https://dx.doi.org/10.22161/ijaers.5.9.4 ISSN: 2349-6495(P) | 2456-1908(O)

Data Mining Technique for Preventional Analysis of Work Accidents

André Gomes Barros[1], Juliano Said[2], Geórgia Regina Rodrigues Gomes[3] and Fabrício Moraes de Almeida[4]

[1]Universidade Cândido Mendes (UCAM) – Campos dos Goytacazes – RJ, Brasil. E-mail: eng.andre1978@gmail.com
[2]Universidade Cândido Mendes (UCAM) – Campos dos Goytacazes-RJ, Brasil.
[3]Universidade Cândido Mendes (UCAM) Campos dos Goytacazes -RJ), Brasil.
[4]PhD in Physics (UFC), with post-doctorate in Scientific Regional Development (DCR/CNPq). Researcher of the Doctoral and Master Program in Regional Development and Environment (PGDRA/UFRO). Leader of line 2, Technological and Systemic Development, and Researcher of GEITEC — Federal University of Rondônia, Brazil. E-mail: dr.fabriciomoraes001@gmail.com

Abstract *— There are many oil companies operating in Macaé-RJ, Campos basin, they value the safety of work and the lives of its employees. These companies do a study to verify the health status of their employees' spine and result in a database with six attributes, such as: Pelvic incidence, Pelvic inclination, Lumbar lordosis angle, Sacral inclination, Pelvic radius and Degree of Spondylolisthesis. For in Brazil, recently, Social Security has released statistical studies that show back pain as leaders in the ranking of departures in the first half of 2016, a fact that directly affects the productivity of companies and health of their employees. This article aims to apply the KDD process, specifically the task of Data Mining classification, ie, classify if the employee will be fit or unfit for the job. The decision tree was the technique chosen through the algorithm J48 to verify the possibilities of treatment of the collaborators in the prevention and improvement in the working environment, and even, a change in the management was made from the results found. It resulted in inadequate staff postures, inadequate service stations, lack of training in equipment handling, lack of knowledge about cargo handling.*

Keywords— Data Mining; KDD; Column; Occupational Accidents.

I. INTRODUCTION

In 1943, the creation of the Consolidation of Labor Laws (CLT) was approved, then sanctioned by the then president of the republic, Getúlio Vargas. One of the chapters dealt with occupational safety and medicine, establishing coordination, orientation, control and supervision of activities related to occupational health and safety throughout the national territory, including the National Campaign for the Prevention of Accidents at Work. In addition, it established as an assignment of the companies "to instruct the employees, through service orders, as to the precautions to be taken to avoid accidents at work or occupational diseases". With CLT, Vargas would go down in history as the benefactor of the working class (FALEIROS, 2002).

The area of work safety directly affects all productivity of a company, and when accidents with time off from work, the sector from which it was deprived of the employee is below its normal production capacity. In Brazil, the National Health Survey indicates that more than 20 million people suffer from some chronic disease in the spine (FALEIROS, 2002).

The pains can worsen from several strands such as stress, overweight or smoking. In more severe cases, it is possible that repetition of movements, overload and poor posture can lead to scoliosis (curving of the spine) or even disc hernias. Depending on the position and function of the worker, it is still possible that other diseases, such as RSI / Dort, Repetitive Strain Injuries / Work-related Musculoskeletal Disorders, are developed.

The International Classification of Impairments and Disabilities of the World Health Organization recognizes low back pain as a compromise that reveals loss or abnormality of the lumbar spine structure of psychological, physiological, or anatomical etiology, or a disability that prevents the full performance of work activities (WORLD HEALTH ORGANIZATION, 1980).

Schilling, in 1984, proposed a classification of work-related diseases divided into three groups:

I. diseases that have work as the necessary cause, such as occupational accidents and occupational diseases legally recognized;
II. diseases that have work as one of the contributing factors;
III. diseases that have work as aggravating or provoking latent or pre-existing disorders.

https://dx.doi.org/10.22161/ijaers.5.9.4 *ISSN: 2349-6495(P) | 2456-1908(O)*

Using the Schilling classification, occupational low back pain can be classified as Schilling II when the work is considered one of the contributing factors for its onset, or Schilling III when the work is considered as aggravating factor of a preexisting disorder or pathology.

