F I S C A L   A F F A I R S   D E P A R T M E N T

For Official Use Only

Republic of Slovenia
Data-Driven Risk Assessment

Ludvig Wier and Shan Zhong

Technical Report

December 2022

Funded by
the European Union

DM 7316699

This technical assistance (TA) was provided with financial support from the European Union.

GLOSSARY

CD	Capacity Development
CRM	Compliance Risk Management
CRMU	Compliance Risk Management Unit
CIP	Compliance Improvement Plan
CIT	Corporate Income Tax
DG	Director General
MNE	Multinational Enterprise
OECD	Organisation for Economic Co-operation and Development
PIT	Personal Income Tax
SFA	Republic of Slovenia Financial Administration
SFO	Special Financial Office
TP	Transfer Pricing
VAT	Value-Added Tax

I. INTRODUCTION

1. **This mission aimed at supporting the Slovenia Financial Administration (SFA) to advance their use of data analytics in risk assessment for corporate income tax (CIT).** Over the course of the mission the team conducted four workshops and five one-to-one sessions to provide guidance on key topical areas, including:

- Ideal organizational arrangements to support data analytics
- How to access and utilize better and more data, including third party data
- Current data workflows and processes with a view to streamlining and improving them
- How to improve data integrity
- Using data science and big data analytics to strengthen risk assessment
- Leveraging the value of country-by-country reports
- Developing several data analytics/risk assessment models

2. **The mission team found that SFA has high-quality data and competent staff to benefit from the yields of data analytics.** The mission team was impressed by the engagement and stamina of the dedicated staff participating in the workshops. Their commitment was demonstrated over the course of this mission by building four new pilot risk assessment models which all showed promising results. E.g., the CIT audit selection model was estimated to increase the strike rate in comprehensive audits (measured as a correction above €5000) from 50-60 percent to roughly 90 percent (while doing the same number of audits).

3. **For SFA to achieve the full benefits of their prior investments in data capabilities, senior management are encouraged to deliver on five actions**: a) fully endorse, prioritize and incentivize the journey towards relying more on data analytics b) allow time for data analytics to grow (sometimes fail) and keep a stable and dedicated team in place c) invest in further recruitment and retention of data analysts and retrain operational staff to join the data analytics team d) prioritize development of high-value projects (deprioritize low-value projects) and vigilantly implement them e) have a constant focus on business needs and rigorously evaluate the success of projects.

II. SFA ACHIEVEMENTS IN ADVANCED DATA ANALYTICS

4. **The mission team found that the SFA has already laid the foundation to reap the benefits of advanced data analytics.** Their established "analysis unit" has data readily available due to timely investments in a data cleaning, manipulation, and storage. The use of data across the SFA is

already widespread with more than 50 data applications in use while an advanced predictive model is under development in the IT department.

5. With an abundance of data available, next steps are to utilize data sources more systematically. In a matter of days, the mission team was able to get high-quality data on audits, tax returns, Country-by-Country-Reports, business risk rules and balance sheets of firms operating in Slovenia. All data easily merged using unique taxpayer identifiers. This information is the bedrock of data analytics and SFA is on solid ground here. The main job ahead is therefore to utilize the existing data in a systematic manner. That is, to automatically identify risks based on systematic data-driven models utilizing all the available data. Especially for the CIT risk assessments, there is scope for further reliance on this type of systematic analysis.

6. The steps needed to expand the usage of systematic risk analysis are in many cases (very) small. The mission team found that the business risk indicators set up for transfer price case selection were accurate in predicting the yield of CIT comprehensive audits. Having a more systematic approach to general CIT audit selection is therefore simply a question of utilizing existing work. Similarly, the transfer price business risk indicators could also be used to deliver on the previous IMF recommendations of setting up risk profiles for the largest taxpayers (in previous capacity development (CD) reports from 2022).

III. SUMMARY OF TECHNICAL ASSISTANCE ADVICE

A. Background

7. "Big Data" does not automatically translate into "Big Improvements".[1] While opportunities exist because of the availability of much more data than in the past, to capitalize on them means that the administration must solve a range of statistical, organizational and technical problems. "Advanced Data Analytics" is the practice of using statistical techniques to make predictions and draw inferences about cause and effect. This approach is not new to revenue administration, so advanced analytics does not aim to achieve anything fundamentally new, rather it seeks to carry out these tasks to make judgements with more and better reliance on data.[2]

8. Advanced data analytics has the power to process incomprehensible amounts of data into accurate predictions. Old-fashioned risk-modelling requires personnel to manually construct sensible risk-rules. To the contrary, advanced data analytics automatically and independently of human intervention builds a model to incorporate all available information. As an example, a machine learning model fed with all available data on taxpayers from CIT, value-added tax (VAT), personal income tax (PIT) and social security registers, can by itself determine which variables are

---

[1] As noted in the Preface to the OECD's publication: OECD (2016), ***Advanced Analytics for Better Tax Administration: Putting Data to Work***, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264256453-en

[2] Ibid. p11.

important when predicting the likelihood of tax evasion. Additionally, when implemented, a machine learning model will learn from past mistakes and self-correct.

9. The benefit of using advanced data analytics is largest where data is available in abundance and the population examined is large. The Organisation for Economic Co-operation and Development (OECD) prepared a detailed report on 16 OECD countries experiences with advanced data analytics[3]. In their report they note the success of implementing advanced analytics being conditional on the next-best alternative. That is, when the amount of data available is overwhelming and the resources available to examine the data are scarce, it makes sense to adopt advanced analytics (examples include customs consignment selection and VAT refund audits). Contrary, when the population examined is small and the data available is difficult to make available to the model, then advanced analytics is less advantageous.

10. Effortless and automatic data integration is the bedrock of data analytics for risk assessment. For this purpose, it's desirable for the SFA to constantly enlarge its dataset, from both internal data integration (tax data and customs data), and external third-party data (automatic) exchange. The target is that all data, which is already being automatically collected (such as through tax returns, bank transfers, financial information, customs data etc.), is automatically uploaded into risk assessment models. It would be especially useful if SFA could collect from reporters of financial data, who are residents in Slovenia, the same data as it is determined with the Directive on Administrative Compliance (DAC) 2 and the Common Reporting Standard (CRS) for foreign financial institutions[4].[5] With the same goal of enhancing data analytics, we would propose that the data managed in the special VAT files (as stated in issued and received invoices) is submitted automatically to the SFA on monthly or quarterly basis. It would also be beneficial that the balance sheets would be submitted to the SFA directly (not just exceptionally but mandatory) every year, with the deadline for submission before the CIT returns are to be submitted. This will in some cases require amendments to data exchange laws (e.g. in the case of bank transfer information).

11. The opportunities for using advanced analytics are endless. As the OECD[6] and others[7] point out, the emergence of advanced analytics, with its ability to examine data or content using sophisticated approaches such as pattern recognition, outlier detection, cluster analysis, experimental design, network analysis, and text mining, has opened new opportunities for the use of intelligence across all aspects of revenue administration (see appendix 1 for a list of suggestions).

