Investigating the Impact of Gender on Rank in Resume Search Engines

Supplementary Materials

FEATURE ENCODING AND NORMALIZATION

In this section, we discuss the details of how we encoded and normalized candidate features. Recall that our dataset includes three broad types types of features: 1) profile data (e.g., experience, education, *etc.*); 2) inferred gender; and 3) rank in search results. We normalize all features to be between 0 and 1 for consistency. Binary features, such as *Authorization*, *Relocate*, and *Skills Match*, are converted to 0 or 1. For other features we apply the normalization procedures described below.

- *Job Title Relevance* and *Skills Relevance* refer to the normalized fraction of keywords in a candidate's current job title and self-reported list of skills that match the terms in our query. For example, if we query for "software engineer" and a candidate's current job title is "software designer," we would assign them a *Job Title Relevance* of 0.5. We convert all words to their root form using the Porter2 word stemmer before computing the scores. Note that Monster allows candidates to enter three skills, hence we use three separate features on that site. In contrast, candidates on Indeed may enter as many skills as they wish, so we calculate the fraction of keyword matches on the aggregated skill set.
- *Education* is normalized using the ordered list of education-levels provided by Monster. The list contains "No Education," "Some High School Coursework," . . . , "Bachelor's Degree," "Master's Degree," "Doctorate," and "Professional," from low to high. We normalize each candidate's education levels uniformly, with 0 being no education, and 1 being professional. Candidates on CareerBuilder must select from the same education-level list as those on Monster. Because Indeed allows free-text input of education-level (e.g., "ba," "b.sc," "be" all mean Bachelors degree), we manually constructed a list of 50 educational keywords that appeared on Indeed and mapped them to Monster's education-level list. These 50 keywords are sufficient to cover 91% of candidates in our Indeed dataset.
- *Job Popularity* and *Skill Popularity* encode the popularity of each candidate's current job title and skills. We normalize these features by applying min-max normalization on the popularity of the current job title or skill, where popularity is computed across all candidates in a given list of search results. For example, candidates with the most popular current job title would have *Job Popularity* of 1.0.
- *Last Modified* and *Experience* encode a candidate's resume modification time and years of experience relative to all the other candidates in a given list of search results. For a given list, we compute the Cumulative Density Function (CDF) of modification times/experience for all candidates. Each candidate's feature value is then their relative rank in the CDF. For example, the candidate(s) with the most years of experience for a given query would have *Experience* of 1.0.

MATCHING

To verify that our regression models are robust, we analyze two populations of candidates: the **Original** population, which includes all candidates, and a **Matched** subset of candidates. To construct our matched subpopulation, we leverage the MatchIt software developed by Ho et al. [2]. MatchIt implements several nonparametric matching methods (e.g., *exact matching*, *coarsened exact matching (CEM)*, *nearest neighbor*, etc.) that make no assumptions about the relationship between rank and visible features. The result of the matching process is a subpopulation of data where the treatment variable is more independent of covariates, i.e., selection bias has been reduced. Because matching uses nonparametric techniques, is does not add additional assumptions to the model, beyond what is already assumed by subsequent, parametric analysis techniques [1].

In this study we use CEM, which is a relaxed version of *exact* matching: the feature values are binned, creating more flexibility to find matches.[1] For categorical features (e.g., *Searched Job Title*, *Searched City*, *Education*, etc.), we set each category as a separate bin and use exact matching. For *Experience* we set seven bins, specified in years of experience: 0–1, 1–5, 5–8, 8–10, 10–15, 15–20, and 20+. For *Last Modified* we set five bins: 1–7 days, 7–30 days, 30–90 days, 90–365 days, 1+ years.[2] We chose to manually specify bins for these two continuous features so that each bin corresponds to a human-interpretable timerange. For the remaining continuous features, we leverage the default binning algorithm in MatchIt. Tables 3 and 2 show the features we match on each hiring website,[3] each feature's type, the number of bins, and whether MatchIt's default binning was used.

[1] We tried to use *exact* matching, but it could only produce subsamples containing <1% of candidates. The problem is that the feature-space in our dataset is large, which makes it very difficult to find exactly matching pairs of candidates.

