

Data Warehousing and Data Mining for improvement of Custom Administration in India – Lessons learnt overseas for implementation in India:

Synopsis: The paper aims to present the usage of Data Warehousing and Data Mining for improvement of Custom Administration in India. It would be an endeavour to examine and explore whether it is possible to extract useful information, pattern, and trend from huge unsupervised data which the systems directorate has in possession in respect of Customs Import and Export transactions. Whether the Data mining technique can dispense the statistical section in Commissionerates, in fact whether at all there is a need to call the data from any field formation, whether certain patterns from unknown data can be formed, whether the large data analysis can throw up patterns, classification models, trends, statistical analysis etc to improve the control over rogue/unscrupulous importer and exporter, organized smuggling, whether it can be used in making Risk Management System More accurate, whether Data warehousing /Data mining can be used to control, predict smuggling, whether data mining can be used for e-governance affects in Custom Administration.

2. The paper would like to suggest certain administrative changes in Man power recruitment to make the systems Directorate a lean and mean organization as well as suggestions to NACEN for conducting courses in Risk Management, Data Analysis and Data Mining.

Data Warehousing and Data Mining for improvement of Custom Administration in India – Lessons learnt overseas for implementation in India:

Introduction:

- Before understanding the usage of Data Mining/Data warehousing in Custom Administration in India, let us first understand in short what is Data Mining/Data Warehousing.

What is Data Mining

- The aim of Data Mining is to make sense of large amounts of mostly unsupervised data, in same domain.

Meaning of word “make sense”:-

The knowledge gained from mined data should be understandable, valid, novel, and useful.

What the term large amounts mean:

The Data Mining (DM) is not about analyzing small data sets. To have a sense of scale of data being good form DM. Let us take some illustrations.

- (a) **Walmart:** In all its stores taken together it handles 21 million transactions a day and stores information of 12 terabytes.
- (b) **NASA:** It generates several gigabytes data per hour.
- (c) **Mobil Oil:** Oil companies like Mobil Oil store hundred of terabytes of data about different aspect of Oil exploration.
- (d) **Homeland Security:** In USA the Homeland Security is collecting petabytes of data on its own & other countries citizen.

What is meaning of mostly unsupervised data

When we have known inputs corresponding to known outputs as determined by domain experts, then data is called supervised data e.g.

e.g. If input is abnormality in Coronary arteries then output is Coronary artery disease. The domain experts here would be Cardiologist.

When the data is open ended, random, unsupervised then DM is used to find “natural” grouping, clusters, relationships in the data. The domain experts viz. Customs officers in this case is very crucial as data miners has to work closely with domain experts (custom officers).

Difference in DM & Statistics

In Statistics, the researchers frequently deal with problem of finding the smallest data size that gives sufficiently confident estimates. In DM, it is exactly the opposite, the data size is very large and our aim is to build the data model that is small/easy to understand, simple but describes the data well. The DM is also known as knowledge extraction, information discovery, information harvesting, data archaeology, data pattern processing, and knowledge discovery.

Data Mining Technologies

DM is an integration of multiple technologies as illustrated below:-

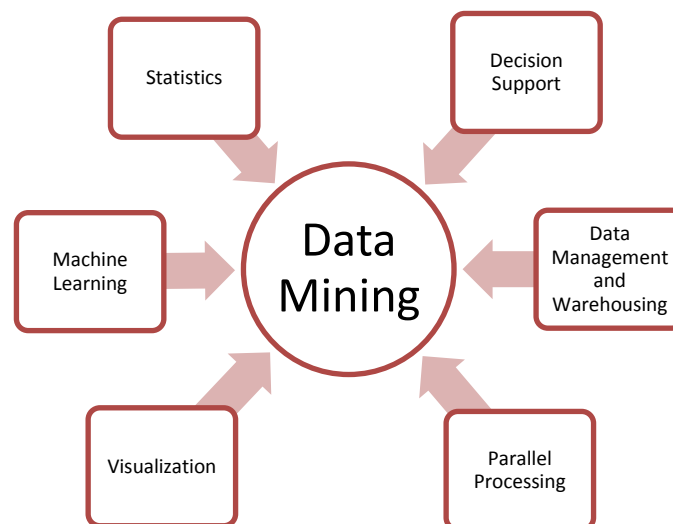


Figure1: Data Mining Technologies

➤ **What is Data Warehousing?**

Data Warehousing is a key data management technology for integrating the various data – Sources and organising the data so that it can be effectively mined.

