© 2015 IJSRSET | Volume 1 | Issue 6 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

# Comparative Analysis of Industrial Mishaps Based on Classified Prediction

**Nahida Parvin*, Ayesha Aziz Prova, Mehnaz Tabassum**
Department of Computer Science and Engineering, Jagannath University, Dhaka, Bangladesh

## ABSTRACT

Industrial accident analysis is a very challenging task and one of most vital issues in the era of globalization. Discovering the attributes becomes more complex because voluminous factors are associated. We have tried to find the specific attributes and made a cumulative dataset depending on the reliable sources allied to Bangladesh. In the study, we have evaluated a meticulous survey on various classification techniques to achieve casualty for textile & garments accidents. We have presented a comparative analysis of accuracy between base and AdaBoost Meta classifier using base classifiers, such as: OneR, J48, REPTree, SimpleCART & Naïve Bayes. The analysis unfurl that using ensemble method with the base classifiers improve accuracy level between 1.8%-6.36%. Depending on the knowledge explored by classification technique which will have the ability to make automated decision that is quite similar to human decision making for reducing the rate of casualty of industrial mishaps.

**Keywords:** Data Mining, Classification Algorithms, Meta Classifier, Textile & Garments Accident Data

## I. INTRODUCTION

Data mining is an extracting process of finding interesting hidden knowledge by analyzing voluminous data of different information repositories. In real life the knowledge of data mining has already used for business intelligent, search engines, financial data analysis, network intrusion detection, fraud detection, bioinformatics, retail & telecommunication industry, health informatics etc. Several authors have been proposed classification methods for constructing the intelligent system with the automated decision making capabilities [2]. It is a complex but effective for different issues like industrial accidents.

Textile & garments industry is the leading sector of manufacturing clothing & occupies a unique position in the economy of Bangladesh. It has emerged as the most important economic sector and accounts for employing one half of the total industrial workforce and contributing about a quarter of the gross value-addition in the manufacturing sector [3]. This sector has been constantly growing which represents about 78% [4] of export earnings of our country and the living of more than five million people directly depend on five thousand factories [5]. The lower labor cost [5] and other facility attract the foreign buyers. About 60% and 20% of the total production of textile and garments industry in Bangladesh respectively export in the European and US markets every year [5]. Along with bringing huge economic progress for our country this sector has experienced some worst industrial accidents. Unfortunately, there is no comprehensive statistics on the current status of accidents. For analyzing social issues like accidents, data mining techniques become more popular [6] [7] as it can discover hidden information that can be used to make intelligent system. Among different techniques of data mining, classification is one by which the prediction of undiscovered information is conceivable. According to the several studies, the causes of casualties during an accident in textile & garments industry of Bangladesh mainly derived by the factors like fire, collapse, false fire alarm, suffocation, stampede, panic, exit [9] [10] [11] [12]. We will take these factors for analysing the

probability of happening casualty as it is the class attribute for our study.

This paper will explore the hidden information through classification techniques to make automated decision that is quite similar to human decision.

## II. METHODS AND MATERIAL

### Literature Review

Data mining is an approach that evolves various techniques to perform tasks including database oriented techniques, statistic, machine learning, pattern recognition, neural network, rough set and others [2]. These techniques are used to extract hidden patterns from large amount of data from the data warehouse [13] that can be used to provide making intelligent decision. Numerous numbers of studies has been found for extracting the undiscovered knowledge of data using data mining classification techniques [6], [7]. For analyzing social issues like accidents data mining techniques become more popular & trustworthy [6]. Textile & garments accidents are one of the major concern issues in Bangladesh [10]. About 78% of the export earnings of Bangladesh come from the textile & garments industry [14] but the safety standard of this sector is one of the most horrible in the world [15]. Many literature analyses the accidental factors like fire, collapse, false fire alarm, suffocation, stampede, panic, exit which causes casualty [9] [10] [11] [12].

Decision tree based classification techniques are used [8] on various applications to introduce a model. Other techniques such as Naïve Bayes & rule based classifiers are also used in various filed for construction classification model [6]. Naïve Bayes is one of the simplest probabilistic classifiers which are based on Bayes theorem with strong naïve independence assumption. This assumption treated each and every word as a single, independent and mutually exclusive. This model can be described as "Independent Feature Model" (18).