---

[3] OECD (2016), ***Advanced Analytics for Better Tax Administration: Putting Data to Work***, OECD Publishing, Paris.

[4] For more information on the DAC2 Reporting model, refer to https://taxation-customs.ec.europa.eu/taxation-1/tax-co-operation-and-control/general-overview/enhanced-administrative-cooperation-field-direct-taxation_en

[5] For more information on CRS, refer to https://www.oecd.org/tax/automatic-exchange/common-reporting-standard/

[6] Ibid. p16.

[7] IOTA (2017), ***Good Practice Guide – Applying Data and Analytics in Tax Administrations,*** IOTA, Budapest; and WCO News (February 2017), ***Data Analysis for Effective Border Management,*** WCO, Brussels.

12. The leap from simple data analytics to advanced data analytics for SFA is small. The SFA already uses data analytics and merges data sources from a wide variety of data sources. This implies that the investment required for SFA to proceed with more advanced data analytics is small.

13. The mission conducted a workshop with key SFA staff that identified an organizationally detached and non-prioritized data analytics unit as a major risk. The dystopia here being a satellite unit out of touch with end-user needs and without the necessary resources/authority to deliver. This risk is very real and emphasized by the OECD[8] in their previously referenced report. To avoid this risk materializing the SFA must vigilantly ensure integration of the data analytics unit into the broader organization through a breadth of measures:

- Ensure employee flow between the broader business and the data science team (e.g. through "tours" where employees are embedded in the data science team).
- Prioritize education of end-users. No application should be developed without setting aside time for education of end-users.
- Aim for rapid deployment of projects to avoid the construction of "white elephants" – i.e. advanced solutions that are not used in practice.
- Rigorously measure the impact of data analytics projects to ensure quality of solutions and organizational buy-in. Currently the Analysis unit have developed >50 applications, along with the SAP system maintained by the IT department, but for many of these there has been no systematic assessment of success from the perspective of the end-user.
- Ensure that model predictions are actionable and transparent.

14. Getting data analytics right is not just a matter of producing valuable insights, it is also about convincing SFA staff of the benefits of using those insights. In the case of successful advanced data analytics, it is crucial that no employees feel left behind. Veterans within the SFA may be heavily invested in methods that they have utilized for years and are now expected to adopt methods they may not even fully understand. There are several ways to mitigate cultural resistance, some of which are:

- Have senior management endorse the mission of the data science team and grant the team the necessary authority to function.
- Get veteran employees to be part of the process and let them understand they have valuable insights that can contribute to the process. Provide training sessions to allow them to understand the basics of analytics.
- Incentivize change: make it clear that technological change is required by rewarding performance and do not punish the many expected failures on the journey ahead

[8] OECD (2016), ***Advanced Analytics for Better Tax Administration: Putting Data to Work***, OECD Publishing, Paris.

- Avoid creating a culture of fear of change by translating efficiencies gained through technological advances into fewer employees, but instead use the additional resources to increase quality of services.
- Prove the value of new approaches by competition with old approaches, e.g. select half of the audit cases using standard risk rules and the other half using more advanced analytics.

15. Data analytics and predictive modelling should *not* follow a "grand master plan". The key to data analytics is a trial and error through rapid deployment and piloting. The process should follow the model to: (1) understand what issues the SFA wishes to target through conversations with the business; (2) get an overview of the data available and use whatever data that is easily accessible before requesting new data; (3) build a statistical model and evaluate its properties with the end-users; (4) refine the model using more/better data; (5) if satisfied with the model's theoretical properties, then deploy in real life; and (6) evaluate the model and refine as necessary.

16. Translating predictions into actionable intel requires model transparency. That is, it is not sufficient to have a black-box prediction that no one understands. To the contrary, the officer on the ground need to know why the model predicted what it did. Recent research provides new tools for inspecting the inner workings of any predictive model. Examples: 1) Make an auxiliary model which is simple, e.g. linear or decision tree, as a proxy for the underlying model (*surrogate models*). 2) For each variable compute its marginal effect on output given all other data matches population (*individual conditional expectation*) 3) Have the model perform *targeted predictions* telling the officer on the ground what to look for, e.g. outcome variable being "adjustment due to un(der)invoicing" instead of just "adjustment".

B. Organizational Structure, Recruitment, and Integration of Solutions

17. SFA staff has been negatively impacted by frequent restructurings in the organization in the last years. When a change in direction occurs, inefficiencies and lost momentum occurs. For example, a Compliance Risk Management Unit (CRMU) was barely in place before being closed again. Organizational stability is critical for progressive reforms that build off each other.

18. In the nascent days of building advanced analytic capabilities, a centralized unit (such as is the case in SFA) is a great way to pool resources. Capable data analysts are an extremely scarce resource in all countries. The skillset/investment is so transferable that it makes little sense to have different units servicing CIT, VAT and customs in a small tax authority. Similarly, the data analytics unit should work on all aspects of data analysis – not just audit selection.

19. The SFA should grant the data analytics unit the necessary authority and develop the synergy among its core functions. The analytics unit, which lies in the Supervision Department of SFA, should work closely with other core functions on some critical issues. E.g., the analytics unit can improve the large business risk profiling by working with the Special Financial Office (SFO) which is in charge of large business (desk) audits. To better maintain the existing data applications and develop even more, the close collaboration with the IT department is critical.

20. The SFA should prioritize retaining/recruiting few but highly competent data analysts. Five good data analysts can function as the core of a data analytic unit. The productivity variance for data analysts is very different from other business lines and it is in general better to have fewer more competent resources than many less qualified. This also implies the SFA should consider a special pay scale for highly specialized staff in this area and should not let language barriers hinder recruitment. Part-time workers studying e.g. statistics at universities can also be a valuable resource.

21. Retraining of operational staff into data analysts is needed to ensure understanding of business needs and to ensure recruitment. In order to have a closer connection to the business, it is advised that selected employees from operations are retrained to join the data analytics team. E.g., retraining already data capable staff from risk-modelling into the data analytics team can help the data analytics team better understand the end-user needs.

22. To reap the full benefits of the data science team a key priority is to train other departments in advanced data skills. The number of tasks the data analytics team can do is limited and the training of other departments will ensure that the team's acquired capabilities spread. One way of doing this is organizing a learning program where relevant staff attend structured multiday workshops with prepared curriculum, homework and assignments.

23. The SFA should consider analytical software packages with low-entry barriers as a way of getting non-trained data scientists onboard. One example being the open-source package KNIME®, which is a point-and-click software with a low-entry barrier. This package comes with substantial online material and the possibility to attend courses across the globe.[9]

C. Low-Cost/High-Reward Initiatives in the Realm of Risk Assessment at SFA

24. The mission team identified several simple adjustments to risk assessment that expectantly will have large returns with little effort. These simple adjustments should be prioritized and implemented before moving towards larger investments in advanced data analytics.