➤ **Step involved in Data Mining:**

Data Mining is process of extracting previously unknown information while knowledge discovery is defined as the process of making sense out of extracted information. The step consists of:

1. **Developing an understanding the application domain:** This involves working closely with domain experts to define the problem and determine the project goals, identifying key people and learning about current solutions to the problem. A description of problem, including restriction is prepared. DM goals are set.
2. **Understanding data:** The sample data is collected and format and size of data is decided. Data is checked for completeness, redundancy, missing values and outlier. The steps are including the verification of the usefulness of data with respect to the DM goals.
3. **Preparation of data:** This involves sampling running correlation and significance test and data cleaning which includes completeness of data records, correction for missing values. The cleaned data may be further processed by feature selection and extraction along with (reduction of dimension), summarization of data (data granularization).
4. **Data Mining:** Various methods are used to derive knowledge from processed data :
 - a) **Classification:** Records being grouped in meaning for subclasses.
e.g.
 - Companies exporting pharmaceutical worth more than Rs. 20 crores per year.
 - Companies importing grain oriented (CRGO) (Transformer Grade) steel sheets more than 100 MT/year.
 - b) **Sequence detection:** observing pattern in data sequences are determined. Importer X importing one consignment of Machine Made Laces from Taiwan at 20% lower value after successful clearance followed by 5 more consignments at Tuticorin and Nhava Sheva Ports. Importer Y importing undervalued Carpets from Belgium at Kandla Port followed by 10 more consignments at Mumbai Port.
 - c) **Data Dependency analysis:** e.g. the importer imports branded purses and footwear at a very low value along with huge consignment of unbranded garments. The other example would be the same importer importing in small quantities under-valued goods at different ports like Kolkata, Tuticorin, Mumbai AirCargo. Here it is noteworthy that import of high valued branded purses

and footwear is sought to be masked by unbranded garments. In other case the importer is keeping the quantum of imports small to remain discrete and avoid detection of his undervalued cargo at 3 geographically distant stations.

- d) **Deviation Analysis**: Anomalous instances and discrepancies found. The garments imported are not in pieces/numbers but dozens.
- e) **Evaluation of discovered knowledge**: Evaluation includes understanding results, checking whether the discovered knowledge is novel and intervening interpretation of result of domain experts and checking impact of discovered knowledge.
- f) **Use of discovered knowledge**: Planning where and have to use the discovered knowledge. A plan to monitor the implementation of discovered knowledge is created entire project documented. The discovered knowledge is used.

Possible use of Data Mining for Indian Customs: The DM techniques can be used in various applications in Custom Administration in India. Some of these are explained below:

i. Generating Centralized reports:

With data Warehousing (EDW) and Data Mining technique the field formation can dispense the Statistical Section. The Directorate of Data Management has already taken steps in this direction and MIS reports project is under implementation. This will make statistical Section in Commissionerates redundant and staff of these sections may be utilized for other revenue generating purposes. In fact after receiving the Revenue realized data from NSDL the Real Time information (RTI) can be given by use of Data Mining technique on daily basis, by the Directorate of Data Management to individual Commissionerates.

ii. Resolving Valuation related issues

The Valuation Directorate is responsible for keeping National Import Database (NIDB) for valuation of different commodities imported into India through air, land and sea routes, for various Customs Tariff Items under Customs Tariff of India. With the reduction in customs duty rates the problem of Under-evaluation has been greatly reduced over a period of time. However, there are certain commodities like ball bearings, artificial fur clothes, polyester/ nylon fabrics garments which are prone to under

evaluation. Data Mining can be used to generate the NIDB data more accurately, easily and uninterruptedly to resolve existing valuation issues.

iii. Supporting Custom Houses in their day to day functions i.e. implementation of export incentive schemes, preventing drawback frauds, monitoring export realisation

The DM can be used to get the export related data for a particular exporter under export incentives schemes and compliance or default can be made available to the assessing officer that will help him take a decision as regards whether EODC can be generated or otherwise. This will help in easier implementation of DEEC, EPCG schemes.

