

# Classification of economic activities in Colombia according to workplace accident and disease rates using a data clustering algorithm.

D. GARAVITO<sup>1</sup>, I. KOCSIS<sup>2</sup>.

<sup>1</sup>University of Debrecen, Faculty of Engineering, Department of Engineering Management and Enterprise, damagadi@gmail.com

<sup>2</sup>University of Debrecen, Faculty of Engineering, Department of Basic Technical Studies, kocsisi@eng.unideb.hu

*Abstract. In 2002 the Colombian Government issued a classification of economic activities according to their level of occupational risk, taking into account the International Standard Industrial Classification of All Economic Activities (ISIC). Since 2008, the federation of Colombian insurance companies (FASECOLDA) has gathered data related to workplace accident and occupational disease rates reported by companies with employees affiliated to the legal labour system. In this study, economic activities have been categorized according to the average rates of workplace accident and occupational disease presented between 2008 and 2016, using some techniques of the Data Clustering Algorithm. The purpose of this research is to evaluate the current classification system, and propose recommendations for a new approach, taking into account the analysis of historic data available.*

## Introduction

All over the world it is widely accepted that there is a huge economic benefit to be gained from improved safety and health in the workplace and a reduction in occupational accidents and diseases. In order to achieve this benefit, there is a need for a better knowledge of the true situation of safety and health in the workplace at the enterprise, industry and national levels in order to make the right decisions [1]. In Colombia, there have been many efforts to achieve an accurate documentation of the workplace conditions. Since 1915, the Colombian government has promulgated several laws related to occupational health and safety, in particular, those indicating the responsibility of the employer with regards of workplace accidents and occupational diseases [2]. It is compulsory for any company situated in Colombia to report timely any accident occurred to their workers or notify the existence of any employee with an illness associated with the job performed.

Despite these efforts, only in 2012 the so called “General System of Labour Risk” was formally created with the aim to protect workers against occupational accidents and diseases through a set of public and private entities that work to protect and assist employees in Colombia. One of the requirements of this system is related to the affiliation of employees to an insurance company by their respective employers, so they are covered in case of accidents or diseases associated to the job performed [3].

The affiliation fee depends on the classification of the company's economic activity according to its occupational risk. The higher the risk, the higher is the amount to be paid. This classification was done back in 2002 without any statistic data available about the performance of companies in terms of occupational health and safety. Therefore, the current study intends to provide a recommendation for a more objective classification taking into account the historic data available on the matter.

## 1. Legal labour system in Colombia

### 1.1. Standard classification system for economic activities

National classifications for occupation and industry were first developed in the last decades of the 19th century. These first classifications started as listings of occupations without any hierarchical structure, and had a tendency to reflect social strata rather than tasks performed. With the increasing need for international and inter-discipline comparisons, the necessity for standardization emerged. International standards were discussed at the International Conferences of Labour Statisticians organized by the ILO in the 1920s, but concrete results only emerged during the 1940s. To date, many countries base their national classifications (to some degree) on standardized classifications. [4].

In 2002, Colombia adopts the International Standard Industrial Classification of all Economic Activities (ISIC) in the field of occupational health and safety, with the purpose of standardizing and generating data that could be compared internationally [5]. The classification is organized in 15 groups as it follows:

- Agriculture, stockbreeding, hunting and forestry: 27 activities
- Public administration and defense: 10 activities
- Wholesale and retail trade: 87 activities
- Education: 12 activities
- Financial and insurance activities: 26 activities
- Accommodation and food service activities: 11 activities
- Manufacturing: 265 activities
- Real estate activities: 50 activities
- Mining and quarrying: 24 activities
- Activities of extraterritorial organizations and bodies: 1 activity
- Fishing: 4 activities
- Housekeeping: 1 activity
- Community, social and personal services: 34 activities
- Health and social services: 11 activities
- Transportation, storage and communication: 41 activities

In total there are 604 activities classified by Colombian law as legal economic enterprises that can be performed in the country. Likewise, for each activity there is an identifying code that indicates the following:

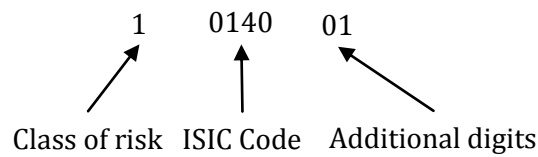


Figure 1. Code structure for economic activities in Colombia

## 1.2. General system of labour risk

For a developing country like Colombia, it is essential to understand and analyze the problems and reality of its labour market, in such a way that the sectorial activities of the economy can be reinforced. Because of this, since 1994 the Government has been developing legal and institutional mechanisms that aim to cover all workers residing in the country [6]. The most important mechanism is the so called General System of Labour Risk or “Sistema General de Riesgos Laborales (SGRL)” formally established in 2012. According to this system, all the working population must be affiliated by their employer to an insurance company specialized in labour risks called “Administradora de Riesgos Laborales (ARL)”.