25. Most of these adjustments relate to the fact that the SFA should have larger emphasis on size. All risks can be decomposed into the likelihood of the event happening multiplied by the consequence if the event happens (size of risk=likelihood*consequence). Concretely, this implies that any type of taxpayer intervention should be prioritized not only based on the likelihood of success, but also by the size of that potential success.

26. In the case of audit case selection, the SFA should appropriately weigh business risk scores according to the size of the taxpayer. The current business risk rules for VAT and transfer

[9] https://www.knime.com/ Other examples of open-source packages are: WEKA, ANACONDA, and RATTLE.

pricing (TP) audit case selection will increase the overall level of risk by at most 10 percent when moving from being a micro firm (e.g. a self-employed) to a large firm (with potentially thousands of employees). This is not an accurate assessment of the variance in risk. The mission team demonstrated how the precision of business risk rules in determining the highest yield cases could easily be increased by simply multiplying the current risk scores with the size of the taxpayer (measured by e.g. sales, employees, taxes paid etc.)

27. SFA should introduce thresholds in their measure of strike rates such that marginal corrections do not contaminate these statistics. A correction of €20 will rarely be perceived as a successful result of an audit given the number of resources spent. If resources are to be spent to maximize success, the measures of success such as strike rates should reflect meaningful adjustments. In particular, this is important when evaluating the accurateness of audit selection models (discussed below). The mission team demonstrated how strike rates from comprehensive audits fall by 10-20 percentage points when applying a €5000 threshold.

28. The 30 largest taxpayers are always a priority and should be continuously observed. As the size distribution of firms is so heavily skewed, the largest taxpayers are always a priority and should be continuously observed (since the consequence of non-compliance is so large). As recommended in the previous CD reports, the SFA should create risk profiles for the 30 largest taxpayers (can be automated to a large extent using the TP indicators).[10] Furthermore, the SFA should establish one-to-one client account managers. These taxpayers should always be educated, able and willing to comply. If the risk profiles suggest otherwise, the SFA should immediately prioritize intervention (in the form of education, guidance or audit depending on the perceived risk).

29. End-user education should be prioritized both to ensure expanded usage of existing applications but also to limit data analytics resources spent on trivial work. E.g., the mission team found that the business risk indicators set up for TP case selection were accurate in predicting the yield of CIT comprehensive audits. Having a more systematic approach to CIT audit selection is therefore simply a question of utilizing existing work. By educating the end-users in existing applications this will allow the end-user to realize such low-hanging fruits. Similarly, end-user education will limit the time spent in the Analysis unit on mundane tasks such as rearranging data in reports.

30. Systematic evaluation of past interventions/strategies. The mission team found that previous audit strategies/business risk rules were not being systematically evaluated. The mission team suggested simple methods to evaluate business risk rules using past audit results. This could be done in simple scatterplots. Going forward the mission team also suggested running actual "horse-races" applying different approaches in parallel and comparing outcomes (discussed below).

31. Using simple benchmarks and standard deviations can help improve existing business risk rules. The mission demonstrated how simple benchmarks such as industry profitability and standard deviations can help root business risk in systematic analysis (instead of intuition).

[10] See Appendix III in John Middleton's 2022 CD report

D. A Multifaceted Approach to Risk

32. The mission team emphasized the need for a multifaceted approach to risk. While the mission team spent most workshops building models for audit selection, it is important to stress that this is only one (small) part of a successful data-driven compliance strategy.

33. As described in previous CD reports, a successful compliance risk management strategy implies a broad range of tools – not just audits. In workshops the mission team presented the "butterfly" approach to risk (see appendix II). This approach consists of three steps:

a. Group non-compliant taxpayers in three clusters: The unaware, the unable and the unwilling.

b. Deter non-compliance through education, service and exemplification targeted at the different clusters

c. Detect non-compliance using risk modelling

d. Deal with detected non-compliance based on the connected risk (likelihood*consequence) and the underlying reason (cluster)

34. To support this multifaceted approach, the mission team recommends reestablishing the Compliance Risk Management (CRM) Unit. This unit will support the CRM Committee and ensure 360º assessment of risk mitigation. The data analytics unit will need to serve the entire CRM spectrum.

E. Proof of Concept: Empirical Models for Risk Assessment in CIT, TP and VAT

35. Due to the easily accessible data, the mission team and the SFA dedicated staff were able to construct four pilot empirical models to inform risk assessment. The ease of getting data and the collaboration between data analysts and business experts in delivering on this is an encouraging sign for future projects. All data/code/workflows were handed over to the SFA (incl. additional models on customs not covered by the mission). The models constructed included:

- An advanced (random forest) CIT audit selection model (incl. deployment workflow)
- An advanced (random forest) VAT audit selection model
- Workflow to evaluate existing business risk rules against historical data (VAT/CIT)
- Country-by-Country-Reporting anomaly detection tool

36. All models had the same target of delivering real-time risk assessment using all available data. In Appendix III we provide an overview of the optimal workflow for using data analytics in tax enforcement. The target is that all data, which is already being automatically collected (such as tax returns, financial information, customs data etc.), is automatically uploaded into the model, that then based on all the available data detects risk. The risks with the largest expected consequence (likelihood*consequence) are dealt with using an appropriate intervention. The result of the intervention is uploaded back into the model, such that it learns and improves.

37. Despite being pilot models with room for improvement, the audit selection models outperformed existing risk rules. Even without using all the available data (and with some short-cuts being made in data manipulation) the model proved relatively accurate in predicting audit yields. E.g., the CIT audit selection model was estimated to increase the strike rate in comprehensive audits (measured as a correction above €5000) from 50-60 percent to 90 percent (while doing the same number of audits)

38. The mission team demonstrated how existing business risk rules could be evaluated in two ways. First, the mission team demonstrated how business risk rules can be correlated with historical audit yields to test for accuracy. In general, the business risk rules proved accurate in predicting yields. However, the mission team demonstrated how the business risk rules could very simply be improved by putting more weight on size. Second, the mission team demonstrated how the advanced audit selection models could inform the relevance of different risk rules. E.g., the advanced models suggested that all variables related to wrongful registration of taxpayer should have larger risk weights than given at present.

39. The Country-by-Country-Reporting anomaly tool showed how simple statistical measures could inform anomaly detection. In particular, the model demonstrated how to compute benchmark profitability for each multinational enterprise (MNE) group and industry and how to compare this to subsidiary profitability in Slovenia.

40. The pilot risk assessment models have plenty of room for improvement and verification is needed. First, all available data should be used and verified and competing models should be tested. Second, it is very important the SFA does not start implementing models they cannot fully comprehend. The end goal is never perfection but instead improvement and even adopting the simplest of the mission recommendations will be a step in the right direction.

41. The real test for all model lies in field implementation. That is, only through actual deployment can a model by rigorously evaluated. The time from the pilot model being built to (small scale) implementation should be minimized.

