Case Study

Topic 3.1

DATA MINING AND OTHER APPLICATIONS IN FINANCIAL AND TAX CRIME INVESTIGATIONS: EXPERIENCE OF INDIA

R.K. Tewari
Chairman of the Central Board of Direct Taxes
Revenue Department
(India)

Contents: *Summary. 1. Introduction. 2. Key information sources.3. Integrated Data Warehouse and Business Intelligence (DW&BI) Platform. 4. Conclusion.*

SUMMARY

Indian Income Tax Department (ITD) has embarked on an ambitious computerization plan to improve taxpayer services, promote voluntary compliance and deter tax evasion.

Availability of information in electronic form provided an opportunity to ITD to develop a wide range of non–intrusive methods for improving compliance. This paper presents the experience of ITD in using data mining methods in following areas:

a. Discovering non-filers with potential tax liabilities
b. Identifying potential under-reporting taxpayers
c. Improving compliance of tax deductors
d. Identification of non compliance in service sector
e. Identifying implicit linkages for effective investigation

ITD is now implementing a comprehensive Data Warehouse and Business Intelligence (DW & BI) Project to develop an integrated platform for effective utilization of information in all areas of tax administration. The design phase of the project commenced in January 2014 and phased implementation rollout is scheduled in 2015-17. The DW & BI platform will integrate enterprise data warehouse, data mining, web mining, predictive modelling, data exchange, master data management, centralized processing, compliance risk management and case analysis capabilities.

1. INTRODUCTION

Data mining is generally defined as processing and analysing data from multiple sources with the purpose of converting it into actionable information which can be used for specified purpose(s). This paper presents the experience of Indian Income Tax Department (ITD) in using data mining methods.

1.1. Background

ITD has embarked on an ambitious computerization plan to improve taxpayer services, promote voluntary compliance and deter tax evasion. Some important initiatives taken by ITD for improving the quality of tax payer services are:

- Permanent Account Number (PAN) is a ten digit alphanumeric number which uniquely identifies a tax payer. PAN has been issued to more than 200 million persons.
- Online tax accounting system (OLTAS) facilitates near real time reporting, monitoring and reconciliation of tax collection.
- E-payment of taxes has been enabled through Net Banking and ATMs and more than 80% of tax is collected through this mode.
- E-filing of Income Tax Return is mandatory for all the corporate taxpayers, taxpayers requiring statutory audit of accounts, and taxpayers with income greater than INR 500,000. The number of e-filed returns has increased to 29.7 million in F.Y. 2013-14. The percentage of e-filed returns now exceeds 75% of the total returns received in a year.
- In 2009, a Centralized Processing Centre for Income tax returns (CPC ITR) became operational for processing of Income tax Returns in an automated environment to determine tax payable or refund of excess taxes paid on the basis of returned income. All e-filed returns are processed at CPC.
- Refund Banker Scheme enabled the system driven process for determination, generation, issue, dispatch, tracking and credit of tax refunds.
- Under the eTDS scheme the tax deductors submit electronic quarterly statements of tax deductions (withholding of tax). In 2013, a TDS Centralized Processing Cell (TDS CPC) became operational for processing, reconciliation, default resolution to enable end-to-end reconciliation of tax payments and tax credits claimed against withholding of tax.
- The online annual tax credit statement (Form 26AS) is generated for each taxpayer (on the basis of PAN) on the ITD portal which shows the details of tax paid, tax deducted/collected and refund issued.
- ITD has now initiated the Income Tax Business Applications (ITBA) Project to rewrite the old applications and develop new interfaces

and process flows to leverage latest technology and meet the requirements of the new operating environment.

2. KEY INFORMATION SOURCES

Availability of information in electronic form also provided an opportunity to ITD to develop a wide range of non–intrusive methods for promoting voluntary compliance and deterring tax evasion. Approximate number of records in key information sources is as under:

S. No.	Information Type	Numbers per year (approx.)
1	Permanent Account Number (PAN)	28 million allotments (200 million cumulative PAN allotment)
2	Income tax Returns	35 million returns
3	Tax Payment	45 million records
4	Tax Deduction at Source	400 million deductee transaction records in 8 million statements
5	Annual Information Return (AIR)	6 million transaction records
6	Centralised Information branch (CIB)	120 million transaction records

Information sources at serial no. 4 to 6 which are specific to India are explained below.