The DM can be used in assessment of Brand rate Drawback for generating Data for imports, its value, wastages, value addition and following of timelines of exports and claims thereof.

By using appropriate data Warehousing tools in conjunction with RBI data, the Real time Information can be obtained from DM as regards BRC default which can help to keep control on the rebates granted in Central Excise.

iv. The fine tuning of Risk Management System (RMS):

Risk Management Division currently under DG(System) and DRI is extensively using data mining techniques, however TARC committee report has suggested for more professional approach and use of in-depth Data Mining tools for their working to make them more effective by using effectively the DM technique for taking decisions on risk based on

- (i) Import export code
- (ii) Name of Exporter/ Importer
- (iii) Name of Custom House Agent
- (iv) Custom Tariff Item of Import
- (v) Name of Importer/ Address of Importer & Exporter
- (vi) Country of origin
- (vii) DRI Alerts
- (viii) Restrictions based on Foreign Trade Policy

v. Profiling of Tax evaders

The smuggling at airports can be effectively curbed by usage of Data Warehousing and Data Mining techniques as same carriers are repeatedly used by the gangs to smuggle the contraband, gold, silver, consumer items etc. The Photographs, Names, Passport Number can be targeted. This will also help in detection of multiple passports by single person. This will also

help in monitoring the COFEPOSA absconders persons against whom red alert and blue alerts have been issued.

vi. Providing Customer Support Services

The status of containerised cargo covered under Bill of Entry/shipping bill can be provided 24X7 services by SMS alerts, e-mails, telephone calls using the DM techniques.

vii. Fraud Detection

The most important application of DM is too predict the commodities prone to smuggling.

The data analysis may throw up certain pattern or cluster of information which on visualisation may lead to prediction of certain offence like smuggling of certain commodity in form of imported goods or as Passenger baggage. This can lead to issuance of Alerts.

Current Scenario of use of Data Mining Techniques in CBEC

Data analysis and to an extent Data Mining techniques in CBEC are mainly developed and implemented by DG (Systems). Though it is difficult to include all the initiatives related to Data Mining, detail of their pilot project i.e. EDW SmartView which is based on Data Mining is given below:

The Directorate of Systems has implemented EDW SmartView, an Enterprise Data Warehouse which aims to fulfill CBEC's need for a decision support system on its enterprise wide business data

EDW SmartView, is:

- A data warehousing portal, accessible through CBEC's Central Application Interface i.e. Citrix (<http://apps.cbec.gov.in>), using your Single Sign-on ID (SSO ID) and password
- Hosted on CBEC's Consolidated IT infrastructure and does not require to be installed locally
- A central repository for Customs, Central Excise and Service Tax data
- Capable of automated extraction of data from various source systems
- User-friendly interface

An **Enterprise Data Warehouse (EDW)** is a **centralized repository** of large amount of related information and is designed for query and analysis rather

than day to day transactions. It is a system to support management's decision making process.

Purpose of this system is to accumulate data from disparate data sources (ACES, ICES 1.5 etc.) and make it accessible for quick and easy analysis for informed investigation and policy making.

Now it is being upgraded to create a complete set of report known as MIS reports

A data warehouse is thus designed:

- For fast retrieval of data to facilitate analytical reporting
- To lighten the reporting load from a transactional system (like ICES 1.5, ACES)
- To enable an organization to consolidate data from several sources (like ACES, ICES 1.5 etc.) and make it easily accessible through the use of simple BI tools

The current source systems for which the data has been brought into the EDW is as under:

Customs: The data source for Customs data is ICES 1.5, the online workflow application for clearance of Imports and Exports in India. EDW Customs reports are currently based on data from electronic Bills of Entry which have been granted Customs' Out of Charge and electronic Shipping Bills which have been granted Let Export Order (LEO) status as of the date of EDW data refresh. This includes the historical data from ICES 1.0 also, for the sites that have migrated to ICES 1.5.

Central Excise and Service Tax: Central Excise and Service Tax data in EDW is sourced from ACES, the online application for various Central Excise and Service Tax related functionalities. At the moment, only Registrations and Returns data from ACES is available in the EDW.

