TRANSFORMING U.S. WORKFORCE DEVELOPMENT POLICIES FOR THE 21st CENTURY

EDITORS

Carl Van Horn, John J. Heldrich Center for Workforce Development at Rutgers University

Tammy Edwards, Federal Reserve Bank of Kansas City

Todd Greene, Federal Reserve Bank of Atlanta

Transforming U.S. Workforce Development Policies for the 21st Century

Carl Van Horn
Tammy Edwards
Todd Greene
Editors

2015

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Part 3

Building Evidence-Based Policy and Practice

19
Doing More with Less

Leveraging Advances in Data Science to Support an Intelligent Workforce System

William Mabe
Scott Powell
Alex Ruder
Rutgers University

In the aftermath of the Great Recession, shrinking budgets and high caseloads all but guarantee that the workforce system of the twenty-first century will have to serve more job seekers with fewer resources. Maximizing the system's efficiency and effectiveness will require the U.S. workforce system to evolve into an intelligent workforce system, where data drive the decisions of all stakeholders—from policymakers to workforce program staff, education and training providers, job seekers, and employers. For the system to be truly intelligent and data driven, state workforce agencies (SWAs) and local workforce areas must be able to extract meaning from multiple types of data, including numeric, location, and text data, stored across multiple state agencies; properly analyze these data to generate accurate insights and integrate them into stakeholder decision making; and foster an organizational culture that values data collection, quality, analysis, and dissemination.

Advances in data science, coupled with the ever-expanding capabilities of open-source and low-cost software, offer the workforce system a genuine opportunity to do more with less. Specifically, developments in two areas—mining information that states have collected for years but examined only infrequently (such as location data and textual data), and analyzing their data in such a way as to generate more accurate insights, especially in the field of prediction—can be harnessed to help states deliver services more effectively to workforce system customers. This chapter describes how SWAs can adopt tools to analyze nontraditional data sources such as geospatial and text data and to improve their predictive practices.

During the past several decades, SWAs have developed tools to analyze more traditional types of data, such as numbers (0, 1, −27.15) and categories (male and female). In addition to numeric and categorical data, however, SWAs also store important geospatial (location) and textual information. Examples of geospatial information include addresses of job seeker customers when they register for services, the addresses of employer customers and the establishments where they have job openings, and the Internet protocol addresses—which can be linked to physical locations—of job seekers who are using state online job boards to search for employment. At the same time, SWA data systems capture vast amounts of textual information. For example, every time a counselor enters a comment or note about a customer into an SWA database, the database records critical qualitative information about the job seeker, such as his skill deficits, the counselor's assessment of his job readiness, and possibly his attitude toward his job search. Although SWAs have made little use of either location or text data, open-source and low-cost software are available to help SWAs extract meaning from them. Incorporating location and textual data can support learning about how SWAs serve their customers, the effectiveness of their programs, and strategies for program improvement.

In an intelligent workforce system, data analysis adds value in many different ways, including performance metrics for tracking program implementation, scorecards for public accountability, rigorous evaluations to identify the programs that most benefit customers, and predictions of which customers are most in need of services and most likely to benefit from them. For SWAs, one of the most widely used data applications is prediction: learning from the data so that when a new customer enters the workforce system, the SWA knows what the experiences of thousands of customers like her have been and can therefore predict how she is likely to fare and what services might benefit her the most. To be more specific, an intelligent workforce system can use prediction to assist SWAs in better serving customers by identifying customers likely to experience an adverse event such as prolonged unemployment, matching customers to the job openings for which they are best suited, or identifying the set of reemployment and job training services that are likely to be the most effective at helping a customer achieve a positive labor market outcome. Of course, prediction cannot foresee the future perfectly. On the contrary, prediction is almost always prone to at least some error. But high-quality prediction can allow us to see the future more clearly than with no prediction at all, and this extra insight can significantly improve program outcomes.

While innovations in data science hold the promise of greatly improving the ability of SWAs to serve their customers, realizing this promise requires the effective use of their resources and capabilities. Fortunately, states already possess the resource that is the most costly and time consuming to develop—namely, detailed customer-level data that they have collected for decades. Effective use of individual-level data begins with high levels of data security to safeguard the privacy and confidentiality of the information the SWAs have collected from the public. Once data security is established, combining data from many different programs affords SWAs a fuller understanding of each customer they serve and allows for more detailed analyses than have generally been possible before. Through the Workforce Data Quality Initiative, the U.S. Department of Labor (USDOL) has funded 32 states to securely link data that have traditionally been housed in separate databases and maintained by multiple state agencies. We aim to introduce SWAs to a number of methods for leveraging this wealth of existing data.

The chapter is organized into two parts. In the first, we examine how location data and then textual data can be analyzed to yield value for SWAs. For each data type, we walk through an application to illustrate how SWAs and local areas can derive insights from these data. In the second part of the chapter, we describe the prediction process and the steps that these agencies need to follow in order to be able to generate accurate predictions and incorporate them into service delivery. We then illustrate how SWAs can improve their predictive practices by applying predictive modeling to identify job seekers who are most likely to experience long-term unemployment.

GEOSPATIAL AND TEXT DATA IN WORKFORCE DEVELOPMENT

Modern analytics involves using a variety of different types of data. The more traditional types, such as numeric and categorical data, are now found alongside data types such as geospatial (Burrough and McDonnell 1998) and text data (Schutt and O'Neil 2014). Geospatial data, which refers to address and location information, and large collections of text—such as online job listings, job seeker profiles, and counselor notes on individual customers—are increasingly available to workforce development professionals. A challenge workforce counselors face is deciding how to make use of these valuable data collections.

Geospatial Data

Spatial data are features—roads, buildings, and addresses—whose locations can be mapped onto the earth's surface along with the feature's descriptive characteristics. Workforce data systems often store data elements on customers and employers that are spatial in nature, such as an employer's address, along with attributes such as current job openings and contact information. Data visualization through geographic information systems (GIS) can be a powerful tool for helping SWAs and workforce boards turn this geospatial data into innovative new service solutions. Specifically, SWAs and local areas can improve their targeting of workforce services to better meet job seekers where they are, including making decisions about where to locate satellite offices and where to concentrate outreach efforts.

