Does Selection Bias Blind Performance Diagnostics of Business Decision Models? A Case Study in Salesforce Optimization

Jun Wang, Moninder Singh, Kush R. Varshney
Business Analytics and Mathematical Sciences Department
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598
Email: {wangjun, moninder, krvarshn}@us.ibm.com

***Abstract*—Modern business decision models are often very complicated due to a deluge of information. Evaluation and diagnostics of such decision models is extremely challenging due to many factors, including the complexity and volume of data. In addition, since there is no ideal data sample to construct a control group for comparison studies, performance evaluation and diagnostics of business actions can easily be distorted by *selection bias*. In this paper, we design a framework to analyze this sample bias issue under a practical business scenario. In particular, we focus on: a) identification of the key factors which drive selection bias during the business decision; b) evaluation of the performance of business actions with consideration of the identified selection bias. We evaluate baseline analytics tools on the worldwide salesforce data of a large global corporation and clearly demonstrate that the selection bias issue makes the usual evaluation very unstable and not trustable. However, by removing such detected sample bias, our framework can generate reasonable diagnostics results across different dimensions. The implemented analysis tool was applied to a worldwide business opportunity dataset of a multinational Fortune 500 corporation; the analytics results clearly show the significance of such a bias detection-based evaluation framework for salesforce optimization.**

I. Introduction

Companies are increasingly turning to the use of business analytics to identify areas where the productivity of their salesforce can be significantly improved, thereby resulting in increased revenues and profits. This is especially true for organizations that have large, often globally-distributed client-facing sales organizations dealing with increasingly complex portfolios of ever-changing products and services.

While anecdotal evidence and case studies provide sufficient evidence to show the benefits of freeing up sellers' time by taking non-productive back-office type tasks off their hands, there is limited data-driven, analytics-based evaluations or insights on this topic and limited quantitation of the effects. Insights from anecdotes and case studies may be contaminated by sample *selection bias*, the systematic error in statistical inference due to a non-random sample of a population. Furthermore, from such studies it is not clear in which situations and for which seller/opportunity characteristics the relief of routine pre-sales activities most helps sellers.

Sellers are naturally apprehensive of turning over their opportunities to other employees who are not only functionally removed, but also often physically distant in central locations where they cover several geographies. As such, sellers feel that the support staff may not be able to give the same kind of attention to their deals as they could themselves. This becomes even more critical as sellers' compensation is often directly dependent on their deals.

Although there is ample evidence that transferring non-selling activity can lead to improved seller productivity and is therefore better for an organizations' performance [1], [2], [3], individual sellers may be reluctant to engage in this process wholeheartedly. Even when sellers do participate in this process, their participation may be less than optimal in that they may pass only certain deals over to the support staff. They may only pass over deals they feel are almost certain to close. On these deals, the seller feels that there is a high probability that support staff taking over the deal will not affect it. They may also only pass over deals that are small in size and keep large deals to themselves, as the large deals have a greater impact on their individual compensation.

In order to have sellers more widely accept this process, it is imperative to provide them with convincing evidence to show that it is of benefit to them, and quantify the benefit to the extent possible. With the availability of huge amounts of data, analytics provides us with the opportunity to do exactly this. The aim of this exercise is thus to explore whether the use of support staff actually helps sellers (in terms of various metrics such as number of deals that the sellers can carry by delegating some of their responsibilities to support staff, rate of winning deals, number of deals won, etc.) as well as, more importantly, quantify the impact of this exercise on various metrics. However, in order to do this, one must first account for the presence of selection bias in the data itself; otherwise, the results may be spurious. Figure 1 illustrates the general analytics framework that we construct and work under in this investigation.

The remainder of the paper is organized as follows. In Section II, we present the business problem we are studying, including the mathematical notation. In Section III, we present

978-1-4673-2401-4/12/$31.00 ©2012 IEEE

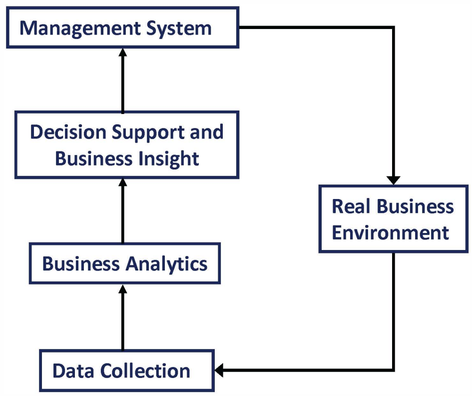

Fig. 1. An illustration of the general framework of analytics-based salesforce optimization.

the analytics framework that we use to detect selection bias and take it into account in inference. In Section IV we present a case study on actual data from a large multinational corporation. Section V concludes and presents directions for future work.

