes that general service quality, accuracy/efficiency and interface were most highly correlated with the overall service quality ratings.

Sylva and Mol [13] considered 1,360 applicants that applied for different positions within a multinational financial organization having more than 100,000 employees. Data was gathered by a questionnaire that appeared right after someone applied online during a period of two months. The collected data included demographical data, perceptions of the online application process and system and the perceived fairness. The researchers found that applicants were generally favorable towards the web-based procedure. Efficiency, user-friendliness, process fairness, and internet selection image were found as main determinants of applicant satisfaction. Candidates being external or highly familiar with the usage of the internet were more positive about the recruitment system.

Jansen and Jansen [14] study job search related queries that were filled into search engines Excite and MSN. From the analyzed queries, 45% was based on location, 17% on industry, 11% on skill set, 8.9% on specific job sites, 7.3% on government and 2.2% on temporal jobs. Furthermore, the search queries were generally long and specific, whereas the actual session duration was relatively short. The words used in the queries also changed over time.

B. Recruitment analytics for managing application excess

Internet recruitment has caused an excess of information spread to both recruiters and job seekers, based on which many studies conclude that these problems are inherent to internet recruitment, i.e., these problems are a natural consequence of internet recruitment and cannot be avoided. However, recent developments in machine learning are increasingly applied to the recruitment process in order to automate parts of this process [15]. For example, [16] found that by fast-tracking candidates who score highly in their pre-selection algorithm through some parts of the selection process, the time to hire can significantly be reduced without negatively impacting the overall quality of hired candidates. Besides automating parts of the selection process, previous research has also focused on automating other recruitment related tasks, such as CV parsing [17].

Although automated (pre-)selection and CV parsing automate part of the recruitment activities, they are more a workaround than an actual solution to the problem of applicant excess. More specifically, by analyzing how the quantity and quality of applications is affected by the online channel and message used by the vacancy, recruitment has the opportunity to control the quality and quantity of applicants before the job seeker applies. This not only may decrease the total number of applications, but also reduce the number of rejected applicants. Interestingly, previous research has not used the increasing amounts of data stored by recruitment departments themselves [18] to study the effect of online recruitment channels and ways of communication on the quality and quantity of applicants.

III. DATA GATHERING AND PREPARATION

This section will give a short overview of how data was gathered for this study, which data was gathered, and how the data was prepared for further analysis.

A. Data gathering

To investigate whether the number applications can accurately be predicted and controlled, a dataset was gathered from a large Dutch company which employs over 30,000 people. This data was gathered over the period 2013-08-26 until 2015-12-31 and contained 5,036 unique vacancies.

The dataset was gathered from three systems: first, from an application tracking system (ATS), in which vacancy characteristics are stored such as work location, required education level, and working hours. Second, data was gathered from the corporate career website's Google Analytics account: how many job seekers visited the corporate career website per week, how frequently job seekers followed different paths from the website's landing page to the application submit page (the web page visited by applicants after submitting an application), and whether job seekers visited the website via a paid hyperlink,

which was part of an online marketing campaign. The latter was used to determine which vacancies had been used in online marketing campaigns. Third, the number of weekly tweets the recruitment department published via their recruitment Twitter account was gathered, along with whether certain vacancies were referred to in a tweet via a hyperlink.

Combining these three data sources (ATS, Google Analytics, and Twitter) yields for each vacancy v and time period t , whether v was used in online marketing campaigns, and how many job seekers navigated from the landing page to the vacancies submit page during time period t . A measurement per vacancy per week is for simplicity denoted as per vacancy-week. This dataset was extended with time related data such as the recruitment lead time at time t , and application rates of a vacancy in weeks prior to week t .

The dataset was split into a test- and training set. The training set contained all values between 2013-08-26 and 2015-09-31, whereas the test set contained all values between 2015-10-01 and 2015-12-31. Note that this split was made only on date: it is possible that a vacancy exists both in the training set and in the test set. The split was chosen for two reasons: first, at the time of splitting the dataset there was no prior knowledge of possible time dependency in the data. A growing popularity of an employer brand might for example cause all vacancies to attract more applicants over time. If the application rate would include this time dependency, then validating the predictive model on the last period of the total dataset would produce the most realistic evaluation. Second, three months is the maximum period for which it is safe to assume that the vacancy portfolio over that period is known.

B. Data preparation

1) *Clustering of categorical variables:* To improve the quality of the data set, multiple operations were performed. Attributes related to work location and job title contained many possible categorical values, which was not practical for analysis. To reduce the number of categorical values, the locations were clustered based on similarities in their application rate probability density.

Let $N_{h,k}^{(l)}$ be the number of times application rate $h = 1, 2, \dots, H$ was found for categorical value $k = 1, 2, \dots, K$, which is a category of attribute $l = 1, 2, \dots, L$. Furthermore, let $N_k^{(l)}$ be the number of times categorical value k of attribute l occurs in the dataset. Then $X_{h,k}^{(l)} = \frac{N_{h,k}^{(l)}}{N_k^{(l)}}$ is the element of the $H \times K$ matrix $X^{(l)}$ at position (h, k) . A single row of $X^{(l)}$ gives the marginal application rate probabilities for categorical value i . Similar marginal probabilities were clustered using a K-means clustering algorithm by Hartigan and Wong [19].

To find the right number of clusters, the Akaike Information Criteria: $AIC = SS + 2CK$, was used for $C = 1, \dots, 10$ clusters, where SS is the sum of squared Euclidean distance between the observations $i = 1, \dots, n$ and the centroid to which it is assigned to, and K is defined as earlier. If a cluster had fewer than 100 observations, we assigned the categorical values of that cluster to the cluster which mean application rate was closest to the overall mean application rate.

Besides location attributes, the job title had even more unique values, which made the usage of the probability density

unpractical. As an alternative, similar job titles were identified and clustered manually.