While workforce professionals have been using maps to improve services for decades, the last few years have produced an exponential increase in mapping possibilities. As a result of innovations in both workforce data and mapping software, powerful maps need not be costly or time-intensive to create. Through programs such as the Workforce Data Quality Initiative, state and local governments are increasingly linking administrative data that are housed across multiple agencies. This allows governments to create powerful maps that display not only workforce information, such as wages and WIA participation, but also data related to education and human services programs.

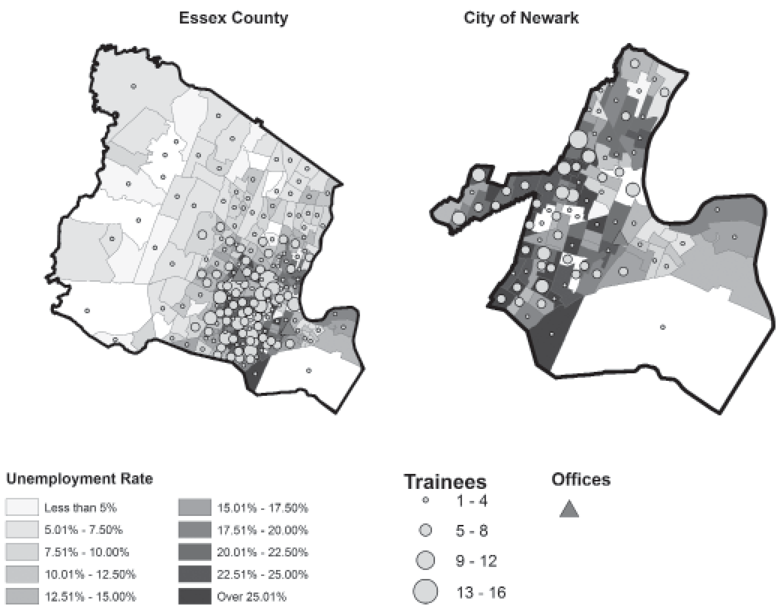
Regarding innovations in mapping software, applications such as ArcGIS easily combine location-based information with workforce data. This software can be preloaded with local census-based labor market information and demographic characteristics, while local infrastructure information, such as roads and public transportation routes, is easily integrated. With so much data already assimilated into the software, workforce agencies need only provide a single piece of information: customer location. Finally, due to the proliferation of geospatial data use in the public sector, trained GIS professionals are often available at all levels of government, as well as in local colleges and universities. Thus, governments frequently already employ all the staff necessary to leverage geospatial data for making workforce policy decisions, making data visualization tools that use geospatial data accessible and affordable, even at the local level.

Application: customer outreach

We illustrate the value of geospatial data by mapping workforce information from Essex County, New Jersey, and the city of Newark. The map below (Figure 19.1) plots the location of occupational training participants, aggregating the information by census tract to protect customer privacy (U.S. Census Bureau 1994). The trainees are represented by circles, with larger circles signifying more trainees within a given census tract. The unemployment rate of each census tract is also represented, with darker-shaded tracts representing higher unemployment rates. Finally, American Job Center (AJC) offices are represented with triangles.

Created for the Newark Workforce Investment Board (WIB) to assist with recent exploration into strategies for customer outreach, these maps quickly convey a large amount of information that is critical to identifying the areas where the WIB can most efficiently target its efforts. For example, the areas with the most customers in need of services are concentrated in close proximity to the city of Newark, with the areas farther out in Essex County benefiting from relatively low levels of unemployment. So while there are currently no offices in the outer tracts of the county, there is also not necessarily a need to increase outreach efforts in this region. Within the city itself, there is substantial variation in unemployment, and many of these areas are underserved. Specifically, the tracts with high unemployment but few trainees could

Figure 19.1 Number of Trainees by Census Tract, Essex County, New Jersey, 2012



be prime candidates for outreach efforts. And, in determining where to place a new outreach center, GIS software can easily overlay roads and public transportation routes onto this map to find a location that would be accessible to the underserved customers in need of assistance.

Perhaps most importantly, the WIB needed to provide only a single piece of workforce information to create this map: the location of trainees. All other data were either publicly available or integrated into the GIS software application. Thus, the maps are not only powerful in their ability to quickly convey information that is critical to developing an outreach strategy but also relatively undemanding to create.

Text Data

Like geospatial data, text information holds a great deal of unlocked potential for improving SWA services. In a workforce system, text data can include titles of job openings, descriptive information on skill requirements and job duties from job postings, counselor comments on job seeker skills and aptitudes, and customer feedback on their satisfaction with the services they have received.

Although many states are moving to apply text analysis algorithms to match job seekers to the jobs with the skill requirements and job duties that most closely align with their experience, nearly all of them use commercial products to do so. A number of companies have developed proprietary algorithms that allow job seekers to use a search function that automatically reviews job postings and notifies them of jobs that match the skills listed in their resumes. Whether organizations analyze text data themselves or enlist the services of a private sector firm, an understanding of the basic tools of text mining aids the use and interpretation of these methods. In addition, advances in computer software have made text mining methods accessible to a wide range of practitioners, increasing opportunities for organizations to conduct "in-house" analyses of text.

Text mining is a collection of analytic methods used to extract useful information from large volumes of text (Sebastiani 2002; Witten 2005). These methods are particularly suited for large text collections whose size makes human reading and coding prohibitively costly. Computer algorithms automate the process of searching the texts for patterns and information. Text mining methods can be used for text summarization and document retrieval, for clustering texts into predefined or previously unknown categories, and for extracting structured information such as Web addresses from texts.

This section reviews several text mining methods that are well suited to workforce development applications.[1] Often, the first challenge is deciding how to summarize the text in a collection. We highlight several text mining methods that can help workforce professionals summarize large text collections and organize similar documents into a set of categories. Then, to give a sense of how these tools might be applied, we analyze open-ended survey responses from a survey of individuals who received services from AJCs in a state in the eastern half of the United States.

Summarization and classification of text

Faced with a large collection of text, an organization may first need a simple method for summarizing the content of the collection.[2] One of the simplest methods that an organization can use is count-based analysis. As the name implies, it involves calculating the most frequently used words in both a text collection and in individual documents. A count-based approach can reveal, for example, that the words *transportation* and *warehouse* are the two most frequently used words in a collection of job ads.