42. A credible evaluation of audit selections models requires a credible benchmark (status quo) comparison. This is done best through a randomized *double-blind study.*[11] The procedure is as follows: (1) the SFA computes business risk rules as they usually do; (2) the SFA picks (initially) 100 cases based on the new model prediction; (3) the SFA *randomly* assigns audits from the new predictive model and the standard SFA model to different inspectors – the inspectors should not be informed of whether the case was picked because of the new or the old model; and (4) the SFA

[11] A double-blind study is one in which neither the participants nor the experimenters know who is receiving a particular treatment. This procedure is utilized to prevent bias in research results.

tracks the adjustment outcomes in the cases picked by the new and the old models.[12] Make sure that the adjustment outcomes and all other taxpayer information related to the top picks of the new and the old model are recorded. This is best achieved by keeping track of the consignment number that relates to the top picks.

F. Prioritization of Future Projects

43. The data analytics team cannot do everything, and prioritization of projects is hence needed. As already discussed, keeping in mind the next-best option is key when determining which projects to conduct – other considerations when include:

- Data available (fit for data analytics?)
- Size of closable tax gap (based on actual analysis or intuition)
- Bidding round (how large resources will each division devote)
- Political decision (importance beyond revenue)
- Next best alternative

44. The mission team recommend the following priorities in the coming year. First, create risk profiles for the 30 largest taxpayers. Second, create business risk indicators for CIT. Third, put larger weight on firm size and redefine strike rates to reflect firm size. Fourth, evaluate existing business risk rules using historical data. Only when these simple adjustments are made, continue with the more advanced modelling. All steps in the list of priorities have been demonstrated in the workshops and the workflows/models have been handed over.

IV. PROPOSED NEXT STEPS

Table 1: Summary of Suggested Actions to be Taken

Actions to take	Deadline	Responsibility	Implementation advice
Topic: Prioritization/endorsement by senior management			
1. Prioritize/endorse high-value work in the sphere of data analytics identified in this report by creating protected time for development of these	Short term (within next 12 months)	Director General (DG)/ Heads of department	To make headway on data analytics. The first step is for senior management to endorse/prioritize the data projects identified in this report. This should happen by dedicating protected time for development/education to the data analytics team

[12] Random assignment of cases from the new and old models can be achieved using the RAND function in Excel. The procedure is as follows: (1) merge the 1,000 top picks from the new model and the 1,000 top picks from the old model into one Excel spreadsheet; (2) assign a random number between 0 and 1 to each using the function RAND; and (3) divide the cases between inspectors based on the random number. For example, if there are 20 inspectors, inspector 1 gets the cases with the five percent highest random number, inspector 2 gets the five percent second highest numbered cases, etc.

projects. Deprioritize current low-value data tasks to free up time.			and deprioritizing current low-value assignments such us preparing reports etc.
2. Move towards "top-down" risk assessments in CIT by defining and using existing business risk indicators	Short term (within next 12 months)	Heads of department/Working Groups (WG)	Begin by building upon the current TP business risk rules.
3. Increase emphasis on size of taxpayers in risk assessments	Short term (within next 12 months)	Heads of department/WG	By 1) risk profiling the 30 largest taxpayers and creating client owners 2) Increasing threshold for "strikes" 3) weighing risk rules according to consequence
4. Endorse a multifaceted approach to compliance risk management	Short term (within next 12 months)	Heads of department/Working Groups (WG)	Ensuring that compliance is improved not just through audits but service/education as well
5. Move towards automatic exchange of third-party information from banks, other regulators and other relevant stakeholders	Medium term (within two to three years)	Heads of department/ Relevant stakeholders	Will in many (most) cases require legislative action.
Topic: Improvements in data analytics and the way-of-work			
1. Readjust risk rules to reflect expected adjustment/risk	Short term (within next 12 months)	Analysis unit and relevant business owners	To begin with simply by multiplying current risk rules with measures of size (turnover, tax paid, employees)
2. Build automated reports for the largest taxpayers	Short term (within next 12 months)	Analysis unit and large taxpayer unit	Begin by building upon the current TP business risk rules. Update according to John Middleton's 2022 report Appendix III. Reports should also indicate suggestive actions such as "contact taxpayer" or "audit"
3. Build CIT business risk indicators	Short term (within next 12 months)	Analysis unit and relevant business owners	Begin by building upon the current TP business risk rules.
4. Evaluate existing risk rules based on historical data and advanced	Short term (within	Analysis unit and relevant business owners	See KNIME workflow for inspiration.

models -> update accordingly	next 12 months)		
5. Expand use of anomaly detection in TP using CbCR-data + other data sources	Short term (within next 12 months)	Analysis unit and TP unit	In particular, this analysis can inform which of the >2,300 cases of firms outside of safe-harbor thin capitalization rules to pursue. Prioritization should, again, also always rely on size of the taxpayer. The Compliance Improvement Plan (CIP) for international tax in John Middleton's 2022 CD report is a good place to start. See KNIME workflow for inspiration on how to do anomaly detection.
6. Advanced data analytics projects	Short term (within next 12 months)	Analysis unit, IT & relevant business owners	After the low-cost investments (1-5) the SFA can proceed with more advanced predictive modelling (such as random forest models etc.). See pilot models from this mission as inspiration.
7. Evaluate implementation of data analytics projects and devise remedying actions if necessary	Short term (within next 12 months)	Impacted divisions & Analysis unit	Aim is here to learn from past mistakes. A common finding is resistance to change, which should be dealt up front.
Topic: Organizational structure			
1. Ensure the Analysis unit supports the entire SFA	Short term (within next 12 months)	DG in consultation with senior management	Potentially by having them refer directly to the DG or as part of a CRMU.
2. Reestablish the Compliance Risk Management Unit		DG in consultation with senior management	As advised in previous IMF reports.
Topic: Staffing and training			
1. Prioritize end-user education in all existing applications	Short term (within next 12 months)		No application should be built without prioritizing end-user education and feedback
2. Establish learning programs from which data scientist will teach non-experts in data analytics	Short term (within next 12 months)	Analysis unit in conjunction with heads of divisions	The intention here is two-fold 1) alleviate the workload of the Analysis unit 2) ensure data science capabilities spread throughout the organization. The software KNIME® can be a good tool to conduct these trainings with its low-barrier of entry.
3. Consider improving terms for data analysts with aim of ensuring recruitment retention	Short term (within	DG/HR	This could include having longer contract terms and/or designing "Golden handcuffs" through time conditioned compensation packages.

	next 12 months)		
4. Retrain 1-3 non-data scientists from SFA to join the Analysis unit	Short term (within next 12 months)	Data Science Team and HR	The intention is two-fold: 1) bring business knowledge into the Data Science Team 2) expand the capacity of the Data Science Team. One possibility is to exploit the KNIME® software and available courses to ease this process.
5. Start "tours" program where non-data analysts from SFA are embedded in the Analysis unit and vice versa.	Medium term (within two to three years)	DG and relevant heads of division in consultation with Analysis units	Make sure this aligns with the projects in the pipeline – e.g. if a customs selection project is taking off, then it make sense to collaborate with customs.

