Data Science for Fundraising: A Review of Analytics in Fundraising

Ashutosh R. Nandeshwar[1]

[1] University of Southern California

Author Note

Correspondence concerning this article should be addressed to Ashutosh R. Nandeshwar, 1150 S Olive St, 25th Floor, Los Angeles, CA 90015. Website: https://nandeshwar.info. E-mail: a7@nandeshwar.info

Abstract

Data analysis, data mining, predictive analytics, machine learning, data science, and artificial intelligence have affected how for-profit organizations make decisions for many years now, many decades for some methods. The nonprofit industry, specifically, nonprofit fundraising has been trying to catch up. This paper lists various applications of analytics in nonprofit fundraising as found in the literature. I present the survey in two ways: a) chronological, by decades and b) by analytical methods. I use the term "analytics" to capture the various methods used in data mining, predictive analytics, machine learning, data science, and artificial intelligence. This paper is structured as follows: 1) a brief review of the analytics methods, 2) review of the literature by decades, 3) review of the literature by methods, 4) summary of the findings, and 5) predictions on future work.

Keywords: data science, predictive analytics, machine learning, fundraising, literature survey, nonprofit

Data Science for Fundraising: A Review of Analytics in Fundraising

Introduction

When computer science is making rapid advances, one may ask "what new knowledge can be gained by reviewing previous work?" Cataloging previous work offers many benefits: a) we can notice the gaps to build upon, b) we can sense the future direction of research, and c) we can learn what has worked and what hasn't. In this paper, I hope to offer an extensive survey of analytics applied to nonprofit fundraising. Using this survey, I note patterns and trends, and present research ideas for future work. The paper has the following structure. First, a brief history of analytics. Then different analytics methods. Followed by a review of the literature in applied analytics in fundraising. A summary of this review and future direction.

Review of Analytics

It is easy to get distracted by the current hype of Artificial Intelligence (AI), but when looked carefully, we can see the meaningful methods and techniques to make sense of the available data and information. Statistical analyses involve collecting, analyzing, drawing conclusions from the available data (Diez, Barr, & Cetinkaya-Rundel, 2012). The field of statistics isn't new. As Fienberg (1992) wrote in his review of statistics article, the classic probability theory was formed in the early 1700s, but the inference methods and statistical models were formed much later in the 1890s.

Advances in statistical research and computational power led to the first hype cycle of AI in the 1960s (Liao, Chu, & Hsiao, 2012). Later, data mining became popular as a means to uncover patterns of significance using modern algorithms. Researchers named it Knowledge Discovery in Databases (KDD): the complete process of finding useful insights (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Machine learning, a computer scientist's way of saying pattern detection, surged in the early 2000s and now AI is back to the future. From a practitioner's perspective, the differences among these terms and fields are now insignificant, but researchers in those fields care about these differences (Mannila, 1996). In the end, as Fayyad et al. (1996) commented, "The unifying goal [of these methods] is extracting high-level knowledge from low-level data in the context of large data sets."

Although the latest developments in natural language processing (NLP), natural language generation (NLG), computer vision, and deep-learning help us with other tasks than solely discovering knowledge (Young, Hazarika, Poria, & Cambria, 2018), we will find that the literature for nonprofit fundraising is focused on KDD. This makes sense because fundraising goes up when the right people are asked for the right amount. But in the future, we will see broader applications of data science, helping us automate tasks and increase productivity.

Methods and Techniques in Analytics

Since the field of analytics is expansive, let's review and categorize the common methods and techniques used in the field. I will use these categories while reviewing the research in nonprofit fundraising.

Descriptive Statistics. Descriptive statistics use standard formulas to calculate measures that reflect the data. Some of these measures include mean, median, standard deviation, frequency, proportions, and other exploratory analyses. These measures give us quick insights into the data. Often, these measures are supported by graphs, such as scatter plots, histograms, and box plots. Such graphs help us see the correlations and patterns in the data (NIST/SEMATECH, 2013).

Regression. Linear regression or least square methods estimate predictions by minimizing the sum of the differences between the actual data points and predicted value. As long as the parameter estimates can be multiplied to a variable (or its function) and these product terms can be added to form a function, we can use a linear regression – even if the function itself isn't a straight line (NIST/SEMATECH, 2013). But when the parameters take a non-linear form, we can't use linear regression and could use non-linear regression. When we estimate parameters to build a model for some data, this approach is called *parametric*. In contrast, in a *non-parametric* approach we estimate a function that follows the data closely (James, Witten, Hastie, & Tibshirani, 2013).

Regression methods can be used both for quantitative prediction (i.e. gift amount) as well as for predicting class probabilities (i.e. yes or no). Many approaches extend or build upon regression methods. In this paper, I have categorized them under regression. Some of these methods include logistic regression, Linear Discriminant Analysis (LDA), Generalized Additive Models (GAMs), Generalized Linear Models (GLMs), Ridge regression, Tobit regression, and Probit regression.

Classification. Classification methods predict the dependent variable into the different values of the dependent variable, such as "Yes" or "No." These values are called *classes*. Although regression methods work on classifications problems, machine learning "divide-and-conquer" and "covering" techniques such as decision trees and rules are better equipped to handle missing values and noisy data (Witten, Frank, Hall, & Pal, 2016).

Clustering. Clustering methods attempt to divide the data into n groups of similar data points. These methods are called unsupervised learning methods as they do not require a dependent variable. They work by finding center points for each of these groups and then mark all the data points close to these centers as part of these clusters (James et al., 2013, p. 385).

Ensemble Methods. There are two types of ensemble methods: a) comparison of many algorithms, and b) using predictions from many algorithms. The comparison of multiple algorithms helps analysts see which methods work well for their data sets. Comparison prevents the potential loss of prediction performance compared to the analyst's preferred method. Predictions from multiple algorithms can outperform a single algorithm by using *stacking* methods or *super learners* (Polley, Rose, & van der Laan, 2011; Polley & van der Laan, 2010).

Polley et al. (2011) argue that *super learners* work well with real-life datasets because no single algorithm can accurately model the data, but combining different algorithms provide us better estimates. As James et al. (2013) note, “there is no free lunch in statistics: no one method dominates all others over all possible data sets.”

Literature Review

Previous Work

Lindahl and Conley (2002) reviewed research and put it into two categories: “Motivational Studies” and “Predicting Alumni Giving.” The first category consists of work that studies why people choose to give. The second category includes research that identifies and test factors that could predict a person’s choice to give.

More recently, Bekkers and Wiepking (2010) reviewed more than 500 articles and categorized these works into eight topical areas. While these reviews summarized methods of philanthropy, this paper focuses on the uses of analytical methods.

Method

I followed the methods and frameworks used in two popular review articles: “Educational data mining: A survey and a data mining-based analysis of recent works” (Peña-Ayala, 2014) and “Data mining techniques and applications–A decade review from 2000 to 2011” (Liao et al., 2012). Both papers used comprehensive methods to collect and review the published works in data mining. Like their approaches, I started with these broad search terms in Google Scholar:

```
("data mining" OR analytics OR "machine learning" OR "data science" OR
 clustering OR statistics OR predictive) AND
 (nonprofit OR fundraising OR fund-raising OR non-profit OR charity OR
 donation OR philanthropy)
```

I filtered the results from these searches and used Google Scholar's citations feature to search for other papers that cited these works. Additionally, I used *Publish or Perish* software (Harzing, 2007) to run searches in Scoups and Microsoft Academic search databases as seen in Figure 1. In the next phase, I looked at other cited works within these results.

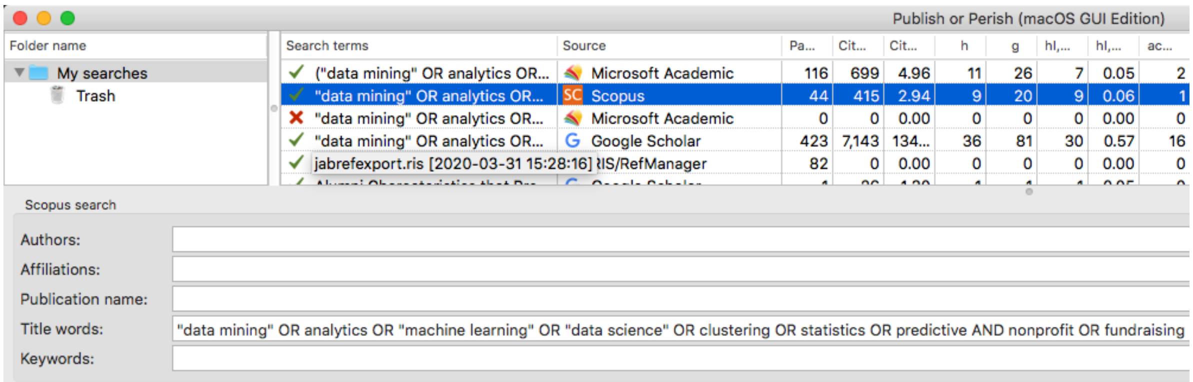

Figure 1. Publish or Perish Search Screen

After reading the results from this search, I decided whether to include the research as part of this review. The excluded work fell into these categories:

- Unpublished work
- Undergraduate thesis
- News articles
- Company white papers
- Research without analytics

I ended up with 145 works. Table 1 shows how the works were published, and you can see that Ph.D. dissertations account for the second most publications.

Limitations

This review and its findings are limited because of my omissions and subjective bias. I omitted any work that I could not find digitally. Although USC library's catalog is extensive and web searches can find many publications, I missed the digitally unavailable research

Table 1

Categories of Published Works

Category	Published Works
Article	78
PhD Thesis	51
In Proceedings	5
Book	4
In Collection	3
Masters Thesis	2
Tech Report	2

(fewer than five). My subjective bias towards what qualifies as a study for this review likely excluded some publications. Also, I may have made errors with the search keywords. Finally, operator error: it is likely that I unintentionally missed some research.