A next step toward summarizing a text collection is to calculate word associations. Word associations reveal which words are highly correlated with the use of a selected word. For example, an organization may calculate associations for both *transportation* and *warehouse*. Word association can reveal that *full-time* and *truck* are strongly associated with the words *transportation* and *warehouse*. In this example, these two simple methods have given the organization preliminary evidence that its collection of jobs ads features many ads for full-time, tractor-trailer truck drivers.[3]

Classification and clustering

Many text mining problems involve grouping documents into natural clusters of similar documents. Consider a scenario in which a workforce organization has a database of thousands of job postings and wants to group them by industry of employment. Human-based coding of these job ads is prohibitively expensive: the organization likely lacks the staff and the time to read and code thousands of job ads. Text mining classification methods offer an automated approach to accomplish this task.

One of the first steps in text classification is choosing the approach that is appropriate for the task. Generally, this choice is determined by the large variety of classification methods, which can be grouped into two general approaches: supervised and unsupervised (Grimmer and Stewart 2013).

Supervised methods

In the phrase "supervised learning methods," the term *supervised* is used to refer to methods where the categories are known in advance. The researcher supervises the automated classification process by providing the computer a training set of documents already labeled with the known categories. The supervised method estimates the words or phrases predictive of the label. The researcher then uses the estimates from the training set to infer the labels for documents in the test set. Popular supervised methods include k-nearest neighbor classification, support vector machines, string kernel clustering, and the Naïve Bayes classifier.[4]

Dictionary methods are a relatively simple and intuitive way to organize texts into known categories (Neuendorf 2002). To assign texts to a given category, dictionary methods use the rate at which certain predefined key words appear in the text. More specifically, a dictionary method takes a list of words (the dictionary) and counts the proportion of words in a text that are also in the dictionary. An organization may use a sample of existing job ads to create a dictionary of keywords that identify the likely industry of new job ads. Another common application of dictionary methods is sentiment analysis, where the goal is to assess degree of positive, neutral, or negative language in text.

When using dictionary methods, organizations must choose dictionaries appropriate for the application, such that the meaning of the words in the dictionary corresponds to the way words are used in the text (Loughran and McDonald 2011). The word *work*, for example, can be positive in many contexts, such as *the machine works*. In workforce context, work is more often a neutral term: *looking for work*, *I worked as a machinist*. Organizations can acquire free text analysis dictionaries on the Web, or construct their own dictionary tailored to the specific application.

Unsupervised methods

In some applications, the categories may not be known in advance, making the application of supervised methods infeasible. Unsupervised learning methods apply when no predefined categories are available and the researcher still seeks to group similar documents into clusters. Unsupervised methods can also help to explore a large collection of text documents by summarizing its thematic content.

Since the methods are fully automated, they can discover both expected categories (e.g., health care jobs) and unexpected categories. For example, the method can reveal that multiple categories define the broader health care industry; one category may feature the words *hospital*, *surgery*, and *nurse*, while another category features *home*, *health*, and *nurse*. In this example, the unsupervised model infers that two broad categories of jobs are prominent in the collection of job ads: hospital-based surgical nurses and nurses employed in home health care services. If the organization were to use a supervised method, it would have to know these two categories in advance. It is possible that the organization may be unaware of the extent of local demand for home health care nurses. If the organization were to rely solely on supervised methods, it would overlook an important piece of information about the local labor market.

Unsupervised methods range from fully automated clustering algorithms (Grimmer and King 2011) to computationally demanding topic models (see Blei [2012] for a review and discussion of topic models). With all unsupervised methods, the goals are generally the same: either explore the categories (or thematic topics) that constitute a text collection, or cluster similar documents together into previously unknown categories.

Application: analysis of open-ended survey responses

Organizations often employ surveys that ask respondents to rate a service along some preset scale, such as poor to excellent. However, these closed-ended responses, while useful, are often too coarse to answer questions such as why respondents selected the rating they did. In contrast, open-ended survey questions allow respondents to elaborate on previous answers, suggest improvements, or offer praise in their own words, rather than in the predefined language of the survey developer.

One challenge that responses to open-ended survey questions present to researchers is how to analyze large amounts of text data. Generally, organizations require a team of human coders to read the responses and code them in a manner consistent with the organization's goals. Human coding is a time-consuming task. An alternative strategy for systematically analyzing open-ended survey responses is to use simple, computationally based text mining tools.

In a recent survey of individuals who received workforce services in a state in the eastern half of the United States, we asked respondents a closed-ended question: *How valuable was this service to you—not at all valuable, somewhat valuable, or very valuable?* We followed this question with an open-ended question:

Is there anything else that you would like to add about your experience, either positive or negative, that could inform the improvement of aspects of the program that did not work as well, or ensure the retention of those things that did work well?

We sought to use the open-ended question to analyze why respondents gave the rating that they chose. In particular, we wanted to know which aspects of the program were prominent in more negative reviews compared to the aspects mentioned in more positive reviews.[5] Rather than human coding of all the responses, our first analysis involved the use of text mining tools provided in the "tm: Text Mining Package" in the open-source statistical software R (Feinerer, Hornik, and Meyer 2008). The tm: Text Mining Package includes tools to download and analyze the data, as well as to implement standard text preprocessing steps such as removing punctuation and numbers, and changing words to reflect their stems or roots.

Even this basic application of text mining revealed several differences across respondents who rated their overall experience negatively compared to those who rated it positively. Respondents who offered a negative rating were more likely to write longer responses and focus their comments on particular aspects of the program: the classes, courses, and the AJC counselors. In contrast, respondents who rated their experiences positively were less likely to identify any particular aspect of the program that they found helpful. Rather, the positive respondents were more likely to use the open-ended question as an opportunity to voice their general satisfaction with the services and the help they received finding a job.[6] The information gained from the open-ended survey responses can help organizational leadership strategically target improvement efforts to the aspects of service that contributed to customers' negative evaluations.

PREDICTIVE ANALYTICS

Although states have been using data to make predictions for over a decade, primarily to implement the Worker Profiling and Reemployment Services (WPRS) system, technological advances in predictive analytics, together with shrinking financial resources and demands for increased performance accountability, have precipitated wider interest in and adoption of predictive analytics for workforce development applications (the White House 2014). Many states and organizations, for example, have contracted with proprietary firms to leverage text data in resumes and job advertisements to make better predictions concerning which applicants are most likely to succeed in a given job.

Applications in predictive analytics generally share a common goal: to generate accurate predictions that contribute to improved organizational performance or service delivery. To meet this goal, SWAs must be able to measure the performance of their predictive analytic applications and design or modify them to improve prediction.

