A DATA MINING FRAMEWORK TO DETECT TARIFF CODE CIRCUMVENTION IN TURKISH CUSTOMS DATABASE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

BURCU BAŞTABAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

JULY 2012

A DATA MINING FRAMEWORK TO DETECT TARIFF CODE CIRCUMVENTION IN TURKISH CUSTOMS DATABASE

Submitted by **BURCU BAŞTABAK** in partial fulfillment of the requirements for the degree of **Master of Science in Information Systems, Middle East Technical University** by,

Prof. Dr. Nazife BAYKAL _____
Director, **Informatics Institute**

Prof. Dr. Yasemin YARDIMCI ÇETİN _____
Head of Department, **Information Systems**

Assist. Prof. Dr. Tuğba TAŞKAYA TEMİZEL _____
Supervisor, **Information Systems, METU**

Examining Committee Members:

Prof. Dr. Sinan KAYALIGİL _____
Industrial Engineering, METU

Assist. Prof. Dr. Tuğba TAŞKAYA TEMİZEL _____
Information Systems, METU

Assoc. Prof. Dr. İnci BATMAZ _____
Statistics, METU

Assist. Prof. Dr. Aysu BETİN CAN _____
Information Systems, METU

Assist. Prof. Dr. Banu GÜNEL HACIHABİBOĞLU _____
Information Systems, METU

Date: 19.07.2012

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Burcu BAŞTABAK

Signature : _____

ABSTRACT

A DATA MINING FRAMEWORK TO DETECT TARIFF CODE CIRCUMVENTION IN TURKISH CUSTOMS DATABASE

Başţabak, Burcu
M. Sc., Department of Information Systems
Supervisor: Assist. Prof. Dr. Tuğba Taşkaya Temizel

July 2012, 81 pages

Customs and foreign trade regulations are made to regulate import and export activities. The majority of these regulations are applied on import procedures. The country of origin and the tariff code become important when determining the tax amount of the merchandise in importation.

Anti-dumping duty is defined as a financial penalty, published by the Ministry of Economy, enforced for suspiciously low priced imports in order to protect the local industry from unfair competition. It is accrued according to tariff code and the country of origin. To avoid such an obligation in order to not to pay tax, a tariff code that is different from the original tariff code may be declared on the customs declaration which is called as "Tariff Code Circumvention". To identify such

misdeclarations, a physical examination of the merchandise is required. However, with limited personnel resources, the physical examination of all imported merchandise is not possible.

In this study, a data mining framework is developed on Turkish customs database in order to detect "Tariff Code Circumvention". For this purpose, four types of products, which are the most circumvented goods in the Turkish customs, have been chosen. First, with the help of Risk Analysis Office, the significant features are identified. Then, Infogain algorithm is used for ranking these features. Finally, KNN algorithm is applied on the Turkish customs database in order to identify the circumvented goods automatically. The results show that the framework is able to find such circumvented goods successfully.

Keywords: Tax Evasion in Customs, Anti - Dumping Tax, Tariff Code Circumvention, Data Mining, K – Nearest Neighbor Algorithm

ÖZ

TÜRK GÜMRÜK VERİTABANINDA GTİP SAPTIRMASI FAALİYETLERİNİ ALGILAMAK İÇİN BİR VERİ MADENCİLİĞİ ÇERÇEVESİ

Baştabak, Burcu
Yüksek Lisans, Bilişim Sistemleri Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Tuğba Taşkaya Temizel

Temmuz 2012, 81 sayfa

Gümrük ve dış ticaret mevzuatları, eşyaların ithalat ve ihracat faaliyetlerini düzenlemek için konulurlar. Bu mevzuatların önemli bir kısmı ithalat usullerine uygulanır. Eşyanın ithalatında vergi tutarı belirlenirken eşyanın menşei ülkesi ve gümrük tarife istatistik pozisyonu (GTİP) önemli olmaktadır. Anti-damping vergisi, Ekonomi Bakanlığınca düzenlenen, yerli sanayiyi haksız rekabetten korumak için düşük fiyatlı ithalata uygulanan para cezası olarak tanımlanır. Tarife kodu ve menşei ülkesine göre tahakkuk ettirilir. Vergi ödememek adına böyle bir yükümlülükten kaçınmak için, gümrük beyannamesinde gerçek tarife kodundan farklı bir tarife kodu beyan edilebilmektedir. Buna "GTİP Saptırması" denir. Böyle yanlış beyanların tespiti için, beyanname konusu ürünün fiziksel muayenesi gereklidir. Bununla

birlikte, kısıtlı personel kaynağı ile tüm ithalat ürünlerinin fiziksel kontrolünün yapılması mümkün değildir.

Bu çalışmada, GTİP saptırmasını tespit etmek için Türk gümrük veritabanında bir veri madenciliği çerçevesi geliştirilmiştir. Bu amaçla, Türk gümrüklerinde en çok GTİP saptırmasının yapıldığının tespit edildiği 4 ürün seçilmiştir. İlk olarak Risk Analizi Dairesinin de yardımlarıyla, anlamlı özellikler belirlenmiştir. Sonrasında, Bu özelliklerin sıralaması için InfoGain algoritması kullanılmıştır. Son olarak, saptırılmış ürünleri otomatik olarak belirlemek için Türk gümrük veritabanında k-en yakın komşu algoritması uygulanmıştır. Sonuçlar, çerçevenin böyle saptırılmış ürünleri başarılı bir şekilde bulabildiğini göstermiştir.

Anahtar Kelimeler: Gümrükte Vergi kaçakçılığı, Anti-Damping Vergisi, GTİP Saptırması, Veri Madenciliği, K – En Yakın Komşu Algoritması

To My Parents, My Brother and Nej

For their endless support,

For their love…

ACKNOWLEDGMENTS

I would like to express my special gratitude to my advisor Assist. Prof. Dr. Tuğba TAŞKAYA TEMİZEL for her invaluable support, suggestions, guidance and patience throughout this study. Without her support, this thesis could not have been completed. I am also grateful to my thesis committee for their suggestions and valuable comments.

I would like to express my sincere gratitude to inspection officers of Ministry of Customs and Trade, Önder DOĞAN and Bekir DURUSOY, for sharing their invaluable knowledge and for their patience to my endless questions throughout this study.

Thanks to Nurcan ÖZYAZICI SUNAY, Head of Department of Information Technology, for supporting me to study on customs processes and allowing me to use data warehouse of the Ministry of Customs and Trade.

Special thanks to my parents and brother who always encouraged me with their love, understanding, patience, and emotional support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EU	: European Union
EFTA	: European Free Trade Association
GÜVAS	: Customs Data Warehouse System of Turkey
KNN	: K - Nearest Neighbor
NTBs	: Non-Tariff Barriers
UNCTAD	: United Nations Conference on Trade and Development
USA	: United States of America
WCO	: World Custom Organization
WTO	: World Trade Organization

TERMS AND NOTATIONS

$V_{i,k,m,t}$: The worth of the *i* good, imported by company *k*, in year *t*, from the country of origin *m*.

$G_{i,k,m,t}$: The gross mass of the *i* good, imported by company *k*, in year *t*, from the country of origin *m*.

$W_{i,k,m,t}$: The net mass of the *i* good, imported by company *k*, in year *t*, from the country of origin *m*.

$U_{i,k,m,t}$: The statistical unit of the *i* good, imported by company *k*, in year *t*, from the country of origin *m*.

V : The declared worth of the good.

G : The declared gross mass of the good.

W : The declared net mass of the good.

U : The declared statistical unit of the good.

CHAPTER 1

INTRODUCTION

1.1. Motivation

Customs and foreign trade regulations are made to regulate the international movements of goods. The majority of these regulations are applied on import procedures. Collected under these regulations, customs duty is one of the major revenue of the state. Apart from customs duty, there are also additional financial obligations and surveillances. The tariff code and the country of origin play an important role on determining the tax rate and the legislation applied to the imports. In order to determine the amount of the tax, the worth of imported merchandise is taken into account.

Taxes collected by customs are classified as tariff and non- tariff barriers, in principle. The upper limit of the tariff is determined by the World Customs Organization (WCO) and enforced by the member states. Countries may enact a number of measures to protect their domestic producers and consumers. Defined as non-tariff barriers, these measures have found global acceptance.

Anti-dumping duty, which is one of the non-tariff barriers, is defined as a financial penalty, published by the Ministry of Economy, enforced for suspiciously low priced imports in order to protect the local industry from unfair competition. They are drawn up and enforced by the Ministry of Economy. The Ministry of Economy initiates an investigation pursuant to the application made by the sector.

As a result of the investigation, if it is determined that domestic industry is damaged, implementation of an anti - dumping duty is decided. Anti - dumping duty is accrued based on the tariff code and the country of origin, and may be charged as a certain rate or ad valorem, based on the type of the product.

To avoid such an obligation in order to not to pay tax, various irregularities on the customs declaration can be made. One of these irregularities is to declare a tariff code that is different from the actual tariff code. It is called as "the tariff code circumvention". Besides tariff code circumvention, in order to avoid the taxes, country of origin circumvention or declaring low worth of the good can be attempted. To detect the false declared information, a physical examination of the product stated in the declaration is required. So, the product can be compared with the information disclosed, corrected if it is wrong, and the tax rates to be applied can be determined. However, with limited personnel resources, it is not possible to physically control all import goods. For this reason, making the physical examination on the high-risk operations which may include irregularities in reality will increase the efficiency, and reduce the crime rate.

Whether a declaration includes a risk factor is determined by running the Risk Analysis System after the registration of the customs declaration. The Risk Analysis System is fed with the risk profiles created as a result of the studies done by the Risk Analysis Office of the Ministry of Customs and Trade. Each of these risk profiles is a combination of the information on the declaration. Declarations matching with the risk profiles of the risk analysis system are directed to the red line for a physical examination.

For the creation of risk profiles, studies are done on the tariff codes that can be circumvented base on the legislation or the tariff code irregularities detected during physical examination at Risk Analysis Office. In this study we modeled the behavior

of the companies, to capture these circumventions on the basis of past declarations of the companies and to create risk profiles attempting the circumventions.

In this study, in order to detect the tariff code circumvention to avoid anti-dumping duty, a data mining framework has been developed for the Turkish customs on customs database. Thus, it is intended to support the work done by the risk analysis department. In this study, based on the submitted information in the declaration and the historical records of the company, by using the declarations detected as circumvented, the declarations submitted to the customs are predicted whether they are risky or not.

1.2. Thesis Overview

By following the introduction chapter, a detailed literature survey is given in Chapter 2. Literature survey consists of three main parts. In the first part, the studies on risk analysis at customs in Turkey are discussed and the studies comparing Turkey with other countries and international unions are presented. In the second part, studies conducted on Non - Tariff Barriers (NTBs) and anti-dumping measures are mentioned. Finally, studies on risk analysis at the customs in other countries are focused on.

In Chapter 3, we explain customs regimes, the work flow of the import process which is one of these regimes, taxes applicable to import procedures and the non-tariff barriers. Then, the anti-dumping duty which is one of the non-tariff barriers, its effects, its objectives and the investigation process are explained. Finally, in the light of this information, the effects of tariff code circumvention and the difficulties experienced at the detection of tariff code circumvention are mentioned.

In Chapter 4, the study made by Risk Analysis Office in order to detect tariff code circumvention explained in Chapter 3 is described in detail. The structure of the dataset, the meaning of the features, feature selection and method selection are described. Summary information about the selected products is also presented in this chapter.

In Chapter 5, the results of the application of the method described in Chapter 4 are interpreted and the comparisons are made between the experiments.

Finally, in Chapter 6, the inferences gained from the study and recommendations for future studies are presented.

Note here that, the tariff code circumvention problem stated in this study was presented in Kuşadası (International Conference on eBusiness & eGovernment (ICEBEG), Kuşadası, Turkey, 16 - 17 April 2011), and published at International Journal of eBusiness and eGovernment Studies (Baştabak & Medeni, 2011).

CHAPTER 2

LITERATURE SURVEY

Customs is an authority in a country responsible for collecting customs duties and controlling the movement of the goods. Carrying out these responsibilities, customs are expected to facilitate legitimate trade and prevent illegal operations.

In Turkey, there are sixteen regional directorates of customs and trade and a hundred and forty-seven customs administrations grouped within the regional directorates. There are six types of administrations, regarding the location and function of the authorities:

1. Sea port customs
2. Airport customs
3. Internal customs
4. Border customs
5. Railway customs
6. Free zones

In customs administration, officers are basically faced with two main illegalities.

1. The illegal movement of the goods that are forbidden to trade.
2. The illegal trade of the goods in order to avoid legislative measures.

The detection of these illegalities is possible only by physical control of the goods and vehicles. However, with limited technological capabilities (X-Ray, etc.) and staff, it is not possible to physically control all the transactions to detect customs fraud. With these competing aims, finding a balance between trade facilitation and control is essential. For this reason, the use of risk analysis methodology at the customs procedures is of great importance in order to determine which transactions should be checked. (UNCTAD & WCO, 2008).

A wide variety of measures is available in our legislation related to trade of goods. Besides the customs duty, there are also trade measures such as additional financial obligations, document supply, and extra control of the goods.

In order to avoid these measures, the trader may apply illegal methods such as;

1. Present an incomplete and / or false declaration.
2. Provide false or misleading additional documents.

Studies on the basis of the subject of this thesis can be handled in three groups:

1. Findings and recommendations for the risk analysis methodology in customs
    - Cross-country comparisons and recommendations for risk analysis
2. NTBs and anti-dumping taxes in Turkey
    - Cross-country comparisons and recommendations for NTBs
3. Data mining applications at risk analysis methodology

2.1. Risk Analysis Methodology in Turkey

Risk analysis and assessment procedures are used to determine the risk of customs transactions. In some expertise theses of Ministry of Customs and the Trade (formerly the Undersecretariat of Customs), by determining the situation; suggestions were made to create a better risk management process.

In Öztürk (2005)'s study, presenting the current state of Turkish Customs Administrations, the importance of using information technologies and exchanging of information with the customs administrations, the traders and other public agencies and organizations were mentioned. In addition, by taking an international approach, the following recommendations were put forward for Turkey;

- Saving the inspection results (positive or negative) to the computer system and using this information to use for creating risk profiles.
- Granting authority to determine the risk profiles and the risk criteria to the local administrations.
- Increasing the functionality of the Customs Data Warehouse System (GÜVAS).
- Enabling the control of the risk areas based on contraventions.
- Statistical analysis of risk.
- The improvement of organizational structure.

One of the studies was about the prevention of smuggling by Can (2005). In this study, the importance of risk management and intelligence to struggle against smuggling was stressed. The traditional customs control methods based on the frequencies and the random selection criteria did not satisfy the objectives of the Customs Administrations. For this reason, the implementation of the risk-based and the intelligence-aided controls were emphasized as major requirements.

2.2. Cross-Country Comparisons for Risk Analysis

The systems of five countries (America, Australia, Netherlands, Germany and France) that use the risk management system effectively at the customs administrations were examined, and a model that proposes a risk management system for the Turkish Customs Administration was developed by Avcı (2001). The proposed risk management system for the Turkish Customs Administration was formed by choosing the similar and the appropriate aspects of the customs risk management systems of the investigated countries. According to this proposed model, the most important and common approach of the customs administrations of the countries successfully implementing the risk management system was "Trade Adjustment" approach. Trade Adjustment refers to the compliance of traders'

commercial system with the legislation of customs and trade. Within the framework of this approach, instead of seeing traders as potential criminals, customs should accept them as business partners. In order to establish an effective risk management system, another suggestion of the model was to prepare the necessary plans and programs for this system by a committee constituted by the top-level managers. These plans and programs should be implemented by expert working groups, and also be monitored by a team of senior managers.

2.3. Non-Tariff Barriers and Anti - Dumping Measures in Turkey

There are a variety of risk factors at the customs. The smuggling and illegal trading in commercial sense are these risk factors. At smuggling, there is a movement of prohibited goods such as weapons and drugs. Illegal trading is to carry out the transactions of the goods having a number of controls, additional financial obligations or document requirements through a variety of ways to bypass the legislation of import/export.

NTBs are for this kind of legislations. The use of this kind of barriers to protect domestic producers and consumers are common in Turkey as it is throughout the world.

In Başaran (2004), on surveillance and safeguard measures on imports, measures were separated into two: those were applied to the textile industry, which was one of our leading sectors, and those ones that were not applied to the textile. In the same study, about the anti-dumping measures, being a widely used protective tool, the importance of the effect of pressures demanding protectionism by domestic companies that cannot compete with foreign companies was discussed. Protectionism is, interfering in the foreign trade by the state, to protect the domestic producers against foreign competition.

Not only positive but also negative effects of anti - dumping measures were mentioned by Beğen (2007). When a country initiates an anti-dumping practice, this situation is followed by other countries. This is called as "domino effect". In other

words, the anti-dumping practices may lead to practices in the entire world. Even if anti-dumping measures serve the purpose of protecting domestic industry of the country, by increasing the price of imported goods or reducing the amount of imports, prevent or delay the structural adjustment of the domestic industry to the free competition market, and reduces competition. Consumers are deprived of the ability to import goods more cheaply.

Unlike other studies, in the study of Cete (1998), the issue of "Tax avoidance by deceptive practices" was mentioned as well as comparisons of legislations.

2.4. The Comparison of the Countries Using Non-Tariff Barriers and Anti - Dumping Measures

Keskin (2011) investigated what led to unfair competition according to the World Trade Organization (WTO) rules on imports. In the same study, the anti-dumping legislation of Turkey and the European Union (EU) was analyzed and compared in detail. According to this comparison, the following differences were identified:

- In EU, anti-dumping measures should come into force a maximum of 15 months after the investigation was opened. This period is 12 months in Turkey in terms of anti-dumping taxes.

- The community interests need to be observed in EU to be put into action of measures. Such a regulation does not exist in Turkey.

Trade diversion, which might be caused by measures against unfair competition in imports, and suggestions about measures against this diversion were also included in the study. Trade diversion is the change of the trade direction and / or flow due to the application of tariff or non-tariff measures in different countries.

Development of protectionism in Turkey and in EU was examined in the historical context, by Dalkıran (1998). The Union developed its own anti-dumping legislation in parallel with the provisions on anti-dumping of the WTO agreement.

Protectionist policies are still applied today. According to Dalkıran (1998), the structural imbalances lay at the origin of the protectionist policies. While these

structural imbalances increase, the protectionist policies will continue to rise. So in the international arena, while the difference occurs in growth rates of productivity in various sectors of trade partners and getting the results of structural adjustment measures extend over time, protectionist measures of quality will change, perhaps, but will continue to exist. Unlike other studies, Dalkiran analyzed the effect of NTBs implemented by America, Japan, Russia, Saudi Arabia and Azerbaijan.

In Karakoç (2008), the types of non-tariff barriers, historical development of the barriers in Turkey and in the world, and national and international legislation about these barriers were examined. Foreign trade practices of Turkey, EU and some countries, such as America, Russia and China, were discussed.

2.5. Data Mining Applications at Risk Analysis Methodology

Data mining has become a frequently referenced area in risk analysis to detect frauds. But, there are still some problems. One of them is a limited number of fraud data present in datasets. As a result, the accuracy of the model may reflect the success of the model. Another problem is the large amount of data that affect the performance of the algorithms. These problems led to the use of hybrid solutions for fraud detection issues. A survey summarizing data mining studies on "Fraud Detection" were made by Phua, Lee, Smith and Gayler (2005). Shao, Zhao and Chang (2002)'s study about customs took part in the survey.

In Shao et al. (2002)'s paper, a hybrid fraud detection strategy was designed for Customs Administration of China. Due to the complex customs data, multi-dimension criterion data model was used in the system. Four phases were defined in this study. The first phase and the third phase were operated on two different systems, and the other two phases were operated by customs experts and checkers. The main process was as follows. The C4.5, decision tree algorithm, generates the fraud detection rules first. Then, customs experts examine and confirm the rules, and sometimes improve the rules by changing them. At last, the customs administration officers check the goods on the basis of the risk information scored by the rules. Figure 2-1 shows this process with four stages and ten steps.