By Decades

The first publication in this field is probably O'Connor's dissertation on characteristics of alumni donors from 1961 (O'Connor, 1961). But you can see from Figure 2 that majority of the works were published between 2010 and 2019. Another noticeable trend, as seen in Figure 3, is the use of a wider set of techniques during the 2010-2019 period – though regression still leads the way.

Table 2 shows the raw numbers of the various analytics methods used over time. You can see regression methods and descriptive statistics total more than 100 studies, followed by 10 ensemble studies. This suggests that researchers feel confident in the results from regression methods. Or, researchers from other fields, especially computer science, have not studied fundraising problems.

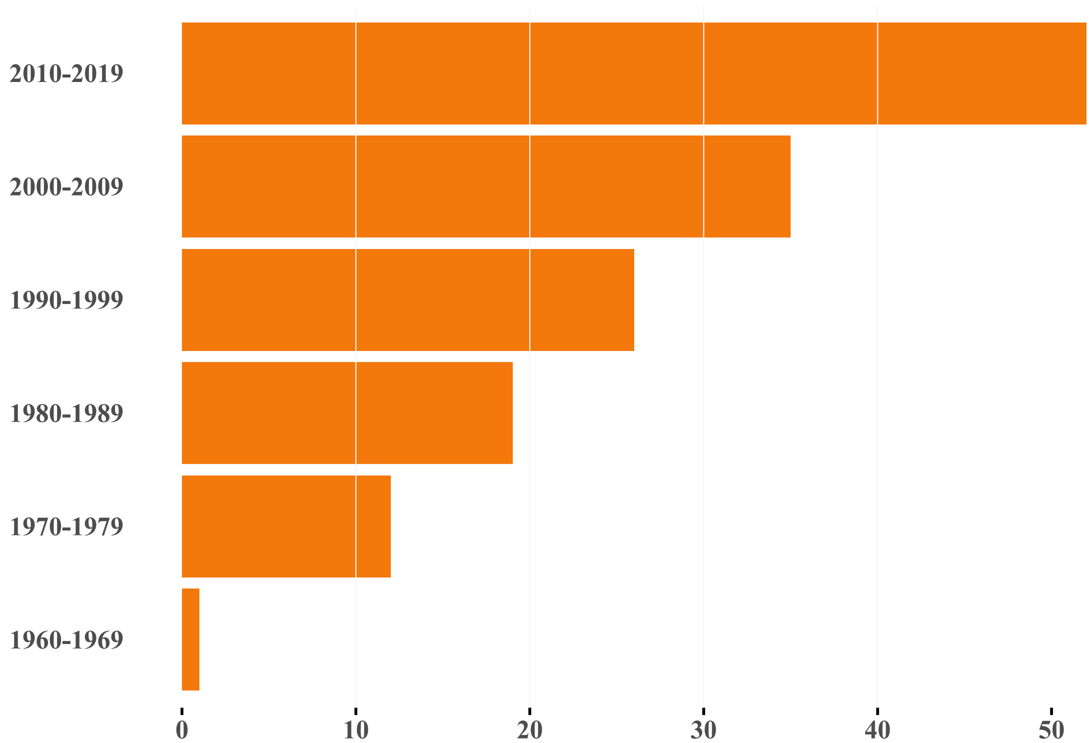

Figure 2. Total Number of Published Works by Decades

By Method

CHAID. CHAID is a decision tree learner, which Liihe (1998) used to study database marketing at UNICEF. Denizard-Ramsamy and Medina-Borja (2008) predicted financial vulnerability in non-profit organizations using CHAID; this is a rare paper as most of the studies in this review focus on donor identification.

Clustering. Segmentation via clustering has a good use case in fundraising for customized marketing as well as prospect identification. Various researchers have applied segmentation at university settings (Blanc & Rucks, 2009; Cermak, File, & Prince, 1994; Durango-Cohen & Balasubramanian, 2015; E. J. Durango-Cohen et al., 2013a; P. L. Durango-Cohen et al., 2013b; Luperchio, 2009; Zhang, 2014).

Descriptive Statistics. Descriptive statistics include mean, percentage distribution, correlations, Chi-squared tests, and Analysis of Variance (ANOVA). Most of the research in this category studied the effects of alumni characteristics to predict giving (Anderson, 1981;

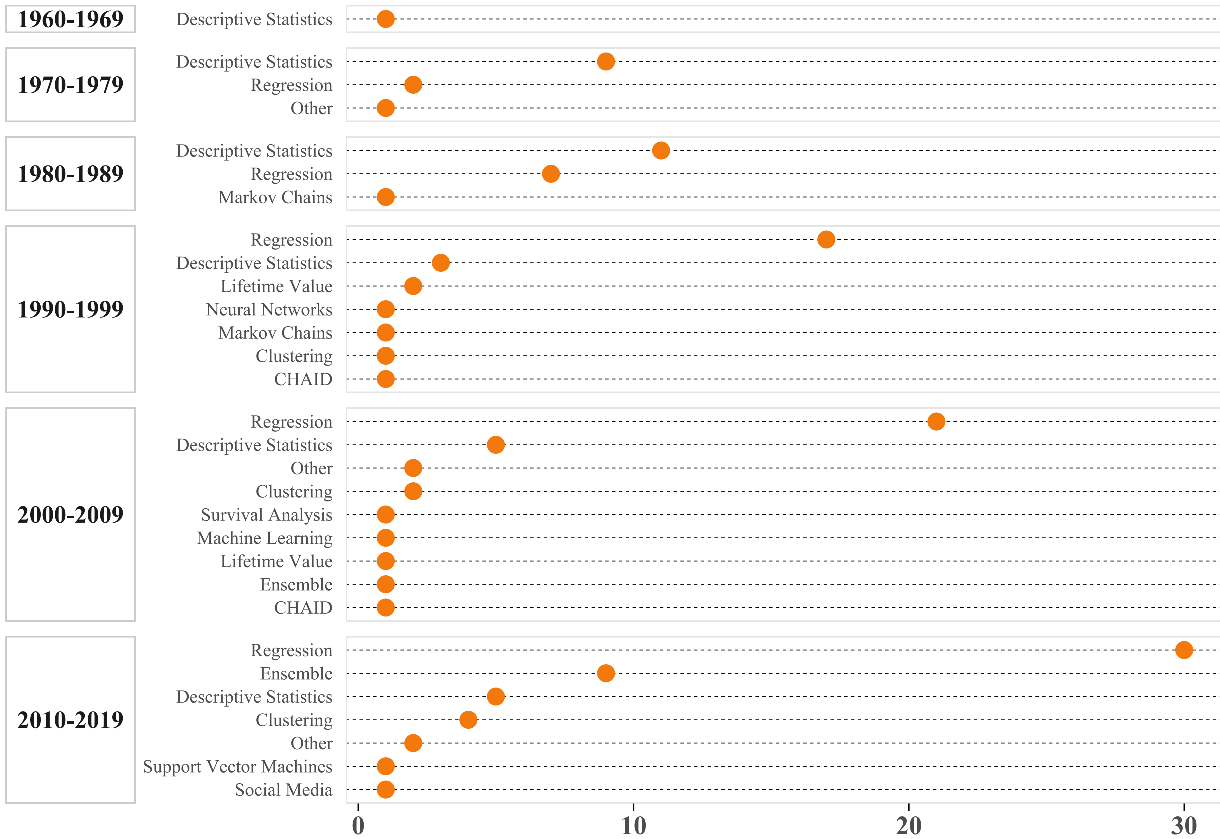

Figure 3. Methods by Decades.

Note: Although regression techniques are oft-used methods, ensemble methods are finding greater use.

Bingham Jr, Quigley Jr, & Murray, 2003; Blumenfeld & Sartain, 1974; Bruyn & Prokopec, 2013; Caruthers, 1973; Chewning, 1984; Dietz, 1985; Gardner, 1975; Gunsalus, 2005; Haddad, 1986; Hunter, Jones, & Boger, 1999; Johnson, 2013; Keller, 1982; Korvas, 1984; Loveday, 2012; Markoff, 1978; McKee, 1975; McKinney, 1978; McNally, 1985; Miller, 2013; Morris, 1970; Nelson, 1984; Newman, 2011; O'Connor, 1961; Oglesby, 1991; Riecken & Yavas, 1979; Schlegelmilch & Tynan, 1989; Smith & Beik, 1982; Sundel, Zelman, Weaver, & Pasternak, 1978; Wylie, 2004).

A few notable exceptions were:

- Frederick (1984) studied football success with institutional giving.

Table 2

Trends of Methods Used by Decade

Analytics Method	1960-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2019	**Total**
CHAID	0	0	0	1	1	0	**2**
Clustering	0	0	0	1	2	4	**7**
Descriptive Statistics	1	9	11	3	5	5	**34**
Ensemble	0	0	0	0	1	9	**10**
Lifetime Value	0	0	0	2	1	0	**3**
Machine Learning	0	0	0	0	1	0	**1**
Markov Chains	0	0	1	1	0	0	**2**
Neural Networks	0	0	0	1	0	0	**1**
Other	0	1	0	0	2	2	**5**
Regression	0	2	7	17	21	30	**77**
Social Media	0	0	0	0	0	1	**1**
Support Vector Machines	0	0	0	0	0	1	**1**
Survival Analysis	0	0	0	0	1	0	**1**
Total	**1**	**12**	**19**	**26**	**35**	**52**	**145**

- Berger and Smith (1997) analyzed the effects of framing the direct mail appeals.
- Quigley, Bingham, and Murray (2002) measured the effects of gift acknowledgments on giving.
- Magson and Routley (2009) looked at planned giving fundraising.