List of Documents Produced and Provided to the Authorities

PowerPoint Presentations:

- High-level introduction
- Workshop presentation
- Exit-meeting presentation

Models (KNIME workflow + data + .xls):

- CIT audit selection model (KNIME workflow plus data)
- CIT audit selection model with deployment (KNIME workflow plus data)
- VAT audit selection model (KNIME workflow plus data)
- CbCR anomaly detector model (KNIME workflow plus data)
- CbCR overview (Excel model)
- Business risk indicator evaluation model (KNIME workflow plus data)
- Customs consignment selection model (KNIME workflow plus data)

Appendix I. Gross List of Possible Projects

Registration

- Predict who should be registered.
- Predict revenue associated with a non-registered person.

Filing

- Predict who will file late before the event.
- Predict who will file once they are late (self-finalize).
- Predict revenue associated with late or non-filing.

Reporting

- Predict who is non-compliant (likelihood) for each tax type.
- Predict size of potential adjustment (consequence).
- Predict high risk refunds.
- Predict who will object to an amended assessment.
- Support text and social network mining in audit cases.

Payment

- Predict who will pay late before the event.
- Predict who will pay late but before intervention (self-finalize).
- Predict who will pay given alternative interventions (phone, mail, visit, court action etc.).
- Predict capacity to pay and propensity to pay.
- Predict business viability (see BVAT[13] model on the ATO website).

Service

- Taxpayer/trader channel use to inform design decisions and identify self-service opportunities.
- Improve service delivery using proactive messaging, calling, and other interventions.

13 https://www.ato.gov.au/calculators-and-tools/business-viability-assessment-tool/

Customs

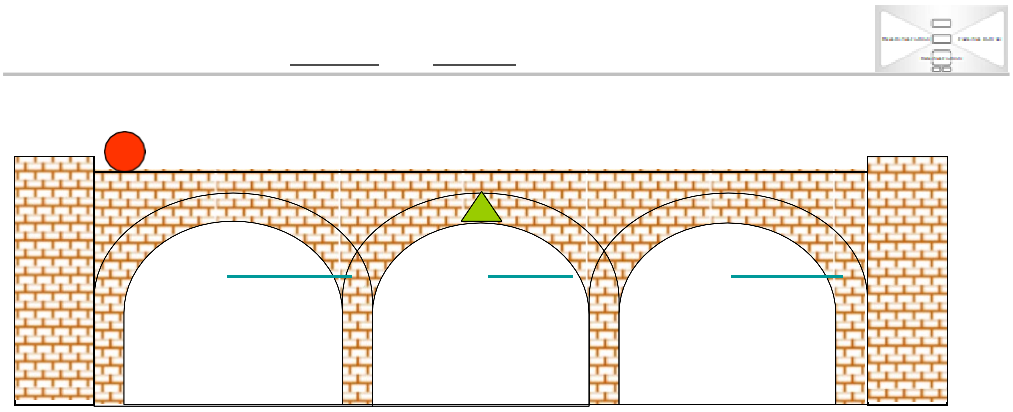
- Predict the likelihood and size of upliftment
- Predict likelihood of smuggling contraband, drugs etc.
- Predict behavior of traders and passengers.

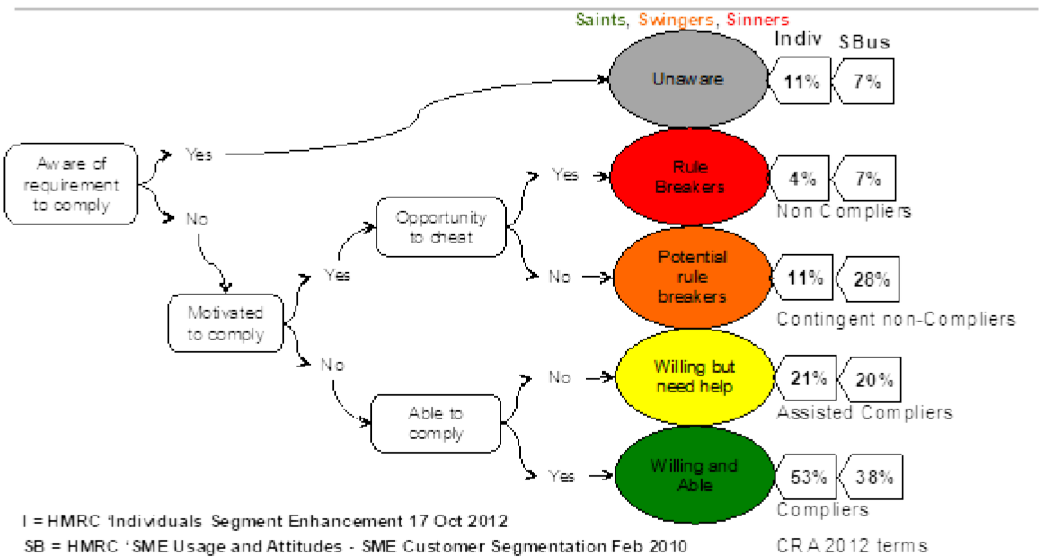
Policy

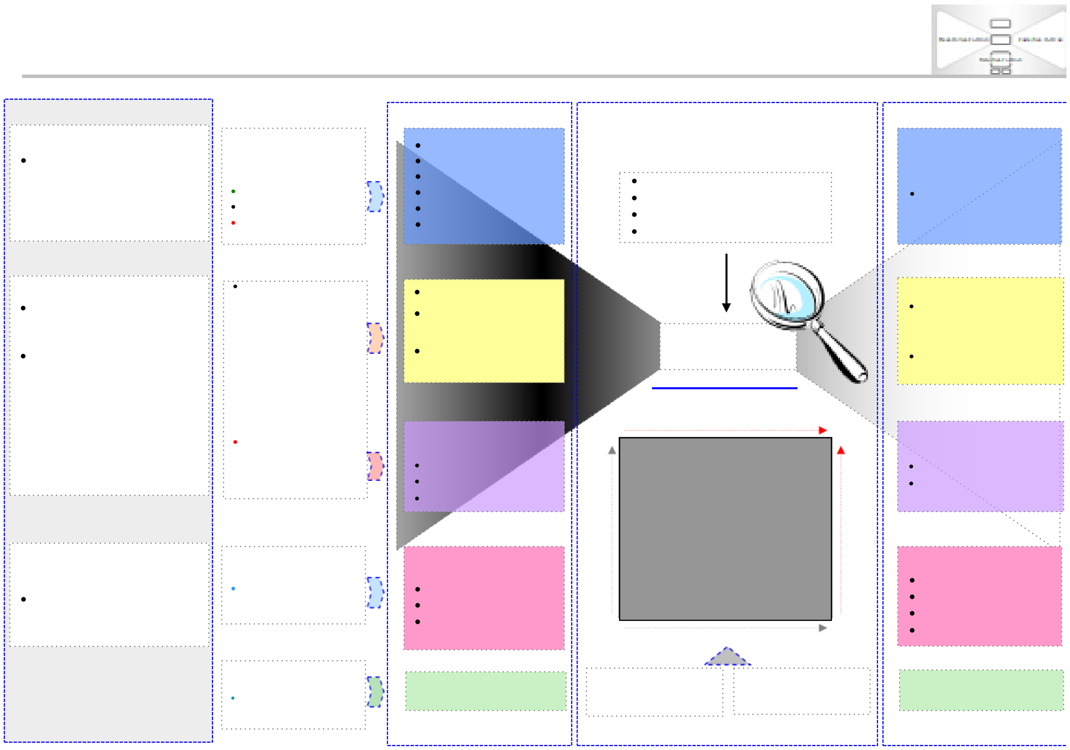
- Tax gap measurement.
- Assessing or forecasting the impact of changes in tax policy.