Because fraud data was less than 1% in the total data, total accuracy was not the significant parameter of the system. The significant parameter was defined as the ratio of the number of truly estimated fraud data to the number of data estimated as fraud. The parameter was specified as 10%. At the tests made at Qmgdao Dagang Customs, in China, in the first week, 34 reports were selected as fraud by the system, and 4 of them were discovered as fraud. And also in the second week, 31 reports were selected as fraud by the system, and 4 of them were discovered as fraud. The ratios were higher than the specified target parameter.

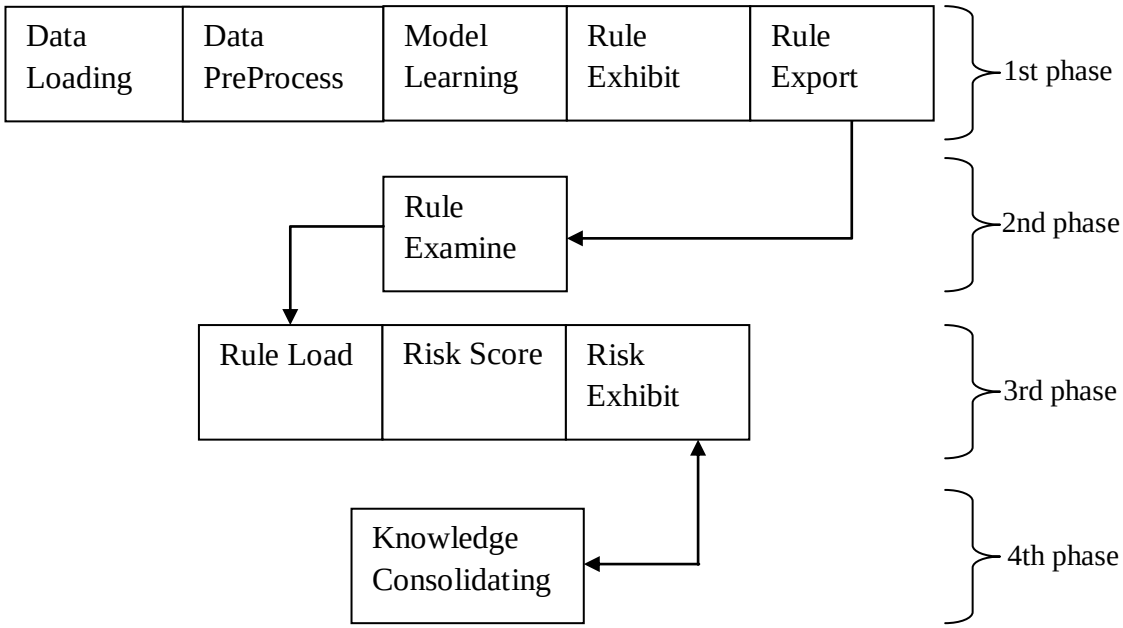

Figure 2-1 : Data mining process in customs[1]

In order to solve the conflict between the number of total transactions and the number of inspection officers, a study was carried out on risk analysis of customs cargo declaration (Yan-Hai and Lin-Yan, 2005). The Pareto 80/20 rule was applied for focusing on the most used types of cargo declarations. Then Q-type cluster method was used in the paper to separate the declarations into groups, thus customs could assign the inspection officers to the high risk transactions. About more than 400 attributes in the customs database were evaluated and for the selection of variances, the experiences of experts were taken into consideration. In the study, 8615 kinds of declarations were first applied 80/20 rule. 17.5 % of the declaration

[1] Shao et al. (2002)

types, contributing to 80 % of the declarations, was selected for the next steps. Then Q-type factor analysis was used for clustering purpose. The analysis resulted with seven clusters. If these clusters were under control by monitoring, the majority of the declarations would be under control.

Valuation fraud is one of the fraud types in customs. The customs duty and additional financial obligations are usually calculated based on the percentage of the worth of the imported goods. In order to evade these obligations, in some declarations, the worth of imported merchandise is declared lower than the actual worth. This is called as valuation fraud. It is a big challenge for customs officers to determine the correct worth of imported merchandise within the limited time for all transactions.

In literature, there is a study on valuation fraud applying data mining methods. In Singh, Sahu and Ujjwal (2003) a decision support system was formed. They first conducted a survey and interviewed the customs officers to use their expertise, and identified the most critical factors as importer, supplier, item type and rate of customs duty. In order to specify sensitivity of a transaction, some attributes were determined depending on the critical factors. For importer, importer type such as manufacturer, trader and government, assets, turnovers in previous years and the number of previous imports were evaluated as important. While type and country were important for supplier, type (consumer, industrial) was the important attribute for item. Aggregate sensitivities of the transactions were calculated by Hybrid Hierarchical Fuzzy Controller. The customs officers were required to control only if the sensitivity was high. The test results showed that it was possible to detect all frauds by controlling less than 10 % of the import transactions.

In Kumar and Nagadevara (2006)'s paper, a hybrid classification methodology was proposed for dealing with skewed data sets in customs database. When used for skew data, consisting of minority (less than 1-2%) and majority classes, standard classifiers tend to be overwhelmed by the majority class and ignore the minority class. The classification tree technique was formed in order to improve the accuracy of the model. Then, the predictions of this classification tree model were fed into an

artificial neural classification model. This latter model gave the flexibility to adjust the accuracy of a particular class label.

Digiampietri, Roman, Meira, Filho, Ferreira and Kondo (2008) discussed two systems for dealing with customs problems in Brazil. First one was an outlier based detection system for customs officers in order to identify suspicious operations which may include fraud, and the second one was a product and foreign exporter information system for the importers to assist them at the registration and classification of their products and exporters. Outlier based detection system was developed as a visual decision support system for Brazilian Customs with the assumption of majority of the international commerce operations were correct. This visual system allows user to see historical data about a tariff code, a transporter, a country of origin or an importer. The product and foreign exporter information system was developed for identifying foreign exporters and products. With this system, a foreign exporter could be identified with a unique number and it would be possible to search and find the exporter in every export declarations for importers.

There are many studies in the literature applied to China customs. In Ye, Zhou and Lu (2007), on the basis of the Risk Management System Data Warehouse, a back propagation neural network model, combined with Levengberg-Marquardt algorithm, was adopted for the ventures risk evaluation in China Customs. It was reported that the model was in test with a customs and certain achievements had been gained in the risk evaluation of it.

In another paper on China, Hua, Li and Tao (2006) developed a rule-based approach. In the proposed approach, a two-step clustering method was employed to classify objects into categories. First, k-means clustering technique was applied to compose preliminary clusters. Then a hierarchical clustering was used to group these preliminary clusters. The decision rules for each category were established by solving an optimization model, which was to minimize the total decision risks. For objects in each category, occurrence probabilities were estimated by logistic regression. The proposed rule-based approach can significantly improve China's customs inspection efficiency.

Yaqin and Yuming (2010) proposed a classification method based on association rules. The purpose of the study was to estimate the risk of goods by using association rules. Apriori algorithm was used to find frequent item sets. Based on the historical inspection data, inconsistencies between customs declaration and actual goods were tried to estimate. The accuracy of the model was not very high. In the paper, this result was associated with lack of some important attributes, for instance, commodity type.

Unlike other studies, a study in China, was not based on customs administrations' problems, but based on the exporters problems (Zhao and Chang, 2006). In this study, the early warning of the anti-dumping investigations in the textile industry was discussed. To foresee the anti-dumping investigations, leaving the companies in the textile industry in a difficult situation, was aimed. A Neuro-Fuzzy Decision Tree (ID3) was used for this purpose. The system could estimate whether a product is dumped or not in eighteen months.

CHAPTER 3

PROBLEM DEFINITION AND BACKGROUND

Customs and foreign trade regulations are made to regulate import and export activities. They intend to differentiate the nationalized goods from non-nationalized goods. In order to move in the international arena, the goods must be subject to a customs regime.

3.1. Customs Regimes

The goods come to the customs zone should be subject to one of the following customs regimes.

- Release for free circulation (Import regime)
- Transit regime
- Customs warehouse regime
- Inward processing regime

- Processing under customs control regime
- Temporary admission regime
- Outward processing regime
- Export regime

These regimes provide functions such as releasing for free circulation of the goods, exportation of the goods in free circulation, movement of the goods from one custom to another or warehousing of the goods without releasing for free circulation. Thus, the national goods are distinguished from the goods that are not national.

Release for free circulation, export and transit regimes are mostly used regimes. However, customs taxes are only applied for the import regime from these three regimes. So, economically, import process is the most important process in customs.

Import regime is also important in terms of risk. As well as tax evasion, people, weapons and drug smugglings should be evaluated by the customs authorities.

3.2. Import Procedure

The main steps of import process are as follows:

- Trader provides an import declaration form, indicating features, dimensions, attributes of the goods which will be imported, to a Customs Administration.
- An approval officer at the customs administration controls the information on the declaration and accepts the declaration. A registration number is assigned to the declaration.
- After submitting the declaration, the risk analysis system works in the background. According to the information contained in the declaration, an inspection / control line is assigned for that declaration. There are four types of lines.

    1. Blue Line: On the blue line, the customs procedures of the goods mentioned in the customs declaration are finalized without documents

or physical examination (the documents will be controlled later). In order to be assigned to the blue line for a declaration, the owner of the customs declaration should have a document called “Certificate of Authorized Economic Operator”.

2. Green Line: On the green line, goods are allowed to pass by not being subject to the document or physical examination. It is allowed for rapid transition. This line is used generally for military purposes or for humanitarian aid.

3. Yellow Line: Based on the risk assessment, assigning a customs declaration to the yellow line means that there is no need for physical examination of the goods. A comparison of the declaration with the supplementary documents, such as bill of goods attached to the customs declaration, is adequate. The other name of the yellow line is the document check.

4. Red Line: On the red line, the status of the goods is controlled. The other name of this line is the physical examination. When a customs declaration of any goods is assigned to the red line, a physical examination, such as weighing and counting should be made by the assigned inspection officer, if necessary. Not only physical but also document checks are carried out at this line. There are three types of physical examination methods.

    - External visual inspection
    - Partial inspection
    - Full examination

- If the inspection line is yellow or red, an inspection officer is assigned for checking the documents and/or goods.

- Inspector carries out checks on the documents and makes the required changes on the declaration in case of incompatibility between the information on the declaration, supplementary documents and the goods on the vehicle.

- If there is a difference occurred on taxes, the payment of this difference is requested.
- After the payment of taxes, the vehicle may leave the customs zone.

3.3. Taxes and Funds Applied to Import Regime in Turkey

With customs control, it is intended to detect inappropriate circumstances by examining the declaration, attachments and the goods. Tariff code, goods description, country of origin and total invoice amount fields on the declaration are the first fields cared at examination. These fields have priority in determining the taxes accrued on the goods.

There are mainly nine types of taxes and funds that should be considered by control officer while controlling.

1. Customs duty (tariff)
2. Special consumption tax
3. Anti-dumping tax
4. Housing development fund (share of agriculture)
5. Tobacco fund
6. Resource utilization support fund
7. Failure of Intellectual and Artistic Works Act
8. Value added tax
9. Stamp tax

3.4. Tariff, Tariff Code and Non-Tariff Barriers

Tariff or customs duty is a tax levied by governments on the worth of imported products. The tariff is assessed at the time of importation along with any other applicable taxes/fees. Tariffs raise the prices of imported goods, thus making them

less competitive within the market of the importing country. In order to determine which rate of customs duty should be used, the tariff code and the origin of the product should be known.

The need to classify products in tariff schedules stems from the fact that tariff rates for products differ. Classifying products in separate categories and sub-categories facilitates determination by customs authorities of the appropriate level of duty to be applied to individual products. Before deciding what ad valorem or other duty is applicable to an imported product, customs officials must first establish the exact category in the national tariff schedule under which that product is classified. (Forrester and Kaul, 2005)

A tariff code is a product-specific code as documented in the Harmonized System (HS) maintained by the WCO. Tariff codes exist for almost every product involved in global commerce. Required on official shipping documents for tax assessment purposes, a tariff code ensures uniformity of product classification worldwide.

A tariff code is a number assigned to each type of product sold internationally. The Harmonized System features 21 sections and 97 chapters of product codes that must be used by WCO members to stay compliant with trade policies. Tariff codes are used for taxation, customs and statistical purposes by WCO member countries. (Başţabak and Medeni, 2011)

The 12-digit code used in Turkey is called Customs Tariff Statistics Position. The first 6 digits of the code show the Harmonized System Nomenclature Code used by all countries that are the members of the WCO. 7-8th digits illustrate Combined Nomenclature Code used by the countries of the EU. 9-10th digits indicate positions formed due to different tax practices. 11-12th digits illustrate the statistical codes.

For example, 4418.90.80.90.11 is the tariff code used for “Wooden Ladder”. A small portion of 44th chapter of tariff code list can be found in Appendix A. It starts with “Wood and Articles of Wood” and branches up to “Ladder”.

Apart from tariff, there are also NTBs used to regulate the international trade. NTBs to trade are trade barriers that restrict imports but are not in the usual form of a tariff.

Although they are called "non-tariff" barriers, they have the effect of tariffs once they are enacted. (Evans and Newnham, 1998)

NTBs can be grouped under the following titles:

- Specific Limitations on Trade
    - Quotas, import licensing requirements, proportion restrictions of foreign to domestic goods, minimum import price limits, embargoes
- Customs and Administrative Entry Procedures
    - Valuation systems, anti-dumping practices, tariff classifications, documentation requirements, fees
- Standards
    - Standard disparities, intergovernmental acceptances of testing methods and standards, packaging, labeling, and marking
- Government Participation in Trade
    - Government procurement policies, export subsidies, countervailing duties, domestic assistance programs
- Charges on imports
    - Prior import deposit subsidies, administrative fees, special supplementary duties, import credit discrimination, variable levies, border taxes
- Others
    - Voluntary export restraints, orderly marketing agreements

3.5. Anti-Dumping Tax[2]

Anti-dumping tax is a tax collected by the Ministry of Economy General Directorate of Imports, for the goods and countries stated in the Communications on the Prevention of Unfair Competition in Importation issued in accordance with the Prevention of Unfair Competition Legislation (Law, Decision and Regulation).

[2] Cete, E. (1998).

In order to put into action an Anti-dumping tax, determination of the presence of the dumping is required. For dumping, the following conditions must be formed.

- Two distinctly separate market for exporters
- The lack of the opportunity of re-import
- The lack of Anti-Dumping Measures at Country of Import
- Price Discrimination at Markets of Country of Import and Export

3.5.1. Effects of Dumping

The effects of dumping are not only exclusive to importing and exporting countries.

✓ The effects on the importing country in terms of Economic Units

- The effects regarding purchaser: The direct effect of dumping on purchasers is positive in the importing country. This is because they are able to get the same or similar goods by paying a lower price. If the dumping is medium or long term, the person / people who dump can win the monopoly power and by dominating on the importing country's industry can raise the prices of the goods much more than before.
- The effects regarding manufacturers: If the export price of the dumped goods is much lower than the price level in the importing country and there is an increase that cannot be ignored in the level of imports, domestic producers may be forced to quit the market or may choose to stop production.
- The effects in terms of production factors: If the producers in the importing country have to offer significantly lower prices due to the dumped imports, this may lead to bad consequences such as layoffs, closing the business or working with idle capacity.

✓ Effects on exporting country

Effects of dumping on the exporting country are concentrated more on units purchasing the goods for the purpose of production or consumption, because the exporter will compensate the losses by rising domestic sales prices.

✓ Effects on other countries

Depending on the import amount of the dumped goods of importing country, other countries are affected by the dumped imports. This effect may occur in the form of distortions of competition at the international level such as market loss, reduction in production, the emergence of unemployment, reduced profits.

3.5.2. Dumping Practices on Legislation Regarding the Prevention of Unfair Competition in Importation

The practices are carried out pursuant to the Legislation of Prevention of Unfair Competition in Importation and the provisions contained in the annex Anti-Dumping Agreement, WTO. The numbers of anti – dumping taxes enforced by Turkey are listed according to the countries in Appendix B.

Under the provisions of the legislation, there are 117 anti-dumping measures in practice. 24 anti-dumping investigations have been closed without any action up to now. There are also 17 ongoing investigations of which 16 are the review investigation and one of them is dumping investigation.

Ten anti-dumping measures, of which three are by the United States of America (USA), are currently being enforced to Turkey. And six investigations are still in progress.[3]

According to data for the period of 01.01.1995-30.06.2011 gathered from the WTO, twenty seven anti-dumping measures were enforced to Turkey by twelve different countries in this period. During the same period, a total of 143 anti-dumping measures were taken by Turkey, of which 56 are applied to China.[4]

3.5.3. The Procedure of Dumping Investigation

Pursuant to the legislation in question, the procedure during a dumping investigation can be summarized as follows.

1. A written application is submitted to the Ministry of Economy, General Directorate of Imports by the natural or legal persons who claim to suffer from imports subject to dumping.

[3] Source: www.tpsa.gov.tr
[4] Source: http://www.wto.org/english/tratop_e/adp_e/ad_meas_rep_exp_e.pdf

2. The presence of the imports subject to dumping and of the damage caused by these imports and the application made by the production branch or on behalf of production branch is determined.

3. It is decided to open the investigation by Unfair Competition in Importation Review Board and the notification regarding the investigation is published in the Official Gazette. During the investigation, the studies are done on the export price, normal worth, dumping margin and the damage occurring on the production branch. Dumping investigation period is one year.

4. In order to identify the export price of the goods, the normal worth in the country of origin and the other features on sales conditions properties, information is requested from relevant importers and exporters.

5. A fair comparison is made between export price and normal worth.

6. Determined by the Board and approved by the Ministry, the final measure, as an anti-dumping tax, is applied to a sufficient extent to eliminate the dumping margin or loss.

7. Final measure remains in force a long time enough to eliminate the effect of the damage caused by imports subject to dumping.

3.6. Risk Analysis and Assessment

Risk is defined as the emergence of an event or probability that would jeopardize the implementation of national legislation or other arrangements for customs procedures of the goods.

In terms of customs controls, risk is the possibility of violating Customs Administration relevant laws, regulations and instructions intentionally or unintentionally. According to the definition of WCO, risk is the potential of non-compliance with customs legislation.

Risk analysis is a method that aims customs' human and financial resources to be used for minimizing risk. The following figure shows the Risk Analysis's position in

customs procedures. By evaluating the information on the declarations, system decides the declaration to be risky or not.

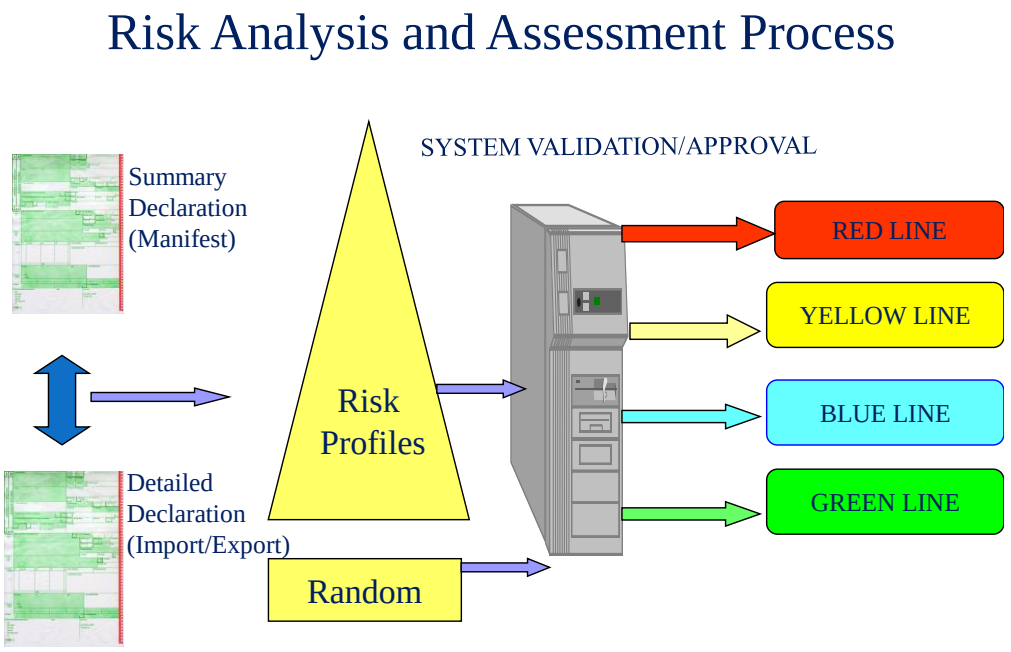

Figure 3-1: Risk Analysis and Assessment Process

3.7. Tariff Code Circumvention

In order to avoid measures taken to protect domestic consumers and producers, the owner of the declaration may divert the information disclosed. Merchandises having taxes, measures and controls may be declared with another tariff code that is similar in terms of features, instead of the real tariff code. It is called as "Tariff Code Circumvention". The declared tariff code is not subject to these taxes, measures or controls, or is subject to taxes and measures with lower rates.