A comparative analysis based on classification algorithms have been used by many researchers for prediction in different fields [8] [16] [17] [18]. In [8], J48 & Naïve Bayes are used to maximize the true positive rate for banking dataset. It reveals J48 as a good classifier in case of efficiency and accuracy [8]. Bagging, boosting & random forests are some significant ensemble method used to decrease incorrectly classified instances [1].

AdaBoost Meta classifier is a popular boosting algorithm among many scholars which is used to increase the accuracy of the base classifiers [1] [6]. In [6] the authors used AdaBoost as ensemble method to get high accurate result to find gender based patterns for road accidents. Their research revealed that RndTree using AdaBoost provide best performance in case of accuracy based on precision, recall & ROC curves.

In some cases researchers have incorporated cross validation evaluation metric technique for performance and accuracy measures as to overcome the problem [18] of classify testing data.

### Methodology

Figure 1 shows the entire step related in our study.

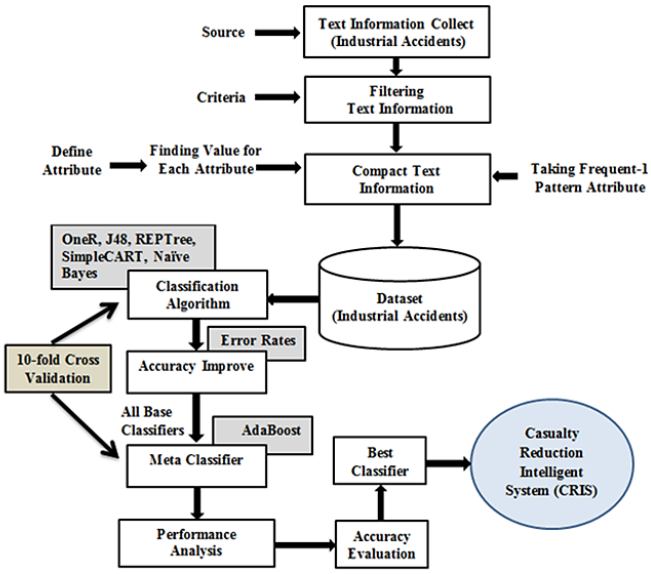


**Figure 1:** Steps Related in the Study

### A. Dataset Formation

The formation of the dataset has been formatted by the following steps:

#### A. Text Information Collect

We have collected text information from different sources of online. The details about the sources are given below.

### A.1. Source

The study has covered the area of garments & textile industry accidents in Bangladesh. Unfortunately there is no cumulative information about the accidents of this sector. To overcome this limitation we have collected text information about real life garments & textile industry accidents in Bangladesh from different reliable secondary sources e.g. research papers, online and daily newspapers, articles, magazines. The dataset of the study consists of textile and garments industry accident information of Bangladesh from 17th December 1990 to 11th March 2015. It consists of 110 records and 8 attributes.

Some of those accidents of garments & textile industry in Bangladesh that we have considered for our study listed in Table I.