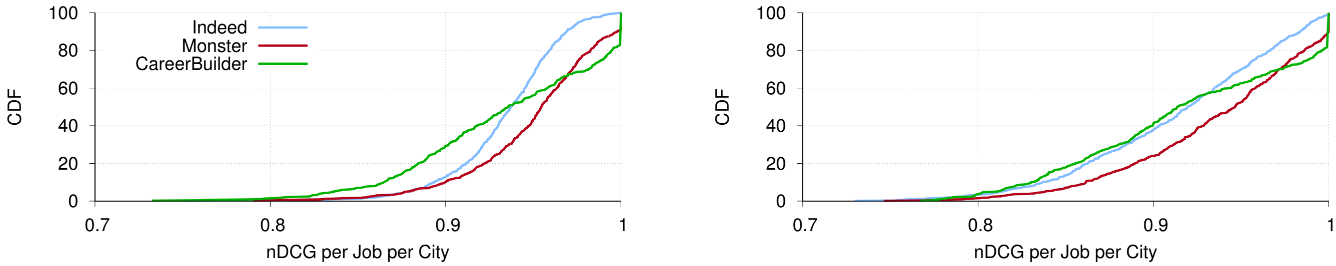
[2] We reduce the number of bins for *Experience* and *Last Modified* to four and three, respectively, when examining the top 100 candidates, to compensate for the reduced range of feature values that we observe in this restricted population.

[3] We do not include the *Job Popularity* and *Skill Popularity* features, as these are not directly observed from the candidates' information.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ISBN 978-1-4503-5620-6/18/04. . . $15.00

DOI: https://doi.org/10.1145/3173574.3174225



(a) Top 1000 candidates, **Matched**

(b) Top 100 candidates, **Matched**

Figure 1: nDCG comparison of the predicted rankings $\hat{\mathbf{R}}$ produced by the mixed linear models versus the original rankings $\mathbf{R}$, computed using the **Matched** candidates. All subfigures share the same key given in Figure 1a.

	Dependent Variable: $\log_2(\text{rank})$											
	Indeed				**Monster**				**CareerBuilder**			
	Top 1000		**Top 100**		**Top 1000**		**Top 100**		**Top 1000**		**Top 100**	
Feature	**Original**	**Matched**	**Original**	**Matched**	**Original**	**Matched**	**Original**	**Matched**	**Original**	**Matched**	**Original**	**Matched**
Fixed Effect Intercept	7.125***	7.281***	4.803***	4.826***	6.991***	7.461***	5.938***	6.275***	4.47***	5.528***	4.129***	4.601***
Job Title Relevance	-1.196***	-0.783***	-0.518***	-0.295***	-1.628***	-1.449***	-1.3***	-1.199***	-1.57***	-1.522***	-1.258***	-1.204***
Skills Relevance (1)	-0.14***	-0.275***	-0.051	-0.006	-0.268***	-0.065**	-0.31***	-0.046				
Skills Relevance (2)					-0.143***	0.002	-0.109***	0.17*				
Skills Relevance (3)					-0.096***	0.053	-0.108***	0.151				
Education level	0.086***	0.117***	0.042**	0.106***	-0.079***	-0.011	-0.061*	-0.009	-0.038*	-0.074**	-0.027	-0.034
Job Popularity	-0.084***	-0.098***	-0.115***	-0.184***	-0.163***	-0.177***	-0.004	0.009	-0.193***	-0.242***	-0.147***	-0.175***
Last Modified	-2.02***	-1.631***	-2.053***	-1.738***	-0.203***	-0.213***	-0.197***	-0.202***	-0.139***	-0.109***	-0.149***	-0.181***
Experience	0.217***	0.279***	0.116***	0.065*	-1.041***	-0.853***	-1.303***	-1.359***	-0.106***	-0.104***	-0.185***	-0.226***
Relocate					-0.019***	-0.015	-0.021	-0.048				
Skills Popularity (1)					-0.08***	-0.099***	-0.048**	-0.093**				
Skills Popularity (2)					-0.086***	-0.103***	-0.062**	-0.103*				
Skills Popularity (3)					-0.103***	-0.113***	-0.017	-0.018				
Bio Relevance	0.046***	0.073**	0.041	0.135								
Information Relevance	-0.32***	-0.251***	-0.255***	-0.134								
Skills Match	-0.034	0.066	-0.072	-0.313								
Information Match	-0.042**	0.037	-0.093*	-0.034								
Bio Match	-0.189	-0.086***	-0.262***	-0.36***								
Random Effect (s.d.)	0.072	0.062	0.01	0.008	0.27	0.231	0.106	0.066	0.229	0.196	0.018	0.118
Prob. of Being Masculine	-0.019***	-0.012*	-0.042***	-0.051**	-0.043***	-0.025**	-0.028*	-0.051*	-0.039**	-0.02	-0.071***	-0.055*
Observations	521783	179630	67410	18630	265172	83862	50813	11836	67580	30502	28289	13205

