Executive report

Improving tax administration with data mining

Daniele Micci-Barreca, PhD, and Satheesh Ramachandran, PhD
Elite Analytics, LLC

Table of contents

Introduction

Both federal and state tax administration agencies must use their limited resources to achieve maximal taxpayer compliance. The purpose of this white paper is to show how data mining helps tax agencies achieve compliance goals and improve operating efficiency using their existing resources. The paper begins with an overview of data mining and then details an actual tax compliance application implemented by Elite Analytics, LLC, a data mining consulting services firm and SPSS Inc. partner, for the Audit Division of the Texas Comptroller of Public Accounts (CPA). The paper concludes with an overview of additional applications of data mining technology in the area of tax compliance.

Tax agencies primarily use audits to ensure compliance with tax laws and maintain associated revenue streams. Audits indirectly drive voluntary compliance and directly generate additional tax collections, both of which help tax agencies reduce the "tax gap" between the tax owed and the amount collected. Audits, therefore, are critical to enforcing tax laws and helping tax agencies achieve revenue objectives, ensuring the fiscal health of the country and individual states.

Managing an effective auditing organization involves many decisions. What is the best audit selection strategy or combination of strategies? Should it be based on reported tax amounts or on the industry type? How should agencies allocate audit resources among different tax types? Some tax types may yield greater per-audit adjustments. Others may be associated with a higher incidence of noncompliance. An audit is a process with many progressive stages, from audit selection and assignment to hearings, adjudication, and negotiation, to collection and, in some cases, enforcement. Each stage involves decisions that can increase or reduce the efficiency of the overall auditing program.

Audit selection methods range from simple random selection to more complex rule-based selection based on "audit flags," to sophisticated statistical and data mining selection techniques. Selection strategies can vary by tax type, and even within a single type. Certain selection strategies, for example, segment taxpayers within a specific tax type, and then apply different selection rules to each segment. Texas categorizes taxpayers that account for the top 65 percent of the state's sales tax collections as Priority One accounts, and audits these accounts every four years. Texas also audits Prior Productive taxpayers—accounts that yielded tax adjustments greater than $10,000 in previous audits.

As in many states, Texas' taxpayer population has risen steadily over the last decade, without any proportionate rise in auditing resources. As a result, Texas and several other states, as well as tax agencies in the United Kingdom and Australia, rely on data mining to help find delinquent taxpayers and make effective resource allocation decisions. Data mining leverages specialized data warehousing systems that integrate internal and external data sources to enable a variety of applications, from trend analysis to non-compliance detection and revenue forecasting, that help agencies answer questions such as:

- How should we split auditing resources among tax types?
- Which taxpayers are higher audit priorities?
- What is the expected yield from a particular audit type?
- Which SIC codes are associated with higher rates of noncompliance?

Why data mining?

Tax agencies have access to enormous amounts of taxpayer data. Most auditing agencies, in fact, draw information from these data sources to support auditing functions. Audit selectors, for example, search data sources for taxpayers with specific profiles. These profiles, developed by experts, may be based on a single attribute, such the taxpayer industry code (SIC), or on a complex combination of attributes (for example, taxpayers in a specific retail sector that have a specific sales-to-reported-tax ratio). Data mining technologies do the same thing, but on a much bigger scale. Using data mining techniques, tax agencies can analyze data from hundreds of thousands of taxpayers to identify common attributes and then create profiles that represent different types of activity. Agencies, for example, can create profiles of high-yield returns, so auditors can concentrate resources on new returns with similar attributes. Data mining enables organizations to leverage their data to understand, analyze, and predict noncompliant behavior.

What is data mining?

Data mining is the exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules [1]. Organizations use this information to detect existing fraud and noncompliance, and to prevent future occurrences.