In order to detect Tariff Code Circumvention, the control of the goods is necessary. Inspector, opening the truck, controls the goods, and if he decides that the declared Tariff Code is different from Tariff Code of the goods in the vehicle, he performs a

change on the declaration form and demands from the declarative to pay taxes for the goods. After the payment of customs taxes, the vehicle can leave the customs area.

At this point two issues arise:

1. If the vehicle is assigned to the Yellow Line, because of not doing the control of the goods at Yellow Line, Tariff Code Circumvention cannot be caught.

2. Even if the vehicle assigned to the Red Line, it may be difficult to detect that the Tariff Code has been set to a wrong Tariff Code. It is sometimes very difficult for an inspector to be able to determine the real Tariff Code of the goods.

However, if the inspector gets a warning that says there could be a Tariff Code Circumvention, he may physically control the goods or may request control of the goods in the laboratory, and know which tariff code must be used for this product.

So, it is clear that we need a system that warns the inspectors. It is not possible to predict which tariff codes will be used for the product. Tariff code list is very long, and each product can use one of the large number of similar tariff codes. Without any mainstay, making such an inference may cause all operations to lead to the red line control. So, it may cause the process to last too long and even to lock the customs administration. In fact, it may also lead to disruption of works of traders who have a clear declaration.

By data mining, the circumvention can be determined through the company's past records and with a warning displayed to inspector.

The reasons of tariff code circumvention can be classified as follows:

1- Taxes and Financial Obligations: Any merchandise which is subject to expenses such as customs duty, value added tax and housing fund on imports.
2- Foreign Trade Measures: Any merchandise which is subject to foreign trade measures such as surveillance, anti-dumping tax and additional financial obligation on imports.

3- Obligations of Document Provision or Fulfillments of Some Conditions and Standards: Any merchandise which is subject to the assents, the controls or the permissions of certain agencies or institutions such as the Ministry of Health, the Ministry of Economy, Turkish Standards Institution, on imports.

The trader may mislead or misdeclare information on import declarations in order to minimize the taxes to be paid and/or conditions or standards to be met.

The tariff code and country of origin become important in determining the tariff (tax rate) of the merchandise. The amount of taxes to pay is determined by the worth of the merchandise. In determining the taxes on declaration, attention is paid to the following fields;

a. The declared tariff code of merchandise
b. The country of origin of merchandise
c. The commercial definition of merchandise
d. The worth of merchandise
e. The type of delivery of merchandise
f. The payment method
g. Gross and net weight: In some cases, surveillance, anti-dumping tax or additional financial obligation is charged based on gross weight.

Depending on the information in these fields, the types of taxes and obligations to be applied are determined.

CHAPTER 4

RESEARCH METHODOLOGY

In Turkey, the tariff code circumvention processes are carried out by the Risk Analysis Office within Directorate General of Risk Management and Control of the Ministry of Customs and Trade. The Risk Analysis Office carries out works to ensure the capture of irregularities in declarations at customs processes, by using risk management methodologies.

4.1. Motivation

Tariff code circumvention to avoid non-tariff measures, causes both loss of tax revenue and loss of its nature of the safeguard measure. Circumvention of the measures enforced to protect the sector's producers eliminates the effectiveness of the measures. For this reason, to monitor the functioning of the introduced measures is necessary during the period of the measure. Thus, determination of existence of deceptive acts such as the tariff code circumvention and country of origin circumvention will be possible. This way it will also be possible to identify the precautions to be taken against them.

These studies should be done at regular intervals for each anti-dumping measure. The necessity of a system, which runs automatically every time a declaration registered and determines the risk based on historical data, emerges. Such a system based on the research methods used within the Risk Analysis Office and the historical data of the companies will also eliminate the human factor effects.

4.2. Risk Analysis Methodology

Risk Management is a logical and systematic method that identifies, analyzes, resolves, and monitors the risks involved in any activity or process. Risk analysis is the systematic use of the available information in order to determine how often the identified risks can occur and the size of possible outcomes.

As a result of risk analysis, risk indicators and depending on these indicators risk profiles are created. Risk indicators are specific criteria that if they are evaluated together, they allow selecting targets by identifying the entries with a potential risk of violating customs laws. For example; a risk indicator can be defined as: if the country of origin is China, Malaysia, Indonesia or Vietnam, this indicates risk.

Risk profiles are pre-determined combination of risk indicators that are based on collected, analyzed and classified knowledge. For example, a risk profile can be specified as: “If the country of origin is China, Malaysia, Indonesia or Vietnam, and the tariff code starts with “5407” and net mass is bigger than 5000 kg, the transactions is risky. Physical control must be done.”

The declarations matching risk profiles are assigned to the red line. A certain rate within declarations that do not comply risk profiles, are also assigned to the red line randomly. The remaining declarations are assigned to blue line if the company has an authorized consignor/consignee status. If not, they are assigned to the yellow line.

In order to determine risk profiles, a study on data warehouse is performed. As well as based on legislation changes, the study is also based on the rectified customs declarations.

As an example of changes in legislation, new anti - dumping measures or surveillances can be applied. Up to 6 months after the amendment of legislation

(important to consider the date of the notification), a study related to the change is done on the tariff code. The change of behavior of the companies, imported merchandise with the tariff code in question before the legislation change, is examined. If while the amount of import products are decreasing, other product/products amount is/are increasing, the later tariff codes are considered as "shelter tariff code" and it is evaluated based on the tariff schedule.

Generally anti-dumping measures are based on the country. If the country is specified, declaring a false country of origin of the goods is attempted deliberately in some cases. In this case, while examining the behavior of companies, the situation of declaring false origin country is also investigated.

The proof of the existence of origin circumvention is possible by taking into consideration the difference between the imports of the product of our country and the exports of the product of the other countries to our country. Figure 4-1 shows the difference between Turkey's import amount from China and China's export amount to Turkey for the product category "Slide fasteners other than those fitted with chain scoops of base metal" (960719). After 2005, as China's export amount increases, Turkey's import amount decreases. The gap between amounts gets larger in time.

Anti-dumping measure of 3$/Kg was taken between 12.03.2005 and 29.10.2010 for "Plastics Slide Fasteners" (9607.19.00.00.11). Figure 4-2 illustrates the import amounts of this product from China, and Indonesia. According to China (Figure 4-1), Turkey should have increased imports of this product. In contrast, according to our data, there is a decrease in the amount of imports from China and an increase from Indonesia.

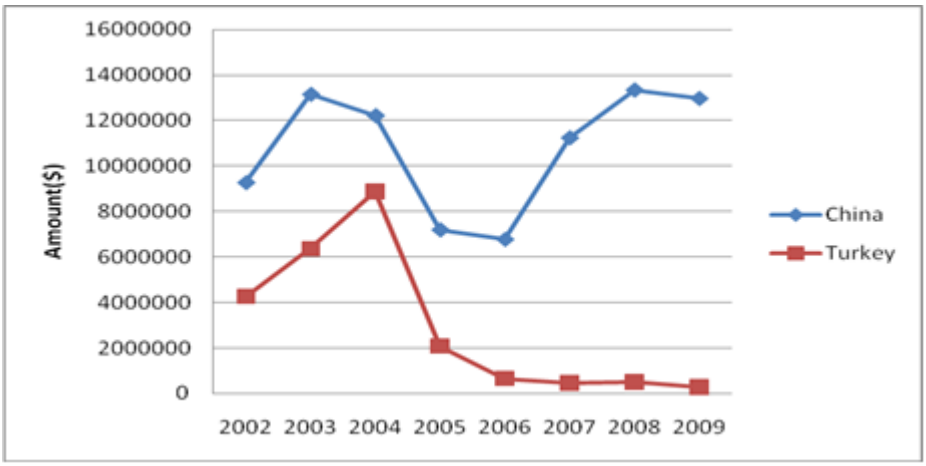

Figure 4-1: Comparison of China's Exports and Turkey's Imports[5] of "Slide Fasteners Other Than Those Fitted with Chain Scoops of Base Metal" (9607.19) between 2002 and 2009,

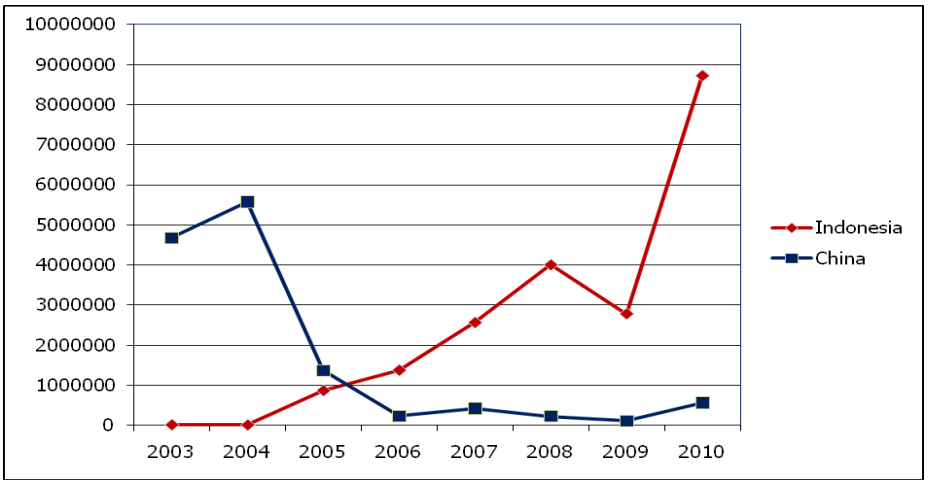

Figure 4-2: Imports of "Plastics Slide Fasteners" (9607.19.00.00.11) Between 2003 and 2010 from China and Indonesia[6]

While creating the risk profiles for the tariff codes and/or the origin countries considered as attempting circumventing, the share of imports of that product in total imports should also be assessed. For the products imported frequently, in order not to

[5] Source: http://comtrade.un.org
[6] Source: GÜVAS

cause intensity on the customs operations, the risk profiles can be defined based on specific company/companies.

Another method is to go through the declarations that have been rectified because of "Tariff Code". The former and latter tariff codes of these declarations are evaluated, and it is investigated whether there is a significant condition indicating a circumvention. (Such as, while the former tariff code has surveillance, additional financial obligation or anti-dumping measure, the latter does not.)

In the explanation of declarations that are decided to be penalized, the article of law which constitutes a basis on penalty is indicated. For example, subparagraph 1-A of Article 234 of the Customs Law includes criminal action taken related to declaration of a false tariff code, 234/1-B includes criminal action taken related to the worth of merchandise.

4.3. Dataset

It was intended to automate the studies done by Risk Analysis Office, only for anti-dumping duty, by using data mining methods. As a result of the research on the database, four anti-dumping taxed goods groups mostly detected as circumvented were determined and selected for this study. The tariff codes and descriptions of these merchandises can be found at Table 4-1. Detailed information on tariff classification is at Appendix C.

Table 4-1: Descriptions of Selected Tariff Codes

Group	Tariff Code	Description
6001	6001.10.00.00.11	Long pile fabrics of synthetic fibers
	6001.92.00.00.00	Others of man-made fibers for blankets
9105	9105.21.00.00.00	Wall clocks (battery accumulator or main powered)
8302	8302.10.00.00.00	Hinges of base metal
	8302.42.00.00.00	Similar articles suitable for furniture
	8302.50.00.00.00	Hat-racks, hat-pegs, brackets and similar fixtures of base metal & base metal mountings, fittings
9608	9608.10.10.10.00	Ball point pens of plastics
	9608.50.00.10.00	Set of ball point pens of plastics

The information about enforced anti-dumping taxes of the goods is shown in Table 4-2 to Table 4-5.

Table 4-2: Anti - Dumping Taxes (for branch of 8302)

TARIFF CODE	ORIJIN	ANTI DUMPING TAX	START DATE	FINISH DATE
8302.10.00.00.00	Indonesia, Malaysia, Taiwan	1.64 $ / Piece	27.08.2008	27.08.2013
8302.42.00.00.00	Indonesia, Malaysia, Taiwan	0.75 $ / Piece	27.08.2008	27.08.2013
8302.50.00.00.00	Indonesia, Malaysia, Taiwan	1.64 $ / Piece	27.08.2008	27.08.2013
8302.10.00.00.00	China	1.39 $ / Piece	07.02.2004	20.07.2015
8302.42.00.00.00	China	0.508 $ / Piece	07.02.2004	20.07.2015
8302.50.00.00.00	China	1.39 $ / Piece	07.02.2004	20.07.2015

Table 4-3: Anti - Dumping Taxes (for branch of 9105)

TARIFF CODE	ORIJIN	ANTI DUMPING TAX	START DATE	FINISH DATE
9105.21.00.00.00	China	2.1 $/Piece	07.11.2001	24.10.2012

Table 4-4: Anti - Dumping Taxes (for branch of 6001)

TARIFF CODE	ORIJIN	ANTI DUMPING TAX	START DATE	FINISH DATE
6001.10.00.00.11	China	4 $/Piece	08.12.2002	01.08.2013
6001.92.00.00.00	China	4 $/Piece	08.12.2002	01.08.2013

Table 4-5: Anti - Dumping Taxes (for branch of 9608)

TARIFF CODE	ORIJIN	ANTI DUMPING TAX	START DATE	FINISH DATE
9608.10.10.10.00	China	0.066 $/Piece	02.03.2004	12.11.2014
9608.50.00.10.00	China	0.066 $/Piece	02.03.2004	12.11.2014

The total numbers of declarations detected and remaining as circumvented are shown in Table 4-6. The distribution of these numbers over the years can be seen in Appendix D.

Table 4-6: The Number of Declarations by Tariff Codes

Tariff Code	Before Tax	After Tax		Total
		Detected	Remaining	
9105	118	81	4884	5083
6001	1455	72	9829	11356
8302	16329	72	113748	130149
9608	22863	82	521357	544302

4.4. Features of Data

In examining the data on these goods, the issues, such as the availability of records, standard data entry and so on, are taken into account. While deciding the features, the studies done by Risk Analysis Office were also considered. As a result of evaluation, it was decided to use the features available in Table 4-7 .

Table 4-7: Features in the Dataset

	TITLE	DESCRIPTION	TYPE
1	INTCONV	International convention	Nominal
2	ORIGCOUNTRY	Country of origin	Nominal
3	CONTCOUNTRY	Contracting country	Nominal
4	TAXIDENUM	Tax identification number of consignee	Nominal
5	ANTIDUMPING	Anti - dumping tax is enforced	Nominal
6	TARIFFCODE	Declared tariff code	Nominal
7	DEVUNITGROSS	Deviation from the average of the ratio of worth to gross mass	Numeric/ Calculated
8	DEVUNITNET	Deviation from the average of the ratio of worth to net mass	Numeric/ Calculated
9	DEVUNITSTAT	Deviation from the average of the ratio of worth to statistical unit	Numeric/ Calculated
10	DEVVALUE	Deviation of the worth from average	Numeric/ Calculated
11	DEVGROSS	Deviation of the gross mass from average	Numeric/ Calculated
12	DEVNET	Deviation of the net mass from average	Numeric/ Calculated
13	DEVIST	Deviation of the statistical unit from average	Numeric/ Calculated
14	CLASS	Detection of tariff code circumvention	Nominal

The explanations and the calculations of these features can be summarized as follows:

1. INTCONV: At customs, a set of special privileges has been granted to the countries and the members of associations in which bilateral agreements have been conducted, such as Albania, Bosnia and Herzegovina, Morocco, European Free Trade Association (EFTA), and EU. In general, anti-dumping measures are not enforced to these countries. INTCOV stores whether an international convention is valid for the declaration.

2. ORIGCOUNTRY: It stores the code of the country of origin.

3. CONTCOUNTRY: It stores whether the contracting country is different from the country of origin. If they are different, it is set to 1; otherwise it is set to 0.

4. TAXIDENUM: It stores the tax identification number. It is used for differentiation of consignees.

5. ANTIDUMPING: If there is an anti-dumping tax for the declared tariff code and country of origin, it is set to 1, else it is 0.

6. TARIFFCODE: It stores the declared tariff code.

The following seven attributes include calculated values. These values reflect the behavioral differences between a declaration's worth with the worth of past declarations' with similar characteristics. Before explaining these attributes, we need to examine on the behavior of data.

As an example, the datasets including gross mass and worth of three different goods imported by three different consignees from China were examined. The histograms related to these data can be seen at Figure 4-3 to Figure 4-8.

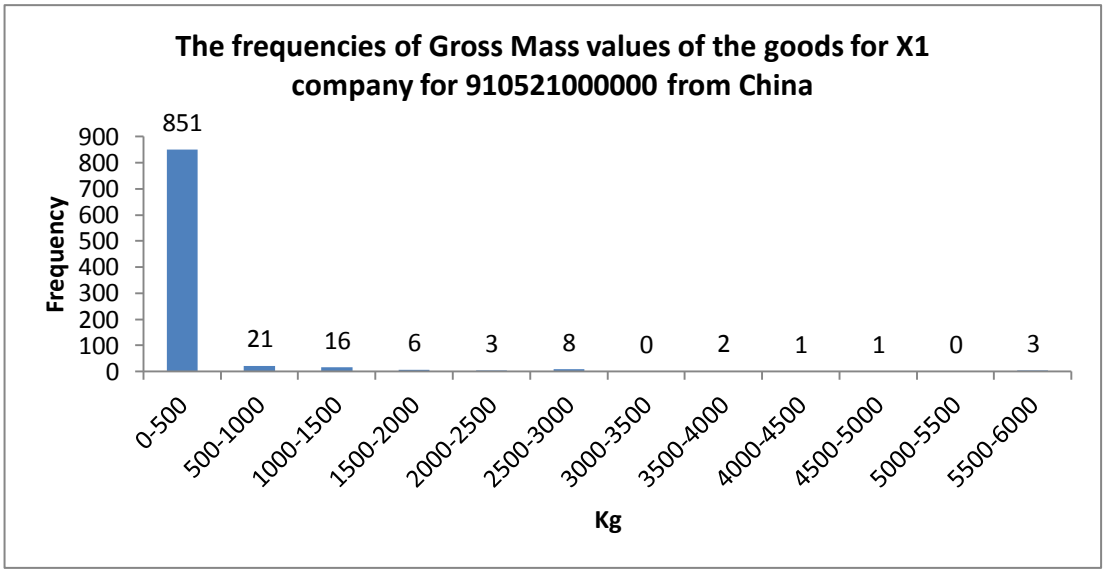

Figure 4-3: The frequencies of Gross Mass values of the goods for X1 company for 910521000000 from China

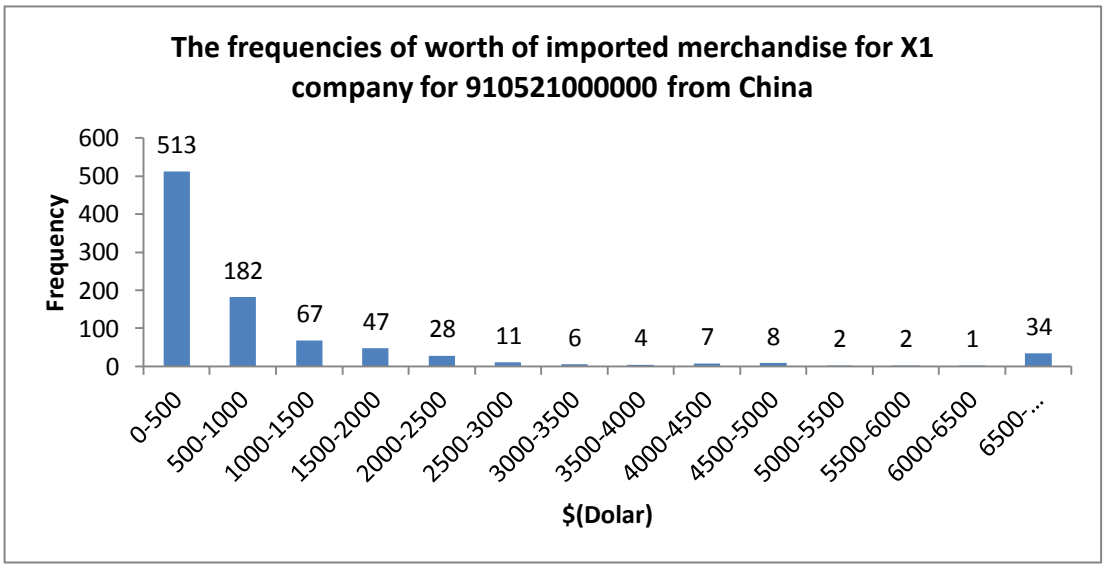

Figure 4-4: The frequencies of worth of imported merchandise for X1 company for 910521000000 from China