An alternative to avoid this kind of problem in companies is to establish a Health and Safety Management (TSS) that considers several aspects, such as the direct and indirect risks to which the worker is exposed. In parallel to this, encouraging other actions, such as Internal Week of Prevention of Work Accident (SIPAT) or Labor Gymnastics, can prevent the incidence of these diseases. In addition to the decrease in the rate of remoteness, OSH-related activities broaden the perceptions of the worker, who tends to become more aware and apply the knowledge also in personal life.

The study in question is the use of KDD (Knowledge Discovery in Databases) in the database of a multinational oil company that operates in Macaé-RJ, in the Campos basin, where six attributes were used: Pelvic incidence, Pelvic inclination, Lumbar lordosis angle, Sacral inclination, Pelvic radius and Degree of Spondylolisthesis.

The KDD process involves several steps ranging from understanding the problem to be solved to extracting knowledge through data mining techniques. KDD is a process proposed in 1989 that according to Fayyad et al. (1996, quoted by Liebstein, 2005) is not trivial, identifying patterns that are valid, new, potentially useful and understandable. This involves finding and interpreting patterns in the data, iteratively and interactively, by repeating the algorithms and analyzing their results.

In order to work with the prevention of spinal diseases and time involved with withdrawal that generate a very high cost for the company, the purpose of this article is to use the task of classification of data mining to define whether the employee is fit or not to the work.

What all companies have in common, regardless of the method used, is the need to analyse the economic efficiency of a particular project, by grouping all costs, the value obtained must be less than the estimated value that it can generate as revenues or benefits.

II. METHODOLOGY

The study in question is a database of a multinational oil company operating in Macaé-RJ, in the Campos basin, related to the column, having six main parameters, such as: Pelvic incidence, Pelvic inclination, angle of lordosis lumbar, sacral inclination, pelvic radius and degree of spondylolisthesis.

The open source data mining tool used was WEKA version 3.8 and the implemented algorithm was J48. The tool was developed by the University of Waikato in New Zealand. It can be defined as a collection of machine learning algorithms to perform data mining tasks.

WEKA has been increasingly applied and some interesting features help to explain its success (MURASSE and TSUNODA apud MARKOV and RUSSELL, 2006):

- Contains several algorithms for data mining, web mining and machine learning;
- Has open source and is available on the Web for free;
- It is relatively easy to use, even by people who are not experts;
- Provides flexible resources for experiments;
- It is kept updated, since new algorithms are added as soon as they appear in the literature.

The data mining process comprises the following (MURASSE and TSUNODA apud MARKOV and RUSSELL, 2006):

- Raise data sources (databases, reports, etc);
- Perform a data cleaning to "load" to WEKA;
- "Upload" to WEKA the post-cleaning data file;
- Search patterns relevant to the problem in question using the algorithms embedded in the software.

The Ministry of Labor and Employment and Social Security (MTPS) contains a website, a friendly and interactive environment for the user, which contains a set of data on the main causes of withdrawal from work throughout the country. In this, was downloaded the information referring to the main causes having column (low back pain) the record holder in the first place.

The discovery of the knowledge on the database was given by the results of the X-ray and Magnetic Resonance examinations together with the implantation of a database with 6 attributes in order to better monitor one of the main causes of low productivity in the company the withdrawal due to in the period from June 2, 2014 to October 31, 2016, specifically aimed at Colbal (low back pain), where it is worth mentioning that it is a record holder of accidents due to remoteness, according to data from the Ministry of Labor and Employment and Social Security, of a total of 310 employees investigated, totaling the members of the company's staff and released in the dataset, requested by the Work Doctor together with a colleague Orthopedist and the Labor Safety Engineer, who is responsible , together with the occupational physician, for the implementation of mitigating measures that may reduce the scenario of remoteness, where the occupational physician who works as coordinator of the Medical Control and Occupational Health Program (NR07) of the company studied, is a member of the company's Specialized Medical and Occupational Safety Engineering (NR04) sector.