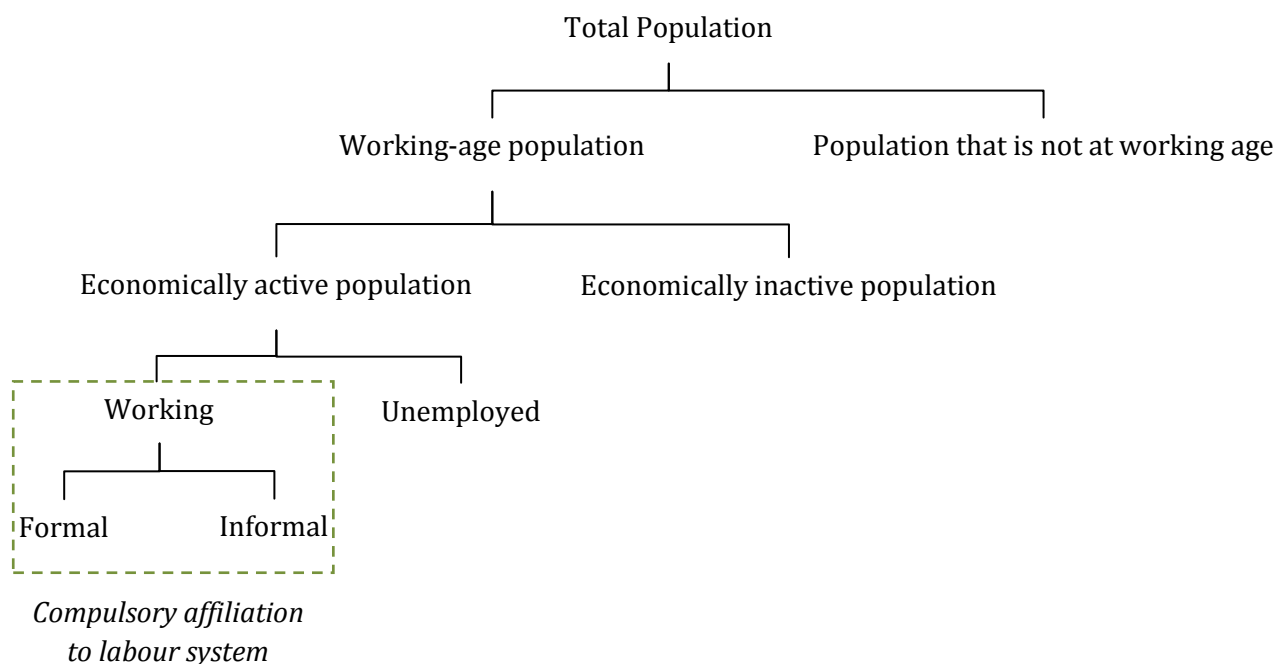


Figure 2. Labour Market Organization in Colombia

The distribution of workers affiliated to the legal system according to each economic activity shows that the 5 main groups are real state (50 activities), manufacturing (250 activities), commerce (87 activities), transport, storage & communication (41 activities) and education (12 activities).

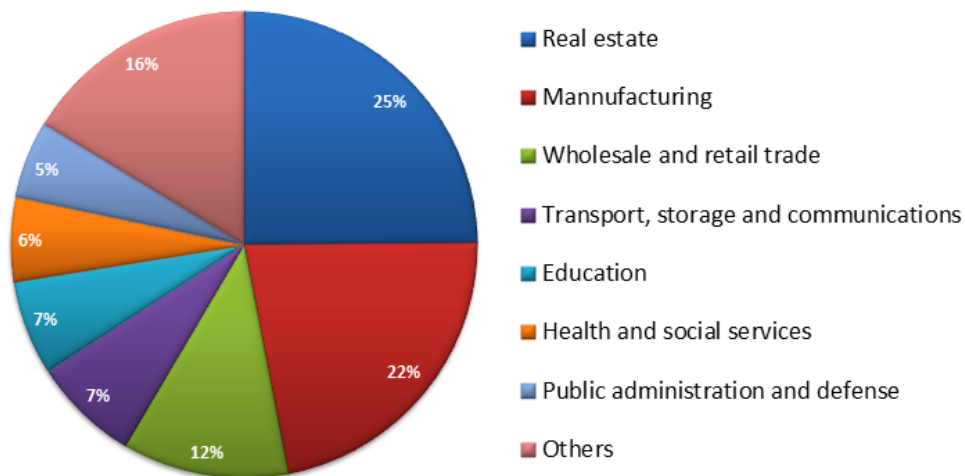


Figure 3. Distribution of affiliated workers per economic activity in Colombia

Now, according to Colombian legislation, every company has to be classified according to the occupational risk of its main economic activity. This risk class determines the value of the employee affiliation to the General System of Labour Risk (SGRL) paid by the employer, the higher the risk, the higher the value. For each class there is an initial value every company has to pay when it enters the system for the first time. After one year, the value can change according to the behaviour of workplace accident and disease rates and the compliance of the company with local legislation as well, always between a certain range established by the government [7].