Table 1: Estimated coefficients and standard deviation of mixed linear regressions on the top 1000 and top 100 **Original** and **Matched** candidates in search results from each hiring website, grouped by city and job title. Significance level is unavailable for *Random Effect*. *Note:* $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Validation. Before regressing on the matched subpopulations, we must first verify that they are large and representative. Tables 3 and 2 show the descriptive statistics of all candidates in the top 100 and top 1000 populations, respectively, versus the corresponding matched subpopulations produced by CEM for Indeed, Monster, and CareerBuilder. We make two key observations: *first*, CEM identifies tens of thousands of matching pairs on each website (unlike *exact* matching), which means the populations are large enough to analyze. *Second*, we observe that the differences in means for the control features in the matched pairs all close to zero, which demonstrates that the matched pairs have extremely similar features. Furthermore, the improvement to the mean difference is almost always $>$90%, which shows that the matched subpopulation is substantially more balanced than the overall population. Taken together, these observations demonstrate that our matched subpopulations are suitable for analysis.

Figure 1 evaluates the goodness of fit of mixed linear models fit to our matched subpopulation of candidates. As before, we use nDCG as our metric to compare the original search results produces by the three websites, to new search results produced by our trained models. We see that our models produce nDCG scores $>$0.9 for the majority of job title/city pairs, which is extremely strong. This demonstrates that our matched models are an even better approximation of the actual ranking algorithms than our models trained on the entire candidate population.

Results. Table 1 presents a complete view of our regression models, covering top 100 and top 1000 candidate models fit to the original and matched populations, for all three hiring websites. The top 100, original models are identical to those presented in the main paper; we include them here to ease comparisons across all models.

We observe that the gender coefficient is negative in all 12 models, and significant in 11 of 12 models. We further note that the magnitude of the gender coefficient is essentially the same across all models. These results strongly highlight the consistency and robustness of our regression results.

Category	Men	Women
All	231001	233575
Matched	91336	88294
Unmatched	139665	145281
Discarded	0	0

(a) Sample Size for Indeed, Top 1000

Feature	Binning	All Data Means Men	All Data Means Women	All Data Mean Diff	Matched Data Means Men	Matched Data Means Women	Matched Data Mean Diff	Improvement Mean Diff
distance	–	0.5176	0.4877	0.0299	0.4924	0.4925	-0.0002	99.4323
Job Title	cat, 35, A	17.4159	14.3647	3.0512	16.0758	16.1012	-0.0254	99.1674
City	cat, 20, A	9.6171	9.4842	0.1329	9.2912	9.2934	-0.0021	98.3945
Job Title Relevance	num, 7, A	0.3004	0.2727	0.0277	0.2078	0.2078	0	100
Skills Relevance (1)	num, 7, A	0.0496	0.0415	0.0081	0.0043	0.0043	0	100
Education	cat, 10, A	0.4679	0.4627	0.0053	0.4301	0.4301	0	100
Last Modified	num, 5, M	0.5073	0.5019	0.0054	0.5004	0.5007	-0.0003	94.4605
Experience	num, 7, M	0.5173	0.5001	0.0172	0.4999	0.4993	0.0006	96.5701
Bio Relevance	num, 7, A	0.135	0.114	0.0211	0.0244	0.0244	0	100
Information Relevance	num, 7, A	0.3579	0.3501	0.0077	0.2444	0.2444	0	100
Skills Match	cat, 2, A	0.0086	0.0075	0.0011	0.0004	0.0004	0	100
Information Match	cat, 2, A	0.0239	0.0219	0.002	0.002	0.002	0	100
Bio Match	cat, 2, A	0.0741	0.0919	-0.0178	0.0445	0.0445	0	100
Rank	–	457.881	454.6293	3.2516	484.0105	480.9978	3.0128	7.346