In this algorithm, the decision tree is modeled based on the most significant attribute, which appears as the root of the tree. From this root, branches are generated,

https://dx.doi.org/10.22161/ijaers.5.9.4 *ISSN: 2349-6495(P) | 2456-1908(O)*

which represent the relevance of this connection. These branches can also generate other branches that would work the same way. Such a structure would then have the capacity to represent, intuitively, where knowledge could be extracted.

Goldshmidt (2005) says that decision trees are also known by the names of regression trees, or even classification trees and that they are graphical representations of a set of rules, consisting of roots, branches and knots, similar to a tree , where the analysis of these representations must be performed from the top to the leaves. These decision trees have as nodes the leaf values of the attributes of the base and the leaf nodes as the instances of these, that is, each of the decisions taken to carry out this classification are pertinent to a single node.

The J48 algorithm generates decision tree models from the top to the bottom, so that on each node other attributes are evaluated individually to determine their significance in the connection or even existence in it.

III. RESULTS AND DISCUSSION

It is noticed that with the degree of spondylolisthesis less than or equal to 19.85º and Pelvic Radius is greater than 125,21º, the collaborator is considered fit.

If the degree of spondylolisthesis is greater than 19, the employee is considered as unfit.

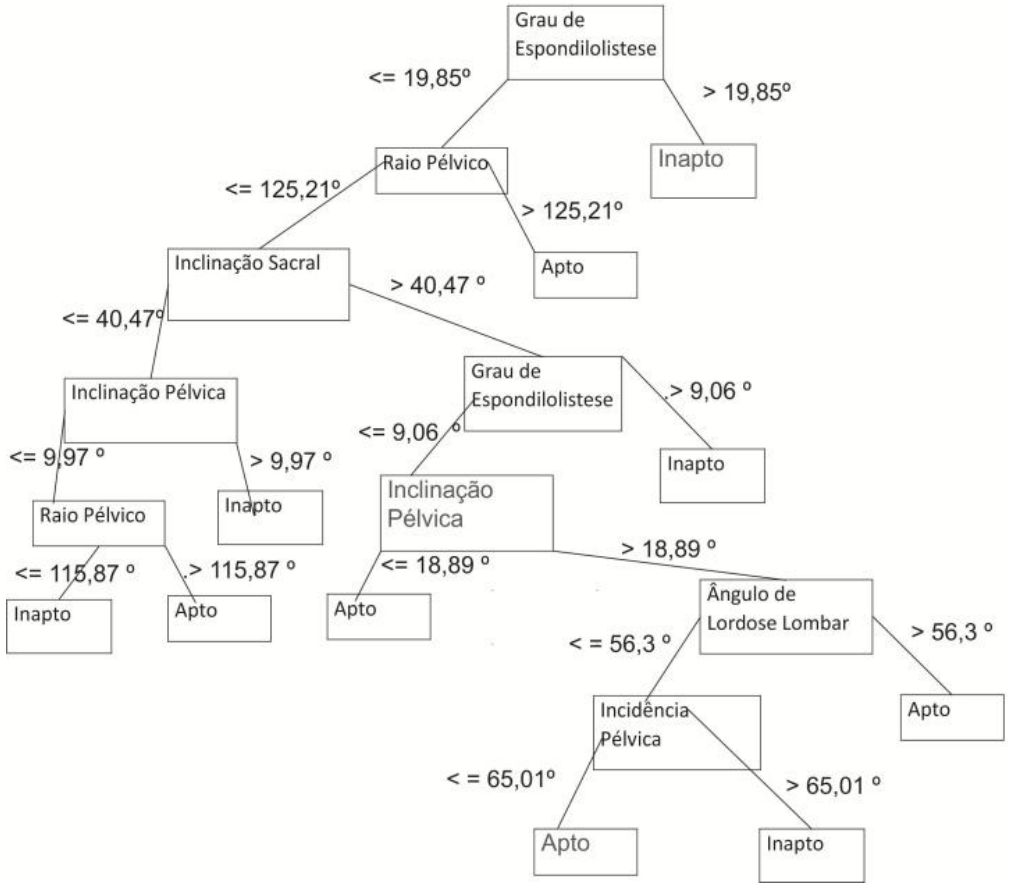

Fig.2: Decision tree. Source: WEKA - Algorithm J48.