Class of Risk	Type of Risk	Minimum Value	Initial Value	Maximum Value
I	Minimum	0.348%	0.522%	0.696%
II	Low	0.435%	1.044%	1.653%
III	Low	0.783%	2.436%	4.089%
IV	High	1.740%	4.350%	6.960%
V	Maximum	3.219%	6.960%	8.700%

Table 1. Employee affiliation fee to SGRL according to occupational risk.

### 1.3. Labour risk indicators

The indicators that evaluate the improvement of the health and job conditions of Colombian workers defined by the Colombian government are accident rate, labour illness rate and workplace mortality. The aim of this measure is to gauge the impact of the promotion and prevention activities held in the companies, according to local legislation [8].

Indicator	Formula	Frequency
Accident Rate	$(\text{Number of workplace accidents}/\text{Number of workers affiliated to legal system}) * 100$	Annual
Occupational Disease Rate	$(\text{Number of occupational legally certified}/\text{Number of workers affiliated to legal system}) * 100.000$	Annual
Mortality rate because of workplace accident and occupational disease	$(\text{Number of mortal events}/\text{Number of workers affiliated to legal system}) * 100.000$	Annual

Table 2. Labour risk indicators established by Colombian law.

## 2. Methodology

This research was conducted using data provided by the federation of Colombian insurance companies (FASECOLDA), which was processed at University of Debrecen using the programs Microsoft Excel® and Matlab®.

### 2.1. Sample

The data obtained corresponds to the number of workers affiliated to the General System of Labour Risk, number of accidents reported by companies and number of occupational diseases legally recognized for all economic activities in Colombia from 2008 to 2016. All the data is categorized according to the standard classification system adopted by the government.

However, it is suitable to clarify that not all working population in Colombia is affiliated to the General System of Labour Risk, despite being a requirement of the local legislation. Therefore, the data obtained can be assumed as a sample of the actual situation. On average, between 2008 and 2015 only 37.79% of working population was affiliated to the system.

Indicator	2.008	2.009	2.010	2.011	2.012	2.013	2.014	2.015
Working age population	33,597,448	34,154,700	34,706,228	35,247,948	35,781,056	36,306,640	36,826,516	37,342,468
Occupation rate	51.90%	53.90%	55.40%	56.80%	57.80%	58.00%	58.40%	59.00%
Working population	17,437,076	18,409,383	19,227,250	20,020,834	20,681,450	21,057,851	21,506,685	22,032,056
Working population affiliated to the system	27.60%	36.40%	35.41%	37.46%	40.77%	39.28%	41.55%	43.83%
Employees affiliated to the system	4,812,838	6,700,836	6,809,221	7,499,480	8,430,796	8,271,915	8,936,929	9,656,781

Table 3. Labour risk indicators established by Colombian law.

### 2.2. Applied method: data clustering algorithm

Clustering algorithms are used extensively not only to organize and categorize data, but are also useful for data compression and model construction. Clustering partitions a data set into several groups such that the similarity within a group is larger than that among groups. Achieving such partitioning requires a similarity metrics that takes two inputs vectors and returns a value reflecting their similarity. [9]

One algorithm used for data clustering is the so called K-means clustering. The K-means algorithm partitions a collection of  $n$  vector  $X_j$ ,  $j=1, \dots, n$ , into  $c$  groups  $G_i$ ,  $i = 1, \dots, c$ , and finds a cluster centre in each group such that a cost function (or an objection function) of dissimilarity (or distance) measure is minimized. When the Euclidean distance is chosen as the dissimilarity measure between a vector  $X_k$  in group  $j$  and the corresponding cluster center  $C_i$ , the cost function can be defined by:

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{k, X_k \in G_i} \|X_k - C_i\|^2 \quad (1)$$

Taking into account that the purpose of this study is to find a logical classification system based on quantitative criteria, the algorithm was adjusted in a way that the groups were formed around fixed centerpoints  $C_i$ . Therefore, in this research the cost function was not minimized as indicated in the K-Means algorithm. The final steps performed were the following:

1. Calculate average rates of workplace accident and occupational disease rates among all data set.
2. Determine upper and lower levels of average rates for each class, according to suggested classification (See point 2.3).
3. Assign a centerpoint for each class, taking into account middle or average rate per class.
4. Calculate the distances from all points of data with respect of all centerpoints.
5. Categorize data according to the minimum distance from all centerpoints.