EASIEST: The Electronic Accounting System in Excise and Service Tax (EASIEST) application makes available accurate tax payment data for revenue and tax payer accounting. EASIEST data is received in the Data Warehouse from the banks through the NSDL gateway and is loaded into the Data Warehouse.

In **EDW SmartView:**

- Data is stored at multiple levels of summarization for faster reporting at the summary as well as detail level

For example: Bill of Entry level data from ICES 1.5 / Returns data from ACES is summarized at Commissionerate level for Commissionerate wise reporting. This data is also available item wise for Bill of Entry / Return level reporting.

In a data warehouse, the data elements are identified as **Measures** and **Dimensions**.

Measures are:

- Important business indicators of an organization
- Factual data about the subject area
- Generally numeric (values) that represent a specific business aspect or activity in a summarized form

Example: Quantity, Assessable Value, PLA, CENVAT, Duty Paid, Duty Foregone etc.

Dimension:

- Is a data element that categorizes each item in a data warehouse
- Qualifies the measures

Example: Customs Tariff Head, Time (Financial Year or Month), Importer Exporter Code, Assessee, Customs or Central Excise Commissionerate, Budget Head etc.

The reporting and analytical functionalities in EDW SmartView are enabled by a business intelligence software called **IBM Cognos**.

Apart from above, Risk Management Division which was earlier under Systems Directorate and now has been placed under Directorate of Revenue Intelligence is extensively using Data Mining techniques for risk management purpose. Though their risk architecture is not disclosed due to its sensitivity, it is understood that for identification of risk they are using various tools related to data mining.

DRI is another organization which use basic data mining tools for fraud detection. They took the support of system database for detection of fraud cases. Looking into the sensitive nature, their internal details have not been released.

Directorate of Valuation is another organization where data mining techniques are used for generating National Import Database giving the transaction values of various commodities imported at various ports and Air Cargo.

Recently Directorate General of Audit has started using Data Mining tools for Audit Planning. They are selecting the units to be audited on the basis of risk parameters using database of EDW. They are selecting 25,000 units out of more than 2 lakhs assesses on the basis of risk parameters using the Data Mining Techniques.

Last but not the least, Tax Research Unit of CBEC use the Data Mining techniques for policy formation. They are using DM techniques for revenue forecasting in case of increase or decrease of Tax rates, impact study due to rate changes, impact of exemption notifications etc.

International perspective/Lessons Learnt:

During the visit abroad, the Chief Intelligence Officer of HMRC Mr. Bob Mathewson has interalia had stated that they in HMRC go for strategic picture of Risks (SPR). The HMRC takes action depending upon tolerance level accepted, resources deployed, the costs involved and geographical locations involved. The HMRC also takes into consideration the Return on Investment in action sought to be taken. It also takes into consideration to see how well their Department is changing the behaviour in respect of Customer satisfaction.

The Customers according to HMRC are:-

- (i) Large Business
- (ii) Mass Market
- (iii) Organized Crime
- (iv) Small Business
- (v) Micro Segment/Individual Imparters/passengers etc.

The Departmental response of Strategic picture of Risk is multitiered & is as follows:-



Analysis of Data generated

- (i) The HMRC has sought to understand the Gaps in their understanding of the situation.
- (ii) What would be commodity in which fraud is likely to be committed?
- (iii) Is Cyber Crime a threat/problem. Internal assessment & safety measures.
- (iv) What are main geographical locations of Oil/Petrol smuggling.
- (v) What is next move/ next area of strike by Organizational Crime.

The HMRC deploys 100 specially trained Data Analysts for the job. In Netherland Customs, it was revealed that along with Data Analysts, they also employ psychologist. The HMRC gives due importance to reporting. The reporting has following characteristics & time lines.

Response reports

These reports are one page reports issued basically for early warning/Alerts, issued when there is imminent threat of fraud/ terror threat & needs quick response. The reports may be of good source information and high confidence. If evidence is open to interpretation then report is of moderate confidence & others of lesser confidence. Thus they explicitly classify their reports being of high confidence level, moderate confidence and low confidence.