There are three ways in which SWAs could generate more accurate predictions. First, they could increase the accuracy of their predictions by comparing the performance of predictions based on multiple different predictive algorithms.[7] Second, SWAs could improve the predictive power of their models by regularly evaluating the accuracy of their predictions and adjusting their models over time.[8] Finally, they could improve predictive accuracy by including more diverse sets of predictors in their models.

The Prediction Process

When most people think about prediction in the context of workforce development, they probably think about something like the following example. John has worked for 10 years as an accountant at a retail store. He loses his job and files for UI. In filing the claim, he provides information about his occupation and industry, how long he worked for the company, and why he lost his job. John also lists his age, race and ethnicity, and level of education. The SWA might then use this information to estimate such items as how likely he is to suffer prolonged unemployment, the jobs for which he is the most qualified, and/or which services are likely to afford the most help in returning to work.[9] Although this example illustrates an important part of the prediction process—the assignment of a prediction to a current SWA customer—it is incomplete because it omits other parts of the process.

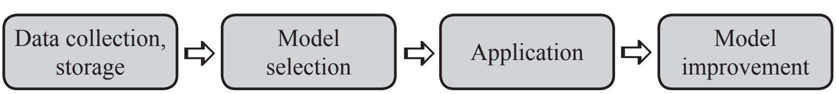
The prediction process actually begins with the identification of a substantive problem to which the application of predictive modeling might help the SWA overcome (Finlay 2014). In the case of workforce development, these problems largely revolve around identifying at-risk customers, matching customers to open jobs, and matching them to the most appropriate services.

After identifying a problem suitable for predictive modeling, the task of using prediction to improve service delivery involves a four-step process: 1) collecting, storing, and preparing for analysis data on the individuals whom the SWA serves; 2) testing many different predictive models on the data and selecting the one that generates the most accurate predictions;[10] 3) using the best model to generate predictions for each new customer and applying the predictions to serve customers better; and 4) assessing and improving the predictive model over time. Figure 19.2 depicts this process.

The more complete the data on workforce system customers, the more diverse the predictors that SWAs can include in their models and the more accurate their predictions are likely to be. Preparing data for analysis involves extracting data from diverse data systems, transforming the data so they can be analyzed using statistical software, and loading them into a database for analysis.

During model selection, researchers learn from the data by engaging in retrospective prediction (Siegel 2013). A SWA may want, as in the example we present below, to be able to predict which newly unemployed individuals are likely to remain unemployed for an extended period. SWA researchers would begin by examining a *subset* of the SWA's existing data, looking only at what was known about the unemployed individuals at the time they became unemployed, and use this information to "predict" who is likely to be unemployed a year later. The challenge is to find patterns that hold not just with the available data, but also in new data. So the researchers then test several predictive models for accuracy on a second subset of data, validate the results on a third subset of the data, and select for deployment in the field the predictive model that emerged from the validation phase with the highest accuracy. While this phase may provide the greatest challenge for

Figure 19.2 Predictive Modeling Process



SWAs in terms of developing new expertise, we demonstrate below that these challenges are not as large as they appear. Additionally, this phase of the prediction process does not need to be repeated on a regular basis, providing SWAs with the opportunity to hire an outside party to perform model selection if they are not equipped to perform this task internally.

In the application phase, the organization uses the winning predictive model to predict which new customers are likely, in our example, to have long spells of unemployment and takes some action based on the predictions. This means that when a newly unemployed worker enters personal information on a UI claim application or an AJC intake form, a predictive model examines the worker's characteristics and predicts how likely the worker is to have a long spell of unemployment. The SWA could then target services to this customer based in part on the predictive score.

Finally, because economic conditions change over time, predictive models must be updated regularly to remain accurate. In addition, the effect of assigning services based on the predictions of the model needs to be rigorously evaluated to ensure that the predictive system not only makes accurate predictions but also positively affects the outcomes it was designed to improve.

Current SWA Uses of Prediction

In this section, we review the substantive problems to which SWAs currently apply prediction and examine how SWAs engage in prediction. To date, SWAs have used predictive models to assist in addressing two substantive problems. First, nearly all states apply predictive modeling to identify the newly unemployed workers who are most likely to remain unemployed for so long that they exhaust their UI benefits. Individuals are then assigned to various services, a process known as worker profiling (USDOL 2000). SWAs can also use predictive models to target services and place customers into programs that are most likely to assist them with labor market reintegration. As discussed in the first section, SWAs are also starting to mine text data and combine it with other data on job seekers in order to develop predictive job matching systems.

Worker profiling

In 1993, Congress passed the Unemployment Compensation Amendments, establishing a federal mandate for the WPRS initiative (Wandner 1997). The law requires SWAs to develop either characteristic screening processes or statistical models to identify the individuals who have been permanently laid off and who are most likely to exhaust their UI benefits, for the purpose of referring them to reemployment services. This process, known as worker profiling, produces a prediction of a UI claimant's probability of exhausting his or her UI benefits based on a set of personal and economic variables that differs from state to state, though five variables are recommended by USDOL—education, job tenure, industry, occupation, and unemployment rate (USDOL 2000).[11] The legislation, as well as subsequent guidance from USDOL, requires states to use data on the outcomes of individuals referred through WPRS to update their models over time. The WPRS Policy Workgroup called on states to "update and revise their profiling models regularly, as well as add new variables and revise model specifications, as appropriate" (WPRS Policy Workgroup 1999, p. 16).

Identifying optimal services

In 2001, with support from USDOL, the W.E. Upjohn Institute built and pilot-tested the Frontline Decision Support System (FDSS) in two Georgia workforce centers with the objective of improving customer and workforce staff decision making with respect to reemployment. The system consists of a series of tools to provide customers with better information on their employment prospects, their job search, and the services that would be the most effective at helping them to return to work. The system generates the probability of a worker being reemployed in the same industry, a list of occupations related to the job seeker's previous occupation, and the services that are likely to be the most effective at helping the job seeker return to work (Eberts and O'Leary 2002). Because FDSS was not implemented on a statewide basis, a rigorous evaluation of the program's effect on reemployment has not been conducted. The FDSS is discussed in more detail elsewhere in this volume.