Tax deduction and collection at source

Tax Deduction at Source (TDS) is one of the modes of collection of taxes, by which a certain percentage of amounts are deducted by a person at the time of making/crediting certain specific nature of payment to the other person and deducted amount is remitted to the Government account. The deductor is also obliged to file quarterly TDS statement giving details of transactions and tax deducted. The concept of TDS envisages the principle of “pay as you earn”. TDS not only ensures regular inflow of cash resources to the Government, it also acts as a powerful instrument to promote voluntary compliance and deter tax evasion. Provisions of TDS are applicable on various payments including:

- Salary
- Interest
- Dividends
- Payment to non-residents
- Payments to contractors
- Commission or brokerage

- Rent
- Fees for professional or technical services, royalty etc.

Under the provisions of Tax Collection at Source (TCS), the seller has to collect tax (in addition to the purchase price) from the person who has purchased following items/rights:

- Alcoholic liquor
- Timber
- Scrap
- Parking Lot
- Toll plaza
- Mining and quarrying

Annual Information Return (AIR) Scheme

Under the Annual Information Return (AIR) Scheme, which was introduced in 2003-04, specified entities are required to report the following transactions to ITD:

S. No.	Class of Person	Nature and Value of transaction
1.	Banking Company	Cash deposits aggregating to INR 1,000,000 or more in a year in any savings account of a person
2.	Banking Company or any other Company or institution issuing credit card	Payments made by any person against credit card bills raised aggregating to INR 200,000 or more in the year
3.	Trustee of a Mutual Fund	Receipt from any person of an amount of INR 200,000 or more for acquiring units of mutual fund
4.	Company or institution issuing bonds or debentures	Receipt from any person of an amount of INR 500,000 or more for acquiring bonds or debentures
5.	Company issuing shares through public or rights issue	Receipt from any person of an amount of INR 100,000 or more for acquiring shares
6.	Registrar or Sub Registrar	Purchase or sale by any person of immoveable property valued at INR 3,000,000 or more.
7.	Officer of Reserve Bank of India	Receipt of INR 500,000 or more in a year for RBI bonds

The Annual Information Return is required to be submitted by 31st August following the financial year in which transaction is registered or recorded.

Centralised Information Branch (CIB) scheme

Besides the AIR scheme, ITD also collects need based information about specific financial transactions under the CIB scheme. Some important sources of information covered under this scheme are:

- Sale of immovable property valued at INR 500,000 but less than INR 3,000,000
- Transfer of capital asset at a value lower than value declared for the purpose of stamp duty
- Time deposit exceeding INR 200,000 with a banking company
- Deposit in cash aggregating INR 200,000 with a banking company on a day
- Payment in cash in connection with travel to any foreign country of an amount exceeding INR 100,000 at one time
- Payment to hotels and restaurants exceeding INR 100,000 at one time

Use of data mining methods

ITD has been able to use data mining methods for the following:

a. Discovering non-filers with potential tax liabilities
b. Identification of potential under-reporting taxpayers
c. Improving compliance of tax deductors
d. Identification of non-compliance in service sector
e. Identification of implicit linkages for effective investigation

The experience of ITD in the above areas is given in following paragraphs.

a. Discovering non-filers with potential tax liabilities

The Non-filers Monitoring System (NMS) was implemented to prioritise action on non-filers with potential tax liabilities. Salient features of this initiative are:

- Data analysis was conducted to identify PAN holders who had not filed Income tax returns despite conducting high value transaction as reported in AIR, CIB data and TDS/TCS Returns.