2) *Low variance removal, normalization and dummification*: In order to identify attributes having a small variance, the frequency cut off from the *nearZeroVar* function of the *caret* package was used [20]. Since all predictors are either binary, categorical, or discrete, it was possible to apply this procedure to all predictors. Let $N_{(z)}^{(l)}$ be the z th order statistic of $N_1^{(l)}, \dots, N_K^{(l)}$, with $N_k^{(l)}$ defined as in III-B 1, then we have frequency ratio:

$$F_l = \frac{N_{(K)}^{(l)}}{N_{(K-1)}^{(l)}}. \quad (1)$$

Thus, F_l gives the ratio of the most frequent and second most frequent value of attribute l . Attributes were removed from the data set if $F_l > 19$.

During the last data preparation step, categorical attributes were dummified into binary vectors. Numerical attributes were normalized using:

$$\tilde{x}_{il} = \frac{x_{il} - \bar{x}_l}{s(x_l)}. \quad (2)$$

Here \bar{x}_l and $s(x_l)$ are the mean and standard deviation over the values $i = 1, 2, \dots, n$ of attribute l respectively.

IV. RESULTS EXPLORATORY DATA ANALYSIS

This section discusses two subjects. First, it will discuss the probability distribution of the application rate, which affected further modeling considerations that will be discussed in Section V. Second, it discusses the predictability of the total number of weekly applicants over the entire vacancy portfolio, using other website traffic indicators, time, characteristics of the vacancy portfolio, and the usage online marketing campaigns as predictors.

A. The application rate

When considering possible probability distributions of the application rate, a Poisson distribution would come first to mind. This follows from assuming that each vacancy has a large population of potential applicants, who each have a small probability of applying in a given week. However, as Fig. 1 suggests, the Poisson distribution does not seem to fit the data well: the application rate's distribution is more zero inflated and overdispersed than a Poisson distribution. Dependent on the nature of the vacancy, a log-normal or negative binomial distribution is more appropriate. The distribution also confirms previous research stating that some vacancies can attract a large number of applications [8]. In fact, 10% of the rates account for 53% of all applications.

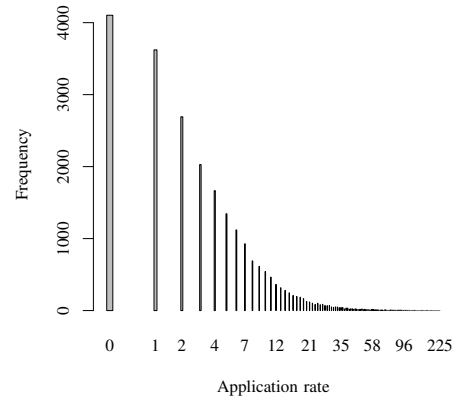


Figure 1. Histogram application rate

B. Total number of applications vs sessions and number of vacancies

Besides considering the distribution of the application rate, also the predictability of the total number of applications per week was considered. In particular its dependency on the total number of sessions on the website, number of vacancies in the vacancy portfolio, usage of online marketing campaigns, and time.

Fig. 2 shows a scatter plot of the total number of weekly sessions against the total number of weekly applications, whereas Fig. 3 shows the size of the vacancy portfolio compared to the total number of applications. From these figures it can be derived that, without considering their interaction or taking into account other covariates, increasing the number of sessions or increasing the size of the vacancy portfolio has a positive effect on the total number of applications (R^2 of 0.67 and 0.51 for the number of sessions and number of applications respectively).

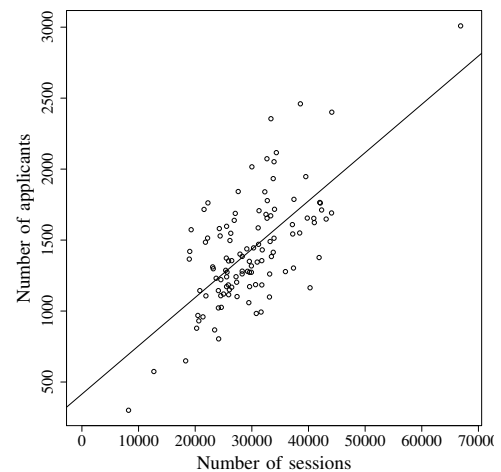


Figure 2. Sessions vs applications

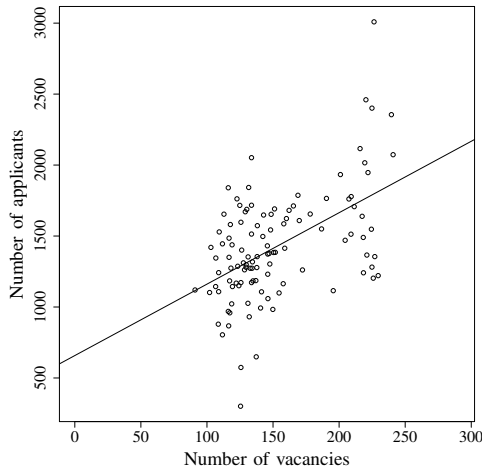


Figure 3. vacancies vs applications

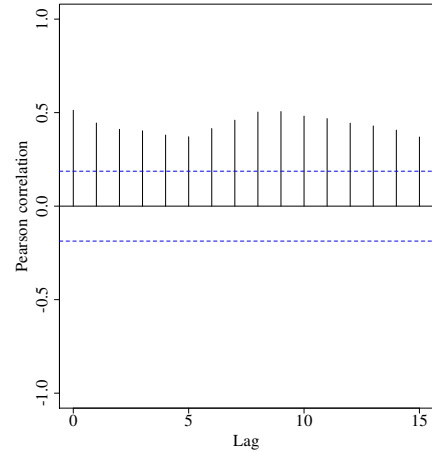


Figure 5. Cross correlation vacancies and applications

Fig. 4 shows the cross-correlation at different lags between the weekly sessions and weekly applicants, the blue lines represent the 5% significance level. The figure indicates that this cross-correlation peaks at lag zero (with a Pearson correlation of 0.67), after which it rapidly drops and becomes insignificant after 2 weeks. Hence, the job seekers are most likely to apply within the same week they first visit the website.