II. Problem Formulation

Assume the given business opportunity data is $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$, where each data sample is represented by a set of D business characteristics $\mathbf{x}_i = \{x_{i1}, \cdots, x_{iD}\}$. Accordingly, $\mathbf{O} = \{o_i\}_{i=1}^n$ denotes the set of business operations applied to $\mathbf{X}$. A typical example of binary business operation in salesforce optimization is the decision of whether to send the business opportunity to a third-party for handling, where $o_i = 1$ indicates the existence of the involvement of the third-party and $o_i = 0$ means no third-party is engaged for $\mathbf{x}_i$. Finally, a set of critical metrics z is predefined to evaluate the business performance. In our case study for salesforce optimization, the evaluation metrics z often involve some key factors about the salesforce performance. As mentioned in [4], these metrics can be defined in different levels of granularity, e.g. business opportunity, seller, territory, or business unit levels, each of which can be measured from different perspectives, such as winning rate and revenue size.

If there is no clear selection bias in the assignment of business actions across opportunities, i.e. the business actions are simply assigned independent of the characteristics of a business opportunity, an easy and straightforward performance evaluation can be performed to measure the impact of such business operations. For instance, one can simply compare the evaluation metrics like winning rates between the opportunity set with $o_i = 1$ and that $o_i = 0$. This is equivalent to estimate winning rates conditional to action $\mathbf{O}$ [5], [6]. However, such assumption often violates the real-world situation since people tend to assign business actions to the opportunities with certain properties, as discussed in Section I. Therefore, such evaluation easily falls into "a comparison of apples and oranges," resulting in invalid business justification.

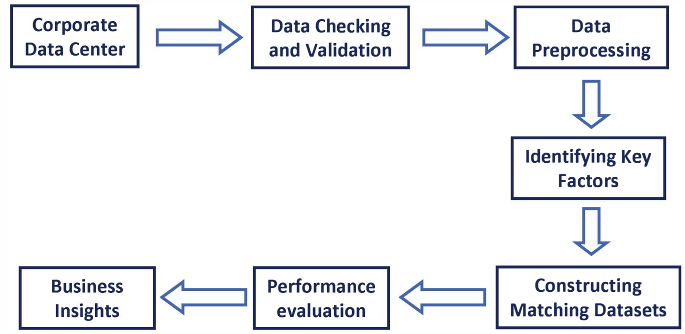

Fig. 2. The proposed framework for salesforce performance evaluation under selection bias.

Realizing the performed business operation tends to be linked to some key characteristics instead of randomly being assigned, we propose a framework herein to identify and handle selection bias to derive valid performance evaluation.

III. Performance Evaluation under Selection Bias

Here we first introduce the overview of the proposed performance evaluation framework and then detail each of the key components.

A. Overview

Figure 2 illustrates the conceptual diagram of the proposed framework for selection bias-based salesforce performance evaluation. Briefly speaking, the processing component of *data checking* removes invalid data points, such as those with significant number of missing values and those with inconsistencies. Since the characteristics of business opportunities can contain various type of values, ranging from continuous, to string, to categorical types, it is necessary to convert all these types into a uniform format for easy processing in the subsequent steps.

There are three key components in the proposed framework, i.e. *identifying key factors*, *constructing matching set*, and *performance evaluation*. To alleviate selection bias during the decision making procedure, the first step is to identify those key factors which bring significant bias conditional to the business actions. We applied the conditional mutual information based features selection technique to achieve this goal. After extracting the key features, the next step is to partition and group the opportunity data to construct comparable matching sets. Finally, the measurement metrics are estimated over the matched datasets to derive justifications and insights. In the following subsections, we will describe the details of each step.

B. Identifying Key Business Factors of Bias Selection

As mentioned earlier, the characteristics of the business opportunity contain multiple data types, including binary, continuous, and categorical. Some examples of the opportunity features are listed in Table I.

To adapt to a uniform feature selection procedure, we first encode all the opportunity characteristics into binary codes.

TABLE I
SOME EXEMPLAR FEATURES AND THE CORRESPONDING DATA TYPES OF THE BUSINESS OPPORTUNITY.