- Bulk Data matching exercise was carried out with the Financial Intelligence Unit (FIU) to include non-filers who had conducted high value cash transactions.
- The first NMS Processing Cycle (January 2013) identified 1.22 million non-filers with potential tax liabilities.
- Rule based algorithms were applied to classify the cases as P1, P2, P3, P4 and P5 priority ratings (P1 being the highest priority) for graded monitoring.
- Compliance Management Cell (CMC) was set up for sending letters and capturing responses from the non-filers.
- Bulk letters were sent to PAN holders communicating the information summary and seeking to know the submission details of Income tax return.
- An online monitoring system was implemented to ensure that information related to non-filers is effectively used by the field formation.
- Standard Operating Procedures (SOP) were issued to ensure that the field formations maintain consistency in their approach.

The second NMS Processing Cycle (January 2014) identified additional 2.21 million non-filers with potential tax liabilities. 'Compliance' module was developed on the e-filing portal and information related to non-filers was made available to the specific PAN holder. SMS and email were sent to the target segment asking them to access e-filing portal. The PAN holder is able to provide details electronically and keep a printout of the submitted response for record purposes.

As a result of this initiative, a large number of taxpayers have submitted their Income tax returns and significant amount of self-assessment tax and advance tax has been collected.

b. Identification of potential under-reporting taxpayers

ITD has implemented the computer aided scrutiny selection (CASS) system to select cases for audit using a centralised rules-based system. The salient features of this approach are as under:

- The rules and parameters for selection are reviewed and fine-tuned every year
- Likelihood and quantum of addition made in previous year is computed to suggest modifications in the rules and parameters
- New rules are introduced to cover a broad range of risk scenarios. Third party information is increasingly being used to select cases for audit
- Different financial limits for different geographical areas are used in some rules for equitable distribution of work

- The practice of manual discretionary scrutiny selection by the assessing officers was discontinued from 2013
- Data analysis is conducted to flag 10% of the cases selected under CASS as high priority cases for enhanced follow-up and monitoring

This automated system has brought efficiency and transparency in the process of selection of case for audit.

c. Improving compliance of tax deductors

ITD has used data analytics to improve the compliance of tax deductors in following manner:

- The errors in return and peer group wise data quality analysis are shown to the deductor to improve compliance
- The details of mismatches between tax credit claimed by the taxpayer and tax deduction reported by the tax deductor are analysed to identify high risk deductors for follow up and monitoring

This initiative has resulted in increase of the number of TDS statements filed and the quantum of tax deducted (withheld) at source.

d. Identification of noncompliance in service sector

ITD has conducted a bulk data analysis exercise with the service tax department to identify following categories of persons:

- Persons not registered with service tax department although declaring receipts from services (exceeding the threshold) in income tax return.
- Persons registered with service tax department but not paying service tax although declaring receipts from services (exceeding the threshold) in income tax return
- Persons registered with service tax department but not paying service tax although receiving amount from specified services (exceeding the threshold) in TDS return
- Persons where receipts from services in income tax return is more than 10% of the gross value of service provided in service tax return
- Persons where receipts from services in income tax return is less than 10% of the gross value of service provided in service tax return
- Persons paying service tax but not filing income tax return

While persons in category i) to iv) are being monitored by service tax department, ITD intends to integrate the details of persons in category v) and vi) with the NMS and CASS projects

e. Identification of noncompliance in service sector

ITD has conducted a bulk data analysis exercise with the service tax department to identify following categories of persons:

- Persons not registered with service tax department although declaring receipts from services (exceeding the threshold) in income tax return.
- Persons registered with service tax department but not paying service tax although declaring receipts from services (exceeding the threshold) in income tax return
- Persons registered with service tax department but not paying service tax although receiving amount from specified services (exceeding the threshold) in TDS return
- Persons where receipts from services in income tax return is more than 10% of the gross value of service provided in service tax return
- Persons where receipts from services in income tax return is less than 10% of the gross value of service provided in service tax return
- Persons paying service tax but not filing income tax return

While persons in category i) to iv) are being monitored by service tax department, ITD intends to integrate the details of persons in category v) and vi) with the NMS and CASS projects.

f. Identification of implicit linkages for effective investigation

ITD has implemented the Income Tax Data Management System (ITDMS) which is a two tier distributed system to enable linking of non-PAN data through use of alternate common identifiers. This system is used by the Investigation wing at 20 centres. These linkages have been found to be very useful in identifying family members and other persons related to the individual or entity under investigation.