### 2.3. Proposed risk classification system

One of the shortcomings of the current legal classification system is that it does not provide detailed description on how to measure of occupational risk. According to Colombian law, risk is a combination of the probability of the occurrence of one or more exposures or dangerous events and the severity caused by these [7]. At the moment of this study, the only data available was associated to the probability of occurrence of these events (accident and disease rates). Therefore, in this study we suggest a new method for categorizing economic activities taking into account these rates as indicated in figure 4.

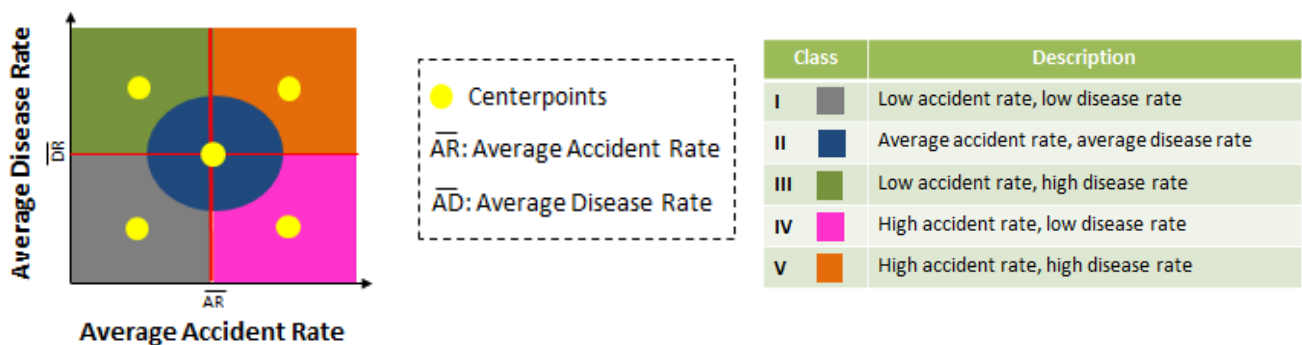


Figure 4. Suggested classification for economic activities according to occupational risk

### 3. Results

#### 3.1. Legal classification

If we take a look at the distribution of economic activities per risk class according to Colombian legislation, we find that most activities in class I (135 activities) are concentrated in the area of a *low average accident and disease rates*, in class II (118 activities) there are some outstanding activities with high accident rate but a low average disease rate, in classes III (146 activities) and IV (110 activities) there is a high dispersion of data and in class V (95 activities) it is also possible to find outstanding activities with high accident rate but low disease rate. This could indicate that the legal classification indeed does not follow the actual behaviour of companies in each risk class, and that it is appropriate to evaluate this organization.