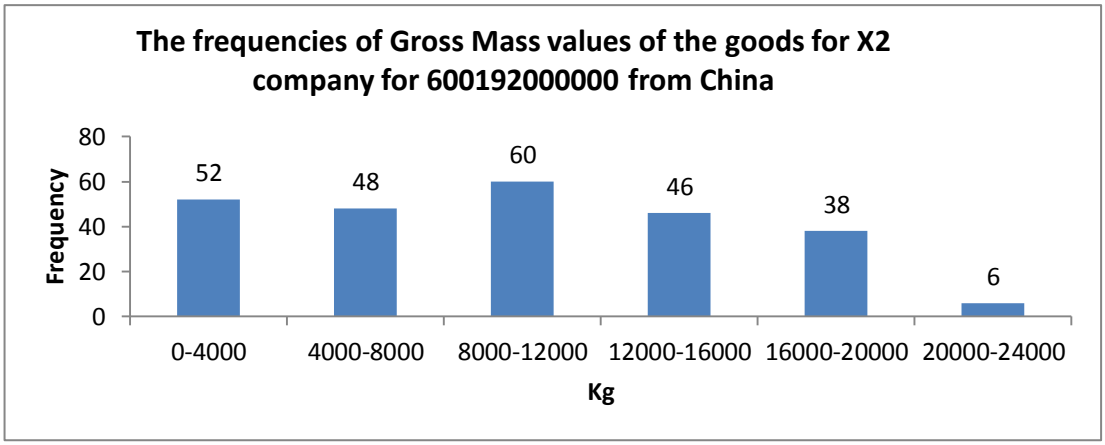

Figure 4-5: The frequencies of Gross Mass values of the goods for X2 company for 600192000000 from China

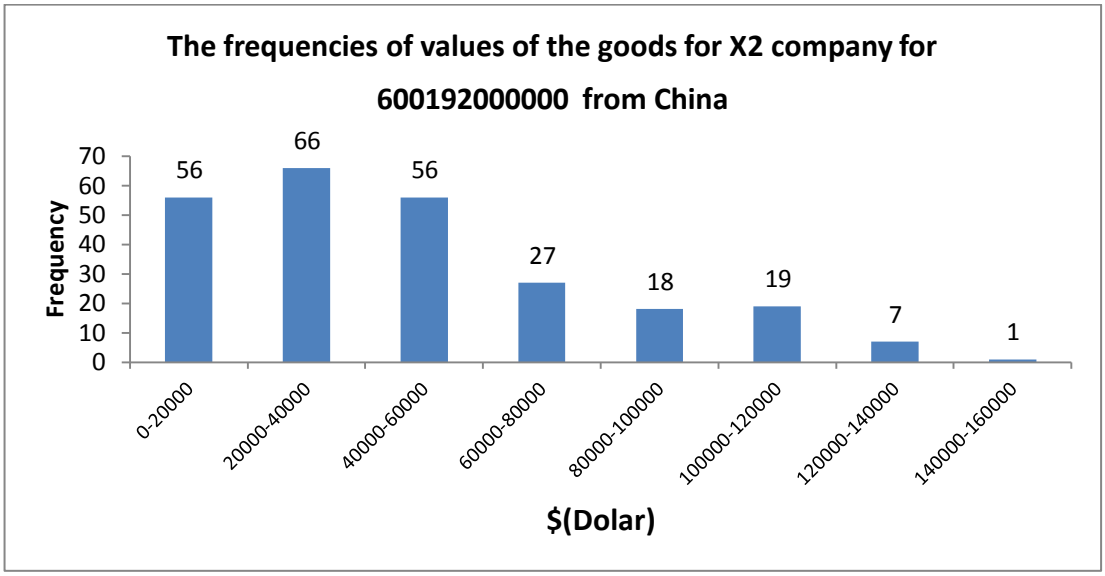

Figure 4-6: The frequencies of worth of imported merchandise for X2 company for 600192000000 from China

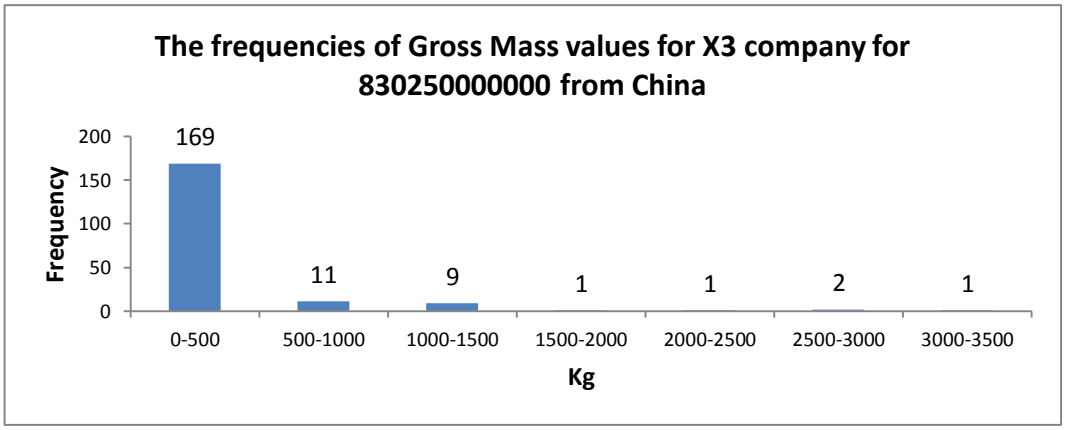

Figure 4-7: The frequencies of Gross Mass values of the goods for X3 company for 830250000000 from China

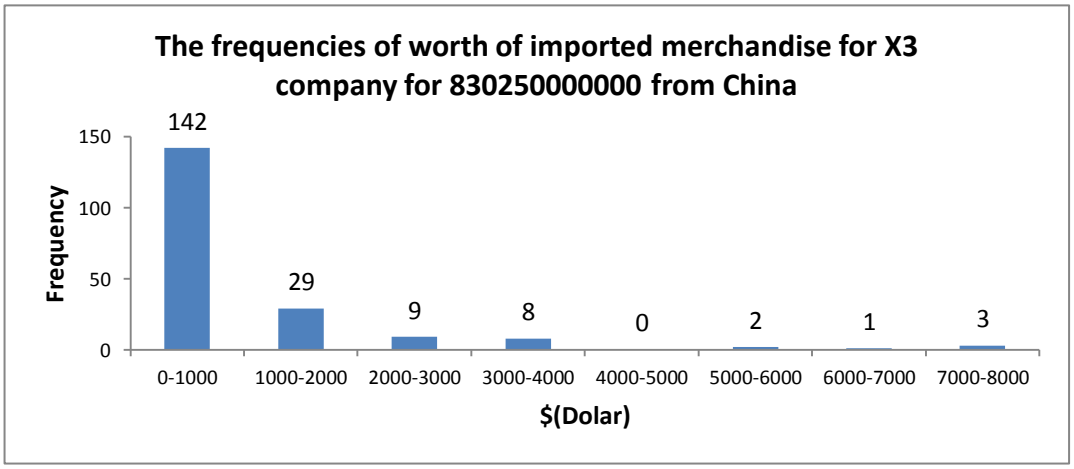

Figure 4-8: The frequencies of worth of imported merchandise for X3 company for 830250000000 from China

In order to use mean and standard deviation values for the data, the data must be normally distributed. The Interquartile Range Test is used to find out whether the distribution of numeric data, such as gross mass and worth of the good, is normally distributed or not.

According to this test, If P is normally distributed, then the standard score of the first quartile, z_1, is -0.67, and the standard score of the third quartile, z_3, is +0.67. Given mean = X and standard deviation = σ for P, if P is normally distributed, the first quartile

$$Q_1 = (\sigma\, z_1) + X \qquad \text{(Equation 4-1)}$$

and the third quartile

$$Q_3 = (\sigma\, z_3) + X \qquad \text{(Equation 4-2)}$$

If the actual values of the first or third quartiles differ substantially from the calculated values, *P* is not normally distributed.

The result of the Interquartile Range Test can be seen in Table 4-8 below.

Table 4-8: The Interquartile Range Test Results

China						
	910521000000		600192000000		830250000000	
Statistics	X1-Gross Mass	X1-Worth	X2-Gross Mass	X2-Worth	X3-Gross Mass	X3-Worth
min	4.07	25.85	7	163	5.2	1.3
Actual Q1	30.34	249.88	4426.25	21636.28	33.77	2.98
median	56.64	420.25	9136.05	40597.24	96.21	4.91
Actual Q3	138.61	913.6	13806.18	67137.5	274.74	6.56
max	5697.12	63882.01	23253.3	154700	3440.64	19.65
Calculated Q1	-166.55	-2046.78	5791.28	25983.24	-48.35	3.23
Calculated Q3	566.08	5214.83	13488.81	70713.61	607.15	7.37

The calculated Q_1 and Q_3 values are significantly different from actual ones. Accordingly, it can be concluded that the distribution of gross mass, net mass, statistical unit and worth attributes does not fit with normal distribution.

Therefore, we use the quartile information of data instead of mean and standard deviation. First the quartiles are determined. Then the attributes are calculated based on this data.

7. DEVUNITGROSS: It is a calculated field that gets integer values from the set {1, 2, 3, 4}. It specifies the quartile of the ratio of the declared worth to the declared gross mass in the set of the ratio of the worth to the gross mass for all declarations made by same consignee, from the same country of origin, for the same good, in the earlier years.

This description can be formulized as in Equation 4.4.

$$DEVUNITGROSS = Quartile(\frac{V}{G}, \frac{V_{i,k,m,t}}{G_{i,k,m,t}})$$ (Equation 4-3)

In this formula, Quartile finds that in which quartile $\frac{V}{G}$ corresponds in the set of all $\frac{V_{i,k,m,t}}{G_{i,k,m,t}}$.

8. DEVUNITNET: It is a calculated field that gets integer values from the set {1, 2, 3, 4}. It specifies the quartile of the ratio of the declared worth to the declared net mass in the set of the ratio of the worth to the net mass for all declarations made by same consignee, from the same country of origin, for the same good, in the earlier years.

This description can be formulized as in Equation 4-4.

$$DEVUNITNET = Quartile(\frac{V}{W}, \frac{V_{i,k,m,t}}{W_{i,k,m,t}})$$ (Equation 4-5)

In this formula, Quartile finds that in which quartile $\frac{V}{W}$ corresponds in the set of all $\frac{V_{i,k,m,t}}{W_{i,k,m,t}}$.

9. DEVUNITSTAT: It is a calculated field that gets integer values from the set {1, 2, 3, 4}. It specifies the quartile of the ratio of the declared worth to the declared statistical unit in the set of the ratio of the worth to the statistical unit for all declarations made by same consignee, from the same country of origin, for the same good, in the earlier years. Statistical unit is a field that stores a unit of measure. It may vary depending on the type of the good. For example, it stores the volume of liquids in cubic meter, the area of textile goods in square meter, or the weight of plastics in kilogram.

This description can be formulized as in Equation 4-6.

$$DEVUNITSTAT = Quartile(\frac{V}{U}, \frac{V_{i,k,m,t}}{U_{i,k,m,t}})$$ (Equation 4-7)

In this formula, Quartile finds that in which quartile $\frac{V}{U}$ corresponds in the set of all $\frac{V_{i,k,m,t}}{U_{i,k,m,t}}$.

10. DEVVALUE: It is a calculated field that gets integer values from the set {1, 2, 3, 4}. It specifies the quartile of the declared worth in the set of the worth for all declarations made by same consignee, from the same country of origin, for the same good, in the earlier years.

This description can be formulized as in Equation 4-8.

$$DEVVALUE = Quartile(V, V_{i,k,m,t}) \quad \text{(Equation 4-9)}$$

In this formula, Quartile finds that in which quartile V corresponds in the set of all $V_{i,k,m,t}$.

11. DEVGROSS: It is a calculated field that gets integer values from the set {1, 2, 3, 4}. It specifies the quartile of the declared gross mass in the set of the gross mass for all declarations made by same consignee, from the same country of origin, for the same good, in the earlier years.

This description can be formulized as in Equation 4-10.

$$DEVGROSS = Quartile(G, G_{i,k,m,t}) \quad \text{(Equation 4-11)}$$

In this formula, Quartile finds that in which quartile G corresponds in the set of all $G_{i,k,m,t}$.

12. DEVNET: It is a calculated field that gets integer values from the set {1, 2, 3, 4}. It specifies the quartile of the declared net mass in the set of the net mass for all declarations made by same consignee, from the same country of origin, for the same good, in the earlier years.

This description can be formulized as in Equation 4-12.

$$DEVNET = Quartile(W, W_{i,k,m,t}) \quad \text{(Equation 4-13)}$$

In this formula, Quartile finds that in which quartile W corresponds in the set of all $W_{i,k,m,t}$.

13. DEVIST: It is a calculated field that gets integer values from the set {1, 2, 3, 4}. It specifies the quartile of the declared statistical unit in the set of the statistical unit for all declarations made by same consignee, from the same country of origin, for the same good, in the earlier years. Statistical unit is a field that stores a unit of measure. It may vary depending on the type of the good. For example, it stores the volume of liquids in cubic meter, the area of textile goods in square meter, or the weight of plastics in kilogram.

This description can be formulized as in Equation 4-14.

$$DEVIST = Quartile(U, U_{i,k,m,t}) \quad \text{(Equation 4-15)}$$

In this formula, Quartile finds that in which quartile U corresponds in the set of all $U_{i,k,m,t}$.

4.5. Feature Selection

Feature selection was performed to determine the level of importance of individual features. Thus, at the next stage, the runs could be made with different parameters and different numbers of features.

For attribute selection, InfoGain Attribute Evaluation Algorithm (Mitchell, T.M., 1997) in Weka, was used. InfoGain evaluates the worth of an attribute by measuring the information gain with respect to the class attribute. In general terms, the expected information gain is the change in information entropy from a prior state to a state that takes some information as given:

InfoGain(Class,Attribute) = H(Class) - H(Class | Attribute). (Equation 4-16)

As a result of the algorithm, the attribute ranking in Table 4-9 was obtained. The numbers in the table correspond to the sequence numbers of the attributes used in Chapter 4.4. As can be seen from the table, the ranks vary according to the products. One reason of this is the products with different trader's profile.

Table 4-9: The Rankings of the Features

	Ranking											
Tariff Code	**Min**											**Max**
9105	3	1	2	12	11	10	13	7	8	9	5	4	6
6001	3	5	1	12	11	10	13	8	9	7	2	4	6
8302	3	5	12	11	10	9	8	7	13	1	2	4	6
9608	5	3	10	9	13	8	11	12	7	1	2	6	4

Min and Max labeled attributes are the least and most worthy attributes respectively. For example; for the merchandise with tariff code "9105", CONTCOUNTRY and INTCONV are the two least worthy attributes, and ANTIDUMPING, TAXIDENUM and TARIFFCODE are the most gainful ones.

As the results of the first three merchandise, 6001, 9105 and 8302, are similar to each other, the result of the last merchandise, 9608, is a little different from them. For example, the most valuable attribute is TAXIDENUM, due to the variety of companies.

With the help of this attribute rank, the algorithms would be run with different numbers of attributes, at a later stage. For instance, considering 9105, the first run is done by using all the attributes. Subsequent runs are carried out by discarding one attribute each time, starting from the "min" labeled attribute such as 3 for product 9105 and continue decreasing the number of attributes one by one according to the attribute significance level. Therefore, this ranking is very important.

4.6. Method Selection

As a classification method, first, J48 decision tree algorithm was applied. The decision tree obtained by the algorithm was much spread and pruned at an earlier stage that causes low performance values for detected records. This occurs because of limited number of detected declarations (class=1) and high accuracy obtained by pruning the first branches. Despite sampling, a satisfactory result could not be obtained.

Therefore, to obtain a better performance, K - Nearest Neighbor (KNN) algorithm (IBk in Weka) is used. Because, general rules could not be extracted for these datasets and the similarities between records in the datasets are important, we have chosen to use KNN (Cover & Hart, 1967).

For nearest neighbor searching, Linear Nearest Neighbor Search method was used and as the distance metric Euclidean Distance was used. In order to observe the behavior of the algorithm, four different runs were accomplished by changing the k values in 2, 3, 4 and 5.

Attributes are selected from the dataset based on the ranking of the attributes as a result of feature selection algorithm. For example; in the experiments executed with four attributes for 9105 product, DEVUNITSTAT, ANTIDUMPING, TAXIDENUM and TARIFFCODE are used, and with three attributes, DEVUNITSTAT is discarded and remaining three attributes, ANTIDUMPING, TAXIDENUM and TARIFFCODE, are used.

CHAPTER 5

RESULTS

KNN was used for classification on the datasets of the selected tariff codes, and the accuracy and specificity values in Table 5-2 – 5-5 were obtained. The detailed performance values are presented in the Appendix E.

As the size of the datasets of the goods with tariff code 8302 and 9608 are too large, subsampling was applied and the sizes of the datasets were reduced to about 10,000. SpreadSubsample algorithm was used to subsample in Weka. This algorithm produces a random subsample of a dataset, allowing specifying the maximum "spread" between the rarest and most common class.

For calculations and performance measures, the non-circumvented transactions were treated as “Positive”, and the circumvented transactions were treated as “Negative”. Table 5-1 shows the necessary definitions.

Table 5-1: Definitions of terms

		Actual	
		Positive	***Negative***
Predicted	***Positive***	True Positive(TP)	False Positive (FP)
	Negative	False Negative(FN)	True Negative(TN)

According to the definitions in Table 5-1, the specificity and the accuracy values were calculated as shown in Equation 5.1 and 5.2;

$$Specificity = \frac{\sum TN}{\sum TN + \sum FP} \quad \text{(Equation 5-1)}$$

$$Accuracy = \frac{\sum TP + \sum TN}{\sum TP + \sum FP + \sum TN + \sum FN} \quad \text{(Equation 5-2)}$$

The experiments were done for the k = 2, 3, 4 and 5 values in the KNN Algorithm, and for each value of k, 13 experiments were carried out according to the number of the attributes used. Because the attribute rankings for the goods were different, the rankings presented in Table 4-9 should be taken into account in the Table 5-2 – 5-5. For example, in the experiments executed with four attributes, the dataset of the good 9105 included DEVUNITSTAT, ANTIDUMPING, TAXIDENUM and TARIFFCODE. But the dataset of the good 6001 included DEVUNITGROSS, ORIGCOUNTRY, TAXIDENUM and TARIFFCODE attributes. The details of the results of the experiments are available in Appendix E.

5.1. The Experiment Results on 9105

If the results of the experiments are examined in Table 5-2, it can be seen that the maximum accuracy value is 0.9988, and the corresponding specificity value is 0.9750, when k=2 and the number of attributes is 11, 12 or 13 and when k=3 and the number of attributes is 11, 12. When we look at the Table 5-2, the accuracy values greater than 0.9925 draw attentions. And the minimum specificity value is 0.6.

As a characteristic feature of this study, the maximization of the specificity value is important in order to ensure the effective use of the limited resources. High specificity values ensure the red line examinations (physical control) to be made on

the transactions with high fraud risk. So, the irregularities detected in the controls can be used for in risk assessments of the subsequent operations.