Focus Report

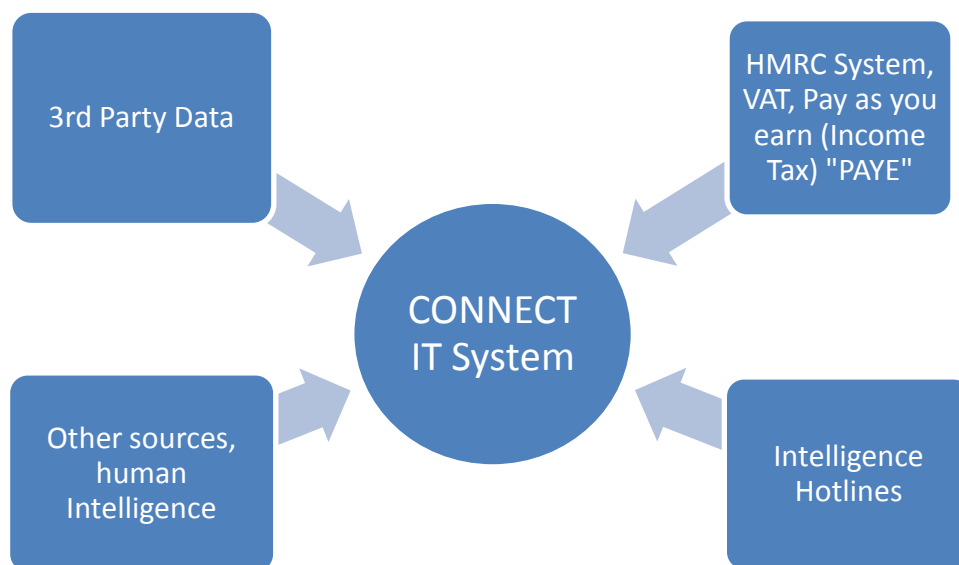
These reports are issued when in-depth assessment of any specific aspect under focus of Customs Administration.

Strategic report

These are quarterly reports covering all aspects of Customs Administration covering import/export, intelligence, Revenue, Adjudication, Administration, passengers cargo seized etc.

Identification of Risks by HMRC by system called 'CONNECT'

The flexible risk analysis based on network:-



Integrated compliance Environment (ICE), Analytical Compliance Environment (ACE), Automatic Output of cases and Platform for Data Analysis of mined data.

The custom entities are interconnected in a “Supernetwork”. This is given name of “Soup”. The data is mined from the Supernetwork for building Risk profile using

- (i) Offshore evasion data
- (ii) False invoicing data
- (iii) Hidden economy data (Income Tax evasion)
- (iv) Refund Fraud
- (v) VAT Fraud

The Bulk data is mined centrally for gathering information by HMRC with Integrated Compliance Environment (ICE) with Human Intelligence and as a response the importer/assessee is also visited where need be.

Recommendation:

- 1) Presently, Data Warehousing/Data Mining is hardly used in Customs or Central Excise field formation. 2 or 3 licenses per Custom Commissionerate are given for access to Enterprise Data Warehouse (EDW). However at field level these are seldom used. There is same usage of DMT in System Directorate, however it is not extensive. In our opinion, there is no need of gathering data from Custom Houses for monthly revenue, as this is available in Commissionerate wise on NSDL and the Directorate of Data Management can be mandated to do the Data mining work for getting monthly report on Revenues. The field formation can only gather the revenues for verifying correctness of Data Mined centrally by Directorate of Data management. Similarly, the total duty foregone, drawback paid also can be mined by Directorate of Data Management. Hence monthly Revenue reports from Commissionerate can be dispensed with using Data Mining technique.

- 2) It was gathered during our visit to UK that HMRC employs more than 100 highly qualified Data Analysts to analyze the data mined and do the Risk Analysis. It appears the efforts in cases of Indian Customs are not adequate. The Tax Administrative Reform Commission (TARC) constituted by Government of India under Chairmanship of Shri Parthasarathy Shome, noted economist, has submitted the second report to the Government in Sept. 2014. The TARC has interlia observed:-
 - (a) The growth in volume as well as Complexity of International trade has naturally lead to an increase in work load to Customs. The furnished by CBEC shows that from 1997 to 2014. The number of Import Documents processed by Customs has gone up 7 times. There has been 20 times increase in the value of imports from Rs. 1,54,000 crore to Rs. 30,00,000 crore and Customs revenue has increased from Rs. 41,000 crore in 1997-98 to Rs. 1,75,000 crore in 2013-14. The world has become ever more interconnected and interdependent through expanded crore broader flows of goods, services, people, transport, capital, information & technology. Globalization makes it easier to conduct international business

than in the past and provides economies with the opportunity to fast track development goal through increased international trade.

