

RESUMEN

El presente trabajo de investigación desarrolla modelos matemáticos descriptivos y predictivos a partir de una data que contiene los resultados de exámenes clínicos de pacientes, la cual fue suministrada por un laboratorio de la ciudad de Quito. Este estudio se centró en el análisis de las variables que constituyen el denominado Síndrome Metabólico (Índice de Masa Corporal, Colesterol, Triglicéridos, Glucosa y Presión Arterial); además, se consideran las variables Edad, Sedentarismo y Tabaquismo. Las variables antes citadas se analizan bajo dos enfoques: el primero es la determinación de un Índice de Salud, que se basa en la categorización y posterior evaluación de las variables cuantitativas de los resultados de exámenes clínicos, de acuerdo al nivel de gravedad que es proporcionada por la literatura médica. Este índice permitirá ubicar a los pacientes en las categorías de salud Baja, Normal, Sobre y Crítica. El segundo enfoque es la determinación de factores de riesgo. En este enfoque se transforma las variables numéricas en categóricas dicotómicas de acuerdo al límite de normalidad de cada variable, es decir, dependiendo si la variable constituye o no un riesgo a la salud del paciente, para luego clasificar a la población según el número de factores de riesgo; esta clasificación permite determinar grupos poblacionales con diferentes niveles de riesgo. Adicionalmente, se aplica la técnica del escalado multidimensional, la cual permite visualizar la influencia de los diversos factores de riesgo en la salud del paciente.

PALABRAS CLAVE:

- Big Data
- Análisis multivariante
- Síndrome metabólico
- Escalado multidimensional

ABSTRACT

This research develops descriptive and predictive mathematical models about the results of clinical exams of patients from a big data base. This information was provided by a clinical laboratory located in Quito city. This study has focused on the analysis of the variables which constitute the so-called Metabolic Syndrome: body mass index, cholesterol, triglycerides, glucose and blood pressure; in addition, other variables as age, sedentary, and smoking are considered. The variables mentioned above are analyzed through two approaches: the first one is the determination of a health index based on the categorization and subsequent evaluation of the quantitative variables from the clinical examination results according to the level of severity provided by the medical literature. This index will allow locating patients in the categories of ‘low normal’, ‘normal’, ‘over normal’, and ‘critical’. The second one is the determination of risk factors. In this approach, the numerical variables are transformed to dichotomous categorical variables according to the normal limit of each one. In consequence, it depends on whether the variable constitutes a risk to the patient’s health or not, and then classifies the population according to the number of risk factors. This classification allows determining population groups with different risk levels. Additionally, it is possible to visualize the influence of the various risk factors on the patient’s health by applying a technique known as multidimensional scaling.

KEY WORDS:

- Big Data
- Multivariate analysis
- Metabolic syndrome
- Multidimensional scaling