How states conduct predictive modeling

Although SWAs have applied predictive modeling to various substantive issues, they most commonly use prediction in implementing WPRS. Through WPRS, nearly every SWA in the nation uses a predictive model on a daily or weekly basis to assign a probability of UI benefit exhaustion to newly unemployed UI claimants and to refer individuals to services based on their scores. Since WPRS is the biggest predictive modeling enterprise that the SWAs undertake, we sought to learn how states engage in predictive modeling by surveying them about their WPRS predictive modeling practices. Specifically, we were interested in learning about three aspects of how they engage in predictive modeling: 1) the variables they include in their predictive models, 2) the algorithms they use to calculate predictions, and 3) the frequency with which they update their predictive models.

In April 2014, we e-mailed the survey to the UI directors in the SWAs of all 50 states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. We received 34 responses, which enabled us to draw three primary conclusions with respect to how SWAs engage in predictive modeling.

First, states primarily include in their models the variables recommended by USDOL (education, job tenure, industry, occupation, and unemployment rate). Of the 34 responding states, 27 use at least the variables recommended by USDOL. The majority of states, however, include few variables beyond this list. The results of our survey are consistent with what others have previously learned about how SWAs conduct predictive modeling. The U.S. Government Accountability Office (2007); Sullivan et al. (2007); and Black, Smith, Plesca et al. (2003) find that many states do not include in their models a number of variables, such as the number of previous employers, past wages, and previous UI receipt, that might improve the predictive power of their worker profiling models. In their reanalysis of Kentucky's UI claims data, Black, Smith, Plesca et al. (2003) conclude that states could improve the predictive power of their models by incorporating more variables, including whether the customer received welfare benefits, the office where the individual received services, and whether the customer was enrolled in postsecondary education at the time of filing a claim. They note, however, that most states' models do not include these variables, and neither did many of the respondents to our survey.

Second, states primarily use a logit model to predict benefit exhaustion. Of the 34 responding states that use predictive models to assign claimants to services, 23 of them use a logit model. While one state used a neural network model, two states did not use a statistical model at all, and instead assigned customers using a characteristic screen, which selects individuals for services based on a handful of individual attributes.

The third conclusion is that many states do not regularly update their models. Despite the requirements of the original legislation and the guidance issued by USDOL, states are not regularly updating their profiling models. In their survey of state profiling models, Sullivan et al. (2007) find that many states had not updated their models in years. In some cases, states were using models estimated possibly 10 years previously to predict worker employment outcomes in the present day. Our survey from this year finds that updating of profiling models remains infrequent, with 16 of the 34 responding states indicating that they have not updated their models since before 2008. In other words, despite the substantial changes in the U.S. labor market over the past six or more years, these states have used models based on the prerecession period to predict job seeker outcomes during the recession and for the postrecession period.

Many of the states that had not updated models since before the recession cited an inability to update due to a lack of resources. This was particularly the case for states that have no in-house statistical staff and those that had their existing models set up directly by USDOL. Nevertheless, when model coefficients are not updated, it increases the chances that the predictive model misallocates services away from those most in need. Indeed, the U.S. Government Accountability Office (2007) finds that not only were many states not updating their profiling models, but also that neither USDOL nor the states had conducted any recent study to evaluate whether assigning individuals to services based on the predictions of the profiling models was having any positive effects on UI claimants' outcomes. The studies that have been conducted (e.g., Black, Galdo, and Smith 2007; Black, Smith, Berger et al. 2003; and Black, Smith, Pleasca, et al. 2003), although they employ rigorous methodological designs, are using data from the 1990s. Without updated research, it is impossible to know whether the states' profiling models are having the desired effect of reducing the duration of unemployment.

Application

In this section, we present an application of predictive modeling to a substantive workforce problem, predicting which customers are likely to have difficulty finding employment and need more extensive services before falling into long-term unemployment. Although our application addresses a substantively important issue, we have selected this application to illustrate the predictive modeling *process*. In particular, we present three approaches that states can take to improve the accuracy of their predictions using three different predictive algorithms, use the results to show the importance of updating predictive models over time, and describe some steps for diagnosing problems with and improving a model's predictive accuracy.

In our application, we assess the predictive accuracy of three algorithms—logit, regularized regression, and neural network—encountered both in our survey of the states and in the statistical literature on predictive analytics.[12] These algorithms represent three different approaches that states can use to improve the accuracy of their predictive models. We present an example in which the predictive accuracy varies only slightly across the three models, in order to highlight a cautionary point for states acquiring data for predictive applications: big data and sophisticated statistical models are not enough to solve every problem. If the statistical model is a poor approximation of the real-life process (e.g., long-term unemployment) that is being modeled, then neither more data nor more complicated methods will greatly improve predictive accuracy. We discuss this issue in more detail below.

Data

We use two primary data sources from the state of New Jersey to construct the sample for this chapter: America's One-Stop Operating System (AOSOS) and UI Wage Record data. AOSOS records the enrollment of customers in the workforce system, their demographic characteristics, the services they receive, and their exit from the system. AOSOS also tracks the participation of workforce system customers in the three largest welfare programs that serve working-age adults: Temporary Assistance for Needy Families (TANF), Supplemental Nutrition Assistance Program (SNAP), and the General Assistance (GA) program, a state-funded program that serves adults without dependent children. The UI wage data system records the wages of all employees at employers that report wages every quarter in the course of paying their UI taxes.

Sample

The sample for this chapter consists of all individuals who interacted with a New Jersey AJC for the first time in 2012. However, we exclude certain groups of individuals from the sample when they differ significantly from other AJC customers both in how they enter and how they interact with the workforce system. Specifically, we remove individuals who had any interaction (in terms of application for benefits or receipt of benefits) with TANF, SNAP, or GA, as well as customers under the age of 25. For both welfare program recipients and youth customers, it is more appropriate to run a separate predictive model for these individuals. In order to highlight the usefulness of predictive models for smaller geographic units than the state-level, we limit the data to a single state workforce investment area. The results presented below are substantively similar when analyzing statewide data.

Predictors

The predictors for the model consisted of demographic characteristics that appeared in the AOSOS data and wage history variables constructed from the UI wage data. Although AOSOS has the capacity to accommodate the entry of hundreds of different job seeker attributes that could be significant predictors of labor market success, in practice a much more limited set of characteristics is available for most job seekers. These include sex, race/ethnicity, education level, and date of birth.