Ensemble. Ensemble methods often include machine learning techniques, which are either combined to improve performance or used for comparison. Potharst, Kaymak, and Pijls (2002) used neural networks and CHAID to improve direct marketing outcomes. Chen (2010) used regression, neural network, and SVMs on the Direct Marketing Education Foundation (DMEF) data. Ye (2017) used Naive Bayes, Random Forest, and SVM to predict major donors and compared the results from these methods. Other works in this category included: E. J. Durango-Cohen (2013), Moon and Azizi (2013), Udenze (2014), Torres (2014), Chung and Lee (2015), Kakrala and Chakraborty (2015), and Rattanamethawong, Sinthupinyo, and Chandrachai (2018).

Lifetime Value. Commonly used in the for-profit/marketing world, lifetime value calculates the future total profit from a customer. This value is used for segmentation and acquisition strategies. Some researchers have built models to calculate this value for donors (Aldrich, 2000; Hunter & Hill, 1998; Sargeant, 1998).

Machine Learning. Many of the studies in the ensemble category fall in the machine learning category also. There was one study that didn't fit in the ensemble category: Weerts and Ronca (2009) used classification trees to predict alumni giving.

Markov Chains. Markov Chains use probabilities of prior events to predict the probability of next events, and such a chain continues. A donor's lifetime giving can also be structured as a chain of events to predict future giving. Soukup (1983) and Toohill, Mullins, Barclay, and Sadnicki (1997) used Markov chains to predict giving.

Neural Networks. Like the machine learning models that fall under ensemble methods, a few neural network applications were also part of that category. But a standalone implementation of neural networks can be found in Goodman and Plouff (1997).

Other. I placed other publications in this category if I couldn't classify them. These tend to be either overarching frameworks (Birkholz, 2008; Nandeshwar & Devine, 2018), descriptive works (Herzlinger, 1977), or rarely applied techniques for fundraising (Hashemi, Le Blanc, Bahrami, Bahar, & Traywick, 2009).

Regression. Researchers in higher education have applied different flavors of regression techniques, and as mentioned in the earlier section, I am using the term regression liberally. Most of these studies are Ph.D. dissertations from education schools and colleges (Baade & Sundberg, 1996; Baruch & Sang, 2012; Beeler, 1982; Belfield & Beney, 2000; Bennett, 2003, 2006; Bohannon, 2007; Boyle, 1990; Bruggink & Siddiqui, 1995; Brunette, Vo, & Watanabe, 2017; Burgess-Getts, 1992; Christian, 2018; Cunningham & Cochi-Ficano, 2002; Day, 2018; Dickert, Sagara, & Slovic, 2010; Diehl, 2007; Duncan, 1999; Duronio & Loessin, 1990; Faisal, 2017; Gaier, 2005; Greenlee & Trussel, 2000; Grill, 1988; Hanson, 2000;

Holmes, 2009; House, 1987; Hoyt, 2004; Hueston, 1992; Ketter, 2013; Key, 2001; Lara & Johnson, 2013; Lawley, 2008; Lawrence, Kudyba, & Lawrence, 2017; Lertputtarak & Supitchayangkool, 2014; Leslie & Ramey, 1988; Lindahl & Winship, 1992, 1994; Liu, Feng, & Ouyang, 2018; Lowe, 2019; Manzer, 1974; Marr, Mullin, & Siegfried, 2005; Martin, 1993; McDearmon & Shirley, 2009; Meer & Rosen, 2008, 2012; Miracle, 1977; Monks, 2003; Morgan, 2014; Mosser, 1993; Naccarato, 2019; Okunade & Berl, 1997; Okunade, Wunnava, & Walsh Jr, 1994; Oliveira, Croson, & Eckel, 2011; Park, Ko, Kim, Sagas, & Eddosary, 2016; Pearson, 1996; Pinion, 2016; Rau, 2014; Rau & Erwin, 2015; Ropp, 2014; Rosenblatt, Cusson, & McGown, 1986; Saraih et al., 2018; Schlegelmilch, Love, & Diamantopoulos, 1997; Selig, 1999; Shadoian, 1989; Shen & Tsai, 2009; Skari, 2014; Steinnes, 2011; Sun, Hoffman, & Grady, 2007; Taylor & Martin, 1995; Terry & Macy, 2007; Thompson, 2010; Tiger & Preston, 2013; Truitt, 2013; Tsao & Coll, 2005; Veludo-de-Oliveira, Alhaidari, Yani-de-Soriano, & Yousafzai, 2016; Verhaert, 2010; Walcott, 2015; Yavas, Riecken, & Parameswaran, 1981).

Social Media. Vequist IV (2017) studied the use of various social media and giving to various nonprofit organizations. Campaign performance data and other meta-data were used to improve the decision making of the stakeholders and increase social media user donations.

Support Vector Machines. One study using SVM is notable because it dealt with the imbalanced (or unbalanced) classes that we typically observe in the donation data i.e. either the proportion of donor records in the data is low or few major donors exist in the data. Kim, Chae, and Olson (2012) used SVMs to build a response model on imbalanced datasets.

Survival Analysis. Although survival analysis is used in analyzing data for a failure event, such as death, Drye, Wetherill, and Pinnock (2001) used it to predict a donor's status in her giving lifecycle.

Quality Assessment

While reviewing the breadth of the methods used for nonprofit fundraising is useful, more important is assessing the rigor, credibility, and relevancy of the predictions in these published works. Wen, Li, Lin, Hu, and Huang (2012) used a 10-question framework to assess the quality of each work. I used a similar method. I answered questions given in Table 3 for each published work; the possible answers were *Yes*, *No*, or *Somewhat* with weights of 1, 0, and 0.5 respectively. All questions, except for *Q4* and *Q6*, are from Wen et al. (2012). Of course, these questions are suitable only for those works in which the researchers made predictions or built predictive models. It is also unfair to assess older research when obtaining enough computing power was a challenge. Also, my subjective bias can skew the findings.

Table 3

Prediction quality assessment questions

ID	Question
Q1	Are the estimation methods well defined and deliberate?
Q2	Is the experiment applied on sufficient data sets?
Q3	Is the estimation accuracy measured and reported?
Q4	Are the estimates significantly better than the baseline?
Q5	Is the proposed estimation method compared with other methods?
Q6	Can the findings be applied widely?
Q7	Are the findings of study clearly stated and supported by reporting results?

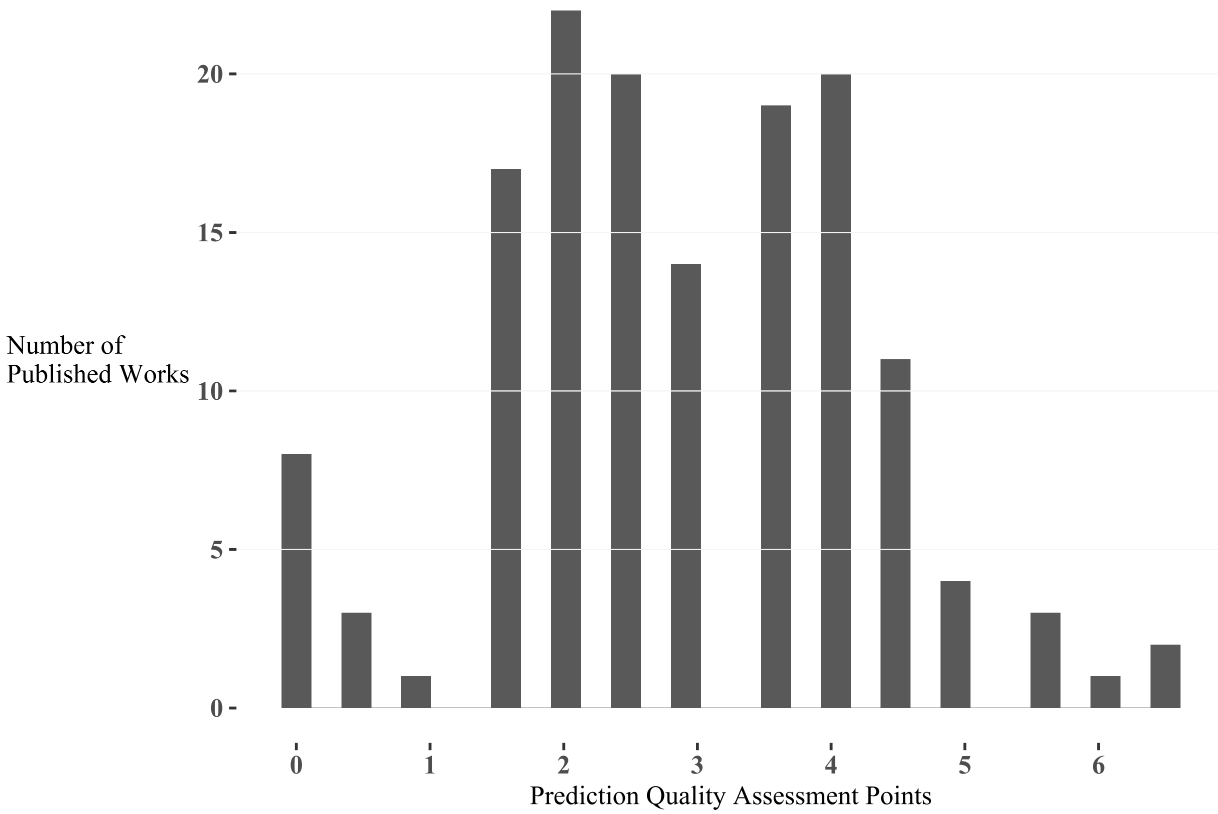

Table 4

Top Published Works by Prediction Quality Assessment

Author	Analytics Method
Shadoian (1989)	Regression
Liihe (1998)	CHAID
Greenlee and Trussel (2000)	Regression
Potharst et al. (2002)	Ensemble
Chen (2010)	Ensemble
Kim et al. (2012)	Support Vector Machines
Moon and Azizi (2013)	Ensemble
Chung and Lee (2015)	Ensemble
Ye (2017)	Ensemble
Liu et al. (2018)	Regression

Summary of Literature

Most of the studies in this review focused on either predicting the likelihood of a person donating or predicting the giving level or amount. An exception was the Greenlee and Trussel (2000) study of the financial stability of institutions. As Brittingham and Pezzullo (1990, p. 39) wrote about the predictive studies in fundraising, "Most of the studies are dissertations, and most are based on a single institution, most often a university. The results … do not support strong conclusions." What was true in the 1990s remains true today. As we saw in the earlier sections, dissertations lare the second-most studies in applied analytics for fundraising.