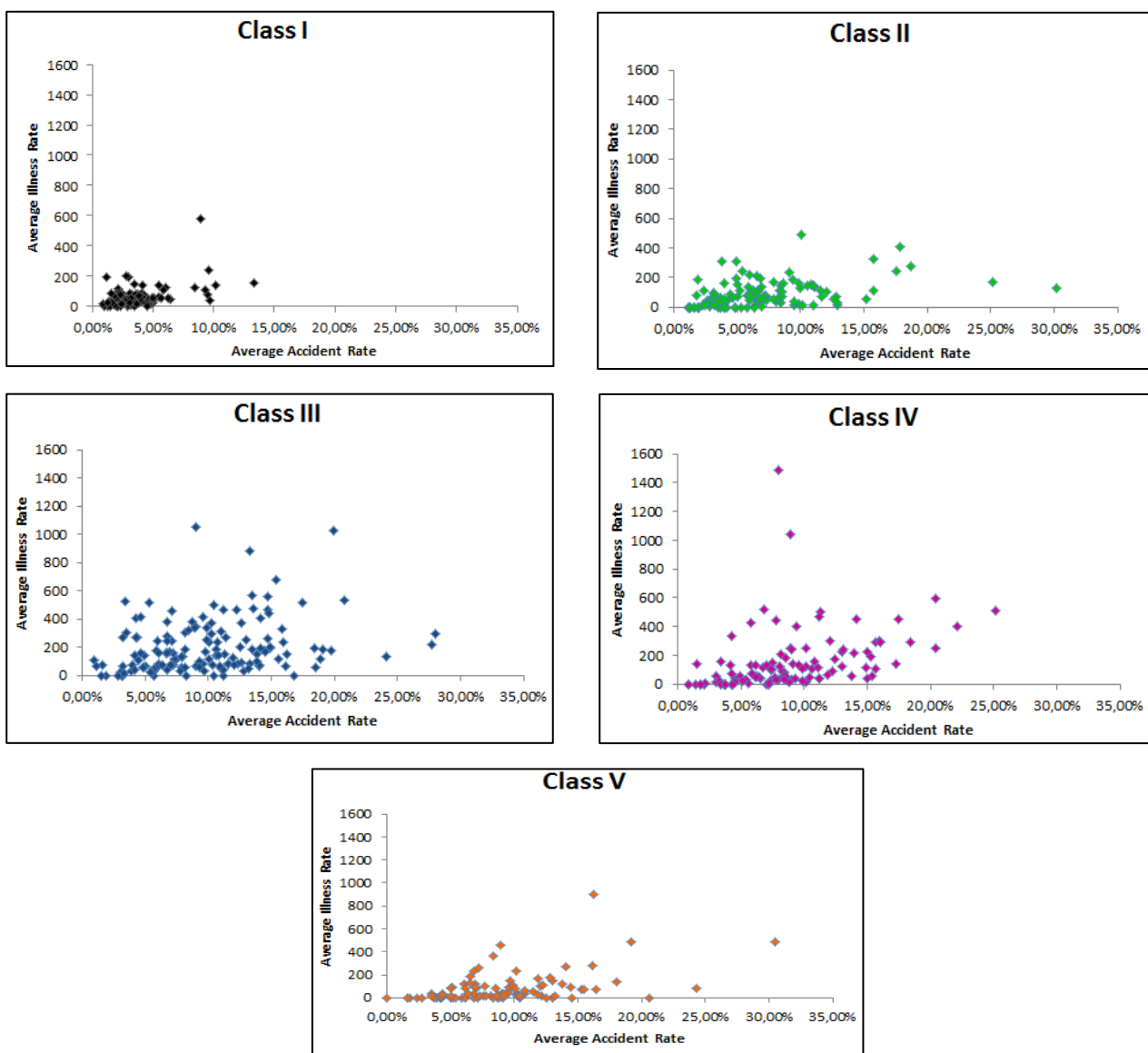


Figure 5. Distribution of original classification of economic activities per occupational risk class

Indeed, if all the data sets are plotted in the same graph, there is no clear distinction between each class in terms of average disease rate and average accident rate.

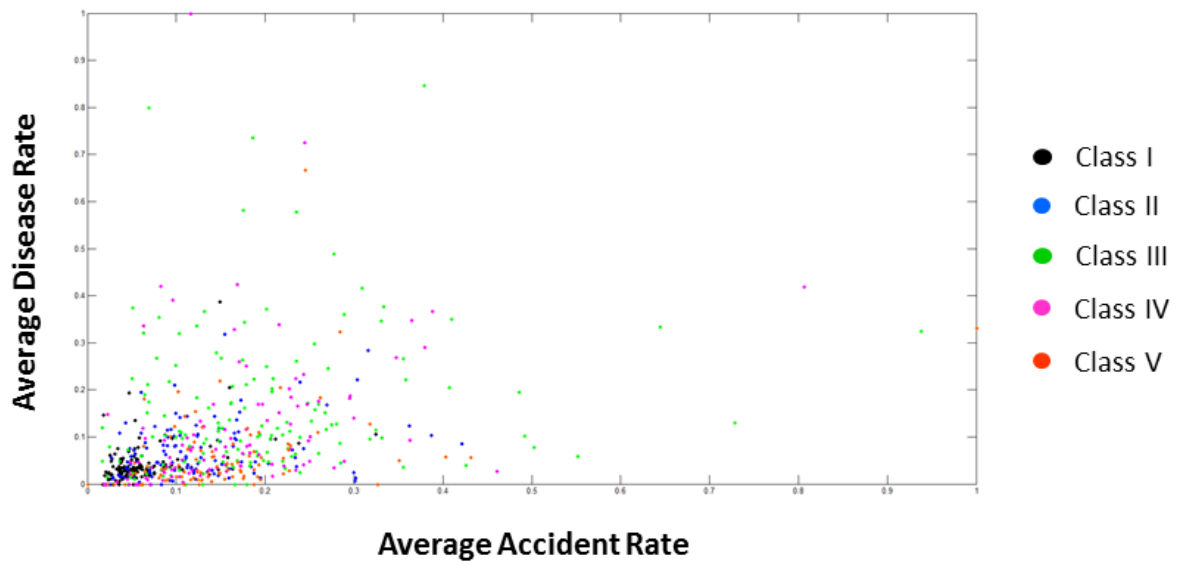


Figure 6. Original classification of economic activities in terms of occupational disease and accident rate

### 3.2. Suggested classification

Applying the suggested classification and criteria, we find a new clustering of all 604 economic activities according to their performance in terms of occupational health and safety from 2008 to 2016. With this re-classification, 252 activities are identified as class I, 267 activities as class II, 7 activities as class III, 61 activities as class IV and 17 activities as class V.

