Preprocesare

Saturday, November 7, 2020   6:00 PM

Institutia sursa:
Procesare:
UPPERCASE

Valori finale:
X
Y
Z
Others

Encoding:
Dictionar

Sex:
Procesare:
UPPERCASE
F -> FEMININ
M -> MASCULIN

Valori finale:
MASCULIN
FEMININ
Others

Encoding:
Dictionar

Varsta:
Procesare:
LOWERCASE
Encoding:
Numar intre 0 -> 200 - OK
Daca contine "an" -> preluare numar dinaintea "an", reverse
Altfel daca contine "lun" || "sap" || "zi" || "or" || "nou -> 0
Altceva -> eliminat

Data debut:
Procesare:
YYYY-MM-dd
Dd.MM.YYY
Dd(./,)MM(./,)YYYY

Encoding:

>>> dt = datetime.datetime(2020, 7, 11)

>>> tt = datetime.datetime.timetuple(dt)

>>> epoch = calendar.timegm(tt)

Simptome declarate:
Procesare & encoding:
LOWERCASE
Daca contine:
"tuse", "dureri musculare", "dispnee", "astenie", " febr", "fatigabilitate", "temp", "cefalee" , "diare",
"covid", "respirat", "transpirati",
"mialgi", "friso", "frisoane", "amete", "disfagi", "afazie", "scaun", "vars", "toracic", "vertij", "palpitatii",
"inapetenta", "odinofagie", "gust", "miros", "corona", "pozitiv", "nazal", "sars", "cov", "somno",
"pneumo", "pulmo" -> 1

Daca contine "cov" si nu contine "asimpto" -> 1

Altele -> 0

Nimic -> -1

Data internare:
- **Data debut**

Simptome raportate la internare:
- **Simptome declarate**

Diagnostic si semne de internare
- **Simptome declarate**

Istoric de calatorie:
LOWERCASE
Contine "nu", "mu", "neag", "ne", "fara", null -> 0
Este "0" -> 0
Altele -> 1

Mijloace de transport folosite:
Ignoram

Confirmare contact cu o persoana infectata:
LOWERCASE
Contine "da", "focar", "pozitiv", "susp", "1"

Data rezultat testare:
Ignoram sau Data debut

Rezultate testare:
LOWERCASE
contine "pozitiv" -> 1
Altfel -> 0