Many dissertations followed a similar pattern: select variables based on literature, study each variable for correlations and significance, include selected variables for an estimation model, reject or accept the null hypothesis, and then present final results.

There are some challenges with this approach.

1. These studies are often limited to one institution; hence the results cannot be generalized.
2. This type of research primarily becomes about the application of a statistical technique to the researcher's dataset and doesn't contribute to knowledge advancement, either through the application of newer and different predictive methods or towards a unified theory of giving.
3. This type of framework can be templatized using a programming language.

While building local predictive models are useful for development offices, we need either groundbreaking research to significantly improve on the donor classification problem, or we need to find different fundraising problems to solve.

Many of these studies used the null hypothesis significance testing (NHST) to infer the answers to research questions. This is problematic for two reasons:

1. As Trafimow (2014) declared in his editorial of the Basic and Applied Social Psychology journal, "The null hypothesis significance testing procedure has been shown to be logically invalid and to provide little information about the actual likelihood of either the null or experimental hypothesis." Then next year, while banning the null hypothesis significance testing procedure from the journal, Trafimow and Marks (2015) said, "$p < .05$ bar is too easy to pass and sometimes serves as an excuse for lower quality research."

2. As Gliner, Leech, and Morgan (2002) noted, "A common misuse of NHST is the implication that statistical significance means theoretical or practical significance." In these surveyed studies, you can find examples of researchers interpreting statistically significant results mistaken for important findings.

While most researchers report on the overall accuracy of their prediction models, very few report on other evaluation measures, such as precision, recall, or specificity. Another challenge is the lack of comparison to baseline proportions. Since such measures or comparisons aren't reported, it is hard to assess whether the new predictive models performed better than guessing.

For example, say our data had 5% donors and 95% non-donors. We built a predictive model that classified donors and non-donors. Let's say that this model had an overall accuracy of 95%. Now, if were to evaluate the model only based on accuracy, we might be satisfied with its performance. But even if we guess every row as a non-donor, we achieve 95% accuracy.

Similarly, if the data has 45% donors and 55% non-donors, and the model had an overall accuracy of 50%, it did worse than the baseline. Even if the predictive models aren't compared to other models, they should at least be compared with the baseline. As my colleagues and I reported in another paper, if the overall accuracy rate is close to the baseline, then the complex analysis can be replicated by a simple majority vote model (Nandeshwar, Menzies, & Nelson, 2011).

One benefit of the research done over decades into the likelihood of a person's donation is that we have a comprehensive list of attributes, attitudes, and values that could go into building new predictive models.

Opportunities and Future Direction

Today's technological advancement offers fascinating paths to study various problems in fundraising. Here are some suggestions and ideas to build on our knowledge of applications of data science in nonprofit fundraising.

- **Establish the baseline**. In classification or numeric prediction models, use a majority vote or the mean value to compare the results against. Witten et al. (2016) call this model is called *ZeroR*. Also, consider using a simple, single-rule classification model known as *1R* or *OneR*. Holte (1993) calculated the results from this simple model on many datasets and compared them with an advanced decision tree model and found that *1R* was only "a few percentage points less accurate."

- **Use and report a wider set of evaluation metrics**. As we saw earlier, reporting accuracy can be misleading. We can consider different evaluation measures shown in the equations below (Branco, Torgo, & Ribeiro, 2016). For example, Rau (2014, p. 30) reported that "76.4% of cases are correctly classified," but you can see in Table 6 that 73% of their study data contained non-donors, so simply predicting everyone a non-donor, our accuracy is 73%. The study predicted only 88 donors who were actual donors, making the recall or true positive rate of 20%. Thus, the model failed at correctly identifying donors. Similarly, the F-measure and balanced accuracy were low at 0.32 and 59%.

Table 5

Confusion Matrix for a Two-class Problem

		Predicted	
		Donor	Non-donor
Actual	**Donor**	True Positive (TP)	False Negative (FN)
	Non-donor	False Positive (FP)	True Negative (TN)

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

$$\text{F} - \text{measure} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

$$\text{BalancedAccuracy} = \frac{Recall + Specificity}{2} \tag{5}$$

- **Consider selecting variables using feature subset selection (FSS)**. In his extensive study of feature subset selectors, Hall (1999) documented compared his feature (or variable) selector with other predictive techniques. He found that FSS removed redundant and irrelevant features, and in some cases, even improved the performance of the underlying predictive algorithms.

- **Consider class balancing methods**. When the number of rows for one class (such as non-donor) is higher than the rows for any other class (such as donor), class imbalance occurs. To overcome this problem, Kim et al. (2012) used undersampling to reduce the number of majority class rows. Some other approaches to achieve class balance: oversampling the minority class rows, synthetic generation of minority class

Table 6

Confusion Matrix from Rau (2014)

		Predicted	
		Donor	Non-donor
Actual	**Donor**	88	342
	Non-donor	32	1126

rows, such as SMOTE and family (Chawla, Bowyer, Hall, & Kegelmeyer, 2002; Han, Wang, & Mao, 2005), and cost-sensitive learning (Domingos, 1999).

- **Consider ensemble methods**. Either combine various models (that is bagging, boosting, or stacking methods (see Witten et al., 2016, Section 8.1)) or compare various models and pre-processors. This type of comparison should be standard. Here's pseudocode to explain this comparison:

```
For each dataset:
    Create P pre-processed datasets
    For each p in P:
        Divide p into ten cross-folds
        For each predictive learning technique t:
            Train t on 9-folds
            Test the model on the remaining folds
            Store results and the resulting model
```

- **Build a large database with data from diverse organizations**. If researchers can collect data from many organizations, they can conduct a large-scale study to build predictive models. For example, Thompson (2010) used data from eight institutions. Such a large-scale study will show either that accurate donor classification is hard, or that a unified, single model can be built and we can research other topics. A related idea is what JOHNSON (1991) attempted: get anonymized data from the Internal Revenue Service (IRS) and build models on it.

- Research other topics and approaches:
  - **Consider modeling methods that work well with long-tail or skewed data**, such as quantile regression (Perlich, Rosset, Lawrence, & Zadrozny, 2007) or HyperSMURF, an ensemble method (Schubach, Re, Robinson, & Valentini, 2017).
  - **Study creation of personalized appeals and communication**. The latest Natural Language Processing and Generation (NLP and NLG) methods are far superior to previous methods (Yang et al., 2019), and they can be used to generate personalized appeals and communication. Ding and Pan (2016), for example, generated gain or risk framed text to increase the text's appeal to the reader.
  - **Study applications of graph theory to learn interests**. Social graphs have value if all the connections in the graph can be known. A better use case for fundraising could be interest graphs, which identify the interests of people and connect people based on these interests (Yu, Chen, Li, & Ma, 2014).
  - **Use NLG and NLP to automate tasks**. Like creating personalized appeals, we can use pre-trained language models to summarize text, among other things, as shown by Liu and Lapata (2019). For example, using a simple Python text summarizer called *sumy*[1], I summarized an article on Bill Gates from biography.com[2].

[1] https://github.com/miso-belica/sumy

[2] https://www.biography.com/business-figure/bill-gates

```
"In 1975, Gates and Allen formed Micro-Soft, a blend of "micro-computer"
and "software" (they dropped the hyphen within a year). Bill Gates Fact
Card Microsoft’s Software for IBM PCs As the computer industry grew,
with companies like Apple, Intel and IBM developing hardware and
components, Gates was continuously on the road touting the merits
of Microsoft software applications. Since stepping down from Microsoft,
Gates devotes much of his time and energy to the work of the Bill &
Melinda Gates Foundation."
```

Conclusion

In this paper, I reviewed the literature of analytics for nonprofit fundraising. Although researchers have applied more sophisticated methods over time, regression methods remain the most-used technique for predicting a donor’s likelihood of giving and her giving amount. Also, dissertations account for second-most published works. Machine learning and ensemble techniques are increasingly in use, and we will see more research using these methods in the future. Researchers will also use natural language processing and generation, along with deep learning.

References

Aldrich, T. M. (2000). How much are new donors worth? Making donor recruitment investment decisions based on lifetime value analysis. *International Journal of Nonprofit and Voluntary Sector Marketing*, *5*(1), 81–89. https://doi.org/10.1002/nvsm.99

Anderson, G. L. (1981). *Self-esteem and altruism perceived as motivational factors for alumni giving, and their relationships to various donor characteristics* (PhD thesis). Western Michigan University.