When the degree of spondylolisthesis is less than or equal to 19.85 °, the pelvic radius less than or equal to 125.21 °, the sacral inclination is less than or equal to 40.47 °, the pelvic inclination is greater than 9.97 °, the employee is considered unfit.

When the degree of spondylolisthesis is less than or equal to 19.85 °, the pelvic radius less than or equal to 125.21, the sacral inclination is greater than or equal to 40.47 °, and the degree of spondylolisthesis is greater than 9.06 °, considered unfit. If, on the other hand, the degree of spondylolisthesis is less than or equal to 9.06 º, the pelvic inclination is less than or equal to 18.89 º is considered fit. If the pelvic tilt is greater than 18.89 degrees and the lumbar lordosis angle is greater than 56.3 degrees, it is considered as Apt. If the angle of lumbar lordosis is less than or equal to 56.3 º and the pelvic inclination is greater than 65.01 º, it is considered as unfit. If the pelvic tilt is less than or equal to 65.01, it is considered fit.

When the degree of spondylolisthesis is less than or equal to 19.85 °, the pelvic radius less than or equal to 125.21 °, the sacral inclination is less than or equal to

https://dx.doi.org/10.22161/ijaers.5.9.4 *ISSN: 2349-6495(P) | 2456-1908(O)*

40.47 ° and the pelvic inclination is greater than 9.97 °, the employee is considered unfit.

As a result of the analysis of the decision tree formed by the algorithm J48, it is possible to emphasize that the collaborators with diagnosis of spondylolisthesis are considered incapable for work activity, within the investigative clinical process, we tried to relate lumbar pain complaints with the cause. Considering this, and considering the attributes analyzed and the so-called normal parameters, it is perceived that the low grade spondylolisthesis associated with a sacral tilt out of normality also makes the collaborator incapable.

When the clinical investigation is the initial diagnosis is inconclusive for spondylolisthesis, analysis begins from the pelvic radius when out of normality, it is advanced to the analysis of the sacral inclination that when associated with spondylolisthesis or pelvic tilt is also considered inapt.

Collaborators who during the clinical investigation showed a sacral inclination associated with pelvic tilt were also unable to perform their activities, as well as those who presented lordosis associated with pelvic incidence out of normality.

This analysis becomes interesting for a company, since from it can develop a profile for physical evaluation compatible with the requirements necessary for the desired position, avoiding sick leave of the employee, also avoiding loss of productive capacity and consequently loss to the company.

IV. FINAL CONSIDERATIONS

This study presents an analysis of the factors that lead to work withdrawal, decreasing the productivity of companies and increasing the overload on social security agencies such as INSS. To identify these factors, a database with several attributes was used, which could influence the removal of the work, through spinal pain. This was possible using the KDD process, and with it was extracted rules that show possible causes that present the highest probability of the work spreads by the column through six attributes. This result allows the company to take mitigating actions to reduce work-related distress due to low back pain (column) and preventive actions such as an Ergonomic Work Analysis (AET), specific training recommended in MTPS NR-17.

The results obtained were also determinant for the Specialized Service in Medicine and Engineering of Work Safety of the company of the petroleum industry studied, in order to comply with the ergonomic procedures of all the platforms in which it provides service, being also possible to assist in the identification of which functions and areas require greater care, such as, for example, workers who work in the area of cargo handling.

Soon, the oil company may include in its work safety procedures and admission, a rigor in the need of hiring for jobs that allow the change of positions during the work day, with the addition of pauses that allow the body not to wear out too much.

In order to do this, it is necessary to design jobs that are adaptable to the anthropometric variations of the workers, in addition to avoiding the transportation of very heavy loads and for long journeys, considering whenever necessary, the rigor in the admission examinations compatible with the position to be occupied, in order to avoid future damage to the physical and mental health of the employee and the financial company.

From this, this article demonstrates the effectiveness of the J48 classifier in assembling a decision tree with the six attributes, with 81.61% of the instances correctly classified, being a satisfactory result for decision making.