When considering the cross-correlation between the weekly size of the vacancy portfolio and the number of applications (Fig. 5), we obtain a rather different result. Again, the cross correlation peaks at lag zero (Pearson correlation of 0.51). However, it remains more or less constant before lag 0.

Table I shows the number of sessions that originated from a certain source, and used a certain device. Since visitors to the corporate career website can originate from many different sources, only the top four sources causing most traffic were considered, whereas smaller sources were combined in an ‘other’ category. What becomes most apparent is that only a small number of sources contributes to the number of sessions: the top three sources (Google, Indeed, and the corporate site) account for 71% of all traffic.

TABLE I. SOURCE AND DEVICE OF SESSIONS

Source-device	Sessions
Google desktop	634982
Indeed desktop	612637
Corporate site desktop	509441
Other desktop	328395
Google mobile	291040
Direct desktop	284256
Direct mobile	196089
Direct tablet	57666

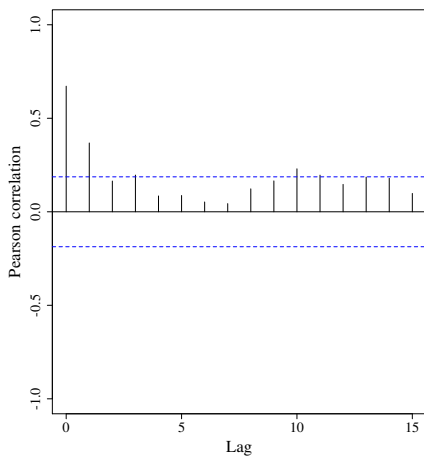


Figure 4. Cross-correlation sessions and applications

C. Predicting the weekly number of applications

From Section IV, we already found that the correlation between the total number of vacancies online in a particular week is significantly correlated with the total number of applications. Therefore, a logical follow-up question would be how the total number of weekly applications depends on vacancy portfolio characteristics and the used online marketing campaigns. Here, the vacancy portfolio’s characteristics are determined by counting the number of vacancies online having certain characteristics, such as vacancies having the same work location or the same job title.

Since the number of observations for the total number of weekly applications was relatively small, especially compared to the number of covariates, a multiple linear regression model was used without interaction terms. Furthermore, to remove non-predictive variables, a backwards AIC algorithm was applied. This algorithm iteratively removes those attributes from the linear regression model resulting in the largest reduction in the AIC value. The algorithm terminates if removing attributes does not result in a smaller AIC value. The resulting model

is shown in Table II, which has an R^2 value of 0.62. From the table it can be observed that especially an increase in the number of vacancies related to certain locations has a positive effect on the total number of applications.

We also observe a negative effect of Google Adwords campaigns and twitter references. It is however difficult to draw solid conclusions from this observation for two reasons. First, the campaigns were frequently used in combination with each other, which makes it difficult to identify the effect of a single campaign. This can easily be seen from their correlation, 0.88 for Google Adwords and Facebook campaigns, and from the condition indices and variance decomposition proportions [21]. The largest condition index (91.68) has, for the number of Facebook and the number of Google campaigns, variance decomposition proportions of 0.636 and 0.620, which are larger than the threshold value for collinearity of 0.5. A possible remedy for this collinearity is to add more characteristics of the campaigns to the data set, such as the profiles used in a Facebook campaign. This was however not considered in this study. Second, online campaigns were most frequently used on vacancies receiving a small number of applications, hence the usage of online marketing campaigns could have been more of a response to a small number of applications rather than that it determines a small number of applications.

TABLE II. Summary linear regression model for predicting number of weekly applications

Coefficient	Estimate	Std. Error	t-value	p-value
Intercept	368.36	222.07	1.66	0.1005
Vacancies	6.47	2.01	3.22	0.0018
Business unit A	-50.55	13.83	-3.65	0.0004
Business unit B	-27.77	8.27	-3.36	0.0011
Business unit C	-32.61	15.43	-2.11	0.0373
Location 1	59.15	18.06	3.27	0.0015
Location 2	27.97	8.24	3.39	0.0010
Job category 1	72.74	18.65	3.90	0.0002
Job category 2	38.89	13.58	2.86	0.0052
Job category 3	2.22	2.38	0.93	0.3537
Location 4	-31.15	8.65	-3.60	0.0005
Location 5	-27.50	17.60	-1.56	0.1215
Location 6	-27.39	8.35	-3.28	0.0015
# Facebook campaigns	18.48	4.61	4.01	0.0001
# Google Adwords campaigns	-9.12	3.38	-2.70	0.0083
# Twitter references	-17.41	12.04	-1.45	0.1517

When considering the residuals of the linear regression model over time (Fig. 6), a Box-Pierce test showed that these residuals were correlated. However, when examining the autocorrelation and partial autocorrelation functions, this correlation turns out to be small: both the autocorrelation and partial autocorrelation show a maximum absolute correlation of 0.21, at lag one and two respectively. Therefore, for simplicity, it was found acceptable to assume that the residuals were uncorrelated. As a result it was assumed that the total number of applications per week was independent of the date of the measurement.