Data mining also enables data exploration and analysis without any specific hypothesis in mind, as opposed to traditional statistical analysis, in which experiments are designed around a particular hypothesis. While this openness adds a strong exploratory aspect to data mining projects, it also requires that organizations use a systematic approach in order to achieve usable results. The CRoss-Industry Standard Process for Data Mining (CRISP-DM) (see **Figure 1**) was developed in 1996 by a consortium of data mining consultants and technology specialists that included SPSS, Daimler-Benz (now DaimlerChrysler) and NCR [2]. CRISP-DM's developers relied on their real-world experience to develop a six-phase process that incorporates an organization's business goals and knowledge. CRISP-DM is considered the *de facto* standard for the data mining industry.

Figure 1

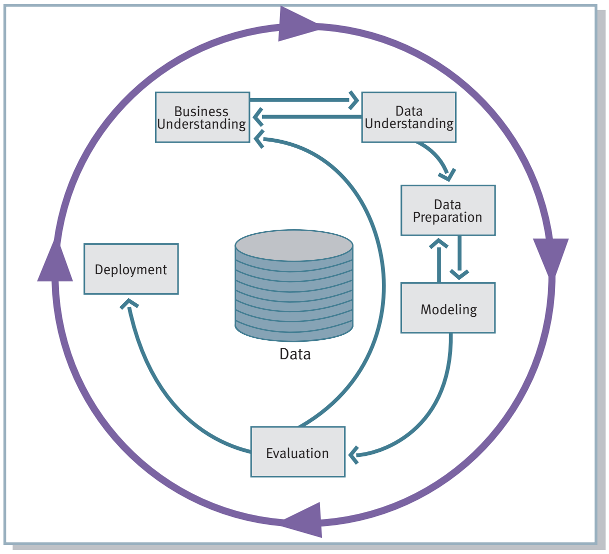

Figure 1: The CRoss-Industry Standard Process for Data Mining (CRISP-DM)

The six phases of CRISP-DM

Business understanding

The first phase ensures that all participants understand the project goals from a business or organizational perspective. These business goals are then incorporated in a data mining problem definition and detailed project plan. For a tax auditing agency, this would involve understanding the audit management process, the role and functions performed, the information that is collected and managed, and the specific challenges to improving audit efficiency. This information would be incorporated into the data mining problem definition and project plan.

Data understanding

The second phase is designed to assess the sources, quality, and characteristics of the data. This initial exploration can also provide insights that help to focus the project. The result is a detailed understanding of the key data elements that will be used to build models. This phase can be time-consuming for tax agencies that have many data sources, but it is critically important to the project.

Data preparation

The next phase involves placing the data in a format suitable for building models. The analyst uses the business objectives determined in the business understanding step to determine which data types and data mining algorithms to use. This phase also resolves data issues uncovered in the data understanding phase, such as missing data.

Modeling

The modeling phase involves building the data mining algorithms that extract the knowledge from the data. There are a variety of data mining techniques; each is suitable for discovering a specific type of knowledge. A tax agency would use classification or regression models, for example, to discover the characteristics of more productive tax audits. Each technique requires specific types of data, which may require a return to the data preparation phase. The modeling phase produces a model or a set of models containing the discovered knowledge in an appropriate format.

Evaluation

This phase focuses on evaluating the quality of the model or models. Data mining algorithms can uncover an unlimited number of patterns; many of these, however, may be meaningless. This phase helps determine which models are useful in terms of achieving the project's business objectives. In the context of audit selection, a predictive model for audit outcome would be assessed against a benchmark set of historical audits for which the outcome is known.

Deployment

In the deployment phase, the organization incorporates the data mining results into the day-to-day decision-making process. Depending on the significance of the results, this may require only minor modifications, or it may necessitate a major reengineering of processes and decision-support systems. The deployment phase also involves creating a repeatable process for model enhancements or recalibrations. Tax laws, for example, are likely to change over time. Analysts need a standard process for updating the models accordingly and deploying new results.

The appropriate presentation of results ensures that decision makers actually use the information. This can be as simple as creating a report or as complex as implementing a repeatable data mining process across the enterprise. It is important that project managers understand from the beginning what actions they will need to take in order to make use of the final models.