REFERENCES

[1] RAINATO, Thalita Alvarenga. (2016). A importância da medicina e segurança do trabalho preventiva. Revista Científica Aprender**,** Varginha, p.01-02, jan. 2007. Disponível em: <http://revista.valeaprender.com/?cat=2>. Acesso em: 12 mar.

[2] ALVARES, Luis Otavio. (2017). **Mineração de Dados:** Classificação: conceitos básicos e árvores de decisão. 2009. Disponível em: <www.inf.ufsc.br/~alvares/INE5644/classificacao_arvores.ppt>. Acesso em: 20 jun.

[3] ANDREATTA, Graziela (Org.). (2017). **Dor nas costas é a principal causa de afastamento do trabalho.** 2017. Disponível em: <http://trabalho.gov.br/noticias/4517-dor-nas-costas-e-a-principal-causa-de-afastamento-do-trabalho>. Acesso em: 29 junho.

[4] **ANUÁRIO BRASILEIRO DE PROTEÇÃO 2015: Edição Anual sobre Saúde e Segurança do Trabalho. (2015).** São Paulo: Proteção Publicações Ltda., Anual.

[5] CHIAVENATO, Roberto; DRUCKER, Peter. (2005). **Administração Geral.** 6. ed. Rio de Janeiro: Record, 400 p.

[6] **DOR NAS COSTAS É A PRINCIPAL CAUSA DE AFASTAMENTO DO TRABALHO. (2107).** São Paulo: Casa Nova, 16 jul. 2016. Disponível em: <http://revistacipa.com.br/dor-nas-costas-e-a-principal-causa-de-afastamento-do-trabalho/>. Acesso em: 11 junho.

[7] FALEIROS, Vicente de Paula. (2002). O trabalho da política: saúde e segurança dos trabalhadores . São Paulo : Makron Books.

[8] **FAYYAD**, U. **PIATETSKY-SHAPIRO**, G. and **SMYTH**, P. (1996). *From Data Mining to*

https://dx.doi.org/10.22161/ijaers.5.9.4 *ISSN: 2349-6495(P) | 2456-1908(O)*

Knowledge Discovery in Databases. AI magazine. P 37-54.

[9] FRAWLEY, W. J.; PIATETSKY-SHAPIRO, G.; MATHEUS, C. J. (1991). Knowledge discovery in databases: an overview. In: PIATETSKY-SHAPIRO, Gregory; FRAWLEY, William J. (Comp.). **Knowledge Discovery in Databases.** Santos: Aaai/mit Press, p. 1-27.

[10] **MARKOV**, Z.; **RUSSELL**, I. (2006). *An Introduction to the WEKA Data Mining System Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education*. P. 367 - 368. Bologna, Italy.

[11] MURASSE, C.M., TSUNODA, D.F. (2017). **Descoberta de conhecimento a partir de uma base de indicadores de desenvolvimento social utilizando WEKA**. Disponível em: http://www.inf.pucminas.br/sbc2010/anais/pdf/wcge/st01_02.pdf. Acessado em: junho.

[12] PASSOS, Ronaldo Goldschmidt/ Emmanuel. (2008). **DATA MINING - UM GUIA PRATICO.** 2005. 10 f. Administraçao-tecnologia Da Informaçao (0) - Campus, São Paulo.

[13] SCHILLING, R. S. F. (1984). **More effective prevention in occupational health practice?. Occupational Medicine**, v. 34, n. 3, p. 71-79.

[14] World Health Organization (WHO). (2017). **International Classification of Impairments, Disabilities and Handcaps (ICIDH). A manual of classification relating to the consequences of disease**. Geneve: WHO; 1980. Link: http://apps.who.int/iris/handle/10665/41003 acessado em junho/.

[15] **Serviços especializados em Engenharia e em Segurança do trabalho. (2010).** 6. ed. São Paulo: Saraiva, 1012 p.

[16] **PROTEÇÃO: Segurança do Trabalho - Estudo de Casos. (2017).** São Paulo: Sge, 2010. Disponível em: <http://www.protecao.com.br/upload/protecao_materiaarquivo/264.pdf>. Acesso em: 16 junho.