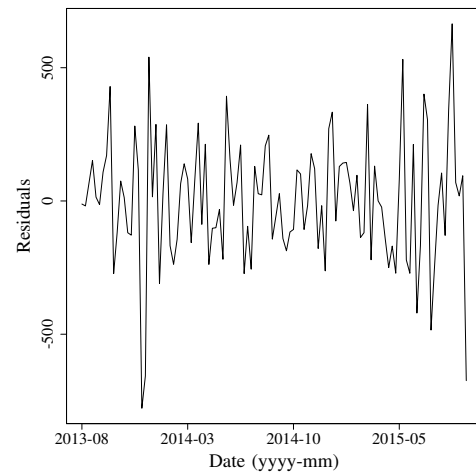


Figure 6. Errors over time

D. Best sources

In Section IV-B, we already considered which source-device combinations contributed most to the number of sessions (Table I). A related question is how the source-device combination contribute to the number of weekly applications. Again a linear regression model without interaction terms was used, combined with the backwards AIC algorithm to remove attributes which did not contribute to the model.

TABLE III. SUMMARY LINEAR REGRESSION MODEL FOR PREDICTING NUMBER OF WEEKLY APPLICANTS

Coefficient	Estimate	Std. Error	t-value	p-value
Direct mobile	0.34	0.06	4.92	0.0000
Corporate site desktop	0.27	0.06	4.93	0.0000
Direct desktop	0.19	0.05	3.99	0.0001
Other desktop	0.07	0.01	5.58	0.0000
Google mobile	0.06	0.02	2.57	0.0116
Direct tablet	-2.00	0.42	-4.73	0.0000

The result is shown in Table III. Interestingly, the source-device combinations causing most traffic to the website did not produce most applications. Where visitors originating from Google on a desktop produced most traffic to the website, changes in direct traffic on either desktop, mobile, or tablet, and traffic from the corporate website were the main drivers for changes in the number of weekly applications. This observation seems intuitive: job seekers will use a search engine the first time they visit a career website, but after that the browser will have stored the vacancy's URL such that the job seeker can easily return to the web page directly.

V. METHODS

This section gives an overview of the methods that were used to predict the application rate. Also, it will describe how the predictive quality of the different methods were compared, and how the effect of online marketing campaigns was computed.

A. Method selection

To determine which methods would be most suited to predict the application rate, six considerations were taken. First, since the application rate is count data, its prediction

is considered to be a regression problem. Second, exploratory data analysis found that the data is more zero inflated and overdispersed than a Poisson distribution. Therefore, predictive models which incorporate zero inflation and overdispersion are preferred. Third, during exploratory data analysis it was found that when predicting the total number of applications per week, the residuals of this model were only slightly correlated. As a result, it was assumed that the total number of applications per week is independent of the date of the measurement. Though it still can be dependent on other time indicators, such as the current recruitment lead time. Fourth, the dataset still contained a large number of attributes, some of which might not be useful for the predictive model. To reduce the number of attributes, methods which included variable selection were preferred. Fifth, since a grid search was applied to find good model parameters, methods which were able to produce good results within reasonable time were preferred (i.e., methods that took more than 1 hour to compute a single predictive model using a 1.6 GHz dual-core Intel Core i5 processor were disregarded). Sixth, methods which have been applied successfully in other regression application were preferred.

Based on these criteria, seven methods were identified: Linear elastic net, Poisson elastic net, Tweedie elastic net, Classification And Regression Trees (CART), random forest (RF), Support Vector Regression (SVR), and Artificial Neural Networks (ANN). For convenience, we write the application rate as $\mathbf{y} = (y_1, \dots, y_N)$, with N the total number of observations, and the covariate matrix as $X = (\mathbf{x}_1, \dots, \mathbf{x}_N)$. All methods were only executed on the training set.

1) *Linear elastic net*: Linear elastic net is a method which attempts to minimize the sum of squared errors, plus a linear combination of the lasso and ridge penalty. Let $SSE(\lambda, \beta, \alpha)$ be the regularized sum of squared errors, with λ the weight of the regularization terms, α the convex combination parameter for ridge vs lasso regularization, and β the vector of main effects. $SSE(\lambda, \beta, \alpha)$ is given by:

$$SSE(\lambda, \beta, \alpha) = \frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - \mathbf{x}_i^T \beta)^2 + \lambda \left[(1 - \alpha) \frac{1}{2} \|\beta\|^2 + \alpha \|\beta\|_1 \right]. \quad (3)$$

To minimize (3), the glmnet R package was utilized, which applies a coordinate descent algorithm to estimate β [22]. To determine good values for λ and α , a grid search was applied. For λ , $K = 100$ uniformly spread values between $\lambda_{\max} = \frac{\max_i |\langle \mathbf{x}_i, \mathbf{y} \rangle|}{N\alpha}$, where \mathbf{x}_l is the l th column of matrix X , and $\langle \mathbf{x}_l, \mathbf{y} \rangle$ is the inner product between vectors \mathbf{x}_l and \mathbf{y} . For λ_{\min} we take $\lambda_{\min} = 10^{-3} \lambda_{\max}$. To find a good value for α , values from 0 up to 1 with increasing steps of 0.2 were used.

2) *Poisson elastic net*: Poisson elastic net is a combination of a generalized linear regression model and elastic net using the link function $g(\mu_i) = \log(\mu_i)$, where $\mu_i = \mathbb{E}(y_i | \mathbf{x}_i)$. Instead of using the sum of squared errors, the log-likelihood is used to estimate β . Let ALL be the regularized adjusted log-likelihood, then β is found by maximizing (4):

$$ALL(\lambda, \beta, \alpha) = \frac{1}{2N} \sum_{i=1}^N [y_i \mathbf{x}_i^T \beta - \exp(\mathbf{x}_i^T \beta)] - \lambda \left[(1 - \alpha) \frac{1}{2} \|\beta\|^2 + \alpha \|\beta\|_1 \right]. \quad (4)$$

To maximize (4), again the glmnet R package was used. In case of Poisson regression, glmnet iteratively creates a second order Taylor expansion of (4) without the penalty, using current estimates for β . This Taylor expansion is then used in a coordinate descent algorithm to update β [22], [23]. To find appropriate values for λ and α , the same grid search as for linear elastic net was applied.