Baade, R. A., & Sundberg, J. O. (1996). What determines alumni generosity? *Economics of Education Review*, *15*(1), 75–81. https://doi.org/10.1016/0272-7757(95)00026-7

Baruch, Y., & Sang, K. J. (2012). Predicting mba graduates' donation behaviour to their alma mater. *Journal of Management Development*, *31*(8), 808–825. https://doi.org/10.1108/02621711211253268

Beeler, K. J. (1982). *A study of predictors of alumni philanthropy in private universities* (PhD thesis). University of Connecticut.

Bekkers, R., & Wiepking, P. (2010). A literature review of empirical studies of philanthropy. *Nonprofit and Voluntary Sector Quarterly*, *40*(5), 924–973. https://doi.org/10.1177/0899764010380927

Belfield, C. R., & Beney, A. P. (2000). What determines alumni generosity? Evidence for the UK. *Education Economics*, *8*(1), 65–80. https://doi.org/10.1080/096452900110300

Bennett, R. (2003). Factors underlying the inclination to donate to particular types of charity. *International Journal of Nonprofit and Voluntary Sector Marketing*, *8*(1), 12–29. https://doi.org/10.1002/nvsm.198

Bennett, R. (2006). Predicting the lifetime durations of donors to charities. *Journal of Nonprofit & Public Sector Marketing*, *15*(1-2), 45–67. https://doi.org/10.1300/j054v15n01_03

Berger, P. D., & Smith, G. E. (1997). The effect of direct mail framing strategies and segmentation variables on university fundraising performance. *Journal of Direct Marketing*, *11*(1), 30–43.

Bingham Jr, F. G., Quigley Jr, C. J., & Murray, K. B. (2003). An investigation of the influence acknowledgement programs have on alumni giving behavior: Implications for marketing strategy. *Journal of Marketing for Higher Education*, *12*(2), 1–14.

Birkholz, J. (2008). *Fundraising analytics: Using data to guide strategy*. John Wiley & Sons.

Blanc, L. A. L., & Rucks, C. T. (2009). Data mining of university philanthropic giving: Cluster-discriminant analysis and pareto effects. *International Journal of Educational Advancement*, *9*(2), 64–82. https://doi.org/10.1057/ijea.2009.28

Blumenfeld, W. S., & Sartain, P. L. (1974). Predicting alumni financial donation. *Journal of Applied Psychology*, *59*(4), 522–523. https://doi.org/10.1037/h0037298

Bohannon, T. (2007). Predictive modelling in higher education. In *SAS global forum*.

Boyle, J. J. (1990). *College quality and alumni giving* (PhD thesis). Syracuse University.

Branco, P., Torgo, L., & Ribeiro, R. P. (2016). A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys*, *49*(2), 1–50. https://doi.org/10.1145/2907070

Brittingham, B. E., & Pezzullo, T. R. (1990). *The campus green: Fund raising in higher education. ASHE-eric higher education report 1*. ERIC.

Bruggink, T. H., & Siddiqui, K. (1995). An econometric model of alumni giving: A case study for a liberal arts college. *The American Economist*, *39*(2), 53–60.

Brunette, C., Vo, N., & Watanabe, N. M. (2017). Donation intention in current students: An analysis of university engagement and sense of place in future athletic, academic, and split donors. *Journal of Issues in Intercollegiate Athletics*, *10*, 78–100.

Bruyn, A. D., & Prokopec, S. (2013). Opening a donor's wallet: The influence of appeal scales on likelihood and magnitude of donation. *Journal of Consumer Psychology*, *23*(4), 496–502. https://doi.org/10.1016/j.jcps.2013.03.004

Burgess-Getts, L. (1992). *Alumni as givers: An analysis of donor-nondonor behavior at a comprehensive i institution* (PhD thesis). The College of William; Mary.

Caruthers, F. A. S. (1973). *Study of certain characteristics of alumni who provide financial support and alumni who provide no financial support for their alma mater* (PhD thesis). Oklahoma State University.

Cermak, D. S. P., File, K. M., & Prince, R. A. (1994). A benefit segmentation of the major donor market. *Journal of Business Research*, *29*(2), 121–130. https://doi.org/10.1016/0148-2963(94)90016-7

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, Q. (2010). *Predictive modeling for non-profit fundraising* (Master's thesis). James Madison University.

Chewning, P. B. (1984). *The attitudes of alumni non-donors, donors, and consecutive donors toward drake university* (PhD thesis). Drake University.

Christian, K. M. (2018). *Identifying demographic variables that can predict alumni giving at a regional comprehensive four-year university in the south* (PhD thesis). Sam Houston State University.

Chung, J., & Lee, K. (2015). A long-term study of a crowdfunding platform. In *Proceedings of the 26th ACM conference on hypertext & social media - HT 15*. ACM Press. https://doi.org/10.1145/2700171.2791045

Cunningham, B. M., & Cochi-Ficano, C. K. (2002). The determinants of donative revenue flows from alumni of higher education: An empirical inquiry. *Journal of Human Resources*, 540–569.

Day, D. A. (2018). *Factors in the undergraduate experience that influence young alumni giving* (PhD thesis). Virginia Tech.

Denizard-Ramsamy, W., & Medina-Borja, A. (2008). Using chaid as a method to predict financial vulnerablity in non-profit organizations. In *IIE annual conference. Proceedings* (p. 487). Institute of Industrial; Systems Engineers (IISE).

Dickert, S., Sagara, N., & Slovic, P. (2010). Affective motivations to help others: A two-stage model of donation decisions. *Journal of Behavioral Decision Making*, *24*(4), 361–376. https://doi.org/10.1002/bdm.697

Diehl, A. G. (2007). *The relationship between alumni giving and receipt of institutional scholarships among undergraduate students at a public, land-grant institution* (PhD thesis). The Pennsylvania State University.

Dietz, L. H. (1985). *Iowa state university alumni contributions: An analysis of alumni giving patterns by selected class years, 1974 and 1979* (PhD thesis). Iowa State University; Digital Repository@ Iowa State University, http://lib. dr. iastate. edu/.

Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2012). *OpenIntro statistics*. OpenIntro.

Ding, T., & Pan, S. (2016). Personalized emphasis framing for persuasive message generation. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 1432–1441).

Domingos, P. (1999). MetaCost. In *Proceedings of the fifth ACM SIGKDD international conference on knowledge discovery and data mining - KDD 99*. ACM Press. https://doi.org/10.1145/312129.312220

Drye, T., Wetherill, G., & Pinnock, A. (2001). Donor survival analysis: An alternative perspective on lifecycle modelling. *International Journal of Nonprofit and Voluntary Sector Marketing*, *6*(4), 325–334. https://doi.org/10.1002/nvsm.158

Duncan, B. (1999). Modeling charitable contributions of time and money. *Journal of Public Economics*, *72*(2), 213–242. https://doi.org/10.1016/s0047-2727(98)00097-8

Durango-Cohen, E. J. (2013). Modeling contribution behavior in fundraising: Segmentation analysis for a public broadcasting station. *European Journal of Operational Research*, *227*(3), 538–551. https://doi.org/10.1016/j.ejor.2013.01.008

Durango-Cohen, E. J., & Balasubramanian, S. K. (2015). Effective segmentation of university alumni: Mining contribution data with finite-mixture models. *Research in Higher Education*, *56*(1), 78–104.

Durango-Cohen, E. J., Torres, R. L., & Durango-Cohen, P. L. (2013a). Donor segmentation: When summary statistics don't tell the whole story. *Journal of Interactive Marketing*, *27*(3), 172–184.

Durango-Cohen, P. L., Durango-Cohen, E. J., & Torres, R. L. (2013b). A bernoulliGaussian mixture model of donation likelihood and monetary value: An application to alumni segmentation in a university setting. *Computers & Industrial Engineering*, *66*(4), 1085–1095. https://doi.org/10.1016/j.cie.2013.08.007

Duronio, M. A., & Loessin, B. A. (1990). Fund-raising outcomes and institutional characteristics in ten types of higher education institutions. *The Review of Higher Education*, *13*(4), 539–556. https://doi.org/10.1353/rhe.1990.0013

Faisal, A. (2017). *An investigation of the relationship of student engagement to alumni giving at an independent technological university* (PhD thesis). Florida Institute of Technology.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37–37.

Fienberg, S. E. (1992). A brief history of statistics in three and one-half chapters: A review essay. *Statistical Science*, *7*(2), 208–225.

Frederick, R. E. (1984). *Intercollegiate football success and institutional private support: A national study of 81 public universities, 1965-1979* (PhD thesis). University of Kansas.

Gaier, S. (2005). Alumni satisfaction with their undergraduate academic experience and the impact on alumni giving and participation. *International Journal of Educational Advancement*, *5*(4), 279–288. https://doi.org/10.1057/palgrave.ijea.2140220

Gardner, P. M. (1975). *A study of the attitudes of harding college alumni with an emphasis on donor and non-donor characteristics* (PhD thesis). *ProQuest Dissertations and Theses*. Ohio University.

Gliner, J. A., Leech, N. L., & Morgan, G. A. (2002). Problems with null hypothesis significance testing (NHST): What do the textbooks say? *The Journal of*

Experimental Education, *71*(1), 83–92. https://doi.org/10.1080/00220970209602058

Goodman, S., & Plouff, G. (1997). Neural network modeling: Artificial intelligence marketing hits the non-profit world. *Fund Raising Management*, *28*, 16–19.

Greenlee, J. S., & Trussel, J. M. (2000). Predicting the financial vulnerability of charitable organizations. *Nonprofit Management and Leadership*, *11*(2), 199–210. https://doi.org/10.1002/nml.11205

Grill, A. J. (1988). *An analysis of the relationships of selected variables to financial support provided by alumni of a public university* (PhD thesis). The Pennsylvania State University.