The six phases described in this paper are integral to every data mining project. Though each phase is important, the sequence is not rigid; certain projects may require you to move back and forth between phases. The next phase or the next task in a phase depends on the outcome of each of the previous phases. The inner arrows in **Figure 1** indicate the most important and frequent dependencies between phases. The outer circle symbolizes the cyclical nature of data mining projects, namely that lessons learned during a data mining project and after deployment can trigger new, more focused business questions. Subsequent data mining projects, therefore, benefit from experience gained in previous ones.

Eighty to 90 percent of a typical data mining project is spent on phases other than modeling, and the success of the models depends heavily on work performed in these phases. To create the models, the analyst typically uses a collection of techniques and tools. Data mining techniques come from a variety of disciplines, including machine learning, statistical analysis, pattern recognition, signal processing, evolutionary computation, and pattern visualization. A detailed discussion of these methods is beyond the scope of this paper. The next sections, however, discuss data mining methods relevant to tax audit applications.

Case study: An audit selection strategy for the State of Texas

As previously mentioned, there are many audit selection methods, each of which results in different levels of productivity. Measured in terms of dollars recovered per audit hour, the relationship between productivity and the audit selection strategy is quantifiable. This section focuses on the Audit Select scoring system implemented by Elite Analytics for the Audit Division of the Texas Comptroller of Public Accounts, and how this approach compares to traditional audit selection strategies used by the division.

Already the data mining tool of choice for the Audit Division, Clementine was used throughout the implementation of the new audit selection strategy. According to Elite Analytics, “Not only did we find the variety of data preparation, modeling, and deployment features unmatched in any other product, but the Clementine workbench made working and categorizing completed work within the CRISP-DM model very intuitive.”

Predictive modeling

Predictive modeling is one of the most frequently used forms of data mining. As its name suggests, predictive modeling enables organizations to predict the outcome of a given process and use this insight to achieve a desired outcome. In terms of audit selection, the goal is to predict which audit leads are more likely to yield greater tax adjustments.

Predictive models typically produce a numeric score that indicates the likely outcome of an audit. A high score, for instance, indicates that the tax adjustment from that audit is likely to be high or above average, while a low score indicates a low likelihood of a large tax adjustment. A real-world example is the Discriminant Index Function (DIF) score developed by the IRS to identify returns with a high probability of unreported income. Use of the DIF score results in significantly higher tax assessments than do purely random audits [3].

Scoring models have become more popular, due in large part to their ability to manage hundreds of variables and large populations. Among other applications, organizations use scoring models to:

- Assess credit risk
- Identify fraudulent credit card transactions
- Determine the response potential of mailing lists
- Rank prospects in terms of buying potential

Using models to predict sales tax audit outcomes

The Audit Division's Advanced Database System (ADS) is a data warehouse designed to support tax compliance applications such as the Audit Select system, which uses predictive models to identify sales tax audit leads. **Figure 2** depicts the process of building and using predictive models for audit selection.

Figure 2

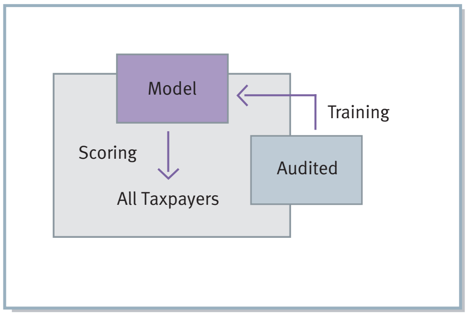

Figure 2: Model training and scoring

The first step is to calibrate or train the model using a training set of historical audit data with a known outcome. This enables the model to "learn" the relationship between taxpayer attributes and the audit outcomes. The Audit Select scoring model uses the following five sources of data to create a taxpayer profile (see **Figure 3**):

- Business information, such as taxpayer SIC code, business type (corporation, partnership, etc.), and location
- Sales tax filings from the most recent four years of sales tax reports
- Other tax filings, primarily for franchise tax
- Employee and wage information reported to another state agency
- Prior audit outcomes