Appendix II. The Multifaceted (Butterfly) Approach to Risk

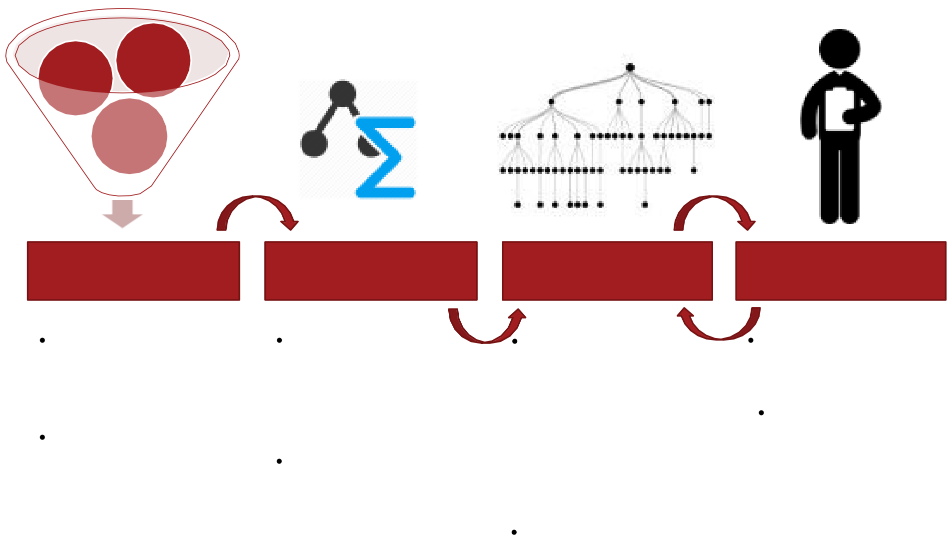
(By Stuart Hamilton, IMF expert)







Appendix III. Suggested Workflow for Data Analytics



Appendix IV. Practical Tips for Predictive Model Development and Use

(By Stuart Hamilton, IMF expert)

1. Data Analysis Process

- Use a formal process, such as CRISP_DM to document data mining efforts for better analysis, consistent reuse and shared learning. https://www.ibm.com/support/knowledgecenter/en/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm

2. Data

- Typical data set for predictive analytics for tax audit case selection:

Identifiers	Demographic	Interaction	Financial	Risk	Adjustment	Penalty
TFN & Name	Region & Industry Age Gender	Files O/S Debt late Last audit	VAT & CIT P&L & Bal Sheet	Score for each rule & overall	Amount	Amount

- Where possible have/use **categorical** data rather than free text fields. You may need to limit the number of categories for data mining purposes (e.g. 1000 categories in a 1000 cases means no category correlation is possible. Reduce by grouping up to a smaller meaningful subsets. (e.g. Use higher ANZIC Codes from industry categorization.) The 'Domain Calculator node' in KNIME can be used to restrict domain numbers.
- Where possible use **ordinal** data (categories that have a meaningful order A>B>C etc) rather than just categorical data.
- Where possible use **interval** data (numeric with meaningful spacing between data) rather than ordinal data.
- Where **time** data may be relevant (e.g. year established for a company or the age of a person) consider converting it into a simple numeric such as Age / Number of Years since establishment. Similarly for overdue accounts – use days overdue rather than date due. Review time data to see if is being considered as a string variable because of its presentation in the data set (e.g. 2015/2016 will be treated as a string. Convert to 2016.)
- Manipulate and transform mirror data sets - never the original data on the database. Even with data cleansing operations maintain the original data set for evidence purposes and create a new / updated dataset.
- Save SQL scripts for consistent data retrieval, reuse and process documentation.
- If faced with extremely large data sets use representative sampling (random or stratified random) to reduce the size of the data set being analyzed.

- If running analytic software on a relatively small device, a laptop or desktop, consider making data transformations on large data sets on the server / mainframe. The trade off is between wait-time for the new data set to arrive v execution time on a small device.
- Always review the data in detail and fully understand how the data was created and input, its possible errors, as well as the distribution of the data.

3. Modeling / Mining

- If the target variable is relatively rare (e.g. <10% strike rate) consider 'oversampling' the target or 'under sampling' the negative class or generating additional 'synthetic' examples using the SMOTE[14] node in KNIME. This is done because the learner needs to 'see' sufficient cases to learn. Similarly if the target variable is almost always present (e.g. >80% strike rate) consider using higher thresholds to reduce the apparent strike rate (e.g. move the threshold towards the median strike and the positive class will reduce towards 40% if originally ~80%). Note: Do not duplicate data (oversample) prior to partitioning data into the training and verification sets as this will result in duplicated data being present in both the learner and verification sets and will thus overstate the models predictive ability.
- Review the data again. Consider restricting the number of categorical variables using the Domain Calculator node. Consider transforming quantitative string data into numeric data using the String to Number node. Consider transforming data via normalization and ratios etc, to better highlight discriminate features. Some modeling approaches work best with normalized interval data. Explore whether such transformations improve the predictions. (e.g. using the KNIME Normalizer node).
- To reduce the number of low value cases selected iteratively raise the threshold for a 'strike' and evaluate the results until a suitable balance between strike rate and caseload is obtained. (The median strike value is often a good starting point.)

  Use a Rules Engine node to set a threshold for a strike at an appropriate value:

  (e.g. $ADJUSTMENT$ >= 50000 => "Y"

  $ADJUSTMENT$ < 50000 => "N"
- Review the data again. Ensure that the variables going into the learner node are not the product of the adjustment.
- Exclude clearly non-relevant variables such as the tax file number (using the learner nodes configuration dialogue in Random Forest or using a Column Filter node before the decision tree learner).
- Use a good 'out of the box' algorithm, such as Random Forest' initially at its default settings. Once a promising model has been found increase the number of trees in the configuration dialogue from the default of 100 to 300 to improve model performance. (Additional

14 https://nodepit.com/node/org.knime.base.node.mine.smote.SmoteNodeFactory

performance tweaks to Random Forest inputs can be obtained by using the Tree Ensemble Learner / Tree Ensemble Predictor node in place of the basic Random Forest pairing).