3) *Tweedie elastic net*: To incorporate the fact that the application rate is more zero inflated and overdispersed than a Poisson distribution, the Tweedie compound Poisson model was used. The Tweedie compound Poisson model can be represented by $Y = \sum_{m=1}^M X_m$, where Y is the response vector, M a Poisson random variable, and X_m are i.i.d. Gamma distributed with parameters α and γ . The regularized negative log-likelihood is given by (5) [24]:

$$NLL(\lambda, \beta, \alpha) = \sum_{i=1}^N \left[\frac{y_i \exp[-(\rho-1)(\mathbf{x}_i^T \beta)]}{\rho-1} + \frac{\exp[(2-\rho)(\mathbf{x}_i^T \beta)]}{2-\rho} \right] + \lambda \left[(1 - \alpha) \frac{1}{2} \|\beta\|_2 + \alpha \|\beta\|_1 \right]. \quad (5)$$

To minimize (5), the HDTweedie R package was used. This method applies an iterative reweighted least squares (IRLS) algorithm combined with a blockwise majorization descent (BMD) [24]. To find appropriate values for λ , the standard procedure from HDTweedie was used. This method first computes λ_{\max} such that $\beta = 0$, and then sets $\lambda_{\min} = 0.001 \lambda_{\max}$. The other $K - 2$ values for λ are found by projecting them uniformly on a log-scale on the range $[\lambda_{\min}, \lambda_{\max}]$. For α the values from 0.1 to 0.9 with an increase of 0.2 were used. Furthermore, $\rho = 1.5$ was used.

4) *Classification And Regression Trees*: To construct a regression tree the rpart implementation in R was used [25]. This implementation first constructs a binary tree by maximizing $SS_T - (SS_R + SS_L)$ in each node, where SS_T is the sum of squared errors of the entire tree, and SS_R and SS_L are the sum of squared errors of the left and right branch respectively. The tree constructions stops when further splits would violate a constraint on the minimum number of observations in each node.

Second, the constructed tree is split into m sub-trees. Let $R(T)$ be the risk of tree T , which is the sum of squared errors in the terminal nodes of T . CART computes the risk of each sub-tree, which is defined by $R_\alpha(T) = R(T) + \alpha|T|$, using K-fold cross validation. The term $\alpha|T|$ is an additional penalty on the size of the tree. The final tree is the sub-tree which minimizes the average sum of squared errors over the K-fold cross validation. The method described here is referred to as the "ANOVA" method.

Alternatively, rpart also has the option to maximize the deviance $D_T - (D_L + D_R)$, where D is the within node deviance, assuming that the response originates from a Poisson distribution. Both the Poisson and ANOVA methods have been applied to the dataset. To find an appropriate value for α , a grid search was applied using $\alpha \in \{0.001, 0.01, 0.1, 0.3\}$, both for the ANOVA and Poisson models.

5) *Random forest*: A random forest model was produced using the RandomForest package in R [26]. RandomForest constructs T unpruned regression trees T_i , where in each split only d randomly chosen predictors are considered. A prediction \hat{y}_i is then created by $\hat{y}_i = \frac{1}{T} \sum_{i=1}^T T_i(x)$, thus the average over all trees. To find the appropriate number of trees, a grid search was applied using 50, 100, and 500 trees. Furthermore, at each split, $d = 61$ randomly sampled attributes were considered.

6) *Support Vector Regression*: Support Vector Regression is the regression alternative for Support Vector Machines. Given the linear regression problem: $y_i = \mathbf{w}^T \mathbf{x}_i + b + \epsilon$, SVR attempts to find the flattest hyperplane $\mathbf{w}^T \mathbf{x}_i + b$ such that, for all data points $i = 1, \dots, N$; we have $|y_i - (\mathbf{w}^T \mathbf{x}_i + b)| < \epsilon$. Also incorporating slack variables ζ_i^+ and ζ_i^- , the problem can be described as (6):

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\zeta_i^+ + \zeta_i^-) \\ \text{s.t.} \quad & y_i - \mathbf{w}^T \mathbf{x}_i - b \leq \epsilon + \zeta_i^+, \quad i = 1, \dots, N \\ & \mathbf{w}^T \mathbf{x}_i + b - y_i \leq \epsilon + \zeta_i^-, \quad i = 1, \dots, N \\ & \zeta_i^+, \zeta_i^- \geq 0. \end{aligned} \quad (6)$$

Since the solution to (6) only depends on inner products between vectors \mathbf{x}_i , the problem can be transformed into a higher dimension without much extra computation using kernels [27]. For the computation of the SVR, the R kernlab package was used [28]. Although in this study initially both a linear kernel (hence no kernel) and the RBF kernel: $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$ were considered, not using a kernel surprisingly had a large negative effect on the runtime and was therefore discarded. To find appropriate values for ϵ and C a grid search was applied using: $\epsilon \in \{0.01, 0.1, 1\}$ and $C \in \{1, 10\}$.

7) *Artificial Neural Networks*: In this study we considered a feed-forward Artificial Neural Network with a single hidden layer. To find the weights the nnet R package was used, which utilizes an L-BFGS algorithm to find the appropriate weights [29], [30]. A grid search was applied to find an appropriate number of units in the hidden layer. During the grid search, 1, 5, 10, 30, and 50 hidden units were considered.

B. Method evaluation

To evaluate the quality of predictive models, two scenarios were distinguished. The first scenario assumes that application rates in weeks prior to the predicting period are known, which is comparable with predicting one week ahead. The second scenario assumes these application rates to be absent, and is more comparable with predicting 2 to 12 weeks ahead. These two scenarios are indicated by including PAR (Previous Application Rates) and excluding PAR respectively.