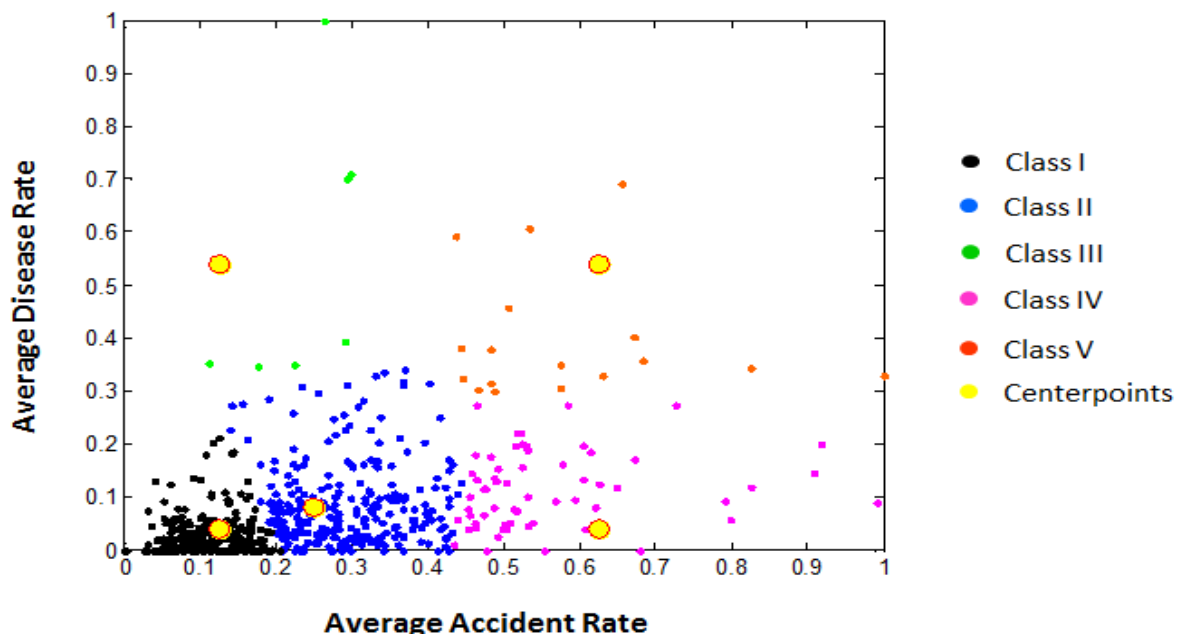


Figure 7. Suggested classification of economic activities in terms of occupational disease and accident rate



### 3.3. Consideration of outliers

As shown in figure 7, the clustering obtained by applying the suggested methodology in this research seems to be clearer and more delimited. However, since the dispersion of the data is relatively high (Standard Deviation of 0.05 for Accident Rate, and 157.48 for disease rate), the result classification is being biased because of the presence of data outliers. An outlier may be defined as an observation in a set of data that appears to be inconsistent with the remainder of that set. Usually, this means a value that is visibly distant from the remainder of the data. [10]

One of the ways to detect outliers in statistics is to use the Standard Deviation as a reference point. For this research, if the deviation of a value from the sample mean is larger than 3 times of the standard deviation then the point may be considered as an outlier [11]. Under these criteria, 16 economic activities were found to have outstandingly high accident and disease rates as indicated in the following table.

<b>Statistically High Accident Rate</b>	<b>Statistically High Disease Rate</b>
Manufacturing of engines and turbines (does not include engines for airplanes, vehicles or motorcycles)	Manufacture of tires and rubber tires
Banana production	Manufacture of fans, shaving machines and hair dryers.
Meat production (slaughter houses)	Production of non-refractory ceramic materials
Manufacture of bodywork for vehicles	Packing of fish and seafood
Specialized production of sugar cane	Non-ferrous metal casting
Production of cereals and oilseeds	Greenhouses and floriculture
Extraction of precious metals	Manufacture of locks, keys and similar
Manufacture of wooden containers	Sugar refinery

*Table 4. Outliers in data set taking into account accident and disease rates*

Now, if we study the sample again extracting all of the outliers from the data set, we obtained a new clustering as indicated in figure 8. With this consideration, the re-classification is composed by 248 activities identified as class I, 181 activities as class II, 30 activities as class III, 77 activities as class IV and 51 activities as class V. Clearly, the effect of taking out the outliers from the main sample is the rearrangement of the 5 centerpoints or averages for each class. This procedure, in consequence, compacts the sample by reducing the distance between each centerpoint.