- To reduce 'noise' in the dataset evaluate and eliminate variables that don't provide predictive ability. (The Meta Node provided for Random Forest 'variable importance' indicates the relative use of variables in the Random Forest learner.)
- As expertise develops explore the use of other modeling approaches to see if they can improve predictions over part of the data. (It is usually hard to beat Random Forest in practice but it can be a useful learning / capability build experience for the team.)
- Consider using 'ensemble approaches' (taking the best predictions from multiple models – those with the highest 'Area Under the Curve') when appropriate. There is a model parameter optimization workflow on the KNIME examples server.
- Economic data is usually very highly skewed, as are the outcomes of compliance activities. As binary classifiers create a view of the likelihood of adjustment and not the consequence, it is crucial that a view of the potential size of the adjustment is brought into the risk equation. Using a view of the median adjustment by turnover range as a useful proxy for the potential consequence and then use this to prioritize cases by predicted risk adjusted value (LxC). Test this set of consequence proxies on the verification data set to establish the best ranking approach that maximizes overall revenue recovery.

4. Deployment

- Once a robust predictive model has been built and rigorously evaluated, for the deployment build push the 'training partition' to 100% and re-execute the learner node to maximize the deployed models' predictive ability.
- Use a single decision tree (CART) to provide a broad explanation of what the more accurate 'black box' model is doing. It won't be exact nor always 'correct' but should provide a useful indication. 'Prune' the tree to an appropriate level to explain most cases by increasing the lowest number of cases for a split point.
- Errors in model development can occur, particularly in the early stages of trialing analytics when building capability. Stage gate model deployment to ensure that the results produced are in line with those expected. E.g. do 100 cases and evaluate the results, if ok then do a further 400 and evaluate, if ok do a further 500 etc...
- Use 'blind' testing when practical so that observer bias (for or against) is less likely.
- A models' predictive ability degrades over time as the underlying economy changes. Revisit models every six months with additional data to see if they need to be rebuilt or enhanced.
- Consider using a small **random case selection component** (e.g. using stratified random sampling) to maintain intelligence on **new risks** and as a means of monitoring model performance over time. A random case selection component can also assist in estimating tax gaps and prioritizing compliance campaigns.

- Build risk rules for segments not previously examined or with low representation in the data set as these will usually be excluded by the predictive model as no or limited prior successful / unsuccessful case data exists for it to learn from.

Glossary of Technical Terms

Audit

A process used to establish whether the correct amount of tax has been assessed. It involves formal evidence gathering to establish the facts and then the application of relevant law to those facts. The time and resources required to appropriately audit a taxpayer depends upon the matter and materiality being audited and one size audit does not fit all. For example, a VAT refund audit generally involves simple fact checking, while at the other extreme an Income Tax TPaudit may involve information exchanges with other tax jurisdictions, taking several months just to establish the functional analysis facts.

CART

Classification and Regression Tree. A decision tree data mining software algorithm. Usually not the optimal data mining method, with a tendency to 'over-fit' the training data, it has the advantage of not being a 'black box'. The single decision tree rules are explainable.

Case Selection

The process (e.g. via data-mining or subject matter expert rules) used to initially identify a set of taxpayers (positives) that may have compliance risks. Ideally should produce a listing of taxpayers prioritized (ranked) by predicted revenue risk = % likelihood multiplied by Rs consequence [aka Risk Adjusted Value].

Risk Filter/Risk Rule

A set of rules used to select cases for a particular risk. Can be created by subject matter experts or from predictive data mining.

- False positives (FP)
- Taxpayers that initially appear to have a tax compliance risk, but on review are found to be compliant. Opposite of true positives (TP)
- False negatives (FN)
- Taxpayers that appear to be compliant but are not. The opposite of true negatives (TN)
- Strike Rate (precision): TP/(TP+FP)
- The ratio of true positives over the number selected. A function of case selection rationale, efficacy and size, auditor detection capability, and the underlying compliance rate.
- Miss Rate: FN/(TP+FN)
- The ratio of false negatives over the total number of non compliant. A function of case selection rationale, efficacy and size, auditor detection capability, and the underlying compliance rate.
- Accuracy: TP+TN/(TP+TN+FP+FN)

- The ratio of correctly determined cases to the total number of cases.

Confusion Matrix

A table setting out True Positives/False Negatives/False Positives/True Negatives from the selection model. Used in case selection model evaluation.

Confusion Matrix example

CONFUSION MATRIX

	Selected	Not Selected
TP	1538	3885
FP	19	

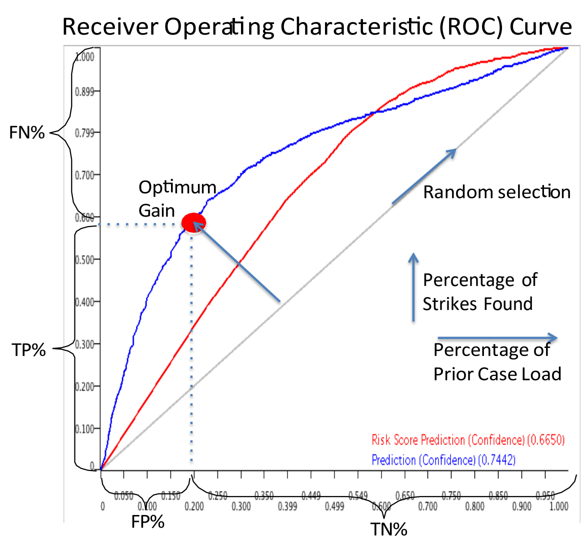
The matrix shows the number of True Positives, False Negatives, False Positives and True Negatives produced by a case selection model being evaluated. It enables the Strike Rate and Accuracy of the selection method to be calculated. E.g:

	MODEL	**PRIOR**	Change
Strike Rate	44%	29%	54%
Miss Rate	72%	0%	-72%
Accuracy	69%	29%	141%
Revenue	95%	100%	-5%
Revenue case	977,991	189,602	416%
Caseload	18%	100%	-82%

Receiver Operating Characteristic (ROC) Curve

A plot of how the ratio of True Positives to False Positives (TP:FP) varies over the sample/caseload. The greater the area under the curve (AUC) the better the selection model. Used in model evaluation.