Gunsalus, R. (2005). The relationship of institutional characteristics and giving participation rates of alumni. *International Journal of Educational Advancement*, *5*(2), 162–170. https://doi.org/10.1057/palgrave.ijea.2140214

Haddad, J., Freddie D. (1986). *An analysis of the characteristics of alumni donors and non-donors at butler university* (PhD thesis). West Virginia University.

Hall, M. A. (1999). *Correlation-based feature selection for machine learning* (PhD thesis). The University of Waikato.

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In *Lecture notes in computer science* (pp. 878–887). Springer Berlin Heidelberg. https://doi.org/10.1007/11538059_91

Hanson, S. K. (2000). *Alumni characteristics that predict promoting and donating to alma mater: Implications for alumni relations* (PhD thesis). University of North Dakota.

Harzing, A.-W. (2007). Publish or perish. Retrieved from https://harzing.com/resources/publish-or-perish

Hashemi, R. R., Le Blanc, A., Louis, Bahrami, A. A., Bahar, M., & Traywick, B. (2009). Association analysis of alumni giving: A formal concept analysis. *International Journal of Intelligent Information Technologies*, *5*(2), 17–32.

Herzlinger, R. (1977). Why data systems in nonprofit organizations fail. *Harvard Business Review*, *55*(1), 81–86.

Holmes, J. (2009). Prestige, charitable deductions and other determinants of alumni giving: Evidence from a highly selective liberal arts college. *Economics of Education Review*, *28*(1), 18–28. https://doi.org/10.1016/j.econedurev.2007.10.008

Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, *11*(1), 63–90. https://doi.org/10.1023/a:1022631118932

House, M. L. (1987). *Annual fund raising in public higher education: The development and validation of a prediction equation* (PhD thesis). *ProQuest Dissertations and Theses*.

Hoyt, J. E. (2004). Understanding alumni giving: Theory and predictors of donor status. In *Annual forum of the association for institutional research*.

Hueston, F. R. (1992). Predicting alumni giving: A donor analysis test. *Fund Raising Management*, *23*(5), 18.

Hunter, C. S., Jones, E. B., & Boger, C. (1999). A study of the relationship between alumni giving and selected characteristics of alumni donors of livingstone college, NC. *Journal of Black Studies*, *29*(4), 523–539. https://doi.org/10.1177/002193479902900404

Hunter, T., & Hill, R. (1998). Prediction of donor lifetime value and the development of true segmented donor strategy. *International Journal of Nonprofit and Voluntary Sector Marketing*, *3*(4), 312–320. https://doi.org/10.1002/nvsm.6090030405

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical* *learning*. Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

JOHNSON, B. (1991). Examining the factors that affect charitable giving. *Compendiumof Studies of of Tax-Exempt Organizations 1986-1992*, *2*, 204–210.

Johnson, E. A. M. (2013). *Factors associated with non-traditional and traditional undergraduate alumni giving to alma maters* (PhD thesis). Mercer University.

Kakrala, R., & Chakraborty, G. (2015). Donor sentiment and characteristic analysis using sas® enterprise miner™ and sas® sentiment analysis studio. In *SAS global forum*. https://doi.org/10.13140/RG.2.1.2716.7842

Keller, M. J. C. (1982). *An analysis of alumni donor and non-donor characteristics at the university of montevallo (alabama)* (PhD thesis). The University of Alabama.

Ketter, J. W. (2013). *Predictors of alumni donor behavior in graduates of the traditional mba and iMBA programs at the pennsylvania state university* (PhD thesis). The Pennsylvania State University.

Key, J. (2001). Enhancing fundraising success with custom data modelling. *International Journal of Nonprofit and Voluntary Sector Marketing*, *6*(4), 335–346. https://doi.org/10.1002/nvsm.159

Kim, G., Chae, B. K., & Olson, D. L. (2012). A support vector machine (SVM) approach to imbalanced datasets of customer responses: Comparison with other customer response models. *Service Business*, *7*(1), 167–182. https://doi.org/10.1007/s11628-012-0147-9

Korvas, R. J. (1984). *The relationship of selected alumni characteristics and attitudes to alumni financial support at a private college* (PhD thesis).

Lara, C., & Johnson, D. (2013). The anatomy of a likely donor: Econometric evidence on philanthropy to higher education. *Education Economics*, *22*(3), 293–304.

https://doi.org/10.1080/09645292.2013.766672

Lawley, C. D. (2008). *Factors that affect alumni loyalty at a public university* (PhD thesis). Purdue University.

Lawrence, K. D., Kudyba, S., & Lawrence, S. M. (2017). Funding analytics: Predictive analysis in a major state research university. In *Advances in business and management forecasting* (pp. 81–85). Emerald Publishing Limited. https://doi.org/10.1108/S1477-407020170000012005

Lertputtarak, S., & Supitchayangkool, S. (2014). Factors influencing alumni donations. *International Journal of Business and Management*, *9*(3). https://doi.org/10.5539/ijbm.v9n3p170

Leslie, L. L., & Ramey, G. (1988). Donor behavior and voluntary support for higher education institutions. *The Journal of Higher Education*, *59*(2), 115–132.

Liao, S.-H., Chu, P.-H., & Hsiao, P.-Y. (2012). Data mining techniques and applications–a decade review from 2000 to 2011. *Expert Systems with Applications*, *39*(12), 11303–11311.

Liihe, M. von der. (1998). How to get more donors: Unicef database marketing and data mining for non-commercial organizations. *WIT Transactions on Information and Communication Technologies*, *22*.

Lindahl, W. E., & Conley, A. T. (2002). Literature review: Philanthropic fundraising. *Nonprofit Management & Leadership*, *13*(1), 91–91.

Lindahl, W. E., & Winship, C. (1992). Predictive models for annual fundraising and major gift fundraising. *Nonprofit Management and Leadership*, *3*(1), 43–64.

Lindahl, W. E., & Winship, C. (1994). A logit model with interactions for predicting major gift donors. *Research in Higher Education*, *35*(6), 729–743. https://doi.org/10.1007/bf02497084

Liu, F., Feng, X., & Ouyang, Q. (2018). Factors exploration on alumni donation: A case study of creighton university. *Journal of Contemporary Management*, *9*(4).

Liu, Y., & Lapata, M. (2019). Text summarization with pretrained encoders. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (emnlp-ijcnlp)* (pp. 3721–3731).

Loveday, C. H. (2012). *An analysis of the variables associated with alumni giving and employee giving to a mid-sized southeastern university* (PhD thesis). East Tennessee State University.

Lowe, L. D. (2019). *Repeated college alumni giving: Application of the commitment-trust theory of relationship marketing* (PhD thesis). Capella University.

Luperchio, D. (2009). *Data mining and predictive modeling in institutional advancement: How ten schools found success. Technical report.* ERIC.

MacDonell, K., & Wylie, P. (2014). *Score data-driven success for your advancement team.* CASE.

Magson, N., & Routley, C. (2009). Using data in legacy fundraising: A practical approach. *International Journal of Nonprofit and Voluntary Sector Marketing*, *14*(4), 333–340. https://doi.org/10.1002/nvsm.374

Mannila, H. (1996). Data mining: Machine learning, statistics, and databases. In *Proceedings of 8th international conference on scientific and statistical data base management* (pp. 2–9). IEEE.

Manzer, L. L. (1974). *Charitable health organization donor behavior: An empirical study of value and attitude structure* (PhD thesis). Oklahoma State University.

Markoff, R. M. (1978). *An analysis of the relationship of alumni giving and level of participation in voluntary organizations: A case study* (PhD thesis). The University of Toledo.

Marr, K. A., Mullin, C. H., & Siegfried, J. J. (2005). Undergraduate financial aid and subsequent alumni giving behavior. *The Quarterly Review of Economics and Finance*, *45*(1), 123–143.

Martin, J., Joseph C. (1993). *Characteristics of alumni donors and non-donors at a research i, public university* (PhD thesis). University of Virginia.

McDearmon, J. T., & Shirley, K. (2009). Characteristics and institutional factors related to young alumni donors and non-donors. *International Journal of Educational Advancement*, *9*(2), 83–95. https://doi.org/10.1057/ijea.2009.29

McKee, D. F. (1975). *An analysis of factors which affect alumni participation and support* (PhD thesis). Indiana University.

McKinney, R. J. (1978). *Factors among select donors and nondonors related to major gifts to a private university* (PhD thesis). University of Miami.

McNally, F. E. (1985). *An analysis of alumni philanthropy related to personal, academic, and social characteristics (fundraising, public institution, university, institutional advancement)* (PhD thesis). University of San Francisco.

Meer, J., & Rosen, H. (2008). *The impact of athletic performance on alumni giving: An analysis of micro data*. National Bureau of Economic Research. https://doi.org/10.3386/w13937

Meer, J., & Rosen, H. S. (2012). Does generosity beget generosity? Alumni giving and undergraduate financial aid. *Economics of Education Review*, *31*(6), 890–907. https://doi.org/10.1016/j.econedurev.2012.06.009

Miller, M. E. (2013). *Why alumni give: How campus environment and sense of belonging shape nontraditional students' intent to give financially to their university* (PhD thesis). Saint Louis University.

Miracle, W. D. (1977). *Differences between givers and nongivers to the university of georgia annual fund* (PhD thesis). University of Georgia.

Monks, J. (2003). Patterns of giving to one's alma mater among young graduates from selective institutions. *Economics of Education Review*, *22*(2), 121–130.

Moon, S., & Azizi, K. (2013). Finding donors by relationship fundraising. *Journal of Interactive Marketing*, *27*(2), 112–129. https://doi.org/10.1016/j.intmar.2012.10.002

Morgan, R. A. (2014). *Factors that lead millennial alumni to donate to their alma mater* (PhD thesis). Southern Illinois University at Carbondale.

