UC Irvine
Machine Learning
Repository

Absenteeism at work

Donated on 4/4/2018

The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil.

Dataset Characteristics
Multivariate, Time-Series

Subject Area
Business

Associated Tasks
Classification, Clustering

Attribute Type
Integer, Real

 # Instances
740

 # Attributes
21

Information ⌃

Additional Information

The data set allows for several new combinations of attributes and attribute exclusions, or the modification of the attribute type (categorical, integer, or real) depending on the purpose of the research.The data set (Absenteeism at work - Part I) was used in ...

SHOW MORE ⌄

Attribute Information ⌃

1. Individual identification (ID)
2. Reason for absence (ICD).
Absences attested by the International Code of Diseases (ICD) stratified into 21 categories (I to XXI) as follows:

I Certain infectious and parasitic diseases

II Neoplasms

III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism

IV Endocrine, nutritional and metabolic diseases

V Mental and behavioural disorders

VI Diseases of the nervous system

VII Diseases of the eye and adnexa

VIII Diseases of the ear and mastoid process

IX Diseases of the circulatory system

X Diseases of the respiratory system

XI Diseases of the digestive system

XII Diseases of the skin and subcutaneous tissue

XIII Diseases of the musculoskeletal system and connective tissue

XIV Diseases of the genitourinary system

XV Pregnancy, childbirth and the puerperium

XVI Certain conditions originating in the perinatal period

XVII Congenital malformations, deformations and chromosomal abnormalities

XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified

XIX Injury, poisoning and certain other consequences of external causes

XX External causes of morbidity and mortality

XXI Factors influencing health status and contact with health services.

And 7 categories without (CID) patient follow-up (22), medical consultation (23), blood donation (24), laboratory examination (25), unjustified absence (26), physiotherapy (27), dental consultation (28).

3. Month of absence

4. Day of the week (Monday (2), Tuesday (3), Wednesday (4), Thursday (5), Friday (6))

5. Seasons (summer (1), autumn (2), winter (3), spring (4))

6. Transportation expense

7. Distance from Residence to Work (kilometers)

8. Service time

9. Age

10. Work load Average/day

11. Hit target

12. Disciplinary failure (yes=1; no=0)

13. Education (high school (1), graduate (2), postgraduate (3), master and doctor (4))

14. Son (number of children)

15. Social drinker (yes=1; no=0)

16. Social smoker (yes=1; no=0)

17. Pet (number of pet)

18. Weight

19. Height

20. Body mass index
21. Absenteeism time in hours (target)

.arff header for Weka:

@relation Absenteeism_at_work

@attribute ID {31.0, 27.0, 19.0, 30.0, 7.0, 20.0, 24.0, 32.0, 3.0, 33.0, 26.0, 29.0, 18.0, 25.0, 17.0, 14.0, 16.0, 23.0, 2.0, 21.0, 36.0, 15.0, 22.0, 5.0, 12.0, 9.0, 6.0, 34.0, 10.0, 28.0, 13.0, 11.0, 1.0, 4.0, 8.0, 35.0}
@attribute Reason_for_absence {17.0, 3.0, 15.0, 4.0, 21.0, 2.0, 9.0, 24.0, 18.0, 1.0, 12.0, 5.0, 16.0, 7.0, 27.0, 25.0, 8.0, 10.0, 26.0, 19.0, 28.0, 6.0, 23.0, 22.0, 13.0, 14.0, 11.0, 0.0}
@attribute Month_of_absence REAL
@attribute Day_of_the_week {5.0, 2.0, 3.0, 4.0, 6.0}
@attribute Seasons {4.0, 1.0, 2.0, 3.0}
@attribute Transportation_expense REAL
@attribute Distance_from_Residence_to_Work REAL
@attribute Service_time INTEGER
@attribute Age INTEGER
@attribute Work_load_Average/day_ REAL
@attribute Hit_target REAL
@attribute Disciplinary_failure {1.0, 0.0}
@attribute Education REAL
@attribute Son REAL
@attribute Social_drinker {1.0, 0.0}
@attribute Social_smoker {1.0, 0.0}
@attribute Pet REAL
@attribute Weight REAL
@attribute Height REAL
@attribute Body_mass_index REAL
@attribute Absenteeism_time_in_hours REAL

SHOW LESS

Features

ATTRIBUTE NAME	ROLE	TYPE	DESCRIPTION	UNITS	MISSING VALUES
					false

false

false

false

false

false

false

false

false

false

Rows per page (10)   1 to 10 of 21

Creators

👤   Andrea Martiniano

👤   Ricardo Ferreira

DOI

[10.24432/C5X882](10.24432/C5X882)

License

This dataset is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given.

THE PROJECT

About Us

CML

National Science Foundation

NAVIGATION

Home

View Datasets

Donate a Dataset

LOGISTICS

Contact

Privacy Notice

Feature Request or Bug Report