3. INTEGRATED DATA WAREHOUSE AND BUSINESS INTELLIGENCE (DW&BI) PLATFORM

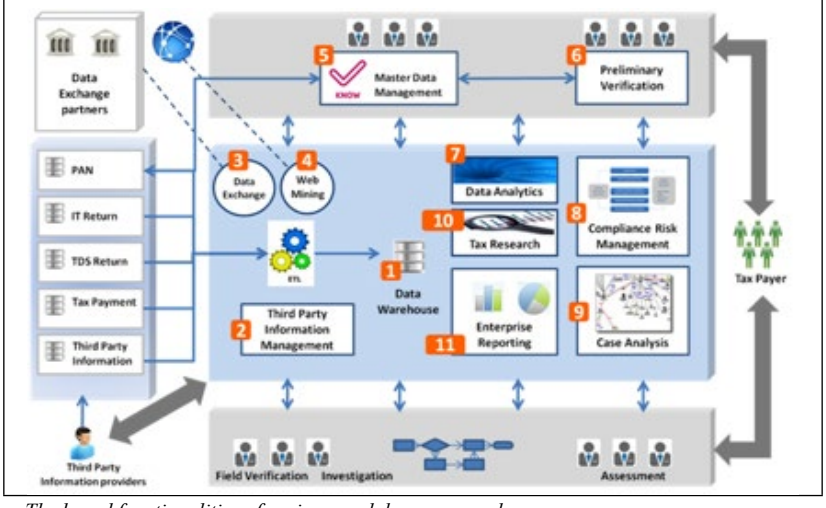
ITD is now implementing a comprehensive Data Warehouse and Business Intelligence (DW & BI) Project to develop an integrated platform for effective utilization of information in all areas of tax administration. The DW&BI platform will integrate enterprise data warehouse, data mining, web mining, predictive modelling, data exchange, master data

management, centralized processing, compliance risk management and case analysis capabilities to achieve the following objectives:

- Widen and deepen tax base
- Improve compliance with tax laws
- Detect fraud and leakage of revenue
- Support Investigation
- Increase effectiveness of tax collection
- Generate enterprise wide reports
- Monitor high risk scenarios
- Provide inputs for policy making
- Tax Research
- Enterprise Reporting

The schematic overview of the proposed DW & BI platform is as under-

Figure: Schematic Overview of Integrated DW & BI platform



The broad functionalities of various modules are as under:

1. Data warehouse: Data warehouse facilitates the collation of data from different sources over the time scale in a manner that is easy for reporting and analysis. The broad functionalities are:

   - Integrate data from multiple source systems, enabling a central view across the enterprise
   - Maintain copy of information from the source transaction systems
   - Improve data quality, by providing consistent codes and descriptions
   - Present the organization's information consistently

- Provide a single common data model regardless of the data's source
- Restructure the data so that it delivers excellent query performance, even for complex analytic queries, without impacting the operational systems
- Maintain data history

2. Third party information management: This module ensures effective collection of third party information. The broad functionalities are:

   - Enlistment of third party information providers
   - Identification of non compliance by information providers (non submission, late submission, poor data quality etc.)
   - Assessment of coverage of reported transactions and assets
   - Monitoring and follow-up

3. Data exchange: Data exchange enables streamlined and secure exchange of information with data exchange partners The broad functionalities are:

   - Enable streamlined process for all types of exchange (request based, spontaneous and automatic) with data exchange partners
   - Develop agreed protocol to ensure inter-operability of data and processes
   - Enable role based Access