We create wage histories for each workforce system customer relative to their date of entry into the workforce system. The wage history consists of each customer's earnings in each of the 24 quarters prior to enrollment in the workforce system, except for the first 2 quarters prior to enrollment, as the six-month lag in the UI wage data means that these quantities would not be available for inclusion in a predictive model at the time a customer enrolled. We then created additional variables, including the total number of quarters worked in the past six years and the number of consecutive quarters the job seeker was employed before entering the workforce system.[13]

Comparison of predictive models

In the predictive models presented below, we operationalize long-term unemployment as collecting zero wages in the four quarters after a customer's initial AJC visit. We then compare the predictive accuracy of three competing models. When the outcome variable is dichotomous, one of the first classification methods that researchers apply is logistic regression, which often achieves high predictive accuracy. However, when the model includes few observations and many variables, some of which may be highly correlated with each other, a statistical problem called *overfitting* may reduce the model's accuracy on new data sets. When a model overfits, it is fitting the random noise in the data and not the underlying relationship between the variables, meaning that it is likely to perform poorly when called upon to make predictions on new data. Numerous and highly multicollinear variables are features of large administrative data sets in workforce development. Regularized regression models, such as the ridge and lasso, were developed to improve predictive accuracy in situations where models are overfitting the data. Thus, in addition to the logit model, we estimate a modified regression model called ridge regression (Hastie, Tibshirani, and Friedman 2009; Kuhn and Johnson 2013).

The third model we show is called a neural network, which at least one state uses for its worker profiling model. The chief advantage of the neural network is its ability to model complex relationships between the predictors and the outcome, which can lead to improved predictive accuracy when compared to competing models. States can implement a neural network, as well as the logit and the ridge regression, without a substantial investment in technical capacity. The models can be estimated using freely available and easy-to-use software such as R (discussed in the Predictive Analytics section on p. 452).

In estimating the models, we follow common practice in predictive analytics by splitting the customer data into three separate data sets: a training set, a test set, and a validation set. The reason we split the data involves choosing models that have high predictive accuracy on new observations. The danger of the overfitting phenomenon mentioned above is that the model estimates may have excellent predictive accuracy on the data set used in estimation while having poor predictive accuracy on any new data. A predictive model should not be assessed on how well it predicts outcomes on the data that were used to estimate the model, but rather on new data for which the outcomes are unavailable. For example, a model may perform well predicting outcomes on past One-Stop customers while poorly predicting outcomes on any new customers. Splitting the data set into a training, test, and validation set helps reduce the possibility that our models overfit the data and thus have poor predictive accuracy on new customers.

Specifically, we follow these four steps:

1) Estimate the logit, ridge, and neural network models on the training data
2) Assess the predictive accuracy of each model on the test data
3) Choose the logit, ridge, and neural network specification with the highest predictive accuracy on the test data[14]
4) Assess predictive accuracy of each model on the validation set to establish final benchmark model accuracy

In practice, a predictive model should produce at least higher predictive accuracy than an alternative strategy of using no model at all. For example, workforce agencies can simply classify all customers as likely to be unemployed. The predictive accuracy of this system will equal the average of the outcome variable for averages above 0.5 and 1 minus the average for values below 0.5. If 60 percent of customers in the data are unemployed, then this system would achieve a predictive accuracy of 60 percent, since it would classify all the 60 percent of unemployed individuals correctly and all of the 40 percent of employed individuals incorrectly. We call this system the *null model*. At a minimum, we want to choose predictive models that have higher predictive accuracy than the null model.

Note that we estimate and validate the model using 2012 customer data. The estimates thus reflect the most current data available for this application. However, as we found in our survey of the states' predictive modeling practices, some states are not updating their models with the most current data. For example, many states are using 2008 customer data to predict 2012 customer outcomes, despite the large differences in the labor market conditions and typical customer profiles across this period of time.

To illustrate the consequences of not updating predictive models, we follow the same steps as those listed above but train and test the models using data from 2008 only. With the estimates from the 2008 data, we measure predictive accuracy using the same 2012 validation set as that used above.

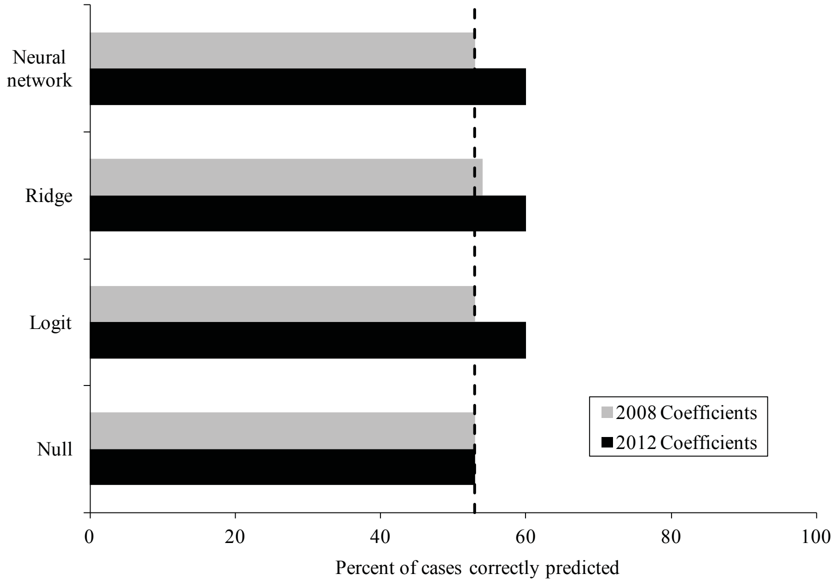
The results are shown in Figure 19.3. The black horizontal bars mark the predictive accuracy of the models that are fit to the 2012 data, with the bottom horizontal black bar representing the null model's predictive accuracy. The three models achieve similar predictive performance on the validation data. The logit, ridge, and neural network models correctly classify 60 percent of customers as experiencing a long spell of unemployment. Each model does significantly better than the null model, which features only 53 percent of customers correctly classified.

The grey horizontal bars in Figure 19.3 represent predictive accuracy for the models estimated using the 2008 data. Recall that the expectation is that the predictive accuracy of a model will decrease when the model's estimates are not updated with more current data. The results confirm our expectation. Across all three models, the predictive accuracy on the validation data is approximately equal to the accuracy of the null model. In other words, when we estimate models using older data, we achieve results no better than simply assuming every customer who enters an AJC will experience a long spell of unemployment.

A natural question to ask is why the performance of the three models is so similar. Why, in other words, do the more sophisticated ridge and neural network models provide little improvement over the logit model? The answer relates to the concepts of the bias and variance of a predictive model.