Figure 3

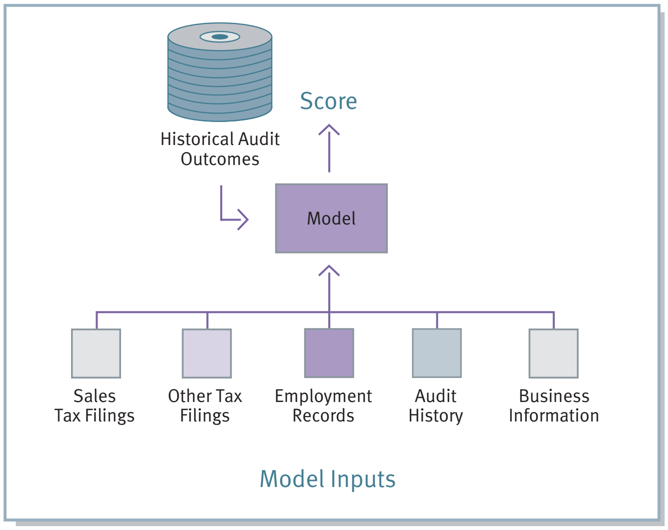

Figure 3: Data sources for the Audit Select scoring model

Once trained, the model can be applied to the entire taxpayer population in a process known as generalization. The model generalizes what it learns from historical audits to analyze new returns and assign Audit Select scores (see examples in **Figure 4**). Human audit selectors then use the Audit Select scores to determine which businesses to audit.

The model relies on extensive data preparation and transformations that map the raw data into more informative indicators. For example, while the actual and relative (compared to gross sales) amount of deduction is a good raw indicator, it is useful to compare the figures to those of similar businesses. This peer group comparison is used throughout the model to develop additional indicators.

Figure 4

Training Set of Historical Audits

Gross Sales	SIC Group	Wages	Receipts	Tax Adjustment
$21,110,288	23	$34,456,345	$988,945	$100,202
$34,234,334	43	$11,476,544	$2,545,251	$434,323
$9,874,556	23	$45,443,343	$4,534,521	$0
$33,421,655	56	$45,433	$4,354,353	$454,352
$254,667,678	55	$445,453	$43,657,176	$82,834

Potential Audits

Gross Sales	SIC Group	Wages	Receipts	Predictive Score
$424,454,762	43	$5,454,362	$22,571,243	760
$44,572,462	32	$45,445,623	$45,653,235	450

Figure 4: Training and scoring examples

In **Figure 4**, the first table represents the training set of historical audits used to calibrate the model. The second table shows the model's predictive scores for the current population of taxpayers.

Results from the Audit Select scoring system

The most effective way to assess a scoring system is to compare the scores it produces to actual outcomes. In the context of audit selection, the goal is to measure the difference between the final tax adjustment and the score assigned by the model.

Figure 5

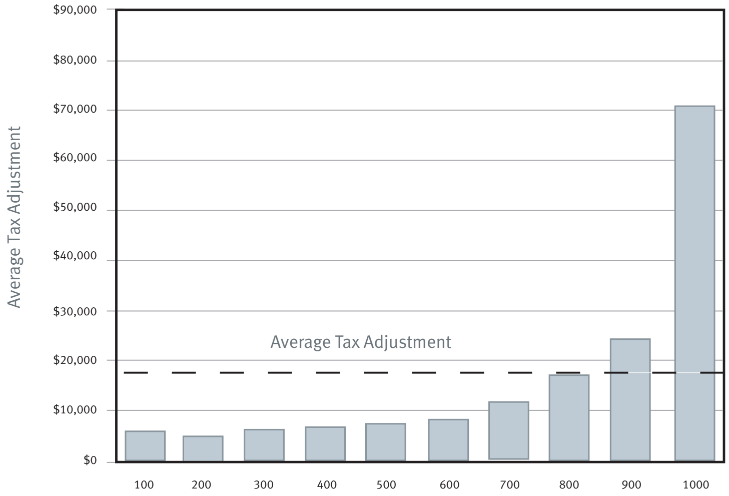

Figure 5: Average final adjustment by score for sales tax audits performed between 1996 and 2002

The chart in **Figure 5** shows the average tax adjustment for Audit Select scores between one and 1000. The average tax adjustment, as expected, increases in proportion to the scores. The chart shows, for example, that audits performed on taxpayers with scores below 100 resulted in an average tax adjustment of $3,300. In contrast, audits performed on taxpayers with scores above 900 resulted in an average tax adjustment of $78,000. While the percentage of taxpayers with scores higher than 900 is smaller than the percentage with scores below 100, additional research in this study revealed that only 57 percent of the current population scoring 900 or greater has been audited at least once. This taxpayer segment clearly represents a significant pool of audit candidates.