Table I: List of Accidents of Garments & Textile Industry in Bangladesh

| *SL* | Date | Place |
|---|---|---|
| ***1*** | 11-03-2015 | Bilash Garments, Dhaka [26] [27] |
| ***2*** | 01-03-2015 | Next Collection, Savar [26] |
| ***3*** | 24-01-2015 | Kader Synthetic and Compact Spinning Mill, Gazipur [26], [28] |
| ***4*** | 01-12-2014 | Legos Apparels, Gazipur [26] |
| ***5*** | 23-11-2014 | Fuji Garments, Savar [26] |
| ***6*** | 07-11-2014 | Warming Wollen Mills, Dhaka [26], [29] |
| ***7*** | 05-10-2014 | Ishrak Spinning Mill, Gazipur [26] |
| ***8*** | 30-09-2014 | Precious Apparels, Chittagong [26] |
| ***9*** | 29-09-2014 | Sagar Garments, Chittagong [26],[30] |
| ***10*** | 28-09-2014 | Mega Yarn Dyeing Mills, Gazipur [26], [31] |
| ***11*** | 14-09-2014 | Northern Fabrics, Dhaka [26] |
| ***12*** | 30-08-2014 | Cordial Design, Dhaka [26], [32] |
| ***25*** | 10-08-2014 | Creasent Fashion of Beximco Industrial Park, Gazipur [33] |
| ***26*** | 10-07-2014 | Amina Exports Wear, Ashulia [26], [34], [35] |
| ***27*** | 10-07-2014 | Mayer Doha, Dhaka [26], [36] |
| ***28*** | 07-07-2014 | S S Sweater, Gazipur [26], [37] |
| ***13*** | 20-06-2014 | Medlar Apparel, Dhaka [40] |
| ***14*** | 20-06-2014 | ZA Sweater Factory, Ashulia [26], [39], [41] |
| ***15*** | 11-05-2014 | Fashion Park International, Dhaka [26] |
| ***16*** | 11-05-2014 | Karnaphuli Knitting, Siddique Knitting Fashion Park International Ltd, Chittagong [26] |
| ***17*** | 09-04-2014 | Rangdhanu Spinning Mills, Ashulia [42], [43] |
| ***18*** | 16-03-2014 | Oporajeo (jute bag factory), Savar[26], [44] |
| ***19*** | 06-03-2014 | Green Leaf Apparel, Dhaka [26], [45] |
| ***20*** | 12-02-2014 | Al-Lima Textile, Savar [42], [46], [47] |
| ***21*** | 13-01-2014 | Al Muslim Garment Factory, Dhaka [46] |
| ***22*** | 10-01-2014 | Swadhin Dyeing Factory, Gazipur [42] |
| ***23*** | 26-11-2013 | Bangla-Japan garment, Dhaka [42] |
| ***24*** | 25-11-2013 | Mondol Group, Dhaka [42], [49] |
| ***47*** | 25-11-2013 | Aman Spinning Mills, Dhaka [49] |
| ***48*** | 25-11-2013 | Riyad Dying, Gazipur[42] |
| ***49*** | 26-10-2013 | Reza Fashion [50] |
| ***50*** | 08-10-2013 | Aswad Composite Mills, Gazipur [42], [51] |

### B. Filtering Text Information

The collected information about textile & garments accident is in text form with multiple piece of news of accidents in a source. So it is significant to filter the collected text information which has done on some criteria basis.

### B.1. Criteria

We have carefully filtered our text information on criteria basis. For avoiding errors and redundancy we have carefully eliminates all unnecessary elements. To find out missing elements we have collected several information for the same incident from different sources in case of most of the available accidents and then alignments all elements with the specific incident.

### C. Compact Text Information

The following steps are involved to compact the text information.

- ✓ Define Attributes
- ✓ Finding Values for Each Attributes
- ✓ Taking Frequent-1 Pattern Attributes

## III. RESULTS AND DISCUSSION

### A. Experimental Results of Base Classifiers

In this phase we have applied OneR, J48, REPTree, SimpleCART& Naïve Bayes algorithm by using 10-fold cross validation model to classify the total dataset of textile & garments accident in Bangladesh under the class attribute "Casualty".

#### i. Cross Validation Results of OneR:

In the cross validation result, Figure 2-4, we have found that after generating one rule for each predictor (fire, collapse, false fire alarm, suffocation, stampede, panic, and exit) in the data of textile & garments industry accident, OneR selected the rule with the smallest total error which is occurred for the attribute "Stampede".

```
=== Classifier model (full training set) ===

Stampede:
        No      -> No
        Yes     -> Yes
(94/110 instances correct)
```

**Figure 2:** OneR Base Classifier's Rule Generation

| | | |
|---|---|---|
| Correctly Classified Instances | 94 | 85.4545 % |
| Incorrectly Classified Instances | 16 | 14.5455 % |
| Kappa statistic | 0.7074 | |
| Mean absolute error | 0.1455 | |
| Root mean squared error | 0.3814 | |
| Relative absolute error | 32.0762 % | |
| Root relative squared error | 80.1489 % | |
| Total Number of Instances | 110 | |

**Figure 3:** Different Statistical Values of OneR

```
=== Confusion Matrix ===

  a  b   <-- classified as
 38  0 |  a = No
 16 56 |  b = Yes
```