(b) Balance Checking Statistics for Indeed, Top 1000

Category	Men	Women
All	112682	127754
Matched	40369	43493
Unmatched	72313	84261
Discarded	0	0

(c) Sample Size for Monster, Top 1000

Feature	Binning	All Data Means Men	All Data Means Women	All Data Mean Diff	Matched Data Means Men	Matched Data Means Women	Matched Data Mean Diff	Improvement Mean Diff
distance	–	0.5408	0.5206	0.0201	0.5522	0.5521	0	99.8889
Job Title	cat, 35, A	16.8625	14.6349	2.2276	17.9649	18.0011	-0.0361	98.3775
City	cat, 20, A	9.8693	9.7387	0.1306	9.8414	9.838	0.0034	97.3949
Job Title Relevance	num, 7, A	0.3377	0.3693	-0.0315	0.2174	0.2174	0	100
Skills Relevance (1)	num, 7, A	0.1241	0.1492	-0.0251	0.0566	0.0566	0	100
Skills Relevance (2)	num, 7, A	0.0847	0.0923	-0.0076	0.0213	0.0213	0	100
Skills Relevance (3)	num, 7, A	0.0693	0.0747	-0.0054	0.0148	0.0148	0	100
Education	cat, 12, A	0.5876	0.5811	0.0064	0.585	0.585	0	100
Last Modified	num, 5, M	0.5103	0.517	-0.0067	0.5236	0.5242	-0.0006	90.7378
Experience	num, 7, M	0.5248	0.4923	0.0326	0.507	0.505	0.0021	93.6805
Relocate	cat, 2, A	0.6541	0.6109	0.0431	0.7054	0.7054	0	100
Rank	–	384.7604	410.1525	-25.3921	448.9949	454.3637	-5.3688	78.8566

(d) Balance Checking Statistics for Monster, Top 1000

Category	Men	Women
All	30362	28762
Matched	16154	14348
Unmatched	14208	14414
Discarded	0	0

(e) Sample Size for CareerBuilder, Top 1000

Feature	Binning	All Data Means Men	All Data Means Women	All Data Mean Diff	Matched Data Means Men	Matched Data Means Women	Matched Data Mean Diff	Improvement Mean Diff
distance	–	0.4977	0.4759	0.0218	0.4908	0.4909	-0.0001	99.6468
Job Title	cat, 35, A	15.6974	13.7541	1.9434	14.6425	14.6774	-0.0349	98.2032
City	cat, 20, A	9.4957	9.1079	0.3878	9.3551	9.3538	0.0013	99.6755
Job Title Relevance	num, 7, A	0.3798	0.3866	-0.0068	0.3259	0.3259	0	100
Education	num, 7, A	0.4632	0.4039	0.0593	0.4672	0.4672	0	100
Last Modified	num, 5, M	0.5004	0.503	-0.0026	0.4958	0.4971	-0.0013	51.2495
Experience	num, 7, M	0.5416	0.5165	0.0251	0.5365	0.5351	0.0014	94.3952
Rank	–	209.7027	247.7418	-38.0391	290.333	291.0539	-0.7208	98.105

(f) Balance Checking Statistics for CareerBuilder, Top 1000

Table 2: Coarsened Exact Matching statistics for all three hiring websites on the top 1000 candidates. Note that *Rank* is not a feature in the matching procedure; we list it here to check the balance statistics. The "Binning" column shows the type of each feature (categorical or numerical), how many bins were used, and whether those bins where chosen **A**utomatically by MatchIt or **M**anually by us.

REFERENCES

1. Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15, 3 (2007), 199–236.
2. Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2011. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software* 42, 8 (2011), 1–28.