Since the purpose of this study does not include the investigation of the main causes for such high accident and disease rates, it is suitable to highlight these outlier activities and recommend further

research in these fields in order to prevent the occurrence of more occupational accidents and diseases.

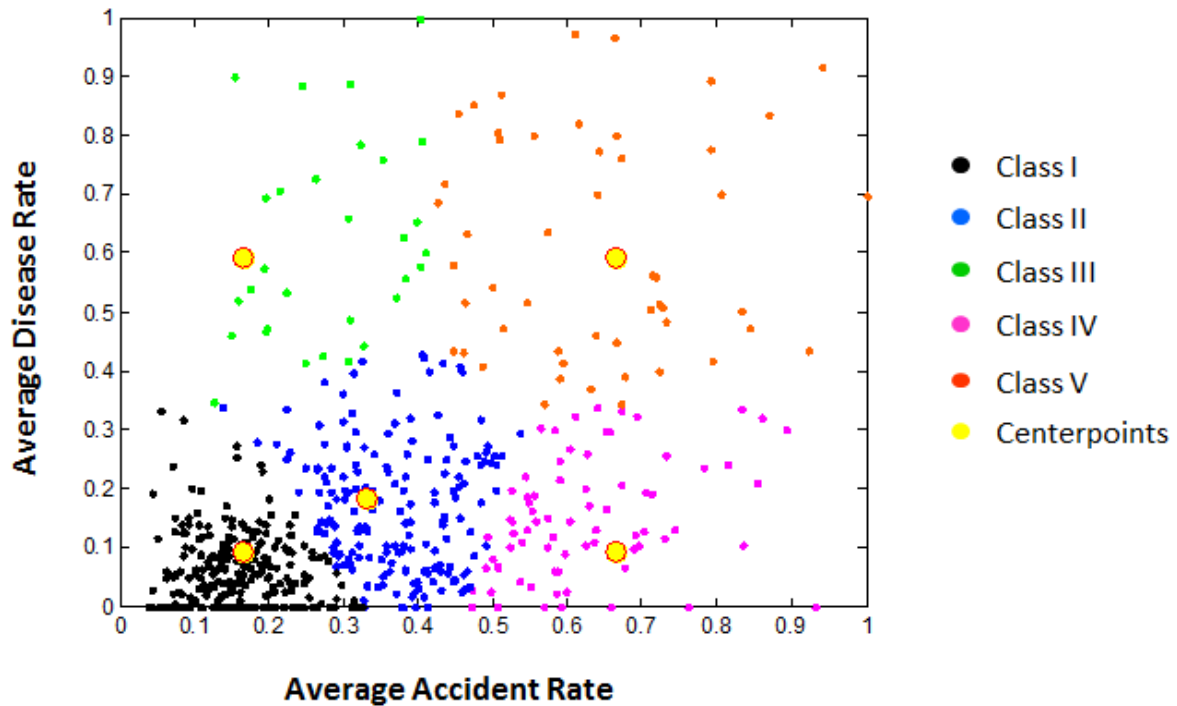


Figure 8. Suggested classification of economic activities in terms of occupational disease and accident rate

#### 4. Discussion and conclusions

The current legal classification system of economic activities in Colombia done in 2012 does not explain the actual behaviour of companies belonging to these activities in terms of occupational accident and disease rates. The classification system proposed in this study is based on the statistic data available and shows a considerable difference with respect of the current system, especially for the classes I and II. If outliers are taken out from the main sample, a more proportionate clustering is obtained, but it keeps proving the need for re-evaluating the current classification system.