Comparison with other selection strategies

As demonstrated in the examples above, the data mining and predictive modeling techniques used by the Audit Select scoring system add a new level of sophistication to the audit selection process. The most accurate measure of a new technology, however, is not technical complexity or theoretical soundness, but the degree to which it improves an existing process. This measurement is of particular importance when less sophisticated, yet well-tested techniques are already in use.

Other selection strategies

Before evaluating the performance of the Audit Select scoring system, therefore, it is important to review the selection strategies that the State of Texas used for many years and continues to use alongside the new system.

- **Priority One**: The State of Texas classifies all businesses that contribute to the top 65 percent of tax dollars collected as Priority One taxpayers. The state audits these taxpayers approximately every four years. The Priority One selection strategy uses a very simple scoring and ranking algorithm. The score is the relative percentage of sales tax dollars contributed to the state's total collection. The state ranks taxpayers by score and then applies a threshold to produce the target list. There are similarities between the Priority One and Audit Select selection strategies, but the main difference is the logic behind the score. The Priority One score is one-dimensional; it considers only the reported tax amount. The Audit Select score, on the other hand, takes into account a number of taxpayer attributes (as depicted in **Figure 3**).

- **Prior Productive**: This strategy automatically selects taxpayers with prior audit adjustments of more than $10,000. As with Priority One, the Prior Productive strategy uses only one selection criterion, which is one of several used by the Audit Select system.

The average tax adjustment for Priority One audits is approximately $76,000 (median $9,600). This outcome is significantly higher than the $12,000 (median $1,300) average for non-Priority One audits and the $18,000 (median $1,600) average for all sales tax audits. Priority One audits also represent approximately nine percent of all Texas sales tax audits.

Audit Select versus Priority One

When comparing selection methods, it is important to consider not only the average tax adjustment of each method, but also the volume of audits. This is particularly important when assessing a score-based solution such as the Audit Select system. In fact, while the score enables ranking of potential targets, each score range produces a different number of candidates. When comparing the Priority One criterion with the score criterion, it is therefore important to use a score range that produces an equal number of audits. In this case, we compare the average outcome of the top nine percent of all audits, based on the score, with the average outcome of Priority One audits (which represent nine percent of the audits).

Based on scores produced by the Audit Select model, the average tax adjustment for the top nine percent of audits is approximately $88,000 (median $16,000) (see **Figure 6**). This is a 16 percent improvement over the Priority One average adjustment (65 percent improvement over the median adjustment) on an equal number of audits. Note that 36 percent of the top nine percent of the Audit Select audits is made up of Priority One audits, while the remaining 64 percent would not have been selected by the Priority One strategy. This demonstrates that the Audit Select strategy is not necessarily orthogonal (or in contrast) to Priority One. The Audit Select score can, in fact, be used to improve on the Priority One strategy by further segmenting the Priority One population and eliminating audits likely to result in small tax adjustments.