The variables included in the application we present are only weakly associated with the outcome variable of unemployment. These variables thus do a relatively poor job representing the complex process that leads individuals to experience long-term unemployment. This phenomenon—the failure of a model to be a good approximation of a real-life process—is called *bias*. Rather than overfitting the data, the logit model is underfitting, so the ridge regression offers little or no gain

Figure 19.3 Predictive Performance of Neural Network, Ridge, and Logit Models



over the logit. In addition, even the complex neural network is unable to model the complexity in the data in a manner superior to the logit and the ridge regression. The result is three models that perform similarly and achieve prediction accuracy at only about 60 percent when, ideally, the model should achieve much higher accuracy. This suggests that additional work needs to be done collecting not more data but more high-quality variables that are associated with the outcome of interest.

A crucial point about a high-bias model is that more data will not substantially improve predictive accuracy. Even when we expand our data set to include hundreds of thousands of additional observations, the results change little. Big data will help primarily when the model has an opposite problem called *high variance*. A high variance model features poor predictive accuracy on data that were not used to estimate the model. Generally, more data can reduce the variance of the model by reducing overfitting, but more data will not reduce its bias.[15] Bias reduction requires the inclusion of additional predictors in the model.

CONCLUSIONS

Building an intelligent workforce system requires high-quality data and the ability to mine insights from all types of data, not just numeric and categorical data, and to analyze that data as accurately as possible. Data science and low-cost software offer SWAs and local workforce areas a series of valuable tools for improving the labor market outcomes of AJC customers. When described using terms such as geospatial analysis, text mining, predictive analytics, or big data analytics, these models can appear new and intimidating. However, despite states' limited experience examining location and text data, the tools for mining these data for insights are within the capabilities of SWA research staff—possibly in collaboration with state university partners or private sector firms. Moreover, many states are already quite familiar with predictive modeling, as nearly every state already implements predictive models through their UI programs. While it is true that the field of predictive modeling offers a wide range of algorithms for predicting workforce outcomes, SWA staff do not need to understand their mathematical intricacies any more than they do the basic logistic regression models currently in use for worker profiling because existing statistical software does most of the heavy lifting.

What SWAs *do* need to ensure is the proper expertise in the application of location and text analysis and in predictive modeling. For location and text data, this requires identifying staff capacity internal to the SWA or available in other agencies of state or local government, universities, or the private sector. In the case of prediction, this may require some training for staff members who currently oversee worker profiling models or hiring an outside party to develop and implement a new predictive model, as setting models up the first time requires careful design and evaluation. But once the models are established, they need to be updated with new data only on an annual basis, which is a much less costly process. In short, while states will need to find resources to develop new models, these resources need not be extensive.

Beyond resource constraints, the much larger and more crucial impediments to an intelligent workforce system are data limitations. If address information is not updated regularly or textual data are collected only sporadically, then these potentially useful sources of information may not be available in an accurate or complete enough form to provide the type of value they could potentially provide. The chapter has also demonstrated that having a large quantity of data is not enough to produce highly accurate predictive models. The *quality* of workforce data is just as important. In order to fully leverage the power of location-based analyses, text analysis, and predictive models, SWAs need not only a large number of observations but also a multitude of variables that are related to workforce outcomes. In the current state of the workforce data, these variables are often not available because state agencies silo their data into separate systems. Furthermore, states often only collect the bare minimum of variables necessary to meet federal reporting requirements.

Data quality is an area where the workforce system needs to strive for improvement, and to some extent this process has already begun. The need for high-quality data is becoming more apparent to public officials, and a limited number of projects are under way at all levels of government to foster improvements in data quality. For instance, the USDOL Workforce Data Quality Initiative has provided grants to 32 states to integrate administrative data systems, breaking down silos and providing the diversity and number of variables that make accurate predictive modeling possible. Other examples of data integration projects include the Workforce Innovation Fund projects in Chicago and Newark, as well as recent efforts to create a federal workforce data system.

In order to derive insights from location and textual data and develop accurate predictive models, the collection of high-quality workforce data must begin now, and an intelligent workforce system should look beyond data integration to further improve the quality of workforce data. For instance, a key component of data quality is data completeness, and in our experience performance metrics have had a significant effect on which fields of data are the most thoroughly recorded and least missing. Those that are required for SWAs to meet their federal reporting requirements are the fields that are the most complete. Data quality improvements may therefore depend on how the federal system holds states and local areas accountable. A system that genuinely incentivizes states, local areas, and workforce counselors to collect and record a greater variety of data elements may be the essential first step to building a truly intelligent workforce system. SWAs can also take other steps to improve data quality, such as designing new customer intake procedures that collect additional variables and provide training to ensure consistent data entry across AJCs. Location data, text data, and predictive models hold much promise for the future of workforce development, and states can capture the benefits that these models provide only by improving data collection in the workforce investment system.

Notes

1. More extensive reviews of the field can be found in Grimmer and Stewart (2013) and Witten (2005).
2. Generally, before any analysis begins, a researcher must preprocess the text for analysis. This step usually involves stemming words, removing punctuation and common stop words such as *the* and *than*, removing numbers, and converting words to lower case. Analysts often apply a weighting scheme to words, such as *tf-idf* weights.
3. For a detailed R example as the count-based method and word associations, see Feinerer, Hornik, and Meyer (2008).
4. Monroe, Colaresi, and Quinn (2008) and Taddy (2013) discuss methods for estimating words that are predictive of category or group labels.
5. At alternative strategy is to look at specific aspects and assess their overall positivity and negativity (Liu 2010). Our research question here is focused on understanding aspects that factor into respondents' overall evaluation of the program rather than understanding variation in ratings across different services.
6. For a discussion of more advanced analyses of open-ended survey items, see Roberts et al. (2014). The result presented here is consistent with the informational negativity effect in psychology whereby individuals are better able to identify more precise justifications to support a negative reaction than a positive one (Lewicka 1997; Peeters and Czapinski 1990).
7. An algorithm is a step-by-step process for making a calculation.
8. A model is a mathematical equation that expresses a formal relationship between variables. In the case of predictive modeling, the model expresses the mathematical relationship between the predictors and the outcome being predicted.
9. To prevent discrimination, federal laws and regulations may prohibit the inclusion of some personal characteristics, such as age, race, sex, and disability status, in models that automatically assign individuals to services.
10. There are many different criteria that a researcher may use to guide her choice of the "best" model. For classification problems where the dependent variable is not skewed, accuracy is a good model evaluation parameter, as is the area under the receiver operating characteristic (ROC) curve. With a skewed dependent variable, it may be necessary to use other metrics, such as precision, recall, the F-score, etc. For models that predict continuous outcomes, the researcher might compare models based on their root mean squared error. For a detailed analysis of model evaluation, see Japkowicz and Shah (2014).

