

Thorogood case study
Insurance

Using Artificial Intelligence to Make Better Hires: A Small-Scale Project Becomes a Full-Scale Solution

Suppose it were possible to predict whether job candidates will be successful hires based solely on their résumés. For data scientists at this financial services firm, it was a tantalizing proposition that they believed could be achieved by harnessing the power of artificial intelligence.

But it was also a logistical challenge. After the data scientists successfully developed an algorithm capable of predicting the future job performance of candidates, the company in question needed to develop the technology that would allow the tool to operate on a large scale. To do so, it enlisted the help of Thorogood, whose consultants responded by delivering a blueprint for a full-scale system that would enable hiring managers across the sprawling company to incorporate the powerful new approach into their recruitment process.

A big first step

Talent recruitment had long been a complex and time-consuming task for this company. The firm features a broad organizational structure that consists of a number of wholly owned subsidiaries spread across the country. Each of these subsidiary brokerages specializes in different products and regularly sees a considerable degree of turnover among the agents responsible for selling these products. To keep pace with its staffing needs, the company is continuously scouring résumés from a variety of different sources and funneling promising candidates to the appropriate hiring managers. But with tens of thousands of new résumés available from a variety of sources each and every day, the process



of picking out the good candidates is a superhuman task - which is why it tasked its data scientists with creating a superhuman solution.

Using sales data, the data scientists defined buckets of job performance by their current and previous insurance agents. Within R scripts, they then applied machine learning and statistical techniques to the analysis of these agents' résumés in order to identify key topics that were most commonly discussed within these documents. Next, the data scientists combined the historical employees' previous job performance buckets with a set of attributes denoting whether the topics were present in those employees' résumés. A subset of this data could be used for both feature selection and model training of a classification algorithm which predicts job performance classification of future candidates based on the presence of the key topics in their submitted or acquired résumés, with the remaining data set aside to test model performance.

After arriving at a satisfactory model, the data scientists wrote additional R scripts to populate the topic presence attributes and apply the classifier to the new observations for new résumés in order to group the corresponding candidates into predicted performance buckets.

The initial exploration was a success, and the client knew that it had the makings of a powerful new human resources tool. But creating such a tool would require taking the code it developed to run locally and manually on data scientists' machines and productionizing it to a point where it could handle the demands of an enterprise-level firm. To do that, it needed help.

Turning a project into a full-scale solution

In collaboratively designing the application, Thorogood carefully took into account the technologies and skillsets that already existed within the company's data landscape. Informatica PowerCenter would pull résumés from a SQL Server database and perform a variety of data cleansing and transformation tasks, both on the résumés themselves and on attributes like the subsidiary for which the candidate's fit is being analyzed. From there, PowerCenter would call the scoring model, which was deployed to and hosted as a web service on Microsoft Machine Learning Server, passing the appropriate dimensional information and the actual cleaned résumé text. The web service would then perform the scoring and return a predicted job performance score and a thresholded value to

PowerCenter for each candidate. There, PowerCenter would complete final transformations and return the results back into the SQL database.

There are a number of challenges in taking a small-scale data project and turning it into a robust, user-friendly solution. When Thorogood's consultants began work on the project, their first step was to examine the company's existing code and identify the lines that could be streamlined in order to perform at scale in an efficient and cost-effective manner. Among the critical considerations in such a transition is the processing power required to perform certain tasks. Running data science applications in the cloud can incur significant costs if not well-designed, making it essential to identify tasks that can be performed in lower-cost environments. Another important consideration is identifying the tasks that need to be performed on a flexible and accessible platform instead of the piece of the solution that does most of the heavy lifting.

In designing the solution, Thorogood's consultants allowed for plug-and-play functionality to accommodate the data scientists' need to continually update the R model and scoring script. In doing so, they had to decide which pieces of data preparation and transformation could be handled ahead of time by Informatica PowerCenter and which should be left within the R scoring script to allow for those pieces to be easily updated by the data scientists as needed. The result was an application that maximized responsiveness, reduced scoring time, and increased the frequency at which the résumés are scored, all while minimizing costs.

New frontiers, new considerations

The realm of Artificial Intelligence poses a set of considerations above and beyond the logistical challenges of developing enterprise-scale solutions. A solution like the one Thorogood designed comes fraught with ethical implications, given the algorithm's reliance on previous hires. Absent deliberate checks and balances, any implicit or explicit human biases that were present in previous hiring decisions will be baked into future ones, despite the offloading of responsibility from man to machine. A key design decision in this solution was therefore giving the data scientists the ability to force their models not to consider a specific attribute in their scoring if they deemed it to be discriminatory. As with any technological endeavor, the right expertise is critical in ensuring that ML-based solutions do not perpetuate preexisting biases.

Microsoft
Partner

Gold Data Analytics
Gold Data Platform
Gold Cloud Platform
Gold Datacenter
Gold Application Development



Find out more:

Contact: Scott.Stieritz@thorogood.com
Scott Stieritz - BI & Analytics Consultant