Figure 6

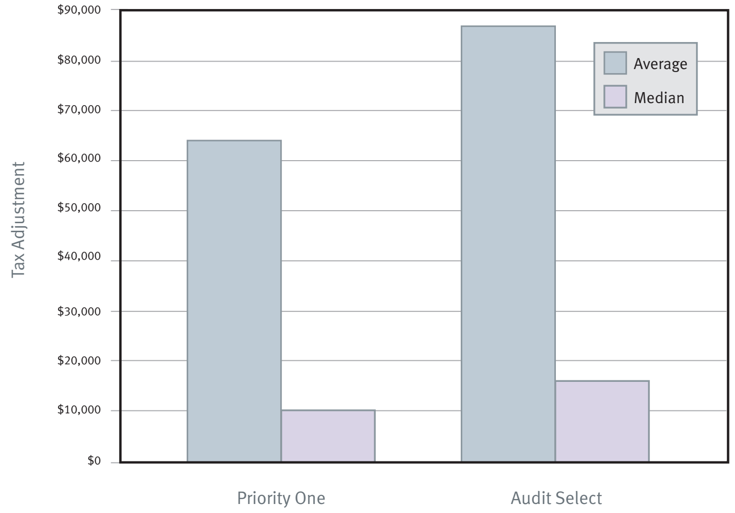

Figure 6: Comparison of Priority One and Audit Select tax adjustment results

Audit Select versus Prior Productive

As with the previous comparison, there are significant differences in the results obtained by the Audit Select and Prior Productive selection strategies (see **Figure 7**). Approximately 20 percent of the audits performed fall within the Prior Productive selection criteria. Five percent of these, however, would also be selected by the Priority One strategy. Excluding those cases, the average tax adjustment for Prior Productive audits was $17,700 (median $4,100). In comparison, the top 20 percent of the audits, based on the Audit Select score (and excluding potential Priority One audits), shows an average adjustment of $27,000 (median $6,500).

Figure 7

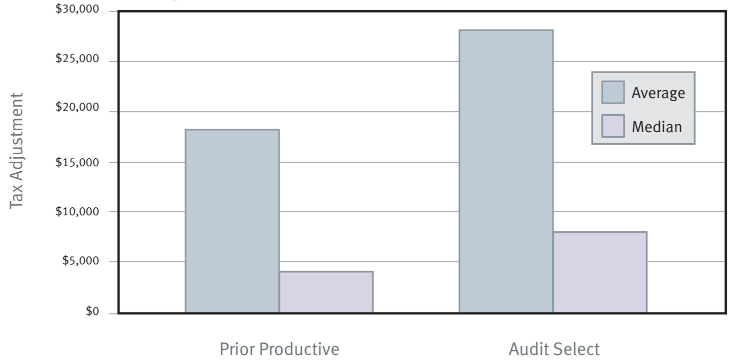

Figure 7: Comparison of Prior Productive and Audit Select tax adjustment results

Model evolution

Data mining has the most impact when it becomes an integral part of any system or application. Data mining models are rarely static. As more data sources are available and processes change, the models evolve. User feedback is also a critical component in model evolution.

Since its implementation, the Audit Select scoring model has evolved significantly, and additional enhancements are currently in progress. Several factors contribute to the evolution of a system of this kind. First and foremost is the evolution of the data sources. The usefulness and accuracy of any model is determined primarily by the quality, quantity, and richness of the data available. The second factor is domain knowledge, which itself evolves to reflect emerging trends and changes in the data generation process. The final factor is technological improvements: More sophisticated modeling algorithms, data transformation strategies, and model segmentation techniques enable significantly improved end results.

Other data mining applications in tax administration

Audit selection is one of many possible data mining applications in tax administration. In the area of tax compliance alone, tax collectors face diverse challenges, from underreporting to nonreporting to enforcement. Data mining offers many valuable techniques for increasing the efficiency and success rate of tax collection. Data mining also leverages the variety and quantity of internal and external data sources available in modern data warehousing systems. Here are just a few of the many potential uses of data mining in the context of tax compliance:

Outlier-based selection

This is an alternative approach to audit selection that uses specific data mining algorithms to build profiles of typical behaviors and then select taxpayers that do not match the profiles. This modeling process involves creating valid taxpayer segments, characterizing the normative taxpayer profile for each segment, and creating the rules for outlier detection. Though this method is similar to that applied by many audit selectors on a daily basis, the added benefit of the data mining approach is its ability to process large amounts of data and analyze multiple taxpayer characteristics simultaneously.