Feature Name	Type
revenue size (in $US)	continuous
won indicator	binary
business name	categorical
cycle time (in days)	continuous
cross brand indicator	binary
country name	categorical
product segment	categorical
market segment	categorical

For a categorical feature with k categories, we represent it using a k-length binary codes, each of which indicates the presence or absence of the selection of the corresponding category. For continuous-valued samples, we segment the values into several categories and then use binary codes for representation. For instance, for the revenue number, we convert it to four categories, i.e. small ($< \$100K$), medium ($\geq \$100K$ and $< \$500K$), large ($\geq \$500K$ and $< \$1M$), and very large ($\geq \$1M$). In the remaining of this paper, we reuse the same symbol $\mathbf{X}$ to denote the finally converted binary features without specific clarification.

The next step is to identify a subset of key features that contribute the most information towards inferring the operation O. Intuitively, for the salesforce optimization problem, people make the decision of whether to apply a certain action to a business opportunity by inspecting several key dimensions. Motivated by [7], here we design a conditional mutual information-based feature selection method to identify key factors. First, let us define the mutual information between a feature x_i and the action variable o.

$$I(o; x_i) = H(o) - H(o|x_i) \tag{1}$$

where $H(o)$ is the marginal entropy and $H(o|x_i)$ is the conditional entropy. If the action o is independent of feature x_i, it is easy to see that $H(o|x_i) = H(o)$ and $I(o; x_i) = 0$. On the other hand, if x_i completely determines o, $I(o; x_i)$ is maximized since $H(o|x_i) = 0$. Through maximizing the mutual information, we can identify the most important features which contribute the most to the decision of business action o.

However, only considering the mutual information between individual features and action variable o might result in the selection of a set of highly correlated features, each of which has high mutual information with o. It is desired to identify a set of features which have high relevance with the action and also has low redundancy [8]. Therefore, we use the conditional mutual information $I(o; x_i|x_j)$, which measures the extra information contributed by x_j given that x_i is already selected.

$$I(o; x_j|x_i) = H(o|x_i) - H(o|x_i, x_j) \tag{2}$$

Obviously, $I(o; x_j|x_i)$ will be small if either x_j does not affect o or the information contained in x_j is overlapped with that in x_i. In other words, x_j won't be selected if either x_j is irrelevant with o or x_j is redundant to x_i. Therefore, the final objective is to select an optimal set of features which can be used to effectively infer o, while maintaining minimum redundancy among these selected features. However, the solution to such an optimization problem can be infeasible due to both data or computation limitations [7]. First, the dataset might not be sufficient for the estimation for all the possible combinations. Second, the selection of a subset of features is essentially a combinatorial problem, which is computationally intractable.

Therefore, an alternative way is used by iteratively selecting the features in a greedy way. Briefly speaking, in each step, we select the feature which can carry the most additional information about o. Assume the selected feature set is denoted as $\mathcal{X}$ and it is initialized as an empty set $\mathcal{X}(0) = \emptyset$ at the beginning. The iterative scheme to select the most informative feature from time t to $t+1$ is defined as the following maximization problem.

$$\begin{aligned} x^*(t+1) &= \arg \max_{x_i \in \mathbf{X} \setminus \mathcal{X}(t)} I(o; x_i|\mathcal{X}(t)) \\ \mathcal{X}(t+1) &= \mathcal{X}(t) \cup x^*(t+1) \end{aligned} \tag{3}$$

where $\mathbf{X} \setminus \mathcal{X}(t)$ is the complementary set of $\mathcal{X}(t)$. Recall some binary features are converted from a single categorical feature and the selection should be exclusive. Therefore, we also exclude those single-categorical element converted binary features if one of them is already selected in previous iterations. Finally, the first selected feature is identified by mutual information instead of conditional mutual information since $\mathcal{X}(0) = \emptyset$, i.e.,

$$\begin{aligned} x^*(1) &= \arg \max_{x_i \in \mathbf{X}} I(o; x_i) \\ \mathcal{X}(1) &= \emptyset \cup x^*(1). \end{aligned} \tag{4}$$

The above iterative scheme ensures that each selected new feature is informative to the division of the action and less relevant to the existing ones, resulting in a set of key factors for inferring o. In additional, the updating rule of maximizing $I(o; x_i|\mathcal{X}(t))$ is equivalent to maximizing $I(x_i, \mathcal{X}(t); o) - I(\mathcal{X}(t); o)$, as proved in [9].

