

Primera entrega de proyecto

Por:

Ana María Montañez Becerra
Joanny Torres Cardona

Materia:

Introducción a la inteligencia artificial

Profesor:

Raúl Ramos Pollan



1. Planteamiento del problema

El sedentarismo y la inactividad física han aumentado los últimos años, tanto en adultos, jóvenes y niños. Esto trae consigo el aumento de obesidad y la aparición de algunas enfermedades crónicas como diabetes y enfermedades cardiovasculares hasta la muerte prematura [1]. En la actualidad existe tecnología que permite tener idea de alguna información que nos da idea del estado de salud, estos datos pueden ser: latidos por minuto, porcentaje de saturación de oxígeno en sangre, frecuencia respiratoria y cardíaca. Y al mismo tiempo permite tener un control sobre el tiempo invertido en ejercicio físico y el número de pasos dados durante el día [2].

Dada información de un grupo de personas obtenida mediante un reloj inteligente y encuestas, se quiere evaluar la salud mental, física y estado de ánimo, teniendo en cuenta valores como frecuencia cardíaca y el porcentaje de saturación de oxígeno se propone diseñar un algoritmo que clasifique personas que puedan ser bradicárdicas o taquicárdicas y quienes estén sufriendo de hipoxia. Por otra parte, con el fin de hallar una relación entre los hábitos que tiene relacionados a las horas de sueño, horas de ejercicio, la escala de estrés que maneja generado por actividades como el trabajo o estudio, con el estado de salud, con valores como temperatura, latidos por minuto, frecuencia respiratoria en el día y noche y el ritmo circadiano.

2. Dataset

A partir de dataset “*Lifesnaps Fitbit data set*” tomado de [Lifesnaps Fitbit dataset | Kaggle](#) se selecciona los datos tomados diariamente para usarlos en este proyecto. Para la construcción de este dataset se contó con la participación de 71 sujetos, el tamaño del dataset a trabajar es de 7410 filas y 63 columnas. Algunos de los datos tomados del usuario por medio del reloj inteligente son:

- **nightly_temperature**: temperatura durante la noche
- **nremhr**: National registry of emergency medical Technicians
- **rmssd**: Medida más relevante y precisa de la actividad del sistema nervioso autónomo a corto plazo
- **spo2**: porcentaje de saturación de oxígeno
- **calories**: Calorías
- **bpm**: Latidos por minutos

Las que el usuario lleno mediante una encuesta son:

- **stress_score**: Clasificación al nivel de estrés
- **sleep_points_percentage**: Porcentaje de puntuación al dormir
- **HAPPY**: Clasificación de nivel de sentimiento de felicidad
- **SAD**: Clasificación de nivel de sentimiento de tristeza
- **TENSE/ANXIOUS**: Clasificación de nivel de sentimiento de ansiedad
- **TIRED**: Clasificación de nivel de sentimiento de cansancio
- **GYM**: Asistencia al gimnasio

Por cada uno de los pacientes existe más de un dato por día, por lo tanto se evaluará la posibilidad de hacer promedio de los datos para cada uno de los pacientes para así comparar los datos de diferentes pacientes y poder hacer estadística descriptiva teniendo en cuenta el género y la edad.

3. Métricas

Se desea evaluar métricas de desempeño como sensibilidad, precisión, exactitud y especificidad, a partir de la matriz de confusión, con el fin de evaluar el comportamiento del modelo debido a que, al tener conocimiento del valor de dichas métricas, se puede observar qué tan confiable es el modelo que se creó.

Para lo anterior, se puede partir de la matriz de confusión, la cual es una herramienta que permite visualizar con mayor facilidad el desempeño de un algoritmo, en donde las columnas de dicha matriz representan las predicciones de cada clase, mientras que las filas hacen alusión a las instancias en la clase real, y con ello, se pueden derivar las demás métricas de desempeño mencionadas (sensibilidad, especificidad, exactitud y precisión); por ende, esta matriz se muestra a continuación:

True Class	Predicted Class	
	True	False
True	True Positive (TP)	False Negative (FN)
False	False Positive (FP)	True Negative (TN)

En donde:

- *True Positive (TP)*: Hace referencia al resultado de una prueba que indica con veracidad la presencia de cierta condición o característica.
- *False Negative (FN)*: Este hace alusión al resultado de una prueba que indica de forma errónea que cierta condición o característica se encuentra ausente.
- *False Positive (FP)*: Es aquel resultado de una prueba que denota erróneamente que una condición o característica en particular se encuentra presente.
- *True Negative (TN)*: Dicho resultado de prueba revela de manera correcta que una condición o característica se encuentra ausente.

Ahora, como se mencionó previamente, con dichos resultados se pueden obtener las métricas de desempeño, como lo son, sensibilidad, especificidad, precisión y exactitud, entonces:

- **Sensibilidad o *true positive rate* (TPR):**

$$TPR = \frac{TP}{TP + FN}$$

- **Especificidad o *true negative rate* (TNR):**

$$TNR = \frac{TN}{TN + FP}$$

- **Precisión o *positive predictive value* (PPV):**

$$PPV = \frac{TP}{TP + FP}$$

- **Exactitud o *accuracy* (ACC):**

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

4. Desempeño

Si a partir de los resultados obtenidos, el modelo presentado no presenta una precisión del 95% y una exactitud del 90% al momento de realizar el diagnóstico de pacientes con padecimientos como la bradicardia, taquicardia o hipoxia, no se ha de llevar a cabo la implementación de este modelo dado que no presentaría la confiabilidad esperada para la predicción; en consecuencia, incluso con los datos recolectados, no sería posible alertar a la persona de una posible alteración en su salud y que consulte de manera oportuna con un especialista, según sea el caso.

5. Bibliografía

[1] J. Ildefonso (2019). "Sedentarismo, la enfermedad del siglo XXI" [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0214916819300543>

[2] S. Yfantidou, C. Karagianni. et al (2023). "Lifesnaps Fitbit dataset" [En línea]. Disponible en: [Lifesnaps Fitbit dataset | Kaggle](#)

[3] F. Demir, "Deep autoencoder-based automated brain tumor detection from MRI data," Artificial Intelligence-Based Brain-Computer Interface, pp. 317–351, 2022, doi: <https://doi.org/10.1016/b978-0-323-91197-9.00013-8>.

[4] J. Niu et al., "Analysis of sensitivity and specificity: precise recognition of neutrophils during regeneration of contused skeletal muscle in rats," Forensic Sciences Research, vol. 7, no. 2, pp. 228–237, Mar. 2020, doi: <https://doi.org/10.1080/20961790.2020.1713432>.

[5] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: <https://doi.org/10.1016/j.patrec.2005.10.010>.