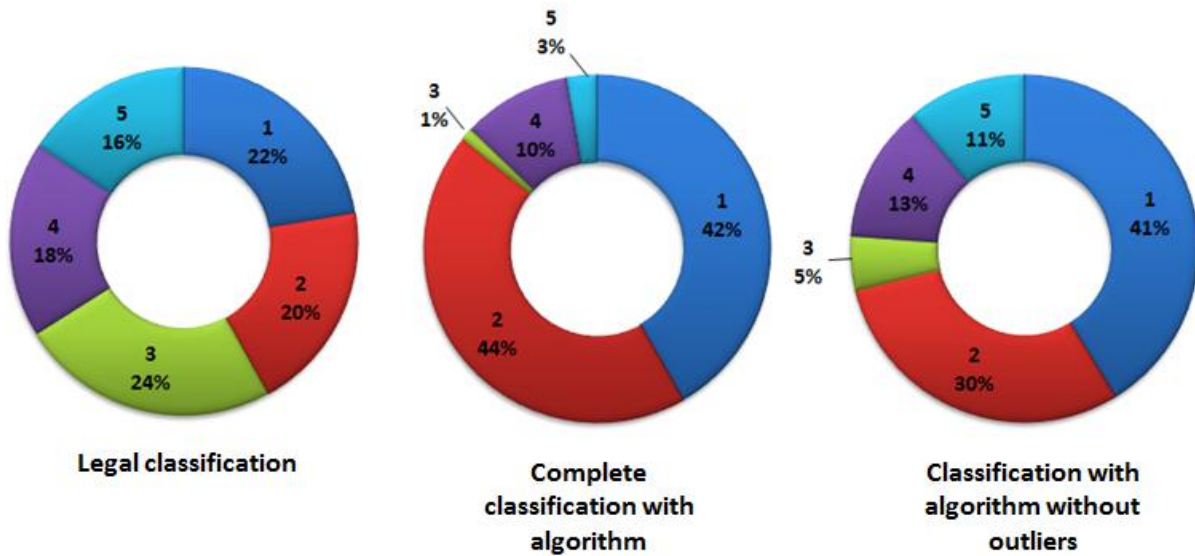


Figure 9. Comparison between current and suggested classification of economic activities

The reliability of the results obtained depends on the data collected by the Federation of Colombian insurance companies (FASECOLDA). The fact that only 43.83% of the working population is affiliated to the General System of Labour Risks imposes a big obstacle in the statistical analysis of the data available. It is crucial that the government increases its efforts in guaranteeing that all working population is part of this system. Also, controls must be in place to ensure that all workplace accidents and occupational diseases are reported as established by the law.

Despite these shortcomings, the use of the data clustering algorithm shows that if we categorize available data according to the average workplace accident rate and occupational disease rate, there is a difference compared to the categorization that has been established by the law. Therefore, it is clear that the actual classification should be evaluated and adjusted to the reality of the companies in the country.

Besides, the economic activities highlighted in this study as outliers must be taken into consideration by the government. It is suggested an in-deep investigation to find the causes for the high accident rate and disease rate in these companies.

Finally, further research could include other important factors in occupational health and safety that were not available at the moment of this study, such as severity and frequency of injuries, mortality rate and percentage of loss of labour capacity in injured or ill employees.

## References

- [1] International Labour Office, *Improvement of national reporting, data collection and analysis of occupational accidents and diseases*, ILO Publications. Web: [http://www.ilo.org/wcmsp5/groups/public/---ed\\_protect/---protrav/---safework/documents/publication/wcms\\_207414.pdf](http://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---safework/documents/publication/wcms_207414.pdf)

- [2] J. Zuleta, *Regulatory framework of the General Labour Risks System in Colombia*, Universidad Nacional de Colombia, Master Thesis 11360.
- [3] Congress of Colombia, *Law 1562 of 2012*
- [4] A. Mannetje, H. Kromhout, *The use of occupation and industry classifications in general population studies*, Int J Epidemiol 2003; 32 (3): 419-428.
- [5] Health and Social Security Ministry of Colombia, *Decree 1607 of 2002*.
- [6] A. Zambrano, *The General Labour Risks System in Colombia*. Fasecolda, 201-244. Web: [http://www.fasecolda.com/files/2813/9101/0320/parte\\_i.captulo\\_5\\_el\\_sistema\\_general\\_de\\_riesgos\\_laborales\\_en\\_colombia.pdf](http://www.fasecolda.com/files/2813/9101/0320/parte_i.captulo_5_el_sistema_general_de_riesgos_laborales_en_colombia.pdf)
- [7] Presidency of Colombia, *Decree 1072 of 2015*.
- [8] Ministry of Health of Colombia, *Labour risks Indicators*. Web: <https://www.minsalud.gov.co/proteccionsocial/RiesgosLaborales/Paginas/indicadores.aspx>
- [9] J. Jang, C. Sun, E. Mizutani, *Neuro-Fuzzy and Soft Computing – A Computational Approach to Learning and Machine Intelligence*. Prentice Hall.
- [10] S. Ellison, V. Barwick, T. Duguid Farrant, *Practical Statistics for the Analytical Scientist*, RSC Publishing, 48.
- [11] X. Yan, X. Gang Su, *Linear Regression Analysis: Theory and Computing*, World Scientific, 134.