C. Constructing Matching Sample Set

Assume the selected key features $\mathcal{X} = \{\tilde{x}_i\}_{i=1}^m$. Now we can partition the opportunity data into subsets, namely matching sets, using these key features. Here, the matching set indicates that each opportunity in the same subset has the same value of the key factors. When we evaluate a business action o_i over such peer groups individually, the comparison results will be valid since the evaluation metrics are estimated over a population of opportunities with identical key factors.

Note that the identified key features are all binary-valued. The number of possible combination of such matching groups is exponential to the number of selected features, i.e. 2^m. To obtain meaningful matching groups with sufficient sample

TABLE II

THE PERCENTAGE OF BUSINESS OPPORTUNITIES ASSOCIATED WITH THE ACTION FOR DIFFERENT YEARS AND QUARTERS. TWO DIFFERENT BRANDS, I.E. A AND B, ARE INVESTIGATED HERE.

$Brand$	**A**					**B**				
Quarter	First	Second	Third	Fourth	Average	First	Second	Third	Fourth	Average
2008	16.00%	17.79%	16.56%	16.97%	16.78%	11.27%	11.06%	11.08%	12.10%	11.37%
2009	20.74%	24.63%	26.03%	26.18%	24.10%	12.45%	12.78%	13.75%	14.93%	13.39%
2010	27.67%	27.12%	25.30%	24.40%	26.21%	18.05%	19.47%	18.69%	19.55%	18.89%
Total	23.74%	24.99%	21.93%	21.97%	-	16.47%	17.01%	15.12%	16.54%	-

TABLE III

THE PERCENTAGE OF BUSINESS OPPORTUNITIES ASSOCIATED WITH THE ACTION FOR THE OPPORTUNITIES WITH DIFFERENT "REVENUE SIZE" AND "CROSS BRAND". CLEAR BIAS SELECTION CAN BE OBSERVED ON BOTH BRANDS.

$Brand$	**A**			**B**		
Cross Brand	NO	YES	TOTAL	NO	YES	TOTAL
Small ($< \$100K$)	15.85%	27.66%	16.09%	12.58%	27.78%	12.71%
Medium ($\$100K \sim \$500K$)	25.95%	45.22%	26.69%	17.69%	39.39%	18.23%
Large ($\$500K \sim \$1M$)	37.45%	58.02%	36.68%	23.50%	42.75%	24.21%
Very Large ($\geq \$1M$)	45.19%	67.62%	46.73%	28.07%	61.74%	29.28%
Total	21.18%	41.27%	-	14.05%	34.96%	-

populations, the number of groups k should be small enough. Furthermore, when performing evaluation on a certain dimension, such as evaluating the effect of the business actions over different territories or business units, we can group such matching sets in a hierarchical way to obtain enough samples.

D. Performance Evaluation Over Matching Sets

Given those matching sets, valid performance evaluation can be conducted. To justify whether a business action is effective, the evaluation metrics should be defined first. As mentioned earlier, each business action can be evaluated from different perspectives and different levels. For instance, if we use the winning rate as an evaluation metric, denoted as w, we can estimate the winning rates for individual business opportunity, seller, business unit, country, territory levels. However, some of the evaluation metrics are only feasible for certain levels. For example, when measuring the productivity of a seller, we can measure the capacity for sellers as the number of carried opportunities.

In our case study, we are particularly interested in the evaluation of winning rate. In particular, we estimate the winning rate in two different levels: opportunity and seller. The evaluation results can be used to justify whether this action effectively increases the winning chance for business opportunities. In addition, the productivity over seller level is also extensively estimated to justify whether the action (sending opportunity to third-parity for handling in our case study) can release certain workforce and improve efficiency for sellers.

IV. CASE STUDY IN SALESFORCE OPTIMIZATION

In this section, we describe the details of our real case study.

A. Data and Material

In the case study, we used samples of business data from two business units (brand A and brand B) of the investigated multinational corporation from 2008 to 2010. Brand A contains $204,004$ and brand B has $203,176$ valid business opportunity records, each of which contains 116 original attributes. There are a total of $27,263$ and $9,867$ unique sellers in brand A and B, respectively. The business decision model launched is to decide whether to take the action of sending the business opportunity to a third-party for processing after the opportunity has been validated. As discussed previously, the underlying motivation is to improve processing efficiency, allow the sellers to spend more time selling, and increase the chance to win such opportunities. Table II shows the percentage of opportunities with such an action in different years and quarters. It is easy to see there is no clear action bias across different quarters in a year for both brands. However, in general, from 2008 to 2010, there is a clear trend for both brands that more and more opportunities tend to be sent for third-parity processing. In particular, the percentage increased from 16.78% to 26.21% for brand A and from 11.37% to 18.89% for brand B. As such, it is imperative to evaluate and quantify the impact of such business actions.