Morris, D. A. A. (1970). *An analysis of donors of* $10,000 Or More To The 55$ *million program at the university of michigan* (PhD thesis). University of Michigan.

Mosser, J. W. (1993). *Predicting alumni/ae gift giving behavior: A structural equation model approach* (PhD thesis). University of Michigan.

Naccarato, S. L. (2019). *Predicting alumni giving at a public comprehensive regional university: Predictive multivariate causal models for annual giving, significant cumulative giving, major giving, and planned giving* (PhD thesis). Saint Louis University.

Nandeshwar, A., & Devine, R. (2018). *Data science for fundraising: Build data-driven* *solutions using r*. Data Insights Partners. Retrieved from https://dl.acm.org/citation.cfm?id=3217650

Nandeshwar, A., Menzies, T., & Nelson, A. (2011). Learning patterns of university student retention. *Expert Systems with Applications*, *38*(12), 14984–14996. https://doi.org/10.1016/j.eswa.2011.05.048

Nelson, J., William T. (1984). *A comparison of selected undergraduate experiences of alumni who financially support their alma mater* (PhD thesis). Indiana University.

Newman, M. D. (2011). Does membership matter? Examining the relationship between alumni association membership and alumni giving. *International Journal of Educational Advancement*, *10*(4), 163–179. https://doi.org/10.1057/ijea.2011.5

NIST/SEMATECH. (2013). NIST/SEMATECH e-handbook of statistical methods. Retrieved from http://www.itl.nist.gov/div898/handbook/

O'Connor, W. J. (1961). *A study of certain factors characteristic of alumni who provide financial support and alumni who provide no financial support for their college* (PhD thesis). State University of New York at Buffalo.

Oglesby, R. A. (1991). *Age, student involvement, and other characteristics of alumni donors and alumni non-donors of southwest baptist university* (PhD thesis). University of Missouri - Columbia.

Okunade, A. A., & Berl, R. L. (1997). Determinants of charitable giving of business school alumni. *Research in Higher Education*, *38*(2), 201–214.

Okunade, A. A., Wunnava, P. V., & Walsh Jr, R. (1994). Charitable giving of alumni: Micro-data evidence from a large public university. *American Journal of Economics and Sociology*, *53*(1), 73–84.

Oliveira, A. C. M. de, Croson, R. T. A., & Eckel, C. (2011). The giving type: Identifying donors. *Journal of Public Economics*, *95*(5-6), 428–435. https://doi.org/10.1016/j.jpubeco.2010.11.012

Park, C., Ko, Y. J., Kim, H. Y., Sagas, M., & Eddosary, M. (2016). Donor motivation in college sport: Does contribution level matter? *Social Behavior and Personality: An International Journal*, *44*(6), 1015–1032. https://doi.org/10.2224/sbp.2016.44.6.1015

Pearson, W. E. (1996). *A study of donor predictability among graduates of a school of education within a research i, public university* (PhD thesis).

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, *41*(4), 1432–1462.

Perlich, C., Rosset, S., Lawrence, R. D., & Zadrozny, B. (2007). High-quantile modeling for customer wallet estimation and other applications. In *Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining - KDD 07*. ACM Press. https://doi.org/10.1145/1281192.1281297

Pinion, T. L. (2016). *Factors that influence alumni giving at three private universities* (PhD thesis). University of Toledo.

Polley, E. C., Rose, S., & van der Laan, M. J. (2011). Super learning. In (pp. 43–66). Springer New York. https://doi.org/10.1007/978-1-4419-9782-1_3

Polley, E. C., & van der Laan, M. J. (2010). Super learner in prediction. *U.C. Berkeley Division of Biostatistics Working Paper Series.*

Potharst, R., Kaymak, U., & Pijls, W. (2002). Neural networks for target selection in direct marketing. In *Neural networks in business* (pp. 89–111). IGI Global. https://doi.org/10.4018/978-1-930708-31-0.ch006

Quigley, C. J., Bingham, F. G., & Murray, K. B. (2002). An analysis of the impact of acknowledgement programs on alumni giving. *Journal of Marketing Theory and Practice*, *10*(3), 75–86. https://doi.org/10.1080/10696679.2002.11501921

Rattanamethawong, N., Sinthupinyo, S., & Chandrachai, A. (2018). An innovation model of alumni relationship management: Alumni segmentation analysis. *Kasetsart Journal of Social Sciences*, *39*(1), 150–160. https://doi.org/10.1016/j.kjss.2017.02.002

Rau, N. (2014). *Predictive modeling of alumni donors: An engagement model for fundraising in postsecondary education* (PhD thesis). James Madison University.

Rau, N. E., & Erwin, T. D. (2015). Using student engagement to predict alumni donors: An analytical model. *The Journal of Nonprofit Education and Leadership*, *5*(2).

Riecken, G., & Yavas, U. (1979). Meeting the solicitation challenge through marketing. *Administration in Social Work*, *3*(3), 327–336. https://doi.org/10.1300/j147v03n03_06

Ropp, C. T. (2014). *The relationship between student academic engagement and alumni giving at a public, state flagship university* (PhD thesis). University of Kansas.

Rosenblatt, J. A., Cusson, A. J., & McGown, L. (1986). A model to explain charitable donation-health care consumer behavior. *Advances in Consumer Research*, *13*, 235–239.

Saraih, U. N., Rahman, N. I. A., Noordin, N., Ramlan, S. N., Ahmad, R., Sakdan, M. F., & Amlus, M. H. (2018). Modelling students' experience towards the development of alumni involvement and alumni loyalty. *MATEC Web of Conferences*, *150*, 05050. https://doi.org/10.1051/matecconf/201815005050

Sargeant, A. (1998). Donor lifetime value: An empirical analysis. *International Journal of*

Nonprofit and Voluntary Sector Marketing, *3*(4), 283–292. https://doi.org/10.1002/nvsm.6090030403

Schlegelmilch, B. B., Love, A., & Diamantopoulos, A. (1997). Responses to different charity appeals: The impact of donor characteristics on the amount of donations. *European Journal of Marketing*.

Schlegelmilch, B. B., & Tynan, A. (1989). The scope for market segmentation within the charity market: An empirical analysis. *Managerial and Decision Economics*, *10*(2), 127–134.

Schubach, M., Re, M., Robinson, P. N., & Valentini, G. (2017). Imbalance-aware machine learning for predicting rare and common disease-associated non-coding variants. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-03011-5

Selig, C. W. (1999). *A study of donor predictability among alumni athletes at the university of virginia* (PhD thesis). University of Virginia.

Shadoian, H. L. (1989). *A study of predictors of alumni philanthropy in public colleges* (PhD thesis). University of Connecticut.

Shen, A., & Tsai, C.-Y. (2009). Are single-gift committed donors different from their multiple-gift counterparts? *International Journal of Nonprofit and Voluntary Sector Marketing*, n/a–n/a. https://doi.org/10.1002/nvsm.387

Skari, L. A. (2014). Community college alumni: Predicting who gives. *Community College Review*, *42*(1), 23–40.

Smith, S. M., & Beik, L. L. (1982). Market segmentation for fund raisers. *Journal of the Academy of Marketing Science*, *10*(3), 208–216. https://doi.org/10.1007/bf02729963

Soukup, D. J. (1983). A markov analysis of fund-raising alternatives. *Journal of Marketing*

Research, *20*(3), 314–319. https://doi.org/10.1177/002224378302000310

Steinnes, D. N. (2011). An econometric analysis of aging and alumni/ae altruism. *International Business & Economics Research Journal (IBER)*, *1*(5). https://doi.org/10.19030/iber.v1i5.3924

Sun, X., Hoffman, S. C., & Grady, M. L. (2007). A multivariate causal model of alumni giving: Implications for alumni fundraisers. *International Journal of Educational Advancement*, *7*(4), 307–332.

Sundel, H. H., Zelman, W. N., Weaver, C. N., & Pasternak, R. E. (1978). Fund raising: Understanding donor motivation. *Social Work*, *23*(3), 233–236.

Taylor, A. L., & Martin, J. C. (1995). Characteristics of alumni donors and nondonors at a research i, public university. *Research in Higher Education*, *36*(3), 283–302. https://doi.org/10.1007/bf02208312

Terry, N., & Macy, A. (2007). Determinants of alumni giving rates. *Journal of Economics and Economic Education Research*, *8*(3), 3–17.

Thompson, L. A. (2010). *Data mining for higher education advancement: A study of eight north american colleges and universities* (PhD thesis). North Dakota State University.

Tiger, A., & Preston, L. (2013). Logged in and connected? A quantitative analysis of online course use and alumni giving. *American Journal of Business Education (AJBE)*, *6*(3), 361. https://doi.org/10.19030/ajbe.v6i3.7816

Toohill, J., Mullins, L., Barclay, J., & Sadnicki, M. (1997). Turning data into donations: A predictive model for individual giving. *International Journal of Nonprofit and Voluntary Sector Marketing*, *2*(2), 136–147. https://doi.org/10.1002/nvsm.6090020205

Torres, R. L. (2014). *Dynamic segmentation modeling: Application of finite mixture models* *to explain the giving behavior of donors in a university setting* (PhD thesis). NORTHWESTERN UNIVERSITY.

Trafimow, D. (2014). Editorial. *Basic and Applied Social Psychology*, *36*(1), 1–2. https://doi.org/10.1080/01973533.2014.865505

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, *37*(1), 1–2. https://doi.org/10.1080/01973533.2015.1012991

Truitt, J. (2013). *The relationship between student engagement and recent alumni donors at carnegie baccalaureate colleges located in the southeastern united states* (PhD thesis). University of Central Florida.

