

UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERÍA

Inteligencia Artificial

Proyecto
Entrega 2

Nombres
Ana María Montañez Becerra
Joanny Torres Cardona

Docente
Raul Ramos Pollan

Documentos
1.152.224.184
1.152.218.692

Medellín - Antioquia
2023

Descripción del progreso alcanzado:

Lo primero que se hizo fue ver el el dataset y tomamos la decisión de agrupar los datos por paciente y sacarle poner el promedio para cada una de los datos de las columnas como se muestra en la *Figura 1* y *Figura 2*. Antes de esto se había hecho una exploración de los datos graficando los que se pueden usar para relacionar el estado físico del paciente y suhábitos diarios.

```
from progressbar import progressbar as pbar
%matplotlib inline
d= pd.read_csv ('/content/rais_anonymized/csv_rais_anonymized/daily_fitbit_sema_df_unprocessed.csv')
d.head() #El dataset tiene 7410 columnas y 63 columnas por lo tanto cumple con los requisitos
```

| Unnamed: 0 | id | date | nightly_temperature | nremhr | rmssd | spo2 | full_sleep_breathing_rate | stress_score | sleep_points_percentage |
|------------|--------------------------|------------|---------------------|--------|---------|------|---------------------------|--------------|-------------------------|
| 0 | 621e2e8e67b776a24055b564 | 2021-05-24 | 34.137687 | 57.432 | 89.603 | NaN | 14.8 | 78.0 | 0.833333 |
| 1 | 621e2e8e67b776a24055b564 | 2021-05-25 | 33.794544 | 57.681 | 94.303 | NaN | 15.8 | 80.0 | 0.833333 |
| 2 | 621e2e8e67b776a24055b564 | 2021-05-26 | 34.611011 | 57.481 | 119.212 | NaN | 14.6 | 84.0 | 0.966667 |
| 3 | 621e2e8e67b776a24055b564 | 2021-05-27 | 34.408304 | 57.493 | 111.709 | NaN | 14.8 | 82.0 | 0.933333 |
| 4 | 621e2e8e67b776a24055b564 | 2021-05-28 | 34.178922 | 56.750 | 103.034 | NaN | 15.2 | 81.0 | 0.866667 |

5 rows x 63 columns

Figura 1. dataset crudo

```
newData=d.groupby('id').mean()
newData
```

| id | nightly_temperature | nremhr | rmssd | spo2 | full_sleep_breathing_rate | stress_score | sleep_points_percentage |
|--------------------------|---------------------|-----------|------------|-----------|---------------------------|--------------|-------------------------|
| 621e2e8e67b776a24055b564 | 34.315168 | 56.571484 | 100.410210 | NaN | 15.270968 | 76.317460 | 0.7951 |
| 621e2eaf67b776a2406b14ac | 34.252140 | 68.988272 | 25.992988 | NaN | 14.767500 | 74.731707 | 0.7421 |
| 621e2ed667b776a24085d8d1 | 33.998974 | 72.814016 | 26.519625 | NaN | 17.680000 | NaN | NaN |
| 621e2ef567b776a24099f889 | 31.975482 | NaN | NaN | 97.700000 | NaN | NaN | NaN |
| 621e2efa67b776a2409dd1c3 | 34.167044 | NaN | NaN | 95.971212 | NaN | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... |

Figura 2. dataset agrupado usando el id de cada paciente

Después de esto se ven la cantidad de datos faltantes para cada una de las columnas y se observa que falta gran porcentaje de datos (*Figura 3*) por lo tanto es necesario tomar la decisión, que puede afectar en gran medida el resultado de este proyecto, y es saber si poner la media, ceros o la normal en los datos faltantes, por lo tanto usamos la función usada en clase para ver cómo afectará esto al promedio total de los datos y la distribución de estos (*Figura 4*), al final se opta por reemplazarlos por datos normales ya que no afecta tanto el promedio y la distribución de los datos. Por último se está viendo cuales pueden ser los mejores hiperparametros a implementar.

```
''' Cantidad de valores faltantes en cada columna '''
k=newData.isna().sum()
k[k!=0]
```

| | |
|----------------------------------|----|
| nightly_temperature | 7 |
| nremhr | 28 |
| rmssd | 28 |
| spo2 | 43 |
| full_sleep_breathing_rate | 28 |
| stress_score | 34 |
| sleep_points_percentage | 34 |
| exertion_points_percentage | 34 |
| responsiveness_points_percentage | 34 |
| daily_temperature_variation | 10 |
| scl_avg | 69 |
| sleep_duration | 2 |
| minutesToFallAsleep | 2 |
| minutesAsleep | 2 |
| minutesAwake | 2 |
| minutesAfterWakeup | 2 |
| sleep_efficiency | 2 |
| sleep_deep_ratio | 10 |
| sleep_wake_ratio | 10 |
| sleep_light_ratio | 10 |
| sleep_rem_ratio | 10 |
| min_goal | 10 |

Figura 3. Datos faltantes del data frame agrupado

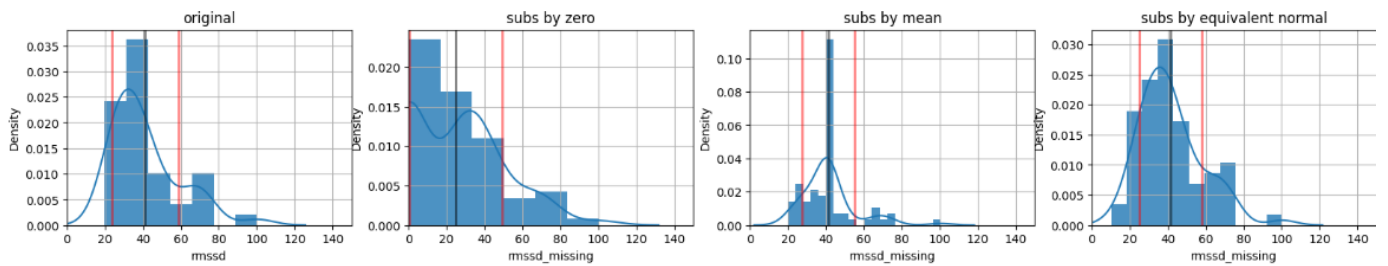


Figura 4. Gráfica cuando se reemplaza por 0, media o normal.