The best specificity values are obtained when k = 2. When the number of attribute is equal to one, i.e. only TARIFFCODE is used, the specificity value of 0.9250 for k=2 is obtained. When the number of attributes is higher up to n=11, lower specificity values are encountered. However, specificity reaches its highest value when 11, 12 or 13 attributes are employed. A high specificity value is obtained using only TARIFFCODE attribute can be explained by the concept of "shelter tariff code". This result occurs because of companies trying to import the good by using a false tariff code. Such code appears when the product with that tariff code has not been used before when importing to our country or has been imported before but in small number of transactions. Thus, the algorithm reaches high accuracy and specificity values, even if it evaluate all transactions as risky of which tariff code is the shelter tariff code.

Table 5-2: Specificity and Accuracy Values for the Experiments of Product 9105

n (# of	k=2		k=3		k=4		k=5	
attributes)	Specificity	Accuracy	Specificity	Accuracy	Specificity	Accuracy	Specificity	Accuracy
1	0.9250	0.9970	0.9000	0.9967	0.8750	0.9963	0.7625	0.9945
2	0.6500	0.9933	0.6500	0.9933	0.6500	0.9933	0.6000	0.9925
3	0.8375	0.9961	0.8125	0.9957	0.8000	0.9955	0.7000	0.9939
4	0.8750	0.9963	0.8500	0.9959	0.8375	0.9957	0.7375	0.9941
5	0.9000	0.9967	0.8750	0.9963	0.8375	0.9957	0.7625	0.9945
6	0.8750	0.9963	0.8500	0.9959	0.8125	0.9953	0.7500	0.9943
7	0.8875	0.9965	0.8625	0.9961	0.8250	0.9955	0.7500	0.9943
8	0.8875	0.9965	0.8625	0.9961	0.8250	0.9955	0.7500	0.9943
9	0.8750	0.9963	0.8375	0.9957	0.8250	0.9955	0.7500	0.9943
10	0.8750	0.9963	0.8375	0.9957	0.8250	0.9955	0.7625	0.9945
11	0.9750	0.9988	0.9750	0.9988	0.9625	0.9986	0.9625	0.9986
12	0.9750	0.9988	0.9750	0.9988	0.9625	0.9986	0.9625	0.9986
13	0.9750	0.9988	0.9625	0.9986	0.9625	0.9988	0.9625	0.9988
Maximum	0.9750	0.9988	0.9750	0.9988	0.9625	0.9988	0.9625	0.9988
Minimum	0.6500	0.9933	0.6500	0.9933	0.6500	0.9933	0.6000	0.9925

However, we could not come to a conclusion with respect to choosing the ideal k and the number of attributes according to the results. Therefore, we apply McNemar test (McNemar, 1967) to see whether there is a statistical difference between the results.

McNemar test is a non-parametric test which is used for comparing any given two experimental results. The tests are applied to a 2 x 2 contingency table, as shown in Table 5-3.

Table 5-3: Contingency Table for McNemar Test

		Prediction of Experiment 2	
		Positive	Negative
Prediction of Experiment 1	Positive	*a*	*b*
	Negative	*c*	*d*

Our null hypothesis is;

$$H_0 : p_b = p_c$$

The test statistic with correction of 1 is;

$$X^2 = \frac{(|b - c| - 1)^2}{b + c}$$

The McNemar test statistics results of the product 9105 are given in the Appendix F. The results are evaluated according to the significance level of 0.05. The p values, greater than 0.05 are emphasized in the table. According to this test, there is a statistically significant difference between (k = 2, n = 1) and (k = 2, n = 2), (k = 2, n = 3), (k = 3, n = 2), (k = 3, n = 3), (k = 3, n = 4), (k = 3, n = 6), (k = 3, n = 9), (k = 3, n = 10) (k = 4, n = 2 to 10) and (k = 5, n = 1 to 10) experiments.

However, for some experiments, we do not have sufficient evidence to reject the null hypothesis as seen in the table. Accordingly, for the k values (k = 2, 3 or 4), if the number of attribute n = 4, 5, 7, 8, 9, 10, 11, 12 or 13, there is no evidence that there is a statistically significant difference. Also there is no significant difference between the results for k=5 n=11, 12 or 13 and k= 2, 3 or 4 n= 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13.

According to the Table 5.2, the best results are obtained for k = 2 and k = 3. The test indicates that there is no significant difference between the results for n=4, 5, 6, 7, 8, 9, 10, 11, 12, 13 when k=2 and k=3. Even though we obtain higher specificity values when k=2 than k=3, according to the McNemar tests, when the experiments are compared for k = 2 and k = 3 and n = 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13, it is seen that there is no significant difference.

5.2. The Experiment Results on 6001

If the results of the experiments are examined in the Table 5-4, it can be seen that the maximum specificity value is 0.6197, and the corresponding accuracy value is 0.9974 when k= 2 and the number of attributes is 3. Although we obtain high accuracy results, the specificity values vary between 0.1127 and 0.6197.

The best specificity values are obtained when k = 2. The best results are acquired when k=2 and n=3 and n=5. However, the specificity results do not improve when n>5. This result shows the tariff code of a good imported to our country is used by the companies which have not imported that good before.

Table 5-4: Specificity and Accuracy Values for the Experiments of Product 6001

n (# of	k=2		k=3		k=4		k=5	
attributes)	Specificity	Accuracy	Specificity	Accuracy	Specificity	Accuracy	Specificity	Accuracy
1	0.3380	0.9941	0.3380	0.9941	0.3380	0.9941	0.3380	0.9941
2	0.2113	0.9944	0.1127	0.9938	0.1127	0.9938	0.1127	0.9938
3	0.6197	0.9974	0.5211	0.9968	0.5211	0.9968	0.5070	0.9967
4	0.5915	0.9973	0.5211	0.9968	0.5352	0.9969	0.5352	0.9969
5	0.6197	0.9973	0.5493	0.9969	0.5775	0.9971	0.5634	0.9970
6	0.5493	0.9969	0.5352	0.9968	0.5352	0.9968	0.5211	0.9967
7	0.5493	0.9969	0.5493	0.9969	0.5493	0.9969	0.5352	0.9968
8	0.5634	0.9970	0.5493	0.9969	0.5493	0.9969	0.5352	0.9968
9	0.5493	0.9969	0.5352	0.9968	0.5352	0.9968	0.5211	0.9967
10	0.5493	0.9969	0.5352	0.9968	0.5352	0.9968	0.5211	0.9967
11	0.5493	0.9969	0.5352	0.9968	0.5352	0.9968	0.5211	0.9967
12	0.5915	0.9971	0.5493	0.9968	0.5493	0.9969	0.5352	0.9968
13	0.5352	0.9968	0.5070	0.9967	0.4930	0.9966	0.4930	0.9966
Maximum	0.6197	0.9974	0.5493	0.9969	0.5775	0.9971	0.5634	0.9970
Minimum	0.2113	0.9941	0.1127	0.9938	0.1127	0.9938	0.1127	0.9938

The McNemar test statistics results of the product of 6001 are given in the Appendix G. The p values, greater than 0.05 are emphasized. According to this test, in general there is no statistically significant difference between k=2, k=3, k=4 and k=5. There is statistically significant difference between (k=2, n=2), (k=3, n=2), (k=4, n=2), (k=5, n=2) and all other experiment results. And there is not enough evidence to say there is a statistically significant difference between the experiments when k=2 and n=1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13.

According to the results in Table 5.4, when the experiment which obtained the best results with k = 2, n = 3 and n = 5 are compared with the other experiments, there is no evidence that experiments are significantly different when k = 2, n = 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13.

5.3. The Experiment Results on 8302

If the results of the experiments are examined in Table 5-5, it can be seen that the maximum specificity value is 0.9286, and the corresponding accuracy value is 0.9995.

Table 5-5: Specificity and Accuracy Values for the Experiments of Product 8302

n (# of	k=2		k=3		k=4		k=5	
attributes)	Specificity	Accuracy	Specificity	Accuracy	Specificity	Accuracy	Specificity	Accuracy
1	0.5000	0.9963	0.4714	0.9961	0.2714	0.9949	0.2286	0.9946
2	0.3143	0.9950	0.3143	0.9950	0.2571	0.9947	0.2143	0.9945
3	0.3571	0.9955	0.3571	0.9955	0.2286	0.9946	0.1571	0.9941
4	0.2714	0.9949	0.2143	0.9945	0.1714	0.9942	0.1143	0.9938
5	0.3143	0.9952	0.2714	0.9949	0.2000	0.9944	0.1143	0.9938
6	0.3571	0.9955	0.3143	0.9952	0.2143	0.9945	0.1286	0.9939
7	0.3429	0.9954	0.2857	0.9950	0.1714	0.9942	0.1000	0.9937
8	0.3429	0.9954	0.2857	0.9950	0.1714	0.9942	0.0857	0.9936
9	0.3286	0.9953	0.2714	0.9949	0.1286	0.9939	0.0429	0.9934
10	0.3286	0.9953	0.2714	0.9949	0.1286	0.9939	0.0286	0.9933
11	0.3286	0.9953	0.2714	0.9949	0.1286	0.9939	0.0286	0.9933
12	0.9286	0.9995	0.9143	0.9994	0.8857	0.9992	0.8714	0.9991
13	0.9000	0.9993	0.8857	0.9992	0.8571	0.9990	0.8429	0.9989
Maximum	0.9286	0.9995	0.9143	0.9994	0.8857	0.9992	0.8714	0.9991
Minimum	0.2714	0.9949	0.2143	0.9945	0.1286	0.9939	0.0286	0.9933

When we look across the table, the specificity values starting from 0.0286 are noteworthy. The accuracy values obtained are always greater than 0.9933.

The best specificity values are obtained when k = 2. The maximum accuracy and the specificity values are reached when the number of attributes is 12. The specificity values obtained from the experiments with fewer numbers of attributes are below 0.5000. The good results are obtained in the experiments with 12 and 13 attributes. This can be explained by using a tariff code of a good that the company has imported before, as shelter tariff code. When combined the information on the declaration and deviation values obtained by the previous declarations of the company, tariff code circumvention could be detected for this situation.

Corresponding probability values of comparative test statistics of the McNemar Test for the experiments of 8302 are shown in the Appendix H. The p values, greater than 0.05 are emphasized. In experiments, regardless of the value of k, it is observed that while the number of attribute is n=12 or 13, the obtained results are successful (table 5.3). In the McNemar tests, these results are compared within, for k = 2, 3, 4, and 5 and for n = 12 and n = 13 tests, there is no evidence of a statistically significant difference between these experiments, except (k=2, n=12) and (k=5, n=13).

5.4. The Experiment Results on 9608

If the results of the experiments are examined in Table 5-6, it can be seen that the maximum specificity value is 0.1463, and the corresponding accuracy value is 0.9917, when k=5 and the number of attributes is 12 or 13. Although we obtain high accuracy results, the specificity values vary between 0 and 0.1463.

Maximum accuracy and specificity values are reached when the number of attributes is 12 and 13. However, the specificity values are very low and very close to each other. The reasons for this are that the declared tariff code is a code of a good that was imported earlier both by the declarant company and by other companies, and inability to differentiate due to the similar features in terms of the worth, and the gross / net weight. If the goods with completely similar features were previously imported by the company, due to inability of differentiation between the two products, it is possible to fail in detecting the tariff code circumvention.

Table 5-6: Specificity and Accuracy Values for the Experiments of Product 9608

n (# of	k=2		k=3		k=4		k=5	
attributes)	Specificity	Accuracy	Specificity	Accuracy	Specificity	Accuracy	Specificity	Accuracy
1	0.0000	0.9918	0.0000	0.9919	0.0000	0.9919	0.0000	0.9919
2	0.0000	0.9918	0.0000	0.9919	0.0000	0.9919	0.0000	0.9919
3	0.0488	0.9921	0.0488	0.9922	0.0000	0.9918	0.0000	0.9918
4	0.0976	0.9915	0.0976	0.9916	0.0976	0.9916	0.0976	0.9916
5	0.0854	0.9914	0.0854	0.9915	0.0854	0.9915	0.0976	0.9916
6	0.0854	0.9913	0.0854	0.9914	0.0854	0.9914	0.0976	0.9915
7	0.0854	0.9913	0.0854	0.9914	0.0854	0.9914	0.0976	0.9915
8	0.0854	0.9914	0.0854	0.9914	0.0854	0.9914	0.0854	0.9915
9	0.0854	0.9914	0.0854	0.9914	0.0854	0.9914	0.0854	0.9915
10	0.0854	0.9913	0.0854	0.9913	0.0854	0.9913	0.0854	0.9914
11	0.0854	0.9913	0.0854	0.9913	0.0854	0.9913	0.0854	0.9914
12	0.1463	0.9915	0.0000	0.9918	0.1463	0.9917	0.1463	0.9917
13	0.1463	0.9915	0.1463	0.9916	0.1463	0.9917	0.1463	0.9917
Maximum	0.1463	0.9921	0.1463	0.9922	0.1463	0.9919	0.1463	0.9919
Minimum	0.1463	0.9921	0.1463	0.9922	0.1463	0.9919	0.1463	0.9919

Because of very low specificity results are obtained in experiments of 9608, McNemar tests are not applied.

5.5. Comparison

In order to allow a comparison between the current system and the proposed framework, the first let's do an assessment of the current system. At current system used in customs business process, the risk profiles constituted by Risk Analysis Office are used to determine the risky import declarations. The declarations matching with these risk profiles are assigned to red line.

For each product groups, the numbers of declarations assigned to the red line (considered risky) and yellow line (considered not risky), and also the numbers of declarations detected as illegal are available in the tables at Appendix İ. According to this information, the ratios of the red line assignments in current system and proposed system can be expressed as in Table 5-7. As seen from the table, while current system ratios are about 60% or above, proposed system ratios are below 2%.

In order to make a comparison between two systems, the current system declarations detected as illegal other than tariff circumvention are excluded and the ratios are also calculated according to these numbers.

Table 5-7: The Red Line Ratios of Current System and Proposed System

Product Group	Current System	Current System (Excluding Declarations Detected as Other Illegalities)	Proposed System (k=2, n=12)
9105	80.29 %	80.30 %	1.62 %
6001	59.38 %	57.80 %	0.41 %
8302	59.54 %	59.53 %	0.65 %
9608	59.15 %	59.15 %	0.28 %

In the system, there are illegalities other than tariff code circumvention, e.g. low-valued declarations. The total number of declarations detected as illegal and the number of declarations detected as tariff code circumvented are shown in Table 5-8. The ratios are quite different from each other. While the ratio of tariff code circumvention of product group 6001 is 9.88 %, the ratio of product group 9105 is 78.64 %. According to this information, we can say about 9105 product group that by assigning only 1.62% of the declarations to red line, we are able to detect 78.64 x 0.9750 = 76.67% of all illegalities. (0.9750 specificity value for 9105 k=2 and n=12). For 6001 product group, by assigning 0.41% of the declarations to red line, we are able to detect 9.88 x 0.5915 = 5.84% of all illegalities. The proposed system will be helpful for all type of goods by decreasing the red line control numbers and allowing the control officers to focus on other illegalities.

Table 5-8: The Number of Illegal Declarations and The Ratio of Tariff Circumvention

	Total Detected Illegalities	Detected as Tariff Code Circumvented	Ratio
9105	103	81	78.64 %
6001	729	72	9.88 %
8302	299	72	24.08 %
9608	162	82	50.62 %

The statistical performance measures for these numbers are shown in the Table 5-9 and Table 5-10.

In the current system, the irregularities other than tariff code circumvention are also identified. Accordingly, we evaluate the system's performance taking into account all the data; first we get the performance values in Table 5-9. For example, we obtain such a high specificity value of 0.9223 for the good 9105. However, the same success is not observed at sensitivity value (0.1996). In addition, it is seen that the accuracy is also low. For other products, there are very similar results. While the good 8302 have the lowest specificity value with 0.6957, corresponding sensitivity value is 0.4048.

Table 5-9: Statistical Performance Measures of the Current System

	Sensitivity	Specificity	Accuracy	F Measure	FP Rate	Precision
9105	0,1996	0,9223	0,2142	0,3323	0,0777	0,9920
6001	0,4234	0,8450	0,4505	0,5906	0,1550	0,9755
8302	0,4048	0,6957	0,4055	0,5761	0,3043	0,9983
9608	0,4086	0,8025	0,4087	0,5801	0,1975	0,9999

In the situation of not considering irregularities other than tariff code circumvention, we face with the results in Table 5-10. Again, for the product of 9105, we reach such a high specificity value of 0.9630. However, it is seen that the sensitivity value is 0.1996 and the value of accuracy is 0.2118. These values are very low when compared with the framework we propose. The sensitivity and specificity values for current system and proposed system are listed at Table 5-11. Although, 4064 declaration directed to the red line of 5061 declaration in current system, only 78 declarations have been identified as illegal. This result makes the performance of the current system lowers.

Table 5-10: Statistical Performance Measures of the Current System Excluding Other Illegalities

	Sensitivity	Specificity	Accuracy	F Measure	FP Rate	Precision
9105	0,1996	0,9630	0,2118	0,3326	0,0370	0,9970
6001	0,4234	0,7917	0,4259	0,5944	0,2083	0,9967
8302	0,4048	0,8194	0,4051	0,5763	0,1806	0,9998
9608	0,4086	0,9024	0,4087	0,5801	0,0976	1,0000

Similar conditions apply to the other three products. Higher number of assignments at the red line cause the accuracy measure decreases, at the same time lead to loss of manpower. In order to use human resources more efficiently, it is possible to reach higher accuracy values with high specificity values in determination of tariff code circumvention by using a data mining framework proposed in this thesis.

Table 5-11: Performance Measures Comparison of Current System and Proposed System

	Current System		**Proposed System**	
	Sensitivity	Specificity	Sensitivity	Specificity
9105	0,1996	0,9630	0,9992	0,9750
6001	0,4234	0,7917	0,9996	0,5915
8302	0,4048	0,8194	1,0000	0,9286
9608	0,4086	0,9024	0,9984	0,1463

CHAPTER 6

DISCUSSION & CONCLUSION

In this study, the tariff circumvention attempted to avoid anti-dumping tax which is one of the NTBs were focused on. Anti-dumping tax is one of the additional financial obligations enforced for low-priced imports to protect domestic industries from unfair competition. In Turkey, the investigation on dumping is conducted and enforced by the Ministry of Economy to regulate the trade. Anti-dumping duty is accrued according to the tariff code and the country of origin of the goods.

It is known that misdeclarations have been made on the information content of a declaration in order to avoid paying the tax in question. Declaring a tariff code that is different from the actual tariff code is one of them. This circumvention eliminates the effect of protecting the domestic producers, as well as causing tax loss.

Risk Analysis Office in the Ministry of Customs and Trade, to be able to detect this kind of risky situations, makes a study examining the customs transactions on GÜVAS six months after the initiation of force of the anti-dumping tax (and on a regular basis throughout the period of time of the tax). In this study, the behaviors of the importers who import the product before tax are examined and in the case of deviation from the past records, the risk profile was prepared for the risk analysis

system. Then, it routes the similar declarations to the red line for the rest. Thus, if there is any difference on the tariff code, it may be detected during inspection.

In this study, on the basis of the risk analysis office's study, the prediction of the tariff circumvention is aimed with data mining methods. For this purpose, four of the most circumvention detected products according to the records in the customs database were discussed. The experiments were carried out with 6 fields on the declaration and 7 calculated attributes depending on the company's past records, a total of 13 attributes. KNN classification algorithm was used in the experiments. The results obtained for each product contains different features, the common points are also available.

Let us take 9105 as the first product coded goods. In the experiments performed subject to the product, the better results were obtained for k = 2. But the McNemar test showed that there was no difference between k=2 and k= 3 or k=4. The 92.50% specificity value obtained when the number of attributes is one, resulted from the shelter tariff code imports from the non-or less than the amount in the case of import of products. Thus, even if all the transactions that import the product are considered as to be risky, a high-performance of the algorithm can be obtained. When the number of attributes was between 4 and 13, the best results were obtained.

Analyzing the results for the 6001 product, it is seen that k=2 gives a better result than k=3, 4 and 5. McNemar test results show us that there is a significant difference in the statistical sense. The best result obtained when the number of attribute is three or five and k=2. It is seen that the results of n=4, 6, 7, 8, 9, 10, 11, 12 and 13 are not different than the result of n=3 or n=5 statistically. Not getting good results as in 9105 product has shown that a tariff code already imported is used for shelter tariff code. However, for the companies not imported shelter tariff coded merchandise never before, the algorithm is able to acquire higher performance values without needing to the historical information of the company.

