**Modelado Multicanal de la Autopercepción del Rendimiento en ESports: Un Enfoque Integrado con Aprendizaje Supervisado, Agrupamiento Jerárquico y AutoML**

Autores: Francisco Antonio Enríquez Cabrera1, Pablo Tadeo Ríos Gallardo1.

Afiliación: Universidad Autónoma de Nuevo León, Facultad de Organización Deportiva1.

Francisco Antonio Enríquez Cabrera, 8vo semestre,

francisco.enriquezc@uanl.edu.mx, 656 832 7606

Mesa 3: Ciencias Aplicadas al Deporte

Participación estudiantil – Proyecto de investigación

En el campo de los deportes electrónicos (ESports), modelar la autopercepción del rendimiento a partir de señales fisiológicas multicanal supone un avance esencial para diseñar sistemas de retroalimentación en tiempo real y optimizar la intervención psicofisiológica de los jugadores. Para ello, empleamos 220 registros tras imputación (KNNImputer, k=3) de un total de 48 variables que incluyeron medias y desviaciones estándar de electromiografía (EMG), frecuencia cardiaca, conductancia de la piel (GSR), temperatura cutánea, acelerómetros (IMU) en casco, silla y manos, movimientos de ratón y eye-tracking, además de índices derivados como razón EMG/teclado, índice de estrés (heart\_rate + GSR – SpO₂) y diferencia GSR–SpO₂. Un análisis de correlación de Pearson reveló que la evaluación subjetiva del desempeño (“report\_performance\_evaluation”) se asoció positivamente con heart\_emg\_ratio (r = +0.299) y de forma inversa con emg\_mean (r = –0.371), spo2\_mean (r = –0.355), imu\_right\_hand\_mean (r = –0.340) e imu\_chair\_seat\_mean (r = –0.315), sugiriendo que mayor esfuerzo muscular y menor oxigenación influyen en juicios más negativos.

Al comparar modelos supervisados en un split train/test, RidgeCV obtuvo R² = 0.143 y MSE = 1.157, Random Forest alcanzó R² = 0.828 y MSE = 0.233, y XGBoost R² = 0.823 y MSE = 0.239. Al validar XGBoost de forma global en 5 folds, se observó una caída a R² = 0.708 ± 0.261, mostrando sensibilidad a la variabilidad interindividual. ElasticNetCV presentó R² = –0.280 ± 0.308, MAE = 0.927 ± 0.168 y correlación de Kendall τ = 0.213 ± 0.101, mientras que XGBoost base logró MAE = 0.838 ± 0.079 y τ = 0.271 ± 0.067. A través de Optuna (20 ensayos), el pipeline Scale + SelectKBest(k=8) + XGBRegressor optimizó hiperparámetros a n\_estimators = 200, max\_depth = 3, learning\_rate ≈ 0.086, subsample ≈ 0.935 y colsample\_bytree ≈ 0.602, alcanzando MAE CV interno de 0.784, aunque su R² en validación cruzada creció solo a 0.729 ± 0.241. Un GridSearchCV exploratorio determinó que (learning\_rate=0.1, max\_depth=4, n\_estimators=200) era la mejor combinación para XGBoost, con R² = 0.7405.

La comparativa de pipelines sobre datos aumentados arrojó que ElasticNetCV se mantuvo en R² = –0.280, MAE = 0.927, τ = 0.213, XGBoost base en R² = 0.140, MAE = 0.838, τ = 0.271, y XGBoost Optuna en R² = 0.188, MAE = 0.788, τ = 0.376. El análisis de importancia por permutación para XGBoost Optuna señaló como principales predictores las características x7 (0.413), x6 (0.379), x0 (0.346), x2 (0.317) y x5 (0.299), que corresponden a señales fisiológicas clave como EMG medio y movimiento de mano. Para mitigar la varianza de cada estimador, diseñamos un StackingRegressor que combinó ElasticNetCV, XGBoost base y XGBoost Optuna, usando RidgeCV como meta-modelo: en validación cruzada global obtuvo R² = 0.161, MAE = 0.813 y τ = 0.305. La importancia por permutación del stacking reveló que emg\_mean (0.438), imu\_left\_hand\_mean (0.326), imu\_right\_hand\_mean (0.194) y heart\_emg\_ratio (0.192) fueron los rasgos más informativos.

Al transformar la evaluación subjetiva en un problema binario (“Bajo” <3 vs. “Alto” ≥3), un XGBClassifier alcanzó accuracidad global de 0.91, con F1-score de 0.59 para “Bajo” y 0.95 para “Alto”, aunque la clase “Bajo” solo contaba con 11 muestras frente a 63 de “Alto”. El análisis de subestimación frente a sobreestimación mostró 100 % de precisión y recall, pero la escasa representación de la clase minoritaria advierte cautela. Al segmentar los datos por rol funcional, el error absoluto medio del stacking osciló entre 0.063 (rol top) y 0.0769 (adc); entrenando pipelines XGBoost Optuna por rol en split test, logramos R² entre 0.809 (top) y 0.999 (mid), con MSE entre 0.127 y 0.002, evidenciando que los patrones fisiológicos que sustentan la autopercepción varían según la función desempeñada.

En el ámbito no supervisado, comparamos DBSCAN (silhouette = 0.154), Gaussian Mixture Models (silhouette = 0.415), KMeans k=4 (silhouette = 0.477), Ward jerárquico y HDBSCAN (silhouette = 0.513). Un ANOVA entre clusters de KMeans mostró diferencias significativas en la evaluación real (F = 51.731, p < 6.77×10⁻²⁸), aunque la habilidad de estos agrupamientos para mejorar la predicción subjetiva resultó limitada. La validación de KMeans arrojó índices de Calinski–Harabasz = 344.92 y Davies–Bouldin = 0.8679, mientras que HDBSCAN identificó 24 clústeres, sugiriendo heterogeneidad fisiológica compleja. Finalmente, el análisis de correlación entre el error absoluto del stacking y los perfiles fisiológicos indicó que gsr\_mean (r = 0.345), stress\_index (r = 0.345) y eye\_tracker\_mean (r = 0.188) se asocian con mayor imprecisión en la autopercepción, lo que refuerza la idea de que la inestabilidad autonómica y sensorial introduce ruido en el juicio subjetivo.

En los 20 primeros casos de prueba, la diferencia absoluta entre valor real y predicho por el stacking osciló entre 0.115 y 0.165 puntos en la escala de 1 a 5, confirmando un MAE práctico de ~0.81 unidades. Todos los modelos y figuras —incluidos los gráficos ROC, curvas precisión-recall y dendrogramas— fueron guardados para replicación. En conjunto, estos resultados demuestran