11. USDOL prohibits states from including some personal characteristics, including age, race, sex, and disability status, from the worker profiling model.
12. We define predictive accuracy as the percent of customers that the model accurately predicts as remaining unemployed.
13. We have no data on individuals who earn wages outside New Jersey. In an effort to partly mitigate the out-of-state employment problem, we delete from our list customers without any recorded education or employment history in New Jersey. Of course, this also removes weaker job seekers who are living in New Jersey but have poor employment histories. The results presented here are substantively similar to the results we obtain when we include those individuals.
14. We choose the ridge regression regularization penalty and the neural network decay parameter and node size to optimize predictive accuracy on the test data.
15. For more information about diagnosing bias and variance, see the concept of learning curves in the statistics and machine learning literature.

References

Black, Dan A., Jose Galdo, and Jeffrey A. Smith. 2007. "Evaluating the Worker Profiling and Reemployment Services System Using a Regression Discontinuity Approach." *American Economic Review* 97(2): 104–107.

Black, Dan. A., Jeffrey A. Smith, Mark C. Berger, and Brett J. Noel. 2003. "Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System." *American Economic Review* 94(4): 1313–1327.

Black, Dan A., Jeffrey A. Smith, Miana Plesca, and Shannon Seitz. 2003. *Profiling UI Claimants to Allocate Reemployment Services: Evidence and Recommendations for States: Final Report*. Washington, DC: U.S. Department of Labor, Employment and Training Administration.

Blei, David M. 2012. "Probabalistic Topic Models." *Communications of the ACM* 55(4): 77–84.

Burrough, Peter A., and Rachael A. McDonnell. 1998. *Principles of Geographical Information Systems*. New York: Oxford University Press.

Eberts, Randall W., and Christopher J. O'Leary. 2002. "A Frontline Decision Support System for Georgia Career Centers." Upjohn Institute Staff Working Paper No. 02-84. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Feinerer, Ingo, Kurt Hornik, and David Meyer. 2008. "Text Mining Infrastructure in R." *Journal of Statistical Software* 25(5): 1–55.

Finlay, Steven. 2014. *Predictive Analytics, Data Mining, and Big Data: Myths, Misconceptions, and Methods*. New York: Palgrave Macmillan.

Government Accountability Office. 2007. *Unemployment Insurance: More Guidance and Evaluation of Worker-Profiling Initiative Could Help Improve State Efforts*. GAO-07-680. Washington, DC: GAO.

Grimmer, Justin, and Gary King. 2011. “General Purpose Computer-Assisted Clustering and Conceptualization.” *Proceedings of the National Academy of Sciences* 108(7): 2643–2650. Washington, DC: The Academy.

Grimmer, Justin, and Brandon M. Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21(3): 267–297.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2d ed. New York: Springer.

Japkowicz, N., and M. Shah. 2014. *Evaluating Learning Algorithms: A Classification Perspective*. New York: Cambridge University Press.

Kuhn, Max, and Kjell Johnson. 2013. *Applied Predictive Modeling*. New York: Springer.

Lewicka, Maria. 1997. “Is Hate Wiser than Love? Cognitive and Emotional Utilities in Decision Making.” In *Decision Making: Cognitive Models and Explanations*. New York: Routledge, pp. 90–108.

Liu, Bing. 2010. “Sentiment Analysis and Subjectivity.” In *Handbook of Natural Language Processing*. 2d ed. Nitin Indurkhya and Fred J. Damerau, eds. London: Chapman and Hall/CRC, pp. 627–666.

Loughran, Tim, and Bill McDonald. 2011. “When Is a Liability Not a Liability?” *Journal of Finance* 66(1): 35–65.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn. 2008. “Fightin’ Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict.” *Political Analysis* 16(4): 372–403.

Neuendorf, Kimberly A. 2002. *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications.

Peeters, Guido, and Janusz Czapinski. 1990. Positive-Negative Asymmetry in Evaluations: The Distinction between Affective and Informational Negativity Effects. In *European Review of Social Psychology*. Vol. 1. Chichester, UK: Wiley.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana K. Gadarian, Bethany Albertson, and David G. Rand. 2014. “Structural Topic Models for Open-Ended Survey Responses.” *American Journal of Political Science* 58(4): 1064–1082.

Schutt, Rachel, and Cathy O’Neil. 2014. *Doing Data Science*. Sebastopol, CA: O’Reilly Media.

Sebastiani, Fabrizio. 2002. “Machine Learning in Automated Text Categorization.” *ACM Computing Surveys* 34(1): 1–47.

Siegel, Eric. 2013. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie or Die*. Hoboken, NJ: John Wiley and Sons.

Sullivan, William F., Lester Coffey, Lisa Kolovich, Charles W. McGlew, Douglas Sanford, and R. Sullivan. 2007. *Worker Profiling and Reemployment Services Evaluation of State Worker Profiling Models: Final Report.* Washington, DC: U.S. Department of Labor, Employment and Training Administration.

Taddy, Matt. 2013. "Multinomial Inverse Regression for Text Analysis." *Journal of the American Statistical Association* 108(503): 755–770.

U.S. Census Bureau. 1994. *Geographic Areas Reference Manual*. Washington, DC: U.S. Department of Commerce.

U.S. Department of Labor (USDOL). 2000. Worker Profiling and Reemployment Services. http://www.doleta.gov/programs/wprs.cfm (accessed September 4, 2014).

Wandner, Stephen A. 1997. "Early Reemployment for Dislocated Workers in the United States." *International Social Security Review* 50(4): 95–112.

The White House. 2014. *Ready to Work: Job-Driven Training and American Opportunity*. Washington, DC: The White House.

Witten, Ian H.. 2005. "Text Mining." In *Practical Handbook of Internet Computing*, Munindar P. Singh, ed. Boca Raton, FL: Chapman and Hall/CRC, pp. 14.1–14.22.

Worker Profiling and Reemployment Services Policy Workgroup. 1999. *Worker Profiling and Reemployment Services Policy Workgroup: Final Report and Recommendations*. Washington, DC: U.S. Department of Labor, Employment and Training Administration.