**Figure 4:** Confusion Matrix of OneR

#### ii. Cross Validation Results of Naïve Bayes:

Figure 5 shows different statistical values of the data of textile & garments industry's accidents.

| | | |
|---|---|---|
| Correctly Classified Instances | 101 | 91.8182 % |
| Incorrectly Classified Instances | 9 | 8.1818 % |
| Kappa statistic | 0.8287 | |
| Mean absolute error | 0.116 | |
| Root mean squared error | 0.2224 | |
| Relative absolute error | 25.5841 % | |
| Root relative squared error | 46.7279 % | |
| Total Number of Instances | 110 | |

**Figure 5:** Different Statistical Values of Naïve Bayes

```
=== Confusion Matrix ===

  a  b   <-- classified as
 38  0 |  a = No
  9 63 |  b = Yes
```

**Figure 6:** Confusion Matrix of Base Naïve Bayes

#### iii. Comparison among all the Base Classifier

The bar chart reflects the correctly classified & incorrectly classified instances of textile & garments industry accident for all the base classifiers which is given below in Figure 7.

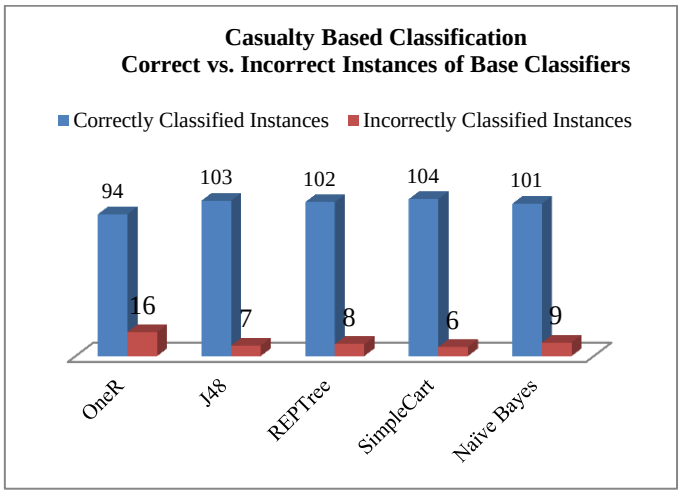


**Figure 7:** Correct vs. Incorrect Instances of Base Classifiers in Cross Validation Testing

### iv. Experimental Results of AdaBoost Classifier using Base Classifier

In our study we have applied ensemble methods with 10-fold cross validation model to improve the number of correctly classified instances for getting higher accuracy for the dataset of textile & garments industry accidents.

### v. Cross Validation Results of AdaBoost Base OneR

Figure 8 shows different statistical values of casualty dependent textile & garments industry's accident data.

| | | |
|---|---|---|
| Correctly Classified Instances | 101 | 91.8182 % |
| Incorrectly Classified Instances | 9 | 8.1818 % |
| Kappa statistic | 0.8245 | |
| Mean absolute error | 0.1327 | |
| Root mean squared error | 0.2459 | |
| Relative absolute error | 29.2571 % | |
| Root relative squared error | 51.6761 % | |
| Total Number of Instances | 110 | |

**Figure 8 :** Different Statistical Values of AdaBoost using OneR

```
=== Confusion Matrix ===

  a  b   <-- classified as
 36  2 |  a = No
  7 65 |  b = Yes
```

**Figure 9 :** Confusion Matrix of AdaBoost using OneR

### vi. Cross Validation Results AdaBoost BaseNaïve Bayes:

Figure 10 shows different statistical values of the data of textile & garments industry's accidents.

| | | |
|---|---|---|
| Correctly Classified Instances | 107 | 97.2727 % |
| Incorrectly Classified Instances | 3 | 2.7273 % |
| Kappa statistic | 0.9408 | |
| Mean absolute error | 0.0525 | |
| Root mean squared error | 0.1657 | |
| Relative absolute error | 11.5823 % | |
| Root relative squared error | 34.8285 % | |
| Total Number of Instances | 110 | |

**Figure 10 :** Different Statistical Values of AdaBoostusing Naïve Bayes

```
=== Confusion Matrix ===

  a  b   <-- classified as
 38  0 |  a = No
  3 69 |  b = Yes
```