Lead prioritization

Many tax agencies flag potential nonfilers by, for example, cross-matching external data with internal lists of current filers. This type of application often produces many leads, each of which must be manually verified and pursued. Given the limited staff available to follow up on leads, it is important to develop criteria for prioritizing the leads. Data mining, however, can predict the tax dollars owed by an organization or individual taxpayer by modeling the relationship between the attributes and reported tax for known taxpayers.

Profiling for cross-tax affinity

Agencies can use data mining to analyze existing taxpayers for associations between the types of businesses and the tax types (more than 30 in Texas) for which the taxpayers file. Co-occurrence of certain tax types may infer liability for another tax type for which the taxpayer did not file.

Workflow analysis

Following up on leads can be a lengthy process consisting of multiple mail exchanges between the tax agency and the organizations classified as potential nonfilers. Tracking this process, however, generates data that is useful for modeling the process. The process models can then be used to predict the value of current pipelines and to optimize resource allocation.

Anomaly detection

Taxpayers self-report several important attributes (SIC, organization type, etc.). Due to unavoidable data entry errors, however, some taxpayers are categorized incorrectly or even uncategorized. Since these attributes and categories drive audit selection and other processes, it is always a good idea to apply data mining's rule induction techniques to detect errors and anomalies.

Economic models for optimal targeting

In most traditional targeting applications, such as target marketing or e-commerce fraud detection, there is a tradeoff between the number of audits to target and the cost per audit. When audit cost information is available, agencies can use data mining to develop targeting strategies that maximize overall collections.

Conclusion

As outlined in this paper, data mining has many existing and potential applications in tax administration. In the case of the Audit Division of the Texas Comptroller of Public Accounts, predictive modeling enables the agency to identify noncompliant taxpayers more efficiently and effectively, and to focus auditing resources on the accounts most likely to produce positive tax adjustments. This helps the agency make better use of its human and other resources, and minimizes the financial burden on compliant taxpayers. Data mining also helps the agency refine its traditional audit selection strategies to produce more accurate results.

About Clementine

Clementine is a data mining workbench that enables organizations to quickly develop predictive models using business expertise, and then deploy them to improve decision making. Clementine is widely regarded as the leading data mining workbench because it delivers the maximum return on data investment in the minimum amount of time. Unlike other data mining workbenches, which focus merely on models for enhancing performance, Clementine supports the entire data mining process to reduce time-to-solution. And Clementine is designed around the *de facto* industry standard for data mining—CRISP-DM. CRISP-DM makes data mining a business process by focusing data mining technology on solving specific business problems.

About SPSS Inc.

SPSS Inc. [NASDAQ: SPSS] is the world's leading provider of predictive analytics software and services. The company's predictive analytics technology connects data to effective action by drawing reliable conclusions about current conditions and future events. More than 250,000 commercial, academic, and public sector organizations rely on SPSS technology to help increase revenue, reduce costs, improve processes, and detect and prevent fraud. Founded in 1968, SPSS is headquartered in Chicago, Illinois. To learn more, please visit **www.spss.com**. For SPSS office locations and telephone numbers, go to **www.spss.com/worldwide**.

About Elite Analytics, LLC

Elite Analytics, LLC is a consulting services firm based in Austin, Texas, that focuses on providing a range of analytical consulting solutions for organizations in the area of noncompliance. Elite Analytics helps organizations leverage the power of data-driven analytical methods to help achieve greater efficiencies in compliance management functions. For more information, visit **www.eliteanalytics.com**.

Acknowledgements

The authors would like to acknowledge the insights and domain expertise provided by Rose Orozco, Lisa McCormack, and Eddie Coats at the Audit Division of the Texas Comptroller of Public Accounts during the course of the data mining project.

References

1. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., eds. (1996), *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press.
2. CRISP-DM Web site: www.crisp-dm.org
3. Alm, J., (1999). “Tax Compliance and Administration,” in *Handbook on Taxation*; eds. Hildreth, W. B., Richardson, J. A., pp. 741-768. Marcel Dekker, Inc.

To learn more, please visit www.spss.com. For SPSS office locations and telephone numbers, go to www.spss.com/worldwide.