Shown that for 8302, k = 2 gives better results than k = 3, 4 and 5. However, successful results cannot be taken for low n values in this product, the best results have been obtained for n = 12 and 13. When the dataset is analyzed, it is seen that the companies have already imported the product with shelter tariff code. But the

deviation attributes calculated from the past records of the company helps to differentiate.

Low specificity values could be obtained for the 9608 product as a result of the KNN. Looking at past records related to the shelter tariff code that used by the company, since it is similar in terms of net/gross weight or worth, the algorithm performance was low.

In light of these assessments, k = 2 and n = 12 parameter values are the common parameters to be used for future classification studies. We do not need to take into account the CONTCOUNTRY attribute, because of the last place taken in ranking at the classification study. In the classification of using all other attributes, by using two nearest neighbor classification (k = 2) good results will be obtained.

The issue of fraud detection in customs has some challenges. First of all, the number of declarations detected as having fraud is very small in accordance to total number of declarations. Another issue is that the probability of the data set containing missed declarations is very high.

Under these conditions, our purpose was to produce more accurate decision about risk of the declarations and to assign only risky declarations to the inspection officers. The proposed solution has led to gain a high rate of success on detecting tariff code circumvention by controlling smaller number of declarations on red line. The framework also provides the ability of detection of circumvention at the time of enforcement of an anti - dumping tax without needing long period of work on historical declarations data.

For future, another technique such as SVM (Support Vector Machines) could be performed on customs data set. Also, considering the data mining system for customs purposes as a whole could be helpful. Thus, the detection of commercial crimes, not only the tariff code circumvention, but also country of origin circumvention and low worth declaration, and the estimation of smuggling crimes could be obtained.

REFERENCES

Avcı, I.E. (2001). Risk Management System for Effective Customs Inspections (Model Proposal for Turkey), Expertise Thesis, Ankara: Ministry of Customs and Trade.

Başaran, B. (2004). Surveillance and Safeguard Measures in Import, Expertise Thesis, Ankara: Ministry of Customs and Trade.

Baştabak, B. & Medeni, T.D. (2011). Application of Informatics Technologies into Customs: Origin and Tariff Code Diversion, Impacts and Identification Problem. *International Journal of eBusiness and eGovernment Studies, vol. 3,* no. 1, pp. 17-27 (Online)

Beğen, Ş. (2007). Analysis and Evaluation of Dumping and Application of Anti-dumping Tax in Our Country. (M.Sc. thesis). Gazi University Institute of Social Sciences.

Can, E. (2005). The Importance of Risk Management and Intelligence Systems on Prevention of Smuggling, Expertise Thesis, Ankara: Ministry of Customs and Trade.

Cete, E. (1998). The Problem of the Anti-Dumping Legislation Compliance in the Framework of Prevention of Unfair Competition In Importation, Expertise Thesis, Ankara: Ministry of Customs and Trade.

Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification, *Information Theory, IEEE Transactions on, vol.13(1),* pp.21-27.

Dalkıran, Y. (1998). Non-Tariff Barriers and Applications in the EU and Turkey, Expertise Thesis, Ankara: Ministry of Customs and Trade.

Digiampietri, L.A., Roman, N.T., Meira, L.A.A., Filho, J.J., Ferreira, C.D. & Kondo, A.A. (2008). Uses of Artificial Intelligence in the Brazilian Customs Fraud Detection System. *Proceedings of the 9th Annual International Digital Government Research Conference, vol.,* no., pp.181-187.

Evans, G., Newnham, J. (1998). *Dictionary of International Relations,* Penguin Books

Forrester, I. S. & Kaul, T. (2005). Tariff Classification. In P. F. J. Macrory, A.E. Appleton & M.G. Plummer (Eds.), The World Trade Organization: Legal, economic and political analysis (pp. 1581-1599). USA, Springer.

Hua, Z., Li, S. & Tao, Z. (2006). A Rule-based Risk Decision-Making Approach and Its Application in China's Customs Inspection. *Journal of the Operational Research Society, vol.57,* no.11, pp.1313-1322.

Karakoç, H. (2008). Non-Tariff Barriers, Legislation, Practice and Trends, Expertise Thesis, Ankara: Ministry of Customs and Trade.

Keskin, Ö. (2011). Measures Taken Against Unfair Competition in Import, Comparison of the EU and Turkey, Recommendations for Prevention of Trade Diversion Caused By Measures Based on Tariff Quotas, Expertise Thesis, Ankara: Ministry of Customs and Trade.

Kumar, A. & Nagadevara, V. (2006). Development of Hybrid Classification Methodology for Mining Skewed Data Sets: A Case Study of Indian Customs Data. *International Conference on Computer Systems and Applications, vol.,* no., pp. 584-591.

Kumar Singh, A., Sahu, R. & Ujjwal, K. (2003). Decision Support System in Customs Assessment to Detect Valuation Frauds. *Engineering Management Conference Managing Technologically Driven Organizations: The Human Side of Innovation and Change, vol.,* no., pp. 546-550

McNemar. Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12(2),* 153-157.

Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.

Öztürk, A. (2005). Risk Analysis Techniques at Import, Export and Customs Crimes, Expertise Thesis, Ankara: Ministry of Customs and Trade.

Phua, C., Lee, V., Smith, K. & Gayler, R. (2005). A Comprehensive Survey of Data Mining-based Fraud Detection Research. vol., no., pp.1-14.

Shao, H., Zhao, H. & Chang, G. (2002). Applying Data Mining to Detect Fraud Behavior in Customs Declaration. *Proceedings of the First International Conference on Machine Learning and Cybernetics, vol.3,* no., pp. 1241-1244.

UNCTAD and World Customs Organization (2008), Technical Note No: 12 Risk management in Customs procedures, http://r0.unctad.org/ttl/technical-notes/TN12_RiskManagement.pdf, [Accessed 15.01.2012]

Yan-Hai, L. & Lin-Yan, S. (2005). Study and Applications of Data Mining to the Structure Risk Analysis of Customs Declaration Cargo. *Proceedings of the 2005 IEEE International Conference on e-Business Engineering, IEEE Computer Society, vol.,* no., pp.761-764.

Yaqin, W. & Yuming, S. (2010). Classification Model Based on Association Rules in Customs Risk Management Application, International Conference on Intelligent System Design and Engineering Application, vol.1, no., pp.436-439.

Ye, F., Zhou, G. & Lu, J. (2007). A BP Neural Network Approach on Risk Evaluation of Ventures in China Customs. *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing, vol.,* no., pp.6175-6178

Zhao, J. & Chang, Z. (2006). Neuro-Fuzzy Decision Tree by Fuzzy ID3 Algorithm and Its Application to Anti-Dumping Early-Warning System. *Proceedings of the International Conference on Information Acquisition, vol.,* no., pp.1300-1304.

APPENDICES

APPENDIX A: A Part of the 44th Chapter of Tariff Code List

TARIFF CODE	DESCRIPTION OF GOODS
4418	Builders' joinery and carpentry of wood, including cellular wood panels, assembled flooring panels, shingles and shakes:
4418.10	- Windows, French windows and their frames:
4418.10.10.00.00	- - Of tropical wood, as specified in additional note 2 to this chapter
4418.10.50.00.00	- - Coniferous
4418.10.90.00.00	- - Other
4418.20	- Doors and their frames and thresholds:
4418.20.10.00.00	- - Of tropical wood, as specified in additional note 2 to this chapter
4418.20.50.00.00	- - Coniferous
4418.20.80.00.00	- - Of other wood
4418.71.00.00.00	- - For mosaic floors
4418.72.00.00.00	- - Other, multilayer
4418.79.00.00.00	- - Other
4418.90	- Other:
4418.90.10.00.00	- - Glue-laminated timber
	- - Other:
4418.90.80.10.00	- - - Cellular wooden panels
	- - - Other
4418.90.80.90.11	- - - - Ladder
4418.90.80.90.12	- - - - Built-in cupboard
4418.90.80.90.13	- - - - Construction castings
4418.90.80.90.14	- - - - Sun blind and shutters

APPENDIX B: The Number of Anti - Dumping Measures Enforced by Turkey

Country	# of dumpings
China, P.R.	**47**
Indonesia	10
Chinese Taipei	8
India	**8**
Malaysia	7
Thailand	7
Vietnam	6
Korea, Rep. of	3
Russian Federation	**3**
Bulgaria	**2**
Saudi Arabia	2
The Philippines	**2**
USA	**2**
Brazil	1
Canada	1
Germany	**1**
Hong Kong	1
Italy	**1**
Kuwait	**1**
Pakistan	**1**
Romania	1
Sri Lanka	**1**
Ukraine	1

APPENDIX C: Tariff Classification of The Selected Goods

Tariff Code	Description of the Good 9105
9105	Other clocks:
	- Alarm clocks:
9105.11.00.00.00	**- - Electrically operated**
9105.19.00.00.00	- - Other
	- Wall clocks:
9105.21.00.00.00	- - Electrically operated
9105.29.00.00.00	**- - Other**
	- Other:
9105.91.00.00.00	**- - Electrically operated**
9105.99.00.00.00	- - Other

Tariff Code	Description of the Good 6001
6001	**Pile fabrics, including 'long pile' fabrics and terry fabrics, knitted or crocheted:**
6001.10	**- 'Long pile' fabrics**
6001.10.00.00.11	- - Of synthetic fibers
6001.10.00.00.12	- - Of wool or fine animal hair, of cotton or artificial fibers
6001.10.00.00.19	- - Other
	- Looped pile fabrics:
6001.21.00.00.00	- - Of cotton
6001.22.00.00.00	- - Of man-made fibres
6001.29	**- - Of other textile materials**
6001.29.00.00.11	- - - Of wool or fine animal hair
	- - - Other
6001.29.00.00.91	- - - - Of silk and silk waste
6001.29.00.00.99	- - - - Other
	- Other:
6001.91.00.00.00	- - Of cotton
6001.92.00.00.00	- - Of man-made fibres
6001.99	**- - Of other textile materials**
6001.99.00.00.11	- - - Of wool or fine animal hair
	- - - Other
6001.99.00.00.91	- - - - Of silk and silk waste
6001.99.00.00.99	- - - - Other

Tariff Code	Description of the Good 8302
8302	Base metal mountings, fittings and similar articles suitable for furniture, doors, staircases, windows, blinds, coachwork, saddlery, trunks, chests, caskets or the like; base metal hat-racks, hat-pegs, brackets and similar fixtures; castors with mountings of base metal; automatic door closers of base metal:
8302.10.00.00.00	- Hinges
8302.20.00.00.00	- Castors
8302.30.00.00.00	- Other mountings, fittings and similar articles suitable for motor vehicles
	- Other mountings, fittings and similar articles:
	- - For buildings:
8302.41.10.00.00	- - - For doors
8302.41.50.00.00	- - - For Windows and French Windows
8302.41.90.00.00	- - - Other
8302.42.00.00.00	- - Other, suitable for furniture
8302.49.00.00.00	- - Other
8302.50.00.00.00	- Hat-racks, hat-pegs, brackets and similar fixtures
8302.60.00.00.00	- Automatic door closers

Tariff Code	Description of the Good 9608
9608	Ballpoint pens; felt-tipped and other porous-tipped pens and markers; fountain pens, stylograph pens and other pens; duplicating stylos; propelling or sliding pencils; pen-holders, pencil-holders and similar holders; parts (including caps and clips) of the foregoing articles, other than those of heading 9609:
9608.10	- Ballpoint pens:
	- - With liquid ink (rolling ball pens)
9608.10.10.10.00	- - - Of plastics
9608.10.10.90.00	- - - Other
	- - Other:
9608.10.92.00.00	- - - With replaceable refill
9608.10.99.00.00	- - - Other
…	…
9608.50	- Sets of articles from two or more of the foregoing subheadings
9608.50.00.10.00	- - Of plastics
	- - Other
9608.50.00.90.11	**- - - Of body lids of precious metals or metals coated with precious metals**
9608.50.00.90.19	- - - Other
9608.60.00.00.00	**- Refills for ballpoint pens, comprising the ball point and ink-reservoir:**
	- Other

APPENDIX D: The Number of Declarations by Year

		After Tax			
Tariff Code	Before Tax	Year	Detected	Remaining	Total
9105	118	2001	0	15	
		2002	5	114	
		2003	6	174	
		2004	12	251	
		2005	5	357	
		2006	5	395	
		2007	13	541	
		2008	8	557	
		2009	3	626	
		2010	11	875	
		2011	13	976	
		N/S	0	3	
		Total	81	4884	5083

		After Tax			
Tariff Code	Before Tax	Year	Detected	Remaining	Total
6001	1455	2002	0	82	
		2003	0	1128	
		2004	3	1167	
		2005	10	1035	
		2006	4	801	
		2007	6	743	
		2008	6	855	
		2009	5	920	
		2010	23	1448	
		2011	15	1650	
		Total	72	9829	11356

		After Tax			
Tariff Code	Before Tax	Year	Detected	Remaining	Total
8302	16329	2004	6	8169	
		2005	7	10320	
		2006	6	12877	
		2007	11	13072	
		2008	7	15468	
		2009	10	15507	
		2010	14	19055	
		2011	11	19277	
		N/S	0	3	
		Total	72	113748	130149

		After Tax			
Tariff Code	Before Tax	Year	Detected	Remaining	Total
9608	22863	2004	12	8214	
		2005	12	10888	
		2006	7	12727	
		2007	11	88135	
		2008	7	99160	
		2009	8	87779	
		2010	13	107404	
		2011	11	107051	
		Total	81	521358	544302

APPENDIX E: Experiment Results

The Product with 9105 Tariff Code								
# of	k=2				k=3			
attributes	Sensitivity	Specificity	Accuracy	F	Sensitivity	Specificity	Accuracy	F
1	0.9982	0.9250	0.9970	0.9985	0.9982	0.9000	0.9967	0.9983
2	0.9988	0.6500	0.9933	0.9966	0.9988	0.6500	0.9933	0.9966
3	0.9986	0.8375	0.9961	0.9980	0.9986	0.8125	0.9957	0.9978
4	0.9982	0.8750	0.9963	0.9981	0.9982	0.8500	0.9959	0.9979
5	0.9982	0.9000	0.9967	0.9983	0.9982	0.8750	0.9963	0.9981
6	0.9982	0.8750	0.9963	0.9981	0.9982	0.8500	0.9959	0.9979
7	0.9982	0.8875	0.9965	0.9982	0.9982	0.8625	0.9961	0.9980
8	0.9982	0.8875	0.9965	0.9982	0.9982	0.8625	0.9961	0.9980
9	0.9982	0.8750	0.9963	0.9981	0.9982	0.8375	0.9957	0.9978
10	0.9982	0.8750	0.9963	0.9981	0.9982	0.8375	0.9957	0.9978
11	0.9992	0.9750	0.9988	0.9994	0.9992	0.9750	0.9988	0.9994
12	0.9992	0.9750	0.9988	0.9994	0.9992	0.9750	0.9988	0.9994
13	0.9992	0.9750	0.9988	0.9994	0.9992	0.9625	0.9986	0.9993
Maximum	0.9992	0.9750	0.9988	0.9994	0.9992	0.9750	0.9988	0.9994
Minimum	0.9982	0.6500	0.9933	0.9966	0.9982	0.6500	0.9933	0.9966
# of	k=4				k=5			
attributes	Sensitivity	Specificity	Accuracy	F	Sensitivity	Specificity	Accuracy	F
1	0.9982	0.8750	0.9963	0.9981	0.9982	0.7625	0.9945	0.9972
2	0.9988	0.6500	0.9933	0.9966	0.9988	0.6000	0.9925	0.9962
3	0.9986	0.8000	0.9955	0.9977	0.9986	0.7000	0.9939	0.9969
4	0.9982	0.8375	0.9957	0.9978	0.9982	0.7375	0.9941	0.9970
5	0.9982	0.8375	0.9957	0.9978	0.9982	0.7625	0.9945	0.9972
6	0.9982	0.8125	0.9953	0.9976	0.9982	0.7500	0.9943	0.9971
7	0.9982	0.8250	0.9955	0.9977	0.9982	0.7500	0.9943	0.9971
8	0.9982	0.8250	0.9955	0.9977	0.9982	0.7500	0.9943	0.9971
9	0.9982	0.8250	0.9955	0.9977	0.9982	0.7500	0.9943	0.9971
10	0.9982	0.8250	0.9955	0.9977	0.9982	0.7625	0.9945	0.9972
11	0.9992	0.9625	0.9986	0.9993	0.9992	0.9625	0.9986	0.9993
12	0.9992	0.9625	0.9986	0.9993	0.9992	0.9625	0.9986	0.9993
13	0.9994	0.9625	0.9988	0.9994	0.9994	0.9625	0.9988	0.9994
Maximum	0.9994	0.9625	0.9988	0.9994	0.9994	0.9625	0.9988	0.9994
Minimum	0.9982	0.6500	0.9933	0.9966	0.9982	0.6000	0.9925	0.9962

The Product with 6001 Tariff Code								
# of	k=2				k=3			
attributes	Sensitivity	Specificity	Accuracy	F	Sensitivity	Specificity	Accuracy	F
1	0.9983	0.3380	0.9941	0.9971	0.9983	0.3380	0.9941	0.9971
2	0.9994	0.2113	0.9944	0.9972	0.9994	0.1127	0.9938	0.9969
3	0.9998	0.6197	0.9974	0.9987	0.9998	0.5211	0.9968	0.9984
4	0.9998	0.5915	0.9973	0.9986	0.9998	0.5211	0.9968	0.9984
5	0.9997	0.6197	0.9973	0.9987	0.9997	0.5493	0.9969	0.9984
6	0.9997	0.5493	0.9969	0.9984	0.9997	0.5352	0.9968	0.9984
7	0.9997	0.5493	0.9969	0.9984	0.9997	0.5493	0.9969	0.9984
8	0.9997	0.5634	0.9970	0.9985	0.9997	0.5493	0.9969	0.9984
9	0.9997	0.5493	0.9969	0.9984	0.9997	0.5352	0.9968	0.9984
10	0.9997	0.5493	0.9969	0.9984	0.9997	0.5352	0.9968	0.9984
11	0.9997	0.5493	0.9969	0.9984	0.9997	0.5352	0.9968	0.9984
12	0.9996	0.5915	0.9971	0.9985	0.9996	0.5493	0.9968	0.9984
13	0.9997	0.5352	0.9968	0.9984	0.9998	0.5070	0.9967	0.9984
Maximum	0.9998	0.6197	0.9974	0.9987	0.9998	0.5493	0.9969	0.9984
Minimum	0.9983	0.2113	0.9941	0.9971	0.9983	0.1127	0.9938	0.9969

# of	k=4				k=5			
attributes	Sensitivity	Specificity	Accuracy	F	Sensitivity	Specificity	Accuracy	F
1	0.9983	0.3380	0.9941	0.9971	0.9983	0.3380	0.9941	0.9971
2	0.9994	0.1127	0.9938	0.9969	0.9994	0.1127	0.9938	0.9969
3	0.9998	0.5211	0.9968	0.9984	0.9998	0.5070	0.9967	0.9984
4	0.9998	0.5352	0.9969	0.9984	0.9998	0.5352	0.9969	0.9984
5	0.9997	0.5775	0.9971	0.9985	0.9997	0.5634	0.9970	0.9985
6	0.9997	0.5352	0.9968	0.9984	0.9997	0.5211	0.9967	0.9984
7	0.9997	0.5493	0.9969	0.9984	0.9997	0.5352	0.9968	0.9984
8	0.9997	0.5493	0.9969	0.9984	0.9997	0.5352	0.9968	0.9984
9	0.9997	0.5352	0.9968	0.9984	0.9997	0.5211	0.9967	0.9984
10	0.9997	0.5352	0.9968	0.9984	0.9997	0.5211	0.9967	0.9984
11	0.9997	0.5352	0.9968	0.9984	0.9997	0.5211	0.9967	0.9984
12	0.9997	0.5493	0.9969	0.9984	0.9997	0.5352	0.9968	0.9984
13	0.9998	0.4930	0.9966	0.9983	0.9998	0.4930	0.9966	0.9983
Maximum	0.9998	0.5775	0.9971	0.9985	0.9998	0.5634	0.9970	0.9985
Minimum	0.9983	0.1127	0.9938	0.9969	0.9983	0.1127	0.9938	0.9969