To evaluate the quality of the predictions two error measures are used: the root mean squared error: $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$, where \hat{y}_i is the predicted value for actual y_i . Second, the determination coefficient: $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$, with $SS_{res} = \sum_{i=1}^N (\hat{y}_i - y_i)^2$, the residual sum of squares, and $SS_{tot} = \sum_{i=1}^N (y_i - \bar{y})^2$, the total sum of squares. A 10-fold cross validation was applied to obtain accurate estimates for the quality of the predictions over the training set in both scenarios. Furthermore, a final prediction over the test set was made using the model showing the best results over the training set to estimate out of sample performance.

C. Effect of online campaigns

To gain more insight into how an increase in the usage of online marketing campaigns might affect the number of applications, two procedures were applied: one for binary indicators of online campaigns, and one for numeric indicators. We first consider the binary indicators, let:

$$z_{ij} = \begin{cases} 1 & \text{if campaign } j \text{ is used for measurement } i \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

and $\alpha_{ij} = \alpha_j = \mathbb{P}(z_{ij} = 1)$ be the probability that a vacancy in some week is stimulated by online campaign $j \in J$ in the training set, with J the set of all campaigns. Then for each online marketing campaign j we iterate over:

$$\alpha_j^{(k)} = 10^{-2}(k-1), \quad k = 1, \dots, K \quad (8)$$

while keeping $\alpha_r^{(k)} = 0$ for $r \in J \setminus \{j\}$, and we compute how the best performing predictive model predicts the application rate for this increasing probability. Hence, we iterate from not using the campaign for any vacancy ($k = 1$), to using the campaign for each vacancy ($k = K$). Note that $\alpha_j^{(k)}$ does not depend on the observation i , i.e., all observations i have the same probability of being used in marketing campaign j . For K we chose $K = 101$.

For numeric indicators, of which there was only one: the number of tweets, a similar procedure was applied. In this procedure we again iterate over $\alpha_j^{(k)}$, $k = 1, \dots, 101$. However, we now impute the usage of this campaign by computing

$$C_j^{(k)} = \left[C_{\min} + \frac{(C_{\max} - C_{\min})}{\alpha_j^{(k)}} \right] z_{ij}^{(k)} \quad (9)$$

where $C_j^{(k)}$ is the frequency campaign j is used in iteration k . $z_{ij}^{(k)}$ are samples from a Bernoulli distribution with probability $\alpha_j^{(k)}$, and C_{\min} and C_{\max} are the minimum and maximum number of times campaign j has been used in the training set for an individual vacancy in an arbitrary week.

VI. RESULTS PREDICTING THE APPLICATION RATE

The following section compares the predictive ability of the methods discussed in Section V. Furthermore, it will give an overview of the variables that provided most predictive value, and discuss the effect of increasing online marketing efforts on the application rate.

A. Method comparison

Table IV shows the best results per model when applying a 10-fold cross validation on the training set. The table indicates that random forest produced the best results, both when predicting with and without PAR. Table IV also indicates that multiple methods, such as artificial neural networks without PAR, Poisson elastic net, and Tweedie elastic net with PAR, did not produce accurate results. Furthermore, Table IV indicates that the added value of including previous application rates into the model is relatively small. Hence, the predictive model would only produce slightly better results when predicting short term (1 week), in comparison with predicting long term (2 to 12 weeks).

TABLE IV. RESULTS 10-FOLD CROSS VALIDATION

Method	Best $RMSE$ including PAR	Best R^2 including PAR	Best $RMSE$ excluding PAR	Best R^2 excluding PAR
Linear elastic net	11.82	0.35	11.87	0.34
Poisson elastic net	15.53	0	NA	NA
CART ANOVA	10.72	0.46	11.12	0.42
CART Poisson	10.17	0.52	10.75	0.46
random forest	9.38	0.59	9.93	0.54
Tweedie elastic net	13.86	0.03	11.62	0.37
SVR	10.58	0.46	11.28	0.41
ANN	11.14	0.42	17.35	0

The $RMSE$ in Table IV is largely influenced by some large application rates, which are difficult to predict. This can also be concluded from the errors of the RF model on the test set (Fig. 7). In fact, 90% of the errors are smaller than 9.63, and the average absolute error over this 90% is 2.43. Also interesting is that adding more trees to the RF model only had a small impact on the predictive quality of the model (Table V).

TABLE V. RESULTS 10-FOLD CROSS VALIDATION FOR DIFFERENT NUMBER OF TREES

Trees	$RMSE$ including PAR	$RMSE$ excluding PAR	R^2 including PAR	R^2 excluding PAR
50	9.50	9.97	0.58	0.53
100	9.40	9.98	0.59	0.53
500	9.38	9.93	0.59	0.54

B. Test set evaluation

Since RF produced the most promising results in a 10-fold cross validation, this model was evaluated on the test set. The results are shown in Table VI, whereas the distribution of the error in the test set is shown in Fig. 7. The quality of the prediction was slightly worse than the average error obtained from 10-fold cross validation. Furthermore, just as in the training set, few vacancies with large application rates account for most of the $RMSE$.

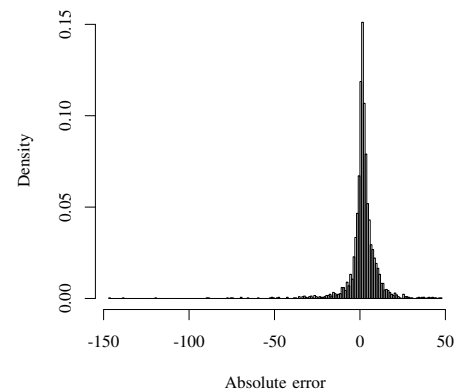


Figure 7. Test set error with PAR

TABLE VI. RESULTS APPLYING RANDOM FOREST ON TEST SET

Performance metric	Value including PAR	Value excluding PAR
MAE	5.25	6.35
MSE	100.44	123.99
RMSE	10.02	11.13
Residual mean	1.53	1.81
Residual sd	9.90	10.98
R^2	0.44	0.32

C. Variable importance

To determine the importance of different variables in the model, the standard method from the R RandomForest package was used [26]. For regression problems, as opposed to classification problems, this method computes the reduction in residual sum of squares when splitting on a certain variable. This amount is summed over all trees to obtain the impurity, which provides an overall picture of the decrease in residual sum of squares of a variable. Table VII shows the ten most important variables found in the model for the case of including PAR.