- (b) At one end of spectrum are clients who are highly compliant in that they exhibit both the capacity and commitment to compliance. The approach towards them has to be highly facilitative and partnership oriented with the full range of facilitation benefits extended. At the other end, there are those who have extreme disregard for law. The goal of strategic Risk Management is to progressively and consistently move the client population towards the complaint end of the spectrum thereby improving the overall compliance environment. The success in such an environment can be achieved through extensive research and analysis with multi-disciplinary skills such as those of Data Analysts, Social Scientists, Customer Service specialists and domain experts. Such research needs to be continually evolving process. This means reconfiguring estimated risk levels, introducing new technologies, create new capacities and sharing more risk with other supply chain participants.
- (c) The TARC has recommended creation of functional vertical in the form of Strategic Planning and Risk Management Directorate with in CBEC to impart a strategic dimension to its efforts by anticipating major challenges and responding ahead of time so that threats to compliance are effectively mitigated. This would require to continuously scan the environment for emerging trends of business practise, technologies etc and prepare organization by planning the required human, organizational and technological capacities to either cope with threats to its mandate or to exploit the potential emerging trends may offer.
- (d) The TARC has also advocated for a Knowledge and Analysis Centre (KAC) comprising of a range of data and data analytical skills to support strategy, policy making and operations.
- (e) The TARC has sought to strengthen the Risk Management Division (RMD) and need of revamping it to assume more active strategic as well as operational role in Customs Risk Management. At strategic level, it seeks that research output should support CBEC in developing programme and policies as outlined earlier. At operational level, it needs to build technological and human capacity to the advanced analytical tools to engage on predictive analysis and to improve risk assessment to such a level of accuracy to allow virtually to legitimate traders to continue their business intervention and to allow for remainder to be targeted. The RMD should develop new algorithms and invest in sophisticated search

that will improve its ability to identify both individuals and Cargo for intervention.

- (f) The TARC has significantly noted that feedback and re-evaluation are critical component of Risk Management Process. However, it has lamented that it is not happening adequately in RMD. It has advised RMD to sharpen the Risk Rules, targets or interventions inserted by national and local Risk Managers to improve the quality of matches with suspect profiles so that large number of consignments is unnecessarily not checked. It has advised CBEC to progressively move away from local approach in Risk Management to a Strong national approach and more forward setting up a national facility in US etc.

Since TARC has advocated for formation of KAC (Knowledge and Analysis Centre) and predictive analysis, it is imperative to greatly improve the Data Mining/Data Analysis skills of the staff in Customs Administration. Hence, following recommendations are being made.

1. There should be thorough revamp of System Directorate including RMD. There should be a lateral entry/recruitment of Data Mining/Customer Specialist, Data Analysts, Domain Experts, Social Scientists, Customs service Specialist, even psychologist to enhance the Data Mining and Research ability of Department to do research and analysis of Mined Data.
2. The Government may at the earliest set up a Knowledge and Analysis Centre (KAC) with major mandate to oversee Data Warehousing and Data Mining. The KAC be headed by Additional Director General level officer and he be called as “Chief Data Officer” of Customs.
3. The KAC be manned by highly qualified and experienced personal with Data Mining and Data Analysis skill. This should act as think tank for policy making as for as Risk Management System is concerned.
4. There should be through revamp of Risk Management Division (RMD). There also should be lateral entry in RMD of Risk Analysts, Data Mining specialist, domain experts, Social Scientist, Customers Service Specialist. The Additional Director General Should renamed as Chief Risk Officer of Indian Customs. This gives a professional flavour to the designation. The RMD should be mandated to do the strategy decisions and development of centralized Risk Management (Enterprise Risk Management) and focus

should be on Centralized Risk Management and should be driven by central offices and not by local Risk Management Committee (LRMC).