The Product with 8302 Tariff Code								
# of	k=2				k=3			
attributes	Sensitivity	Specificity	Accuracy	F	Sensitivity	Specificity	Accuracy	F
1	0.9998	0.5000	0.9963	0.9982	0.9998	0.4714	0.9961	0.9981
2	0.9998	0.3143	0.9950	0.9975	0.9998	0.3143	0.9950	0.9975
3	1.0000	0.3571	0.9955	0.9978	1.0000	0.3571	0.9955	0.9978
4	1.0000	0.2714	0.9949	0.9975	1.0000	0.2143	0.9945	0.9973
5	1.0000	0.3143	0.9952	0.9976	1.0000	0.2714	0.9949	0.9975
6	1.0000	0.3571	0.9955	0.9978	1.0000	0.3143	0.9952	0.9976
7	1.0000	0.3429	0.9954	0.9977	1.0000	0.2857	0.9950	0.9975
8	1.0000	0.3429	0.9954	0.9977	1.0000	0.2857	0.9950	0.9975
9	1.0000	0.3286	0.9953	0.9977	1.0000	0.2714	0.9949	0.9975
10	1.0000	0.3286	0.9953	0.9977	1.0000	0.2714	0.9949	0.9975
11	1.0000	0.3286	0.9953	0.9977	1.0000	0.2714	0.9949	0.9975
12	1.0000	0.9286	0.9995	0.9998	1.0000	0.9143	0.9994	0.9997
13	1.0000	0.9000	0.9993	0.9997	1.0000	0.8857	0.9992	0.9996
Maximum	1.0000	0.9286	0.9995	0.9998	1.0000	0.9143	0.9994	0.9997
Minimum	0.9998	0.2714	0.9949	0.9975	0.9998	0.2143	0.9945	0.9973

# of	k=4				k=5			
attributes	Sensitivity	Specificity	Accuracy	F	Sensitivity	Specificity	Accuracy	F
1	1.0000	0.2714	0.9949	0.9975	1.0000	0.2286	0.9946	0.9973
2	0.9999	0.2571	0.9947	0.9974	1.0000	0.2143	0.9945	0.9973
3	1.0000	0.2286	0.9946	0.9973	1.0000	0.1571	0.9941	0.9971
4	1.0000	0.1714	0.9942	0.9971	1.0000	0.1143	0.9938	0.9969
5	1.0000	0.2000	0.9944	0.9972	1.0000	0.1143	0.9938	0.9969
6	1.0000	0.2143	0.9945	0.9973	1.0000	0.1286	0.9939	0.9970
7	1.0000	0.1714	0.9942	0.9971	1.0000	0.1000	0.9937	0.9969
8	1.0000	0.1714	0.9942	0.9971	1.0000	0.0857	0.9936	0.9968
9	1.0000	0.1286	0.9939	0.9970	1.0000	0.0429	0.9934	0.9967
10	1.0000	0.1286	0.9939	0.9970	1.0000	0.0286	0.9933	0.9966
11	1.0000	0.1286	0.9939	0.9970	1.0000	0.0286	0.9933	0.9966
12	1.0000	0.8857	0.9992	0.9996	1.0000	0.8714	0.9991	0.9996
13	1.0000	0.8571	0.9990	0.9995	1.0000	0.8429	0.9989	0.9995
Maximum	1.0000	0.8857	0.9992	0.9996	1.0000	0.8714	0.9991	0.9996
Minimum	0.9999	0.1286	0.9939	0.9970	1.0000	0.0286	0.9933	0.9966

The Product with 9608 Tariff Code								
# of	k=2				k=3			
attributes	Sensitivity	Specificity	Accuracy	F	Sensitivity	Specificity	Accuracy	F
1	0.9999	0.0000	0.9918	0.9959	1.0000	0.0000	0.9919	0.9959
2	0.9999	0.0000	0.9918	0.9959	1.0000	0.0000	0.9919	0.9959
3	0.9998	0.0488	0.9921	0.9960	0.9999	0.0488	0.9922	0.9961
4	0.9988	0.0976	0.9915	0.9957	0.9989	0.0976	0.9916	0.9958
5	0.9988	0.0854	0.9914	0.9957	0.9989	0.0854	0.9915	0.9957
6	0.9987	0.0854	0.9913	0.9956	0.9988	0.0854	0.9914	0.9957
7	0.9987	0.0854	0.9913	0.9956	0.9988	0.0854	0.9914	0.9957
8	0.9988	0.0854	0.9914	0.9957	0.9988	0.0854	0.9914	0.9957
9	0.9988	0.0854	0.9914	0.9957	0.9988	0.0854	0.9914	0.9957
10	0.9987	0.0854	0.9913	0.9956	0.9987	0.0854	0.9913	0.9956
11	0.9987	0.0854	0.9913	0.9956	0.9987	0.0854	0.9913	0.9956
12	0.9984	0.1463	0.9915	0.9957	0.9999	0.0000	0.9918	0.9959
13	0.9984	0.1463	0.9915	0.9957	0.9985	0.1463	0.9916	0.9958
Maximum	0.9999	0.1463	0.9921	0.9960	1.0000	0.1463	0.9922	0.9961
Minimum	0.9984	0.0000	0.9913	0.9956	0.9985	0.0000	0.9913	0.9956

# of	k=4				k=5			
attributes	Sensitivity	Specificity	Accuracy	F	Sensitivity	Specificity	Accuracy	F
1	1.0000	0.0000	0.9919	0.9959	1.0000	0.0000	0.9919	0.9959
2	1.0000	0.0000	0.9919	0.9959	1.0000	0.0000	0.9919	0.9959
3	0.9999	0.0000	0.9918	0.9959	0.9999	0.0000	0.9918	0.9959
4	0.9989	0.0976	0.9916	0.9958	0.9989	0.0976	0.9916	0.9958
5	0.9989	0.0854	0.9915	0.9957	0.9989	0.0976	0.9916	0.9958
6	0.9988	0.0854	0.9914	0.9957	0.9988	0.0976	0.9915	0.9957
7	0.9988	0.0854	0.9914	0.9957	0.9988	0.0976	0.9915	0.9957
8	0.9988	0.0854	0.9914	0.9957	0.9989	0.0854	0.9915	0.9957
9	0.9988	0.0854	0.9914	0.9957	0.9989	0.0854	0.9915	0.9957
10	0.9987	0.0854	0.9913	0.9956	0.9988	0.0854	0.9914	0.9957
11	0.9987	0.0854	0.9913	0.9956	0.9988	0.0854	0.9914	0.9957
12	0.9986	0.1463	0.9917	0.9958	0.9986	0.1463	0.9917	0.9958
13	0.9986	0.1463	0.9917	0.9958	0.9986	0.1463	0.9917	0.9958
Maximum	1.0000	0.1463	0.9919	0.9959	1.0000	0.1463	0.9919	0.9959
Minimum	0.9986	0.0000	0.9913	0.9956	0.9986	0.0000	0.9914	0.9957

APPENDIX F: McNemar Test Statistic Results for 9105

	k	2													3												
k	n	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
2	1		0.00	0.01	0.13	0.48	0.13	0.25	0.25	0.13	0.13	1.00	1.00	1.00	0.48	0.00	0.00	0.04	0.13	0.04	0.07	0.07	0.02	0.02	1.00	1.00	0.79
	2			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3				0.07	0.02	0.18	0.11	0.11	0.23	0.23	0.12	0.12	0.12	0.07	0.00	0.48	0.45	0.18	0.55	0.39	0.39	0.79	0.79	0.12	0.12	0.19
	4					0.48	0.62	1.00	1.00	0.68	0.68	0.63	0.63	0.63	0.68	0.00	0.02	0.48	0.62	0.68	1.00	1.00	0.50	0.50	0.63	0.63	0.81
	5						0.48	1.00	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.00	0.01	0.13	0.48	0.13	0.37	0.37	0.13	0.13	1.00	1.00	0.80
	6							1.00	1.00	0.48	0.48	0.58	0.58	0.58	0.68	0.00	0.07	0.68	0.62	0.48	1.00	1.00	0.37	0.37	0.58	0.58	0.79
	7								N/A	1.00	1.00	0.77	0.77	0.77	1.00	0.00	0.04	0.45	1.00	0.25	0.48	0.48	0.13	0.13	0.77	0.77	1.00
	8									1.00	1.00	0.77	0.77	0.77	1.00	0.00	0.04	0.45	1.00	0.25	0.48	0.48	0.13	0.13	0.77	0.77	1.00
	9										N/A	0.58	0.58	0.58	0.68	0.00	0.10	0.72	0.68	0.62	1.00	1.00	0.25	0.25	0.58	0.58	0.77
	10											0.58	0.58	0.58	0.68	0.00	0.10	0.72	0.68	0.62	1.00	1.00	0.25	0.25	0.58	0.58	0.77
	11												N/A	N/A	1.00	0.00	0.06	0.36	0.63	0.30	0.42	0.42	0.21	0.21	N/A	N/A	1.00
	12													N/A	1.00	0.00	0.06	0.36	0.63	0.30	0.42	0.42	0.21	0.21	N/A	N/A	1.00
	13														1.00	0.00	0.06	0.36	0.63	0.30	0.42	0.42	0.21	0.21	N/A	N/A	1.00
3	1															0.00	0.01	0.13	0.48	0.13	0.25	0.25	0.07	0.07	1.00	1.00	0.80
	2																0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3																	0.07	0.02	0.18	0.11	0.11	0.39	0.39	0.06	0.06	0.10
	4																		0.48	0.62	1.00	1.00	1.00	1.00	0.36	0.36	0.50
	5																			0.48	1.00	1.00	0.37	0.37	0.63	0.63	0.81
	6																				1.00	1.00	1.00	1.00	0.30	0.30	0.45
	7																					N/A	0.48	0.48	0.42	0.42	0.61
	8																						0.48	0.48	0.42	0.42	0.61
	9																							N/A	0.21	0.21	0.30
	10																								0.21	0.21	0.30
	11																									N/A	1.00
	12																										1.00
	13																										

	k	4													5												
k	n	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
2	1	0.13	0.00	0.00	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.79	0.79	0.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.79	0.61
	2	0.00	N/A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.13	0.36	0.07	0.03	0.05	0.05	0.05	0.05	0.03	0.00	0.00	0.00
	3	0.27	0.00	0.25	0.72	0.77	0.79	1.00	1.00	1.00	1.00	0.19	0.19	0.29	0.48	0.00	0.00	0.21	0.48	0.36	0.36	0.36	0.36	0.50	0.19	0.19	0.29
	4	0.72	0.00	0.01	0.25	0.45	0.18	0.34	0.34	0.34	0.34	0.81	0.81	1.00	0.03	0.00	0.00	0.00	0.03	0.02	0.02	0.02	0.02	0.04	0.81	0.81	1.00
	5	0.68	0.00	0.00	0.07	0.07	0.02	0.08	0.08	0.08	0.08	0.80	0.80	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.80	0.80	1.00
	6	0.72	0.00	0.04	0.45	0.37	0.07	0.22	0.22	0.22	0.22	0.79	0.79	1.00	0.02	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.02	0.79	0.79	1.00
	7	1.00	0.00	0.03	0.29	0.22	0.04	0.07	0.07	0.07	0.07	1.00	1.00	0.79	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.79
	8	1.00	0.00	0.03	0.29	0.22	0.04	0.07	0.07	0.07	0.07	1.00	1.00	0.79	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.79
	9	0.72	0.00	0.06	0.50	0.37	0.07	0.13	0.13	0.13	0.13	0.77	0.77	1.00	0.02	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.01	0.77	0.77	1.00
	10	0.72	0.00	0.06	0.50	0.37	0.07	0.13	0.13	0.13	0.13	0.77	0.77	1.00	0.02	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.01	0.77	0.77	1.00
	11	0.63	0.00	0.04	0.26	0.24	0.10	0.15	0.15	0.15	0.15	1.00	1.00	0.48	0.02	0.00	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.02	1.00	1.00	0.48
	12	0.63	0.00	0.04	0.26	0.24	0.10	0.15	0.15	0.15	0.15	1.00	1.00	0.48	0.02	0.00	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.02	1.00	1.00	0.48
	13	0.63	0.00	0.04	0.26	0.24	0.10	0.15	0.15	0.15	0.15	1.00	1.00	0.48	0.02	0.00	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.02	1.00	1.00	0.48
3	1	0.48	0.00	0.00	0.07	0.07	0.02	0.04	0.04	0.04	0.04	0.80	0.80	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.80	0.80	1.00
	2	0.00	N/A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.13	0.36	0.07	0.03	0.05	0.05	0.05	0.05	0.03	0.00	0.00	0.00
	3	0.07	0.00	1.00	0.22	0.34	0.77	0.58	0.58	0.58	0.58	0.10	0.10	0.15	0.80	0.00	0.01	0.42	0.80	0.63	0.63	0.63	0.63	0.81	0.10	0.10	0.15
	4	0.68	0.00	0.04	1.00	1.00	0.45	0.72	0.72	0.72	0.72	0.50	0.50	0.66	0.07	0.00	0.00	0.01	0.07	0.04	0.04	0.04	0.04	0.10	0.50	0.50	0.66
	5	0.62	0.00	0.01	0.25	0.25	0.07	0.22	0.22	0.22	0.22	0.81	0.81	1.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.81	0.81	1.00
	6	0.68	0.00	0.11	1.00	1.00	0.25	0.62	0.62	0.62	0.62	0.45	0.45	0.63	0.05	0.00	0.00	0.02	0.05	0.01	0.01	0.01	0.01	0.05	0.45	0.45	0.63
	7	1.00	0.00	0.07	0.68	0.62	0.13	0.25	0.25	0.25	0.25	0.61	0.61	0.80	0.03	0.00	0.00	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.61	0.61	0.80
	8	1.00	0.00	0.07	0.68	0.62	0.13	0.25	0.25	0.25	0.25	0.61	0.61	0.80	0.03	0.00	0.00	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.61	0.61	0.80
	9	0.45	0.00	0.27	0.72	0.62	0.48	1.00	1.00	1.00	1.00	0.30	0.30	0.45	0.08	0.00	0.00	0.03	0.08	0.02	0.02	0.02	0.02	0.04	0.30	0.30	0.45
	10	0.45	0.00	0.27	0.72	0.62	0.48	1.00	1.00	1.00	1.00	0.30	0.30	0.45	0.08	0.00	0.00	0.03	0.08	0.02	0.02	0.02	0.02	0.04	0.30	0.30	0.45
	11	0.63	0.00	0.04	0.26	0.24	0.10	0.15	0.15	0.15	0.15	1.00	1.00	0.48	0.02	0.00	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.02	1.00	1.00	0.48
	12	0.63	0.00	0.04	0.26	0.24	0.10	0.15	0.15	0.15	0.15	1.00	1.00	0.48	0.02	0.00	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.02	1.00	1.00	0.48
	13	0.81	0.00	0.07	0.38	0.33	0.15	0.21	0.21	0.21	0.21	N/A	N/A	1.00	0.04	0.00	0.00	0.02	0.04	0.02	0.02	0.02	0.02	0.03	N/A	N/A	1.00

k	n	k = 4													k = 5												
k	n	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
4	1		0.00	0.03	0.37	0.37	0.13	0.22	0.22	0.22	0.22	0.81	0.81	1.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.81	0.81	1.00
	2			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.13	0.36	0.07	0.03	0.05	0.05	0.05	0.05	0.03	0.00	0.00	0.00
	3				0.07	0.18	0.55	0.39	0.39	0.39	0.39	0.07	0.07	0.11	1.00	0.00	0.01	0.58	1.00	0.80	0.80	0.80	0.80	1.00	0.07	0.07	0.11
	4					0.62	0.68	1.00	1.00	1.00	1.00	0.38	0.38	0.52	0.11	0.00	0.00	0.01	0.11	0.07	0.07	0.07	0.07	0.15	0.38	0.38	0.52
	5						0.48	1.00	1.00	1.00	1.00	0.33	0.33	0.48	0.04	0.00	0.00	0.01	0.04	0.02	0.02	0.02	0.02	0.08	0.33	0.33	0.48
	6							1.00	1.00	1.00	1.00	0.15	0.15	0.24	0.22	0.00	0.01	0.08	0.22	0.07	0.07	0.07	0.07	0.22	0.15	0.15	0.24
	7								N/A	N/A	N/A	0.21	0.21	0.33	0.13	0.00	0.00	0.05	0.13	0.04	0.04	0.04	0.04	0.07	0.21	0.21	0.33
	8									N/A	N/A	0.21	0.21	0.33	0.13	0.00	0.00	0.05	0.13	0.04	0.04	0.04	0.04	0.07	0.21	0.21	0.33
	9										N/A	0.21	0.21	0.33	0.13	0.00	0.00	0.05	0.13	0.04	0.04	0.04	0.04	0.07	0.21	0.21	0.33
	10											0.21	0.21	0.33	0.13	0.00	0.00	0.05	0.13	0.04	0.04	0.04	0.04	0.07	0.21	0.21	0.33
	11												N/A	1.00	0.04	0.00	0.00	0.02	0.04	0.02	0.02	0.02	0.02	0.03	N/A	N/A	1.00
	12													1.00	0.04	0.00	0.00	0.02	0.04	0.02	0.02	0.02	0.02	0.03	N/A	N/A	1.00
	13														0.07	0.00	0.00	0.03	0.07	0.04	0.04	0.04	0.04	0.06	1.00	1.00	N/A
5	1															0.00	0.02	0.48	N/A	1.00	1.00	1.00	1.00	0.48	0.04	0.04	0.07
	2																0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3																	0.07	0.02	0.08	0.08	0.08	0.08	0.05	0.00	0.00	0.00
	4																		0.48	1.00	1.00	1.00	1.00	0.62	0.02	0.02	0.03
	5																			1.00	1.00	1.00	1.00	0.48	0.04	0.04	0.07
	6																				N/A	N/A	N/A	1.00	0.02	0.02	0.04
	7																					N/A	N/A	1.00	0.02	0.02	0.04
	8																						N/A	1.00	0.02	0.02	0.04
	9																							1.00	0.02	0.02	0.04
	10																								0.03	0.03	0.06
	11																									N/A	1.00
	12																										1.00
	13																										