**Figure 11:** Confusion Matrix for AdaBoost using Naïve Bayes

### vii. Comparison among all Meta Classifiers

The bar chart reflects the correctly classified & incorrectly classified instances of textile & garments industry accident for Meta classifier i.e., AdaBoost using base classifiers which is given below in Figure 12

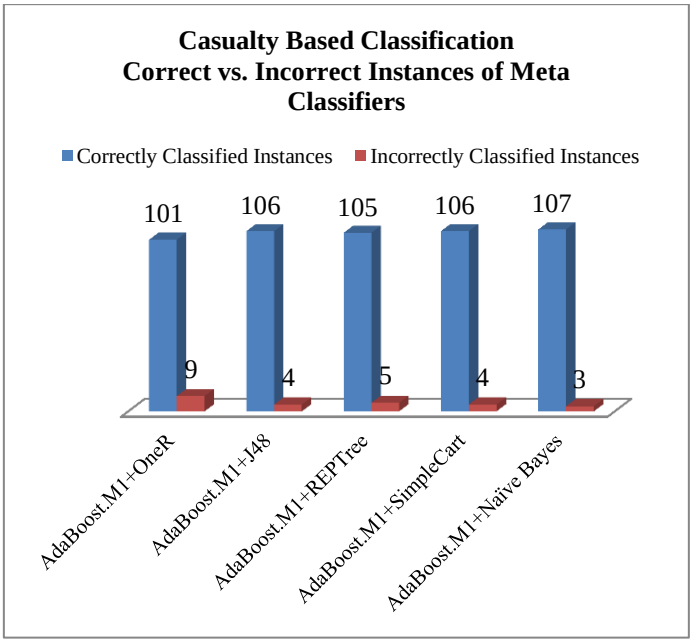


**Figure 12:** Correct vs. Incorrect Instances of Meta Classifiers in Cross Validation Testing

We have found from our study that there is better improvement in Naïve Bayes classifier than other classifiers. The most nearest classifier to Naïve Bayes are J48 and SimpleCART in case of predicting the casualty class from the data of textile & garments industry accidents in Bangladesh.

## B. Performance Analysis Based on Different Statistical Values

In our study the performance analysis is based on different values, found in cross validation testing of base & AdaBoost base classifiers for the data of textile & garments industry accidents. Different statistical value of all the base & AdaBoost base classifiers are given below in the Table IV.

Table IV: Performance Study of Base & AdaBoost Base Classifiers

| Algorithms Implemented | Correctly Classified Instances (%) | Incorrectly Classified Instances (%) | Kappa Statistics | Root Mean Squared Error | Relative Absolute Error (%) | Root Relative Squared Error (%) |
|---|---|---|---|---|---|---|
| OneR | 85.4545 | 14.5455 | 0.7074 | 0.3814 | 32.0762 | 80.1489 |
| AdaBoost (OneR) | 91.8182 | 8.1818 | 0.8245 | 0.2459 | 29.2571 | 51.6761 |
| J48 | 93.6364 | 6.3636 | 0.8651 | 0.2463 | 22.2714 | 51.7692 |
| AdaBoost (J48) | 96.3636 | 3.6364 | 0.9214 | 0.1808 | 15.385 | 38.0005 |
| REPTree | 92.7273 | 7.2727 | 0.8468 | 0.2735 | 28.6774 | 57.4837 |
| AdaBoost (REPTree) | 95.4545 | 4.5455 | 0.9025 | 0.2087 | 18.6359 | 43.8563 |
| SimpleCART | 94.5455 | 5.4545 | 0.8837 | 0.2166 | 18.6059 | 45.5093 |
| AdaBoost (SimpleCART) | 96.3636 | 3.6364 | 0.9215 | 0.1887 | 14.9595 | 39.6623 |
| Naïve Bayes | 91.8182 | 8.1818 | 0.8287 | 0.2224 | 25.5841 | 46.7279 |
| AdaBoost (Naïve Bayes) | 97.2727 | 2.7273 | 0.9408 | 0.1657 | 11.5823 | 34.8285 |

From above table it has been found that AdaBoost using Naïve Bayes shows better result than that of the base classifiers. It is more specific & more sensitive than all of the foresaid classifiers when we considered them for analyzing our accidental dataset of garments sector. The base classifiers are less sensitive & less specific than AdaBoost using base classifiers.