Category	Men	Women
All	29840	30012
Matched	9375	9255
Unmatched	20465	20757
Discarded	0	0

(a) Sample Size for Indeed, Top 100

		All Data			Matched Data			
Feature	**Binning**	**Means Men**	**Means Women**	**Mean Diff**	**Means Men**	**Means Women**	**Mean Diff**	**Improvement Mean Diff**
distance	–	0.5126	0.4903	0.0223	0.4941	0.4951	-0.001	95.5994
Job Title	cat, 35, A	17.9644	15.4478	2.5166	16.7918	16.9282	-0.1364	94.5795
City	cat, 20, A	9.5852	9.431	0.1542	9.6072	9.5984	0.0089	94.2436
Job Title Relevance	num, 7, A	0.4091	0.3923	0.0168	0.3211	0.3211	0	100
Skills Relevance (1)	num, 7, A	0.0544	0.0475	0.0068	0.0054	0.0054	0	100
Education	cat, 10, A	0.466	0.4784	-0.0123	0.4182	0.4182	0	100
Last Modified	num, 3, M	0.3307	0.3345	-0.0038	0.29	0.2874	0.0027	30.6878
Experience	num, 4, M	0.482	0.4683	0.0137	0.4331	0.4334	-0.0003	97.7444
Bio Relevance	num, 7, A	0.1333	0.1118	0.0215	0.02	0.02	0	100
Information Relevance	num, 7, A	0.3863	0.3897	-0.0034	0.2594	0.2594	0	100
Skills Match	cat, 2, A	0.0109	0.0086	0.0023	0.0009	0.0009	0	100
Information Match	cat, 2, A	0.0269	0.0227	0.0042	0.0023	0.0023	0	100
Bio Match	cat, 2, A	0.0932	0.1161	-0.0229	0.0538	0.0538	0	100
Rank	–	49.7075	50.2131	-0.5057	49.2912	50.6214	-1.3302	-163.0531

(b) Balance Checking Statistics for Indeed, Top 100

Category	Men	Women
All	20339	26183
Matched	5652	6184
Unmatched	14687	19999
Discarded	0	0

(c) Sample Size for Monster, Top 100

		All Data			Matched Data			
Feature	**Binning**	**Means Men**	**Means Women**	**Mean Diff**	**Means Men**	**Means Women**	**Mean Diff**	**Improvement Mean Diff**
distance	–	0.5714	0.5518	0.0196	0.5667	0.5668	-0.0002	99.224
Job Title	cat, 35, A	17.679	15.149	2.53	17.38	17.4438	-0.0638	97.4788
City	cat, 20, A	9.7571	9.6203	0.1368	9.7278	9.7089	0.0189	86.1649
Job Title Relevance	num, 7, A	0.5593	0.577	-0.0177	0.5244	0.5244	0	100
Skills Relevance (1)	num, 7, A	0.1743	0.1813	-0.007	0.0963	0.0963	0	100
Skills Relevance (2)	num, 7, A	0.1129	0.1083	0.0045	0.0319	0.0319	0	100
Skills Relevance (3)	num, 7, A	0.0869	0.082	0.0049	0.0147	0.0147	0	100
Education	cat, 12, A	0.6069	0.6091	-0.0022	0.6184	0.6184	0	100
Last Modified	num, 3, M	0.5086	0.5198	-0.0112	0.5302	0.5326	-0.0024	78.8243
Experience	num, 4, M	0.5735	0.5627	0.0108	0.5829	0.5799	0.003	71.8669
Relocate	cat, 2, A	0.6545	0.6127	0.0418	0.716	0.716	0	100
Rank	–	46.1663	46.7084	-0.5421	48.0163	49.7964	-1.78	-228.3505

(d) Balance Checking Statistics for Monster, Top 100

Category	Men	Women
All	11775	13161
Matched	6278	6927
Unmatched	5497	6234
Discarded	0	0

(e) Sample Size for Career-Builder, Top 100

		All Data			Matched Data			
Feature	**Binning**	**Means Men**	**Means Women**	**Mean Diff**	**Means Men**	**Means Women**	**Mean Diff**	**Improvement Mean Diff**
distance	–	0.5394	0.5149	0.0245	0.5333	0.5336	-0.0002	99.0317
Job Title	cat, 35, A	16.7937	14.0781	2.7156	16.0784	16.1792	-0.1008	96.2887
City	cat, 20, A	9.3992	9.118	0.2812	9.3211	9.35	-0.0289	89.7113
Job Title Relevance	num, 7, A	0.4818	0.4967	-0.0148	0.4808	0.4808	0	100
Education	num, 7, A	0.459	0.4246	0.0344	0.4589	0.4589	0	100
Last Modified	num, 3, M	0.4969	0.4988	-0.0019	0.4922	0.4881	0.0041	-118.9678
Experience	num, 4, M	0.5434	0.5143	0.0291	0.534	0.5271	0.0069	76.3071
Rank	–	38.6497	40.4052	-1.7555	41.7055	42.1318	-0.4263	75.7162