4. Web mining: Web mining leverages open source information for improving tax compliance. The broad functionalities are:

   - Crawling and mining of open source information to detect new relationships and identify information relevant to the case
   - Detect latest trends

5. Master data management: This module resolves the mismatches and gaps in the master and transactional data. The broad functionalities are:

   - Identify duplicates in PAN Master / other Masters such as AIN, TAN etc.
   - Populate PAN in records with missing or invalid PAN
   - Reconcile differences in attributes (address, date of birth etc.) of transactional and master data and create authoritative source of master data
   - Identify and resolve relationships between persons
   - Standardize key fields including name and address
   - Process address field to enable geocoding

- Identify records where information in fields (name, address etc.) is insufficient
- Enable online validation of PAN

6. Preliminary verification: This module captures preliminary response of the taxpayer to a compliance issue in an efficient manner. The broad functionalities are:

   - Seek confirmation from taxpayer on resolved identities and information
   - Seek information on whether information relates to taxpayer
   - Seek information about other person to whom the information relates to
   - Promote voluntary compliance
   - Communicate action to be taken by tax payer (e.g. Updation of address in PAN, Quoting of PAN in transactions, submission of return, payment of demand etc.)
   - Communicate list of pending proceedings

7. Data analytics: Data analytics enables identification of useful / actionable information for effective decision making and investigation. The broad functionalities are:

   - Assist in identification and management of risk by using data mining, text mining, neural networks, machine learning, fraud analytics, network analytics, event simulation modelling, subject based mining, spatial data mining
   - Use descriptive analytics to gain insight from historical data with reporting, scorecards, clustering etc. anomaly detection, association rule learning, clustering, classification, regression, summarization
   - Use predictive analytics using statistical and machine learning techniques
   - Use prescriptive analytics to recommend decisions using optimization, simulation etc.

8. Compliance risk management: This module enables effective assessment and management of risk. The broad functionalities are:

   - Identification and assessment of persons not registering, not filing return, not showing appropriate income or not paying appropriate tax
   - Identification and assessment of all types of non-compliance
   - Detection of fraud and leakage of revenue
   - Detection of potential targets (individual or group)

- Analysis of compliance behavior and selection of appropriate treatment strategy
- Monitoring of compliance

9. Case analysis: Case analysis prepares and provides case related information and analysis for effective assessment and investigation. The broad functionalities are:

    - Conduct automated iterative searches to link all available and accessible information
    - Link persons having explicit or implicit relationships with the subject matter of interest
    - Use predefined templates for conducting link analysis, fund flow analysis, ratio analysis
    - Enable linkage of investigation findings, field enquiry results and additional data to the case
    - Ensure integrity / evidentiary value of digital data
    - Present all related information to the user for taking action and capture feedback

10. Tax research: Tax research module facilitates various types of research. The broad functionalities are:

    - Enable research related to economic, structural and revenue aspects of tax policy such as tax buoyancy, revenue stability, revenue foregone analysis, tax gap analysis etc.
    - Facilitate revenue forecasting using regression analysis, predictive modelling, micro simulation models etc.
    - Allow exploration of “what-if” scenarios

11. Enterprise reporting: This module transforms data into meaningful and useful information for decision making, performance management and risk monitoring. The broad functionalities are:

    - Enable adhoc as well as standard multi dimensional analytical (MDA) queries for user in CBDT and field formations with “drill-down” and “roll-up”
    - Publish performance indicators and MIS in standard templates and dashboards
    - Generate report on temporal, sectoral and geographic trends
    - Enable monitoring of high risk scenarios

The design phase of the project commenced in January 2014 and phased implementation rollout is scheduled in 2015-17.

4. CONCLUSION

Availability of information in electronic form provided an opportunity to ITD to develop a wide range of non–intrusive methods for improving compliance.

ITD is now developing a comprehensive Data Warehouse and Business Intelligence (DW & BI) platform to integrate enterprise data warehouse, data mining, web mining, predictive modelling, data exchange, master data management, centralised processing, compliance risk management and case analysis capabilities for effective utilization of information in all areas of tax administration.