### 5.2 Evaluation of Experimental Results

In our study the error rate & accuracy of all the base & AdaBoost base classifiers are calculated from confusion matrix in case of casualty which is given below in Table VI.

**Table VI: Accuracy & Error Rate of Base & Meta Classifiers**

| Algorithms Implemented | Error Rate | Error Rate (%) | Accuracy | Accuracy (%) | Diff. Between AdaBoost Base &Base Classifier's Accuracy (%) |
|---|---|---|---|---|---|
| OneR | 0.1454 | 14.54 | 0.8545 | 85.45 | 6.36 |
| AdaBoost(OneR) | 0.0818 | 8.18 | 0.9181 | 91.81 | |
| J48 | 0.0636 | 6.36 | 0.9363 | 93.63 | 2.73 |
| AdaBoost (J48) | 0.0363 | 3.63 | 0.9636 | 96.36 | |
| REPTree | 0.0727 | 7.27 | 0.9272 | 92.72 | 2.73 |
| AdaBoost - (REPTree) | 0.0445 | 4.45 | 0.9545 | 95.45 | |
| SimpleCART | 0.0545 | 5.45 | 0.9454 | 94.54 | 1.82 |
| AdaBoost (SimpleCART) | 0.0363 | 3.63 | 0.9636 | 96.36 | |
| Naïve Bayes | 0.0818 | 8.18 | 0.9181 | 91.81 | 5.46 |
| AdaBoost (Naïve Bayes) | 0.0272 | 2.72 | 0.9727 | 97.27 | |

From above table it is seen that in case of class level of casualty the Naïve Bayes using AdaBoosthas the highest accuracy of all the foresaid classifiers. The accuracy is about 97.27%. But the improvement in accuracy of OneR using AdaBoost is higher than other classifiers as OneR is a weak classifier. After using AdaBoost, J48 &SimpleCART each in same accuracy level (96.36%).But the improvement in accuracy is not same. The accuracy improved in Naïve Bayes classifier is about 5.46%where its error rate is about 2.72% which is lowest from all of the foresaid classifiers. Figure 5 illustrates the comparison of accuracies of all the classifiers with & without AdaBoost in cross validation testing.

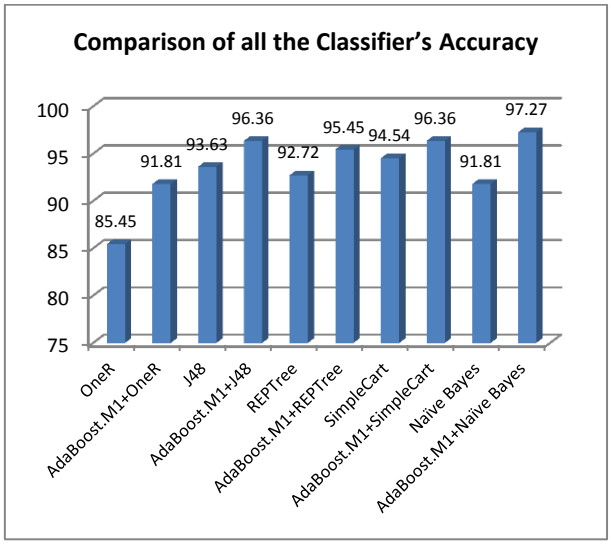


**Figure 13 :** Comparison of all the Classifier's Accuracy

## IV. CONCLUSION

In this paper, we have used five different data mining classification methods for the prediction of reasons of accidental data. We analyzed textile & garments accidents dataset using OneR, J48, REPTree, SimpleCART, Naïve Bayes and a combination of the AdaBoost algorithm with all of this base classifier to find patterns using casualty based classification. But for superior prediction, we focus on accuracy. Among the algorithms AdaBoost using Naïve Bayes gives high accuracy. The accuracy is evaluated based on precision, recall which is calculated from confusion matrix. The result shows that the AdaBoost using Naïve Bayes improved accuracy from 91.81% to 97.27%.

## V. FUTURE WORKS

In future we will find other reasons for textile & garments accidents & compare results with other supervised & machine learning method. In the future we can propose a casualty reducing intelligent system, which can be used for disaster management.