5. The recruitment/posting of ADC/JC, AC/DC & Superintendent in R.M.D. should be on the basis of job profiles and specific skills. They should be liberally trained in Data Mining Technique if the person has aptitude and level of competency.
6. NACEN should permanently employ in its faculty the persons with qualification of Data Mining, Data Analysis, Risk Management Experts. They should not merely arrange the lectures on Data Mining or Risk Analysis/Risk Management for a day or two but NACEN should organise short term courses and may consider giving its own certificate/Diploma in Business Analysis/Data Mining/Data Analysis, Risk Management and only those officer, AC/DC level, JC/ADC level be posted in RMD who has passed certificate level courses of NACEN in Risk Management. This will help in Co-ordination between laterally recruited domain experts and Departmental Officers monitoring RMD.
7. Data Mining technique should be used in big way to do 360^o profile of known-carriers/frequently fliers to Bangkok/Dubai/Singapore/Hong Kong (sensitive cities) who bring in Contraband and predictive Alerts to be used by Air Customs officers at All Airports.
8. More than 80-90% cargo may be going through RMS but Risk Rules need to be sharpened and more effective with help of Data Mining Techniques.
9. The Data mining technique can be used to make much effective Impact analysis in Risk Mapping/Risk Matrix. This may improve R.M.S. which intern may increase clearance of the goods and reduce the dwell time of goods.

Challenges:

The recruitment of Data Mining experts and establishment of Knowledge and Analysis Centre as hub of Data Mining, Analysis and Policy think tank is indeed a daunting task. However, the Tax Administration Reform Commission Report (popularly called as Parshasarathy Shome Committee report) have strongly recommended creating an environment for extensive research and analysis through employing multi disciplinary skills of Data Analysis, Social Scientist, domain experts (Data Mining experts) and Customer Service Specialist and predictive analysis and improve accuracy of Risk assessment and creation of knowledge and Analysis Centre. Hence,

though one feels apprehensive about creating infrastructure for knowledge and Analysis Centre (KAC) recruitment of highly qualified and experienced domain experts in Data Mining appears to be difficult in near future; However, the winds of change are blowing due to probable introduction of GST. The strong recommendation by Shome Committee also may possibly help in giving effect of usage of Data Mining in Custom Administration at an reasonably early date.

Conclusion:

By using Data warehousing (DW) and Data Mining Technique (DMT), the Department will be able to do research in large Data which is generated and Custom Administration will be able to:

- Reconfigure the estimated risk level.
- Introduce new technologies and analytical techniques.
- Create new capacities.
- Share more risk with other supply chain participants.
- Create the Knowledge and Analysis Centre (KAC)
- Create a functional centre for strategic policy planning.
- Improve functioning of Risk Management Division (RMD)
- Predict and anticipate frauds both in import of contraband goods and also goods Smuggled by passengers in conjunction with DRI
- Response ahead of time so that threats to compliance are effectively mitigated.
- NACEN can transform itself into a centre for giving good education in Data Warehousing, Data Mining, Advanced Risk Management techniques by organizing short term (1-2 months) courses for staff for human, organizational and technological capacity building in Customs Administration of Indian Customs.
- Data mining will ensure that RMD will be using advanced analytical tool to engage predictive analysis and to improve the accuracy and effectiveness of Risk assessment and possibly reduce the dwell time to improve the India's rating in ease of doing business.
- Data Mining can be used to accurately collect Commissionerate wise monthly revenue thus monthly revenue reports by Commissionerate can be thing of the past.
- Data mining can be win-win situation for Custom Administration and for trade.

Bibliography:-

- (1) Data Mining- Acknowledge Discovery Approach
Author- Krzysztof J. Cios, Wifold Pedrycz, Roman W. Swinianski , Lukasz A. Kurgan

- (2) Data Mining:- Technologicals, Techniques, Tools and Trends
Author:- Dr. Bhavani Thuraisingham
Publisher:- M/s CRC.
- (3) Second report of Tax Administration Reform Commission, Ministry of Finance, Government of India, Published in September 2014 (popularly called as Shone Committee Report)
- (4) Lecture by Bob-Mathewson- Head of Intelligence Assessment given before MCTP participants o 05/08/2015 at LSE, London. (e-mail:- bob.mathewson@hmrc.gsi.gov.uk