Tsao, J. C., & Coll, G. (2005). To give or not to give: Factors determining alumni intent to make donations as a pr outcome. *Journalism & Mass Communication Educator*, *59*(4), 381–392. Retrieved from http://libproxy.usc.edu/login?url=https://search.proquest.com/docview/215288904?accountid=14749

Udenze, A. (2014). APPLICATION of data mining techniques to problems in fund raising. *International Journal of Current Research and Review*, *6*(22), 1.

Veludo-de-Oliveira, T. M., Alhaidari, I. S., Yani-de-Soriano, M., & Yousafzai, S. Y. (2016). Comparing the explanatory and predictive power of intention-based theories of personal monetary donation to charitable organizations. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *28*(2), 571–593. https://doi.org/10.1007/s11266-016-9690-7

Vequist IV, D. G. (2017). Nonprofit fundraising transformation through analytics. In *Decision management: Concepts, methodologies, tools, and applications* (pp. 1549–1557). IGI Global.

Verhaert, G. (2010). *The role of database marketing in improving direct mail fundraising* (PhD thesis). Ghent University.

Walcott, M. E. (2015). *Predictive modeling and alumni fundraising in higher education* (PhD thesis). Illinois State University.

Weerts, D. J., & Ronca, J. M. (2009). Using classification trees to predict alumni giving for higher education. *Education Economics*, *17*(1), 95–122. https://doi.org/10.1080/09645290801976985

Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology*, *54*(1), 41–59. https://doi.org/10.1016/j.infsof.2011.09.002

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wylie, P. B. (2004). *Data mining for fund raisers: How to use simple statistics to find the gold in your donor database–even if you hate statistics: A starter guide*. Sterling Publishing Company.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems* (pp. 5754–5764).

Yavas, U., Riecken, G., & Parameswaran, R. (1981). Personality, organization-specific attitude, and socioeconomic correlates of charity giving behavior. *Journal of the Academy of Marketing Science*, *9*(1-2), 52–65. https://doi.org/10.1007/bf02723565

Ye, L. (2017). *A machine learning approach to fundraising success in higher education* (Master's thesis). University of Victoria.

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75.

Yu, S., Chen, D., Li, B., & Ma, Y. (2014). A personalized recommendation algorithm based on interest graph. In *The 2014 2nd international conference on systems and informatics (ICSAI 2014)*. IEEE. https://doi.org/10.1109/icsai.2014.7009419

Zhang, W. (2014). *Segmentation modeling: Applications of finite mixture regression models in university fundraising and management of transportation infrastructure* (PhD thesis). Northwestern University.

Appendix

Table 7

CHAID

Authors	Decade
Liihe (1998)	1990-1999
Denizard-Ramsamy and Medina-Borja (2008)	2000-2009

Table 8

Clustering

Authors	Decade
Cermak et al. (1994)	1990-1999
Blanc and Rucks (2009)	2000-2009
Luperchio (2009)	2000-2009
Durango-Cohen et al. (2013b)	2010-2019
Durango-Cohen et al. (2013a)	2010-2019
Zhang (2014)	2010-2019
Durango-Cohen and Balasubramanian (2015)	2010-2019

Table 9

Descriptive Statistics

Authors	Decade
O'Connor (1961)	1960-1969
Morris (1970)	1970-1979
Caruthers (1973)	1970-1979
Blumenfeld and Sartain (1974)	1970-1979
Gardner (1975)	1970-1979

Authors	Decade
McKee (1975)	1970-1979
Markoff (1978)	1970-1979
McKinney (1978)	1970-1979
Sundel et al. (1978)	1970-1979
Riecken and Yavas (1979)	1970-1979
Anderson (1981)	1980-1989
Keller (1982)	1980-1989
Smith and Beik (1982)	1980-1989
Chewning (1984)	1980-1989
Frederick (1984)	1980-1989
Korvas (1984)	1980-1989
Nelson (1984)	1980-1989
Dietz (1985)	1980-1989
McNally (1985)	1980-1989
Haddad (1986)	1980-1989
Schlegelmilch and Tynan (1989)	1980-1989
Oglesby (1991)	1990-1999
Berger and Smith (1997)	1990-1999
Hunter et al. (1999)	1990-1999
Quigley et al. (2002)	2000-2009
Bingham Jr et al. (2003)	2000-2009
Wylie (2004)	2000-2009
Gunsalus (2005)	2000-2009
Magson and Routley (2009)	2000-2009
Newman (2011)	2010-2019

Authors	Decade
Loveday (2012)	2010-2019
Bruyn and Prokopec (2013)	2010-2019
Johnson (2013)	2010-2019
Miller (2013)	2010-2019

Table 10

Ensemble

Authors	Decade
Potharst et al. (2002)	2000-2009
Chen (2010)	2010-2019
E. J. Durango-Cohen (2013)	2010-2019
Moon and Azizi (2013)	2010-2019
Torres (2014)	2010-2019
Udenze (2014)	2010-2019
Chung and Lee (2015)	2010-2019
Kakrala and Chakraborty (2015)	2010-2019
Ye (2017)	2010-2019
Rattanamethawong et al. (2018)	2010-2019

Table 11

Lifetime Value

Authors	Decade
Hunter and Hill (1998)	1990-1999
Sargeant (1998)	1990-1999

Authors	Decade
Aldrich (2000)	2000-2009

Table 12

Machine Learning

Authors	Decade
Weerts and Ronca (2009)	2000-2009

Table 13

Markov Chains

Authors	Decade
Soukup (1983)	1980-1989
Toohill et al. (1997)	1990-1999

Table 14

Neural Networks

Authors	Decade
Goodman and Plouff (1997)	1990-1999

Table 15

Other

Authors	Decade
Herzlinger (1977)	1970-1979
Birkholz (2008)	2000-2009

Authors	Decade
Hashemi et al. (2009)	2000-2009
MacDonell and Wylie (2014)	2010-2019
Nandeshwar and Devine (2018)	2010-2019

Table 16

Regression

Authors	Decade
Manzer (1974)	1970-1979
Miracle (1977)	1970-1979
Yavas et al. (1981)	1980-1989
Beeler (1982)	1980-1989
Rosenblatt et al. (1986)	1980-1989
House (1987)	1980-1989
Grill (1988)	1980-1989
Leslie and Ramey (1988)	1980-1989
Shadoian (1989)	1980-1989
Boyle (1990)	1990-1999
Duronio and Loessin (1990)	1990-1999
Burgess-Getts (1992)	1990-1999
Hueston (1992)	1990-1999
Lindahl and Winship (1992)	1990-1999
Martin (1993)	1990-1999
Mosser (1993)	1990-1999
Lindahl and Winship (1994)	1990-1999
Okunade et al. (1994)	1990-1999

Authors	Decade
Bruggink and Siddiqui (1995)	1990-1999
Taylor and Martin (1995)	1990-1999
Baade and Sundberg (1996)	1990-1999
Pearson (1996)	1990-1999
Okunade and Berl (1997)	1990-1999
Schlegelmilch et al. (1997)	1990-1999
Duncan (1999)	1990-1999
Selig (1999)	1990-1999
Belfield and Beney (2000)	2000-2009
Greenlee and Trussel (2000)	2000-2009
Hanson (2000)	2000-2009
Key (2001)	2000-2009
Cunningham and Cochi-Ficano (2002)	2000-2009
Bennett (2003)	2000-2009
Monks (2003)	2000-2009
Hoyt (2004)	2000-2009
Gaier (2005)	2000-2009
Marr et al. (2005)	2000-2009
Tsao and Coll (2005)	2000-2009
Bennett (2006)	2000-2009
Bohannon (2007)	2000-2009
Diehl (2007)	2000-2009
Sun et al. (2007)	2000-2009
Terry and Macy (2007)	2000-2009
Lawley (2008)	2000-2009

Authors	Decade
Meer and Rosen (2008)	2000-2009
Holmes (2009)	2000-2009
McDearmon and Shirley (2009)	2000-2009
Shen and Tsai (2009)	2000-2009
Dickert et al. (2010)	2010-2019
Thompson (2010)	2010-2019
Verhaert (2010)	2010-2019
Oliveira et al. (2011)	2010-2019
Steinnes (2011)	2010-2019
Baruch and Sang (2012)	2010-2019
Meer and Rosen (2012)	2010-2019
Ketter (2013)	2010-2019
Lara and Johnson (2013)	2010-2019
Tiger and Preston (2013)	2010-2019
Truitt (2013)	2010-2019
Lertputtarak and Supitchayangkool (2014)	2010-2019
Morgan (2014)	2010-2019
Rau (2014)	2010-2019
Ropp (2014)	2010-2019
Skari (2014)	2010-2019
Rau and Erwin (2015)	2010-2019
Walcott (2015)	2010-2019
Veludo-de-Oliveira et al. (2016)	2010-2019
Park et al. (2016)	2010-2019
Pinion (2016)	2010-2019

Authors	Decade
Brunette et al. (2017)	2010-2019
Faisal (2017)	2010-2019
Lawrence et al. (2017)	2010-2019
Christian (2018)	2010-2019
Day (2018)	2010-2019
Liu et al. (2018)	2010-2019
Saraih et al. (2018)	2010-2019
Lowe (2019)	2010-2019
Naccarato (2019)	2010-2019

Table 17

Social Media

Authors	Decade
Vequist IV (2017)	2010-2019

Table 18

Support Vector Machines

Authors	Decade
Kim et al. (2012)	2010-2019

Table 19

Survival Analysis

Authors	Decade
Drye et al. (2001)	2000-2009