TABLE VII. VARIABLE IMPORTANCE

Top including PAR	Impurity ($\cdot 10^5$)
Application rate 1 week ago	5.65
Job category 1	4.94
Current vacancy lead time	4.38
Job type 1	1.75
Application rate 2 weeks ago	1.49
Job type 2	0.91
Min. hours in contract	0.85
Job type 3	0.80
Job type 4	0.67
Business Unit C	0.62

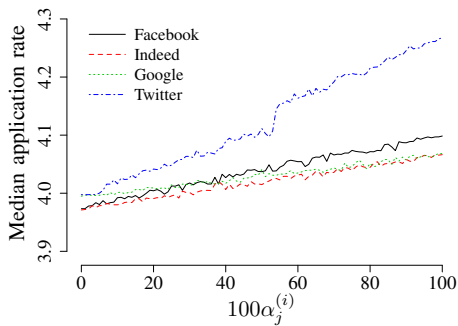


Figure 8. Effect of using online campaigns

From Table VII it can be observed that there is some overlap with the variable importance found when predicting the total weekly number of applications (Table II). In particular, both tables include the factors job category 1 and business unit C. Table VII also shows the importance of time related variables such as the current vacancy lead time and previous application rates of the vacancy, indicating that the application rate of a vacancy is time dependent. Another interesting observation is that the effect of online marketing campaigns in this figure is missing, meaning that their effect is smaller than the effect of the variables in the table.

To obtain more insight into the effect of online marketing campaigns we follow the procedure described in Section V-C, which resulted in Fig. 8. To avoid that some large application rates would have a large influence on the found effect of online campaigns, we consider the median application rate instead of the mean. The figure shows that all campaigns had a positive effect on the median application rate, though this effect is quite small. Especially an increasing usage of Twitter shows a positive effect on the application rate. However, since the campaign data was not obtained in an experimental set-up, it is difficult to draw firm conclusions.

VII. CONCLUSION

This paper considered the predictability of the number of weekly applicants per vacancy on a corporate career website, also referred to as the application rate. Being able to predict and influence this metric can be important to recruitment departments: it provides information about how to either prevent applicant excess from online platforms, or how the number of online applications could be increased. In order to study the predictability of the application rate, a dataset from a large Dutch organization employing over 30,000 employees was considered, which was collected from the organization's Applicant Tracking System (ATS), Google Analytics account and Twitter.

From the preliminary data analysis, which was mainly focused on predicting the total number of weekly applications and determining the distribution of the application rate, we found that the total number of applications could be predicted quite well from characteristics of the vacancy portfolio and the usage of online marketing campaigns ($R^2 = 0.62$). Interestingly, the number of vacancies referenced by Google Adwords campaigns and Twitter was found to have a negative effect on the total number of applications. However, it is difficult to make clear conclusions based on this observation. Marketing

campaigns were found to be highly correlated with each other, and may suffer from reverse causality: they are most frequently used for vacancies which are already under-performing.

When focusing on the probability distribution of the application rate, we found that its distribution is more zero-inflated and overdispersed than what would be expected from a Poisson distribution. A negative binomial or log-normal distribution was found to be a better fit. In an attempt to predict the application rate, multiple machine learning algorithms were used, including linear elastic net, Poisson elastic net, Tweedie elastic net, CART, random forest, Support Vector Regression, and feed-forward Artificial Neural Networks. Predictors included characteristics of the vacancy (e.g., location, work hours, etc.), usage of online marketing campaigns, number of competing vacancies on the same corporate career website, and vacancy lead time up to the time of prediction.

In a comparison between the different machine learning algorithms, random forest showed the best results. Time related attributes, such as the application rates in prior weeks and the vacancy lead time, contributed most to the quality of the predictions. Of the online campaigns, Twitter was found to have the most positive effect on the application rate.

From this analysis a few conclusions can be made. First, even though the predictions are quite accurate in most situations, i.e., have an error of less than five applicants, some vacancies can attract a large number of job seekers (> 50), which the model is unable to predict. On the other hand, both the predictions and insights into the variability of these predictions are helpful to manage the expectations recruiters and hiring managers might have when publishing a vacancy. In particular, recruiters should manage vacancies expecting a large number of applications carefully to avoid excess of applications, especially when applicant excess does not improve the quality of the hire. Also, recruiters could consider how attractive vacancies can be used to market less attractive vacancies, for example, by generalizing the vacancy such that it may refer to both popular and less popular job positions.

Second, we found that the effect of online marketing campaigns on the application rate is positive, though quite small. Furthermore, it is likely that there exists reverse causality between online marketing campaigns and the application rate: online marketing campaigns were only used on already under-performing vacancies. Therefore, we were not able to draw a firm conclusion on the effect of online marketing campaigns on the application rate.

VIII. FURTHER WORK

The research presented here can be extended in a number of directions. First, we have showed that it can be problematic to infer clear conclusions on the effect of influential factors, such as online marketing campaigns, on the number of applications based on historic recruitment data. Since the data is not obtained in an experimental set-up, the usage of such influential factors can be rather a response to too few applications, hence suffer from reverse causality. A question for further research would therefore be how this reverse causality can be efficiently accounted for, either by incorporating it into the model, or by collecting data in an experimental set-up. Besides only considering online marketing, also the effect of other recruitment endeavors could be examined, such as the effect of using job-boards on the application rate.