(f) Balance Checking Statistics for CareerBuilder, Top 100

Table 3: Coarsened Exact Matching statistics for all three hiring websites on the top 100 candidates. Note that *Rank* is not a feature in the matching procedure; we list it here to check the balance statistics. The "Binning" column shows the type of each feature (categorical or numerical), how many bins were used, and whether those bins where chosen **A**utomatically by MatchIt or **M**anually by us.

Variable	**Mean**	**s.d**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**
(1) Job Title Relevance	0.29	0.38											
(2) Skills Relevance (1)	0.05	0.18	0.02										
(3) Education level	0.47	0.35	0.11	0.04									
(4) Job Popularity	0.04	0.19	0.33	0.00	-0.02								
(5) Last Modified	0.50	0.29	0.10	-0.04	-0.03	0.05							
(6) Experience	0.50	0.29	0.04	-0.02	0.00	-0.05	0.18						
(7) Bio Relevance	0.13	0.27	0.09	0.40	0.07	0.02	0.02	0.01					
(8) Information Relevance	0.35	0.38	0.11	0.08	0.08	0.00	0.05	0.07	0.11				
(9) Skills Match	0.01	0.09	-0.01	0.44	-0.03	0.01	-0.01	-0.01	0.17	-0.01			
(10) Information Match	0.02	0.15	0.04	0.17	-0.01	0.04	0.01	0.00	0.43	-0.01	0.42		
(11) Bio Match	0.08	0.28	0.03	-0.01	-0.04	0.03	0.04	0.02	-0.03	0.45	0.05	0.06	
(12) Prob. of Being Masculine	0.50	0.46	0.03	0.02	0.01	-0.01	0.01	0.03	0.04	0.01	0.01	0.01	-0.03

(a) Indeed

Variable	**Mean**	**s.d**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
(1) Job Title Relevance	0.36	0.43												
(2) Skills Relevance (1)	0.14	0.29	0.26											
(3) Skills Relevance (2)	0.09	0.24	0.18	0.08										
(4) Skills Relevance (3)	0.07	0.22	0.15	0.08	0.10									
(5) Education level	0.59	0.24	0.05	0.09	0.08	0.08								
(6) Job Popularity	0.12	0.30	0.38	0.13	0.06	0.05	-0.05							
(7) Last Modified	0.51	0.29	0.02	0.00	0.00	0.00	0.01	0.02						
(8) Experience	0.50	0.29	-0.09	0.00	0.00	-0.01	0.02	-0.07	0.05					
(9) Relocate	0.64	0.48	0.00	-0.01	-0.01	-0.01	0.04	0.00	0.09	-0.07				
(10) Skills Popularity (1)	0.31	0.38	0.22	0.58	0.02	0.02	0.03	0.15	0.00	0.00	-0.01			
(11) Skills Popularity (2)	0.19	0.30	0.15	0.01	0.53	0.03	0.02	0.08	0.01	0.00	0.00	0.02		
(12) Skills Popularity (3)	0.14	0.26	0.10	-0.01	0.03	0.50	0.01	0.06	0.00	-0.01	0.00	-0.02	0.02	
(13) Prob. of Being Masculine	0.53	0.47	-0.04	-0.04	-0.01	-0.01	0.01	-0.03	-0.01	0.06	0.04	-0.06	-0.03	-0.02

(b) Monster

Variable	**Mean**	**s.d**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
(1) Job Title Relevance	0.39	0.43					
(2) Education level	0.44	0.36	0.03				
(3) Job Popularity	0.13	0.30	0.50	-0.02			
(4) Last Modified	0.50	0.29	0.00	0.03	0.00		
(5) Experience	0.52	0.28	-0.13	-0.01	-0.10	0.04	
(6) Prob. of Being Masculine	0.49	0.46	-0.01	0.08	-0.04	0.00	0.04

(c) CareerBuilder

Table 4: Means, standard deviation, and correlations of the variables on Indeed, Monster, and CareerBuilder. These results were computed using the **Original** candidates from our dataset.