Receiver Operating Characteristic (ROC) Curve example

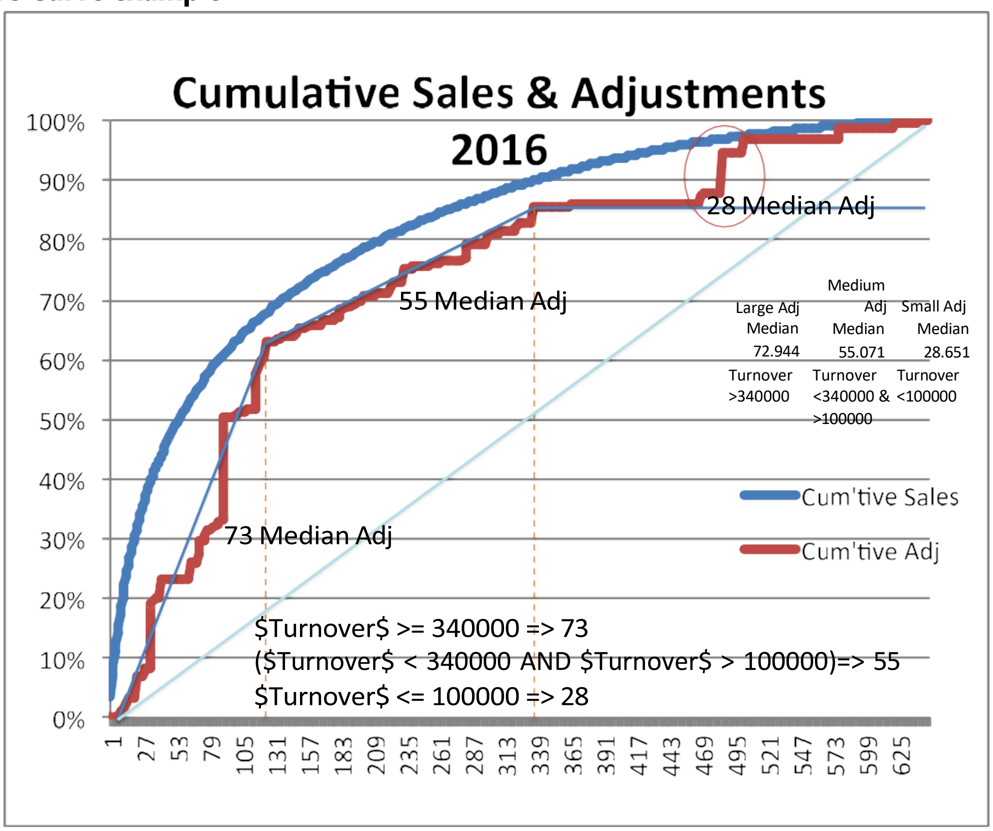


The further the selection models ROC performance line (blue predictive model & red risk score model) is from the random selection line (grey 45 degrees) the better the selection model.

Cumulative Curve Inflection Point Approach

The cumulative curve inflection point approach is a technique that looks for the 'turning points' (or points of inflection) on matched cumulative curves of turnover (to give taxpayer 'size') and the adjusted tax. The median adjustment for the range between turning points is then used as a proxy for **consequence** in the risk calculation likelihood x **consequence** for ranking/prioritizing candidate cases. The approach reduces the overestimation of potential consequence (the adjustment size) in highly skewed populations. (e.g. If a single figure such as the mean, median of the adjusted taxpayers was used as a consequence proxy it would *overestimate* consequence in the majority of cases, while if the modal adjustment was used it would *underestimate* consequence in most cases. By using several consequence proxies based on the differing size of the underlying companies a more accurate stepwise by turnover consequence estimate is produced, improving the reliability of case risk based prioritization.) Missing value turnover cases are dealt with by imputing a value with an appropriate rule.

Cumulative Curve example



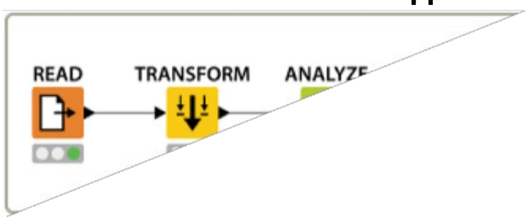
Data mining

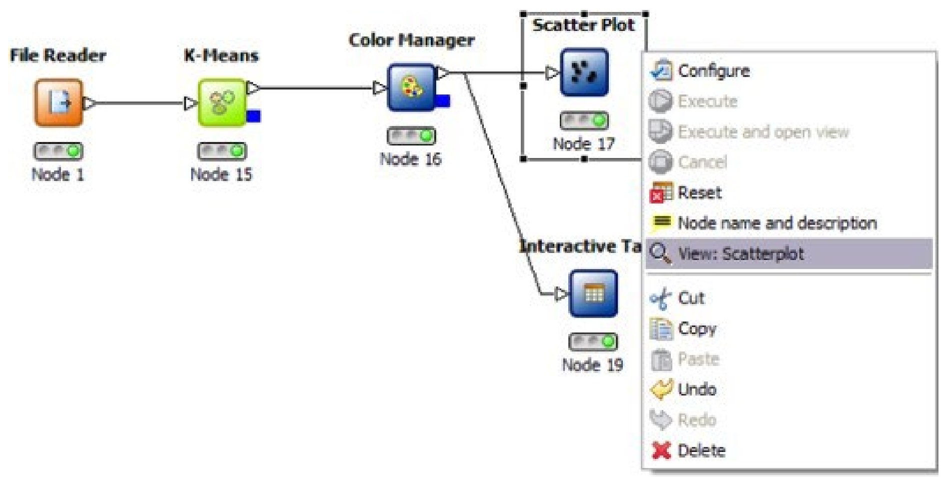
The use computer algorithms (machine learning programs) to identify patterns and associations (knowledge discovery) in data. Data mining can be descriptive or predictive. Predictive data mining takes a set of historic data and attempts to identify rules that best predict the outcome. Data mining approaches can be contrasted with subject matter expert approaches, where a person (a subject matter expert) defines (imposes) how taxpayers should be categorized or selected for compliance actions.

KNIME

KNIME (KoNstanz Information MinEr, pronounced 'Nime') is an open source data analytics, reporting and integration platform that runs on OS, MS and Linux operating systems. **KNIME** integrates various components for exploratory data analysis and data mining through a modular workflow / linked node concept.

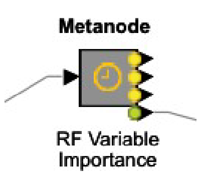
KNIME Node based workflow approach





Meta Node

A user created node in KNIME allowing workflows to be enclosed within it to reduce the complexity of the overall workflow display and enable the easy 'packaging' of reusable processes. In the workflow provided meta-nodes were used to simplify the large case and deployment workflows. (Double click the meta-node to open and edit its internal workflow.)



Random Forest

A data-mining / machine learning algorithm that is usually provides good 'out of the box' performance that is close to optimal. Created by repeated (e.g. 300 times) random sampling of the data and building a decision tree each time. The multiple decision trees (the Forest) then 'vote' on the correct categorization of an instance. The multitude of decision trees makes a simple explanation of a decision more difficult to trace through. Relatively fast, copes with missing values and non-numeric data and is resistant to 'over-fitting' the training data set.