Second, this research can be extended to also incorporate the quality of applicants. Although multiple studies have considered the quality of applicants [15], literature is scarce on the relationship between applicant quality and quantity. Further research could therefore consider not only this relationship, but also how recruitment departments can use this information to optimize their recruitment process.

Third, this study considered that the number of applications on the entire vacancy portfolio was independent of time. It is however likely that other corporate career websites will show a drift or seasonality in the number of applications. Estimation of this seasonality can be problematic, as recruitment data might not have sufficient history for proper estimation. Further research could therefore consider how the methods proposed in this paper should be adjusted to incorporate this time dependency.

REFERENCES

- [1] C. de Ruijt, S. Bhulai, H. Rusman, and L. Willemsens, "Predicting candidate uptake for online vacancies," *DATA ANALYTICS 2016*, pp. 63–68, 2016.
- [2] R. Greenspan, *Job seekers have choices*, 2003, URL: <https://www.clickz.com/job-seekers-have-choices/76679/> [accessed 2017-05-09].
- [3] F. Suvankulov, "Job search on the internet, e-recruitment, and labor market outcomes," DTIC Document, Tech. Rep., 2010.
- [4] E. Galanaki, "The decision to recruit online: A descriptive study," *Career development international*, vol. 7, no. 4, pp. 243–251, 2002.
- [5] S. Lang, S. Laumer, C. Maier, and A. Eckhardt, "Drivers, challenges and consequences of e-recruiting: a literature review," in *Proceedings of the 49th SIGMIS annual conference on Computer personnel research*. ACM, 2011, pp. 26–35.
- [6] W. Cascio and J. Boudreau, *Investing in people: Financial impact of human resource initiatives*. Ft Press, 2010.
- [7] B. Stevenson, "The impact of the internet on worker flows," *The Wharton School, University of Pennsylvania*, pp. 1–24, 2006.
- [8] E. Parry and S. Tyson, "An analysis of the use and success of online recruitment methods in the UK," *Human Resource Management Journal*, vol. 18, no. 3, pp. 257–274, 2008.
- [9] S. D. Maurer and Y. Liu, "Developing effective e-recruiting websites: Insights for managers from marketers," *Business horizons*, vol. 50, no. 4, pp. 305–314, 2007.
- [10] R. E. Petty and J. T. Cacioppo, *The elaboration likelihood model of persuasion*. Springer, 1986.
- [11] P. W. Braddy, A. W. Meade, and C. M. Kroustalis, "Organizational recruitment website effects on viewers perceptions of organizational culture," *Journal of Business and Psychology*, vol. 20, no. 4, pp. 525–543, 2006.
- [12] J. P. Tong, V. G. Duffy, G. W. Cross, F. Tsung, and B. P. Yen, "Evaluating the industrial ergonomics of service quality for online recruitment websites," *International Journal of Industrial Ergonomics*, vol. 35, no. 8, pp. 697–711, 2005.
- [13] H. Sylva and S. T. Mol, "E-recruitment: A study into applicant perceptions of an online application system," *International Journal of Selection and Assessment*, vol. 17, no. 3, pp. 311–323, 2009.
- [14] B. J. Jansen, K. J. Jansen, and A. Spink, "Using the web to look for work: Implications for online job seeking and recruiting," *Internet Research*, vol. 15, no. 1, pp. 49–66, 2005.
- [15] S. Strohmeier and F. Piazza, "Domain driven data mining in human resource management: A review of current research," *Expert Systems with Applications*, vol. 40, no. 7, pp. 2410–2420, 2013.
- [16] S. Mehta, R. Pimplikar, A. Singh, L. R. Varshney, and K. Visweswariah, "Efficient multifaceted screening of job applicants," in *Proceedings of the 16th International Conference on Extending Database Technology*. ACM, 2013, pp. 661–671.
- [17] M. Tosik, C. L. Hansen, G. Goossen, and M. Rotaru, "Word embeddings vs word types for sequence labeling: the curious case of cv parsing," in *Proceedings of NAACL-HLT*, 2015, pp. 123–128.
- [18] M. Burgard and F. Piazza, "Data warehouse and business intelligence systems in the context of e-hrm," *Encyclopedia of Human Resources Information Systems: Challenges in e-HRM: Challenges in e-HRM*, 2008.
- [19] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.
- [20] M. Kuhn, "Building predictive models in R using the caret package," *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008.
- [21] D. A. Belsley, "A guide to using the collinearity diagnostics," *Computer Science in Economics and Management*, vol. 4, no. 1, pp. 33–50, 1991.
- [22] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, vol. 33, no. 1, pp. 1–24, 2010.
- [23] T. Hastie and J. Qian, *Glmnet Vignette*, 2014, URL: http://web.stanford.edu/~hastie/Papers/Glmnet_Vignette.pdf [accessed 2017-05-09].
- [24] W. Qian, Y. Yang, and H. Zou, "Tweedies compound poisson model with grouped elastic net," *Journal of Computational and Graphical Statistics*, vol. 25, no. 2, pp. 606–625, 2016.
- [25] E. J. Atkinson and T. M. Therneau, "An introduction to recursive partitioning using the rpart routines," *Rochester: Mayo Foundation*, 2000.
- [26] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [27] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [28] A. Karatzoglou, A. Smola, and K. Hornik, *The kernlab package*, 2007, URL: <https://cran.r-project.org/web/packages/kernlab/kernlab.pdf> [accessed 2017-05-09].
- [29] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical programming*, vol. 45, no. 1-3, pp. 503–528, 1989.
- [30] B. Ripley and W. Venables, *Package 'nnet'*, 2016, URL: <https://cran.r-project.org/web/packages/nnet/nnet.pdf> [accessed 2017-05-09].