## VI. REFERENCES


[1] Han, J., Kamber, M., Pei, J. (2012), "Data Mining: Concepts & Techniques", third edition, Morgan Kaufmann Publishers an imprint of Elsevier, ISBN: 978-0-12-381479-1.

[2] Jantan, H., Hamdan, A. R., Othman, Z. A. (2009). "Classification for Talent Management Using Decision Tree Induction Techniques." *2nd Conference on Data Mining and Optimization*. 27-28 October 2009. Selangor, Malaysia.

[3] Hoque, M. S., Debnath, A. K., Mahmud, S. M. S. ( ). "Road Safety of Garment Industry Workers in Dhaka City". Unpublished

[4] Claeson, B. (2012). "Deadly Secrets" (Online). Available from: http:// laborrights.org/sites/default/files/publications-and resources/ DeadlySecrets.pdf. (Access on April 1, 2015).

[5] TUDelft (2015). "Working Conditions in the Bangladeshi Garment Sector: Social Dialogue and Compliance" (Online). Available from: http://www.fairwear.org/ul/cms/fck-uploaded/ documents/ country studies/ bangladesh/ Working conditions in the Bangladeshi garment sector Socialdialogueandcompliance.pdf. (Access on April 1, 2015).

[6] Shanthi, S. & Ramani, Dr.R.G. (2012), "Gender Specific Classification of Road Accident Patterns through Data Mining Technique", *IEEE International Conference On Advances In Engineering, Science And Management*, pp.359-365.

[7] Besha, T. & Hill, S.,"Mining Road Traffic Accident Data to Improve Safety: Role of Road Related Factors on Accident Severity in Ethiopia", unpublished.

[8] Zhang, L., Chen, Y., Liang, Y. & Li, N., "Application of Data Mining Classification Algorithms in Customer Membership Card Classification Model", *International Conference on Information Management and Industrial Engineering*, pp. 211-215, 2008.

[9] Firoz, A. (2011). "Design of Readymade Garments Industry for Fire Safety" (Online). Available from: http:// dspace.bracu.ac.bd/bitstream/handle/10361/1818/ ahsan%20firoz.pdf?sequence=1. (Access on April 12, 2015)

[10] Calvert, J. B. (2002). "The Collapse of Buildings" (Online), Available from: http://mysite.du.edu/ ~jcalvert/tech/failure.htm. (Access on April 2, 2015)

[11] MFB (2009). Available from: http:// www.mfb.vic.gov.au/Industry/Managing-False-Alarms.html. (Access on April 3, 2015)

[12] Oxford Dictionaries (2015). Available from: http:// www.oxforddictionaries.com/definition/english/casualty. (Access on April 3, 2015)

[13] Vohra, G., Bhushan, B. (2012). "Data Mining Techniques for Analyzing the Investment Behavior of Customers in Life Insurance Sector in India." Available From: http:// www.euroasiapub.org/pastissue.php?title= IJRIM&vol=Volume%202,Issue%209,September-2012. (Access on 25 February, 2015)

[14] Ahmed, F., "Improving Social Compliance in Bangladesh's Ready-made Garment, Industry" (Online). Available from: http://www. nla.gov.au/openpublish/index.php/lmd/article/view/2269 /3148. (Access on April 10, 2015)

[15] Clean Clothes Campaign (2012), "Hazardous workplaces: Making the Bangladesh Garment Industry Safe" (Online). Available from: http:// www.cleanclothes.org/resources/publications/2012-11-hazardousworkplaces.pdf. (Access on April 10, 2015)

[16] Ahmed, I., Guan, D., Chung, T.C.,"SMS Classification Based on Niave Bayes Classifier and Apriori Algorithm Frequent Itemset", *IJMLC*, vol.4, No. 2, pp. 183-187, 2014.

[17] Shah, M.C., and Jivani, A.G., "Comparison of Data Mining Classification Algorithms for Brest Cancer Prediction", 4th ICCCNT, 2013.

[18] Bala1, S. & Kumar, K., "A Literature Review on Kidney Disease Prediction using Data Mining Classification Technique", *IJCSMC*, Vol. 3, Issue. 7, July 2014, pp: 960 – 967.