

## An Innovative Approach to Sales Lead Generation Using Machine Learning

By

Denis Edwards, University of South Florida

Even as business is lost to predatory cost undercutting from competitors, many staffing firms continue to rely heavily, if not solely, on relationship-based direct sales strategies. In a relationship-based sales approach, sales associates develop and seek to expand deal flow based on an intimate understanding of the employing firm and its hiring needs. The challenge with such an approach is the time and expense required to develop and cultivate

these relationships in an environment where employers view staffing services as a commodity. We posit that newer, more innovative approaches are required to augment relationship based sales strategies. Working with a leading U.S.

staffing firm, we developed a machine learning solution to enable the powerful combination of automation and relationships to gain a competitive advantage in the U.S. staffing market.

The source of new sales leads were Internet job boards where hundreds of employers signal their clear intent to hire employees by placing thousands of ads on these job boards each day. The partner firm believed that identifying new sales opportunities could be accomplished by intercepting and evaluating job ads placed by

employers or competing staffing firms. While job ads placed by employers and competing staffing firms provide potential sales leads, those placed by competitor staffing firms represent the most potential value. Competitor placed job ads represent an opportunity to take market share from the competitor and indicates that the employer the competitor represents has the propensity to use staffing firm services to fill open job vacancies.

**New innovative sales approaches are critical to generating revenue in the U.S. staffing industry. We design and implement a machine learning based sales lead generation solution in-situ with a partner U.S. staffing firm.**

While job ads placed by competitor firms are more valuable than those placed directly by employers, ads placed by staffing firms have a unique challenge in that the identity of the employer represented by these ads is intentionally obfuscated.

Employer identities are obfuscated as a defensive measure to protect the staffing firm's perceived business relationship with the employers they represent. The ability to automatically determine the identity of the employer represented by a competitor placed ad is critical to the partner firm's ability to contact the employer and act on the potential lead. The complexity involved in determining anonymized employer identities represents what researchers term a *sticky wicked problem*.

Keywords: Sales, Innovation, Machine Learning, Stylometry, Elaborated Action Design Research (eADR), Text Obfuscation

We believed our sticky wicked problem could be addressed with machine learning and stylometry, which is a research area that uses writing styles of authors to identify anonymous authors of literature, tweets, emails and many other document types. Our hypothesis was that stylometry could be used to identify employers based on their historic writing styles when creating job ads just as authors of literature, emails and texts have been identified in other studies. Our research questions related to what machine learning approach and stylometric features (words, sentences, etc.) provided the best performance for identifying employers from job ad content alone.

Using the elaborated Action Design Research (eADR) method (Mullarkey & Hevner, 2019) as our research method, researchers and practitioners iterated through numerous eADR design and evaluation cycles until we discovered a machine learning algorithm and stylometric features that met our research goal of 90% or greater prediction of what employer is represented by a job ad.

## Findings

Our research involved multiple experiments using different supervised machine learning solutions. Determining the identity of obfuscated employers was based on the probability the job ad could be attributed to one of  $n$  employers of interest to the partner firm. Hence the scope of supervised machine learning algorithms was limited to those that supported probabilities as outputs. We limited our research to supervised machine learning solutions due their successful use in research studies similar to ours (Stamatatos, 2009). The stylometric feature sets used in research were based on lexical features such as the count of words in a corpus of documents being evaluated as this approach also provides the best results for our research (2009).

Supervised machine learning solutions are able to predict employer identities when trained to identify the employer's job ad writing style. To accomplish this, we accumulated job ads placed directly by employers and trained solutions using these job ads. Since staffing companies typically cut and paste an employer's job ad text into the ads they place after obfuscating text useful to identifying the employer, the machine learning solution is able to compare common stylometric values in known employer ads and obfuscated job ads to make its predictions.

The supervised machine learning algorithms we evaluated included Naïve Bayes, Logistic Regression, State Vector Machines (SVM) and Multilayer perceptrons (MLP). After three eADR design and evaluation cycles we determined that a multilayer perceptron model (MLP) using a stylometric feature set of up to 1,000 words occurring most frequently across the corpus of job ads evaluated resulted in

prediction accuracies greater than our goal of 90% without overfitting. We found that each of the supervised machine learning algorithms performed well, however the MLP model performed with slightly better accuracy (99.5%) while requiring less time and computing resources. We refer to the solution comprised of the trained MLP algorithm and use of lexical features (word counts) as the who.am.i classifier.

To evaluate the predictive accuracy of the who.am.i classifier using production data, we planned and executed an eADR implementation cycle to determine if the classifier was capable of generalizing to new data. The partner firm's sales strategy at the time of our research was to identify from targeted employers in specific geographic areas known to do business with competitor staffing firms. Using job ads posted by ten competitor staffing companies in the specified geographies up to thirty days prior to implementation, we found the who.am.i classifier correctly identified an employer from the firm's target employer list associated with a job ad 95.4% of the time.

A total of fifty one new job leads were generated by the who.am.i classifier, which included all results where an employer was predicted with a probability of .5 or greater.

## Conclusions

Working with a leading U.S. staffing firm, we used stylometry and machine learning to generate new sales leads. The leads were identified by quickly discovering and responding to employers actively seeking new employees by intercepting and evaluating job ads placed on-line by employers or a staffing firm on an employer's behalf. Job ads placed by competitor firms are most valuable to the firm we worked with as they potentially represent an opportunity to take market share, but identifying the employer represented in these job ads proved difficult due to the common staffing industry practice of removing all clues to the employers identities prior to placing the ads. A machine learning solution was developed to identify the employer's identity from these obfuscated job ads. The machine learning solution is able to identify employers even from obfuscated job ads because employers constantly place job ads on their own that are not obfuscated. Using employer based job ads, the machine learning solution was able to learn the employer's unique job ad writing style. The trained solution was then able to identify this style in the obfuscated job ads placed by staffing firms and hence identify the employer.

Following the eADR research method, we designed and evaluated several supervised machine learning algorithms to determine which best achieved our goal of consistently predicting employer identities from obfuscated job ads greater than 90% of the time. We found that an MLP solution using the fre-

quency of up to 1,000 words across all job ads evaluated accurately predicted the employer represented by a competitor placed job ad 95.4% of the time based on at least a .5 probability. An initial test of the solution in production identified fifty-one potential sales leads previously unknown to the partner firm.

## References

- Mullarkey, M. T., & Hevner, A. R. (2019). An elaborated action design research process model. *European Journal of Information Systems*, 1, 6. <https://doi.org/10.1080/0960085X.2018.1451811>
- Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science & Technology*, 60(3), 538–556. <https://doi.org/10.1002/asi.21001>

## Review

This article was accepted under the *constructive peer review* option. For further details, see the descriptions at:

<http://mumabusinessreview.org/peer-review-options/>

## Authors



**Denis Edwards** is an Information Technology professional with more than thirty-five years' experience across a variety of industries including hospitality, movies and music, Big 4 consulting, manufacturing and workforce solutions. For the past fifteen years he has held roles as CIO or CTO in companies ranging in size from startups to global, Fortune 150 firms. Mr. Edwards has earned a reputation as a successful turn-around executive. Mr. Edwards education includes a B.S. in Computer Systems from Nova Southeastern University, M.S. in Information Technology from Concordia University and a Doctorate in Business Administration from the University of South Florida.