APPENDIX G: McNemar Test Statistic Results for 6001

k	n	2													3												
		1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
2	1		0.00	0.77	1.00	0.66	1.00	1.00	0.88	1.00	1.00	1.00	0.77	0.89	N/A	0.00	0.65	0.64	1.00	0.88	1.00	1.00	0.88	0.88	0.88	0.88	0.55
	2			0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.02	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.02
	3				0.72	1.00	0.45	0.45	0.61	0.45	0.45	0.45	0.80	0.38	0.77	0.00	0.02	0.07	0.42	0.30	0.45	0.45	0.33	0.33	0.33	0.61	0.10
	4					0.25	0.72	0.72	1.00	0.75	0.75	0.75	0.77	0.63	1.00	0.00	0.23	0.07	0.72	0.50	0.75	0.75	0.55	0.55	0.55	1.00	0.21
	5						0.07	0.07	0.22	0.13	0.13	0.13	1.00	0.18	0.66	0.00	0.06	0.01	0.07	0.04	0.13	0.13	0.08	0.08	0.08	0.34	0.03
	6							N/A	1.00	0.48	0.48	0.48	0.13	1.00	1.00	0.00	0.58	0.45	0.62	1.00	0.48	0.48	1.00	1.00	1.00	1.00	0.29
	7								1.00	0.48	0.48	0.48	0.13	1.00	1.00	0.00	0.58	0.45	0.62	1.00	0.48	0.48	1.00	1.00	1.00	1.00	0.29
	8									1.00	1.00	1.00	0.25	0.72	0.88	0.00	0.42	0.29	1.00	0.62	1.00	1.00	0.48	0.48	0.48	0.62	0.13
	9										N/A	N/A	0.13	1.00	1.00	0.00	0.61	0.50	0.68	1.00	0.48	0.48	1.00	1.00	1.00	1.00	0.29
	10											N/A	0.13	1.00	1.00	0.00	0.61	0.50	0.68	1.00	0.48	0.48	1.00	1.00	1.00	1.00	0.29
	11												0.13	1.00	1.00	0.00	0.61	0.50	0.68	1.00	0.48	0.48	1.00	1.00	1.00	1.00	0.29
	12													0.07	0.77	0.00	0.12	0.07	0.29	0.13	0.13	0.13	0.07	0.07	0.07	0.25	0.01
	13														0.89	0.00	0.81	0.79	1.00	0.75	1.00	1.00	0.72	0.72	0.72	0.68	0.25
3	1															0.00	0.65	0.64	1.00	0.88	1.00	1.00	0.88	0.88	0.88	0.88	0.55
	2																0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3																	0.68	0.50	0.75	0.58	0.58	0.79	0.79	0.79	0.39	1.00
	4																		0.25	0.62	0.45	0.45	0.72	0.72	0.72	0.29	1.00
	5																			1.00	0.62	0.62	1.00	1.00	1.00	1.00	0.29
	6																				1.00	1.00	0.62	0.62	0.62	0.62	0.45
	7																					N/A	1.00	1.00	1.00	1.00	0.22
	8																						1.00	1.00	1.00	1.00	0.22
	9																							N/A	N/A	0.62	0.45
	10																								N/A	0.62	0.45
	11																									0.62	0.45
	12																										0.07
	13																										

	k	4													5												
k	n	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
2	1	N/A	0.00	0.65	0.76	1.00	0.88	1.00	1.00	0.88	0.88	0.88	1.00	0.45	N/A	0.00	0.54	0.76	0.88	0.76	0.88	0.88	0.76	0.76	0.76	0.88	0.45
	2	0.00	0.02	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.03	0.00	0.02	0.02	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03
	3	0.77	0.00	0.02	0.11	0.77	0.30	0.42	0.42	0.30	0.30	0.30	0.42	0.07	0.77	0.00	0.01	0.11	0.58	0.21	0.30	0.30	0.21	0.21	0.21	0.30	0.07
	4	1.00	0.00	0.23	0.22	0.75	0.55	0.77	0.77	0.58	0.58	0.58	0.77	0.15	1.00	0.00	0.15	0.29	1.00	0.39	0.58	0.58	0.42	0.42	0.42	0.58	0.15
	5	0.66	0.00	0.06	0.05	0.45	0.08	0.18	0.18	0.11	0.11	0.11	0.18	0.02	0.66	0.00	0.04	0.07	0.29	0.05	0.11	0.11	0.07	0.07	0.07	0.11	0.02
	6	1.00	0.00	0.58	0.72	0.68	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.18	1.00	0.00	0.42	0.75	1.00	0.62	1.00	1.00	0.68	0.68	0.68	1.00	0.18
	7	1.00	0.00	0.58	0.72	0.68	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.18	1.00	0.00	0.42	0.75	1.00	0.62	1.00	1.00	0.68	0.68	0.68	1.00	0.18
	8	0.88	0.00	0.42	0.50	1.00	0.62	1.00	1.00	0.62	0.62	0.62	1.00	0.08	0.88	0.00	0.30	0.55	0.72	0.37	0.62	0.62	0.37	0.37	0.37	0.62	0.08
	9	1.00	0.00	0.61	0.75	0.72	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.18	1.00	0.00	0.45	0.77	1.00	0.68	1.00	1.00	0.62	0.62	0.62	1.00	0.18
	10	1.00	0.00	0.61	0.75	0.72	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.18	1.00	0.00	0.45	0.77	1.00	0.68	1.00	1.00	0.62	0.62	0.62	1.00	0.18
	11	1.00	0.00	0.61	0.75	0.72	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.18	1.00	0.00	0.45	0.77	1.00	0.68	1.00	1.00	0.62	0.62	0.62	1.00	0.18
	12	0.77	0.00	0.12	0.11	0.72	0.07	0.13	0.13	0.07	0.07	0.07	0.13	0.01	0.77	0.00	0.08	0.15	0.50	0.04	0.07	0.07	0.04	0.04	0.04	0.07	0.01
	13	0.89	0.00	0.81	1.00	0.55	0.72	1.00	1.00	0.72	0.72	0.72	1.00	0.13	0.89	0.00	0.65	1.00	0.77	1.00	0.72	0.72	1.00	1.00	1.00	0.72	0.13
3	1	N/A	0.00	0.65	0.76	1.00	0.88	1.00	1.00	0.88	0.88	0.88	1.00	0.45	N/A	0.00	0.54	0.76	0.88	0.76	0.88	0.88	0.76	0.76	0.76	0.88	0.45
	2	0.00	N/A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	0.65	0.00	N/A	1.00	0.13	0.75	0.55	0.55	0.77	0.77	0.77	0.55	0.80	0.65	0.00	1.00	1.00	0.29	1.00	0.77	0.77	1.00	1.00	1.00	0.77	0.80
	4	0.64	0.00	0.68	1.00	0.07	0.68	0.45	0.45	0.72	0.72	0.72	0.45	0.77	0.64	0.00	1.00	1.00	0.22	1.00	0.72	0.72	1.00	1.00	1.00	0.72	0.77
	5	1.00	0.00	0.50	0.62	0.48	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.18	1.00	0.00	0.34	0.68	1.00	0.62	1.00	1.00	0.68	0.68	0.68	1.00	0.18
	6	0.88	0.00	0.75	1.00	0.25	0.48	1.00	1.00	0.62	0.62	0.62	1.00	0.29	0.88	0.00	0.55	1.00	0.62	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.29
	7	1.00	0.00	0.58	0.72	0.68	1.00	0.48	0.48	1.00	1.00	1.00	0.48	0.13	1.00	0.00	0.42	0.75	1.00	0.62	1.00	1.00	0.62	0.62	0.62	1.00	0.13
	8	1.00	0.00	0.58	0.72	0.68	1.00	0.48	0.48	1.00	1.00	1.00	0.48	0.13	1.00	0.00	0.42	0.75	1.00	0.62	1.00	1.00	0.62	0.62	0.62	1.00	0.13
	9	0.88	0.00	0.79	1.00	0.45	0.62	1.00	1.00	0.48	0.48	0.48	1.00	0.29	0.88	0.00	0.61	1.00	0.72	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.29
	10	0.88	0.00	0.79	1.00	0.45	0.62	1.00	1.00	0.48	0.48	0.48	1.00	0.29	0.88	0.00	0.61	1.00	0.72	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.29
	11	0.88	0.00	0.79	1.00	0.45	0.62	1.00	1.00	0.48	0.48	0.48	1.00	0.29	0.88	0.00	0.61	1.00	0.72	1.00	0.62	0.62	1.00	1.00	1.00	0.62	0.29
	12	0.88	0.00	0.39	0.45	1.00	0.48	1.00	1.00	0.48	0.48	0.48	1.00	0.04	0.88	0.00	0.27	0.50	0.68	0.25	0.48	0.48	0.25	0.25	0.25	0.48	0.04
	13	0.55	0.00	1.00	0.75	0.08	0.37	0.13	0.13	0.37	0.37	0.37	0.13	1.00	0.55	0.00	0.80	0.77	0.18	0.68	0.37	0.37	0.68	0.68	0.68	0.37	1.00

	k	4													5												
k	n	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
4	1		0.00	0.65	0.76	1.00	0.88	1.00	1.00	0.88	0.88	0.88	1.00	0.45	N/A	0.00	0.54	0.76	0.88	0.76	0.88	0.88	0.76	0.76	0.76	0.88	0.45
	2			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3				1.00	0.13	0.75	0.55	0.55	0.77	0.77	0.77	0.55	0.80	0.65	0.00	1.00	1.00	0.29	1.00	0.77	0.77	1.00	1.00	1.00	0.77	0.80
	4					0.13	1.00	0.68	0.68	1.00	1.00	1.00	0.68	0.55	0.76	0.00	0.68	0.48	0.37	0.68	1.00	1.00	0.72	0.72	0.72	1.00	0.55
	5						0.25	0.62	0.62	0.37	0.37	0.37	0.62	0.05	1.00	0.00	0.08	0.13	1.00	0.13	0.37	0.37	0.22	0.22	0.22	0.37	0.05
	6							1.00	1.00	0.48	0.48	0.48	1.00	0.22	0.88	0.00	0.55	1.00	0.62	1.00	0.48	0.48	1.00	1.00	1.00	0.48	0.22
	7								N/A	1.00	1.00	1.00	N/A	0.07	1.00	0.00	0.39	0.72	1.00	0.48	1.00	1.00	0.48	0.48	0.48	1.00	0.07
	8									1.00	1.00	1.00	N/A	0.07	1.00	0.00	0.39	0.72	1.00	0.48	1.00	1.00	0.48	0.48	0.48	1.00	0.07
	9										N/A	N/A	1.00	0.22	0.88	0.00	0.58	1.00	0.68	1.00	0.48	0.48	1.00	1.00	1.00	0.48	0.22
	10											N/A	1.00	0.22	0.88	0.00	0.58	1.00	0.68	1.00	0.48	0.48	1.00	1.00	1.00	0.48	0.22
	11												1.00	0.22	0.88	0.00	0.58	1.00	0.68	1.00	0.48	0.48	1.00	1.00	1.00	0.48	0.22
	12													0.07	1.00	0.00	0.39	0.72	1.00	0.48	1.00	1.00	0.48	0.48	0.48	1.00	0.07
	13														0.45	0.00	1.00	0.55	0.08	0.37	0.13	0.13	0.37	0.37	0.37	0.13	N/A
5	1															0.00	0.54	0.76	0.88	0.76	0.88	0.88	0.76	0.76	0.76	0.88	0.45
	2																0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3																	0.62	0.13	0.75	0.55	0.55	0.77	0.77	0.77	0.55	1.00
	4																		0.25	0.68	1.00	1.00	0.72	0.72	0.72	1.00	0.55
	5																			0.25	0.62	0.62	0.37	0.37	0.37	0.62	0.08
	6																				1.00	1.00	0.48	0.48	0.48	1.00	0.37
	7																					N/A	1.00	1.00	1.00	N/A	0.13
	8																						1.00	1.00	1.00	N/A	0.13
	9																							N/A	N/A	1.00	0.37
	10																								N/A	1.00	0.37
	11																									1.00	0.37
	12																										0.13
	13																										

APPENDIX H: McNemar Test Statistic Results for 8302

k	n	k = 2													k = 3												
		1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
2	1		0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2			1.00	0.13	0.75	1.00	0.79	0.79	1.00	1.00	1.00	0.00	0.00	0.01	N/A	1.00	0.02	0.30	0.81	0.48	0.48	0.36	0.36	0.36	0.00	0.00
	3				0.04	0.50	0.77	1.00	1.00	0.79	0.79	0.79	0.00	0.00	0.01	1.00	N/A	0.00	0.18	0.63	0.33	0.33	0.24	0.24	0.24	0.00	0.00
	4					0.25	0.04	0.13	0.13	0.29	0.29	0.29	0.00	0.00	0.00	0.13	0.04	0.13	0.72	0.55	1.00	1.00	0.77	0.77	0.77	0.00	0.00
	5						0.25	0.62	0.62	1.00	1.00	1.00	0.00	0.00	0.01	0.75	0.50	0.02	0.37	0.72	0.72	0.72	0.50	0.50	0.50	0.00	0.00
	6							1.00	1.00	0.48	0.48	0.48	0.00	0.00	0.06	1.00	0.77	0.00	0.08	0.37	0.07	0.07	0.04	0.04	0.04	0.00	0.00
	7								N/A	1.00	1.00	1.00	0.00	0.00	0.04	0.79	1.00	0.02	0.18	0.68	0.13	0.13	0.07	0.07	0.07	0.00	0.00
	8									1.00	1.00	1.00	0.00	0.00	0.04	0.79	1.00	0.02	0.18	0.68	0.13	0.13	0.07	0.07	0.07	0.00	0.00
	9										N/A	N/A	0.00	0.00	0.02	1.00	0.79	0.04	0.34	1.00	0.37	0.37	0.13	0.13	0.13	0.00	0.00
	10											N/A	0.00	0.00	0.02	1.00	0.79	0.04	0.34	1.00	0.37	0.37	0.13	0.13	0.13	0.00	0.00
	11												0.00	0.00	0.02	1.00	0.79	0.04	0.34	1.00	0.37	0.37	0.13	0.13	0.13	0.00	0.00
	12													0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.25
	13														0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
3	1															0.01	0.01	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00
	2																1.00	0.02	0.30	0.81	0.48	0.48	0.36	0.36	0.36	0.00	0.00
	3																	0.00	0.18	0.63	0.33	0.33	0.24	0.24	0.24	0.00	0.00
	4																		0.13	0.02	0.13	0.13	0.29	0.29	0.29	0.00	0.00
	5																			0.25	1.00	1.00	0.68	0.68	0.68	0.00	0.00
	6																				0.48	0.48	0.25	0.25	0.25	0.00	0.00
	7																					N/A	1.00	1.00	1.00	0.00	0.00
	8																						1.00	1.00	1.00	0.00	0.00
	9																							N/A	N/A	0.00	0.00
	10																								N/A	0.00	0.00
	11																									0.00	0.00
	12																										0.48
	13																										

	k	4													5												
k	n	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
2	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.27	0.07	0.03	0.00	0.02	0.05	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	0.21	0.11	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.79	0.75	0.45	0.02	0.18	0.34	0.05	0.05	0.01	0.01	0.01	0.00	0.00	0.58	0.34	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.63	0.58	0.11	0.00	0.03	0.07	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.21	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.26	0.21	0.03	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.38	0.33	0.06	0.00	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.38	0.33	0.06	0.00	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	9	0.52	0.48	0.12	0.01	0.04	0.06	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.19	0.10	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	10	0.52	0.48	0.12	0.01	0.04	0.06	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.19	0.10	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	11	0.52	0.48	0.12	0.01	0.04	0.06	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.19	0.10	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.04
	13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.68	0.13
3	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.27	0.07	0.03	0.00	0.02	0.05	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	0.21	0.11	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.34	0.29	1.00	0.25	1.00	0.68	0.37	0.37	0.08	0.08	0.08	0.00	0.00	1.00	0.68	0.13	0.02	0.02	0.08	0.04	0.02	0.00	0.00	0.00	0.00	0.00
	5	0.79	0.77	0.45	0.02	0.07	0.22	0.05	0.05	0.01	0.01	0.01	0.00	0.00	0.58	0.34	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.63	0.61	0.11	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	1.00	1.00	0.34	0.03	0.11	0.18	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.45	0.27	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	1.00	1.00	0.34	0.03	0.11	0.18	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.45	0.27	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	9	0.81	0.80	0.55	0.07	0.23	0.34	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.63	0.42	0.04	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	10	0.81	0.80	0.55	0.07	0.23	0.34	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.63	0.42	0.04	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	11	0.81	0.80	0.55	0.07	0.23	0.34	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.63	0.42	0.04	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.07
	13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.62	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.25

	k	4													5												
k	n	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
4	1		0.72	0.50	0.07	0.27	0.42	0.10	0.10	0.02	0.02	0.02	0.00	0.00	0.25	0.22	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	2			0.37	0.05	0.23	0.39	0.07	0.07	0.02	0.02	0.02	0.00	0.00	0.50	0.13	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	3				0.13	0.68	1.00	0.22	0.22	0.05	0.05	0.05	0.00	0.00	0.72	1.00	0.07	0.01	0.01	0.05	0.03	0.01	0.00	0.00	0.00	0.00	0.00
	4					0.48	0.25	0.48	0.48	0.37	0.37	0.37	0.00	0.00	0.34	0.45	1.00	0.13	0.13	0.37	0.18	0.08	0.02	0.01	0.01	0.00	0.00
	5						1.00	0.62	0.62	0.13	0.13	0.13	0.00	0.00	0.77	1.00	0.37	0.04	0.04	0.07	0.05	0.01	0.01	0.00	0.00	0.00	0.00
	6							0.25	0.25	0.04	0.04	0.04	0.00	0.00	1.00	0.75	0.22	0.02	0.02	0.04	0.01	0.01	0.00	0.00	0.00	0.00	0.00
	7								N/A	0.25	0.25	0.25	0.00	0.00	0.39	0.50	1.00	0.22	0.22	0.45	0.13	0.08	0.01	0.00	0.00	0.00	0.00
	8									0.25	0.25	0.25	0.00	0.00	0.39	0.50	1.00	0.22	0.22	0.45	0.13	0.08	0.01	0.00	0.00	0.00	0.00
	9										N/A	N/A	0.00	0.00	0.12	0.15	0.72	1.00	1.00	0.75	0.68	0.45	0.04	0.02	0.02	0.00	0.00
	10											N/A	0.00	0.00	0.12	0.15	0.72	1.00	1.00	0.75	0.68	0.45	0.04	0.02	0.02	0.00	0.00
	11												0.00	0.00	0.12	0.15	0.72	1.00	1.00	0.75	0.68	0.45	0.04	0.02	0.02	0.00	0.00
	12													0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.25
	13														0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
5	1															1.00	0.13	0.03	0.03	0.07	0.04	0.02	0.00	0.00	0.00	0.00	0.00
	2																0.13	0.02	0.02	0.08	0.04	0.02	0.00	0.00	0.00	0.00	0.00
	3																	0.25	0.25	0.62	0.29	0.13	0.03	0.02	0.02	0.00	0.00
	4																		N/A	1.00	1.00	0.62	0.13	0.08	0.08	0.00	0.00
	5																			1.00	1.00	0.62	0.13	0.08	0.08	0.00	0.00
	6																				0.62	0.25	0.08	0.05	0.05	0.00	0.00
	7																					1.00	0.13	0.07	0.07	0.00	0.00
	8																						0.37	0.22	0.22	0.00	0.00
	9																							1.00	1.00	0.00	0.00
	10																								N/A	0.00	0.00
	11																									0.00	0.00
	12																										0.48
	13																										

APPENDIX İ: Declaration Numbers by Control Line and Detection Status

9105			
	DETECTED	REMAINING	TOTAL
RED LINE	95	3986	4081
YELLOW LINE	8	994	1002
TOTAL	103	4980	5083

6001			
	DETECTED	REMAINING	TOTAL
RED LINE	616	6127	6743
YELLOW LINE	113	4500	4613
TOTAL	729	10627	11356

8302			
	DETECTED	REMAINING	TOTAL
RED LINE	208	77282	77490
YELLOW LINE	91	52568	52659
TOTAL	299	129850	130149

9608			
	DETECTED	REMAINING	TOTAL
RED LINE	130	321812	321942
YELLOW LINE	32	222328	222360
TOTAL	162	544140	544302

9105 - Excluding Declarations Detected as Other Illegalities			
	DETECTED	REMAINING	TOTAL
RED LINE	78	3986	4064
YELLOW LINE	3	994	997
TOTAL	81	4980	5061

6001 - Excluding Declarations Detected as Other Illegalities			
	DETECTED	REMAINING	TOTAL
RED LINE	57	6127	6184
YELLOW LINE	15	4500	4515
TOTAL	72	10627	10699

8302 - Excluding Declarations Detected as Other Illegalities			
	DETECTED	REMAINING	TOTAL
RED LINE	59	77282	77341
YELLOW LINE	13	52568	52581
TOTAL	72	129850	129922

9608 - Excluding Declarations Detected as Other Illegalities			
	DETECTED	REMAINING	TOTAL
RED LINE	74	321812	321886
YELLOW LINE	8	222328	222336
TOTAL	82	544140	544222

METU LIBRARY

TEZ FOTOKOPİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü ☐

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☐

Deniz Bilimleri Enstitüsü ☐

YAZARIN

Soyadı : Baştabak.....
Adı : Burcu.....
Bölümü : Bilişim Sistemleri.....

TEZİN ADI (İngilizce) : A Data Mining Framework to Detect Tariff Code Circumvention in Turkish Customs Database.....
.....
.....
.....
.....

TEZİN TÜRÜ : Yüksek Lisans ☐ Doktora ☐

1. Tezimin tamamı dünya çapında erişime açılsın ve kaynak gösterilmek şartıyla tezimin bir kısmı veya tamamının fotokopisi alınsın. ☐
2. Tezimin tamamı yalnızca Orta Doğu Teknik Üniversitesi kullanıcılarının erişimine açılsın. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☐
3. Tezim bir (1) yıl süreyle erişime kapalı olsun. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☐

Yazarın imzası ..... Tarih .....