B. Results

First, we applied the feature selection algorithm described in Section III-B to detect the key business factors of the

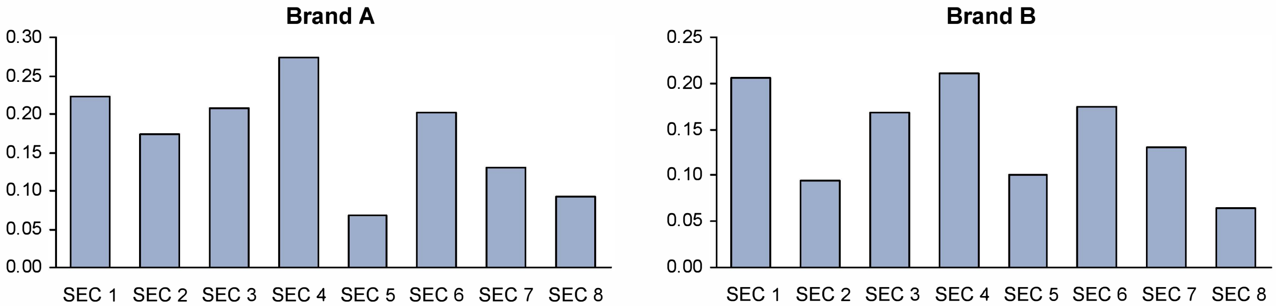

Fig. 3. The bar diagram shows the percentage of opportunities associated with the action across different business sectors, where moderate selection bias can be observed.

TABLE IV
EVALUATION OF WINNING RATE OVER OPPORTUNITY LEVEL FOR DIFFERENT PEER SUBSETS.

Revenue Size	Cross Brand	**Brand A**				**Brand B**			
		Action	No Action	Action	No Action	Action	No Action	Action	No Action
$\geq$ \$500K	YES	723	412	54.22%	30.58%	156	134	36.54%	19.40%
$\geq$ \$500K	NO	7,291	9,665	68.18%	60.00%	1,837	4,666	57.27%	47.56%
$<$ \$500K	YES	1,915	3,075	55.67%	53.37%	804	1,454	39.80%	29.16%
$<$ \$500K	NO	38,467	142,456	73.41%	66.30%	30,363	163,762	58.53%	51.99%

opportunities resulting in significant selection bias . The results from the feature selection indicated that the key factors that result in selection bias are: “revenue size”, and “cross brand”.[1] Table III shows the percentage of business opportunities associated with the action. It is clear that larger size and cross-brand opportunities generally tend to receive this action. Moreover, there were also some other key features that exhibit moderate selection bias, e.g. “sector” (shown in Figure 3). However, since this feature did not show a significant and consistent bias over the action decision, we excluded it from the grouping of matching sets.

Given the two identified key factors, the data was split into multiple peer sets for performance evaluation. To simplify the partitioning, we treated the revenue size as a binary value by setting a threshold of $500K$. Combining with the binary attribute of “cross brand”, we generated four peer subsets, as shown in Table IV. From this figure, we can see the subset with the revenue $\geq 500K$ and “cross brand=YES” has a fairly smaller number of opportunities compared to the other three. However, by taking the action, the winning rates can be significantly improved, e.g. from 30.58% to 54.22% for brand A, and from 19.4% to 36.54% for brand B. Taking such an action can also improve the winning rates for the opportunities falling in other matching sets, albeit much more moderately. As such, large, complex deals can really benefit from the use of third-party support for sellers and result in a significant increase in the win rate of such deals.

[1]If an opportunity is indicated as “cross band”, it often means the opportunity is associated with high business complexity.

The other evaluation we conducted is to justify whether taking such an action can help improve sellers’ performance and productivity. The performance was measured by the winning rate and the productivity was the number of simultaneously carried opportunities by a seller. There were three types of sellers in the opportunity dataset defined based on the functions. Some sellers used the help from third-party quite often, while some sellers tended not to use this service. Table V presents the evaluation on the seller level, where it is clear that the business action clearly improve seller’s winning rate and frees up their time to allow them to carry more opportunities. Although this impact is consistent for different seller types, the improvement for type II seller is even more significant.

Through the evaluation over both opportunity and seller levels, it is clear that the action of involving a third-party to support seller transactions can help improve business performance. Moreover, the impact of such an action is more pronounced in the case of large and complex opportunities.

V. CONCLUSION

In this paper, we studied the selection bias issue in an analytics-based business decision model. We argue that the conventional performance evaluation for such decision procedures can be easily misled due to common bias happening when one decides whether to take certain business actions. To identify such bias and the correlated business factors, we propose a conditional mutual information based feature selection method to identify such factors. Then, these factors can be used to partition data into matching subsets, where

TABLE V

EVALUATION OF WINNING RATE AND PRODUCTIVITY OVER SELLER LEVEL FOR DIFFERENT PEER SUBSETS. IN THE TWO INVESTIGATED BRANDS, THE SELLERS USING THE ACTION CLEARLY DEMONSTRATE HIGHER WIN RATES AND PRODUCTIVITY THAN THOSE NOT USING THE ACTION.

	Brand A						
Seller Type	Total # of Sellers	*Using Action*			*Not Using Action*		
		# of Sellers	Productivity	Win Rate	# of Sellers	Productivity	Win Rate
I	$20,035$	$6,887$	25.5	26.8%	$13,148$	6.4	17.8%
II	$3,849$	562	34.8	28.0%	$3,287$	8.0	14.5%
III	$3,379$	635	9.7	40.3%	$2,744$	3.0	21.8%
	Brand B						
Seller Type	Total # of Sellers	*Using Action*			*Not Using Action*		
		# of Sellers	Productivity	Win Rate	# of Sellers	Productivity	Win Rate
I	$8,699$	$2,785$	8.8	28.6%	$5,914$	4.2	18.3%
II	813	127	15.5	31.9%	686	4.0	12.7%
III	355	126	5.6	39.1%	229	2.8	20.0%

each sample shares the same valued key factors. Performance evaluation over such matching sets significantly alleviates the selection bias issue, thereby resulting in a valid comparison.

A case study of the performance evaluation in salesforce optimization was reported in this paper. In particular, we explored whether it was beneficial to use support staff to help sellers with non-productive, back-office type of activities so as to free up the sellers' time to enable them to spend more time on productive selling activities. The results clearly show the existence of a strong selection bias in terms of the opportunities that sellers choose to pass on to the support staff, thereby clouding the true effect of using such support staff. However, our analysis shows that through the proposed selection bias-based evaluation framework, we can not only show that delegating non-productive activities to support staff is beneficial to business performance, but can also quantify the size of this benefit. The framework can easily be applied to other business activities to not only derive valid business insights and justifications, but can also feed the evaluation results back to the decision model to further improve performance.

ACKNOWLEDGMENT

The authors thank Léa Deleris for providing initial material that motivated this study.

REFERENCES

[1] Alexander Group, "Business intelligence provider gets smarter on sales productivity," 2010.

[2] D. Ledingham, M. Kovac, and H. L. Simon, "The new science of sales force productivity," *Harvard Bus. Rev.*, vol. 84, no. 9, pp. 124–133, Sep. 2006.

[3] Bain's Global Customer Practice, "How leaders get the most out of their salesforce," Mar. 2011.

[4] M. Singh, D. Bhattacharjya, L. Deleris, D. Katz-Rogozhnikov, M. Squillante, B. Ray, A. Mojsilovic, C. Barrera, D. Kakrania, J. Richard, A. Saha, and J. Fu, "The growth and performance diagnostics initiative: A multi-dimensional framework for sales performance analysis and management," *Serv. Sci.*, vol. 3, no. 1, pp. 82–98, Spring 2011.

[5] W. G. Cochran and D. B. Rubin, "Controlling bias in observational studies: A review," *Sankhyā Series A*, vol. 35, no. 4, pp. 417–446, Dec. 1973.

[6] F. Vella, "Estimating models with sample selection bias: A survey," *J. Human Resour.*, vol. 33, no. 1, pp. 127–169, Winter 1998.

[7] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, pp. 1531–1555, Nov. 2004.

[8] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[9] M. Vidal-Naquet and S. Ullman, "Object recognition with informative features and linear classification," in *Proc. IEEE Int. Conf. Comp. Vis.*, Nice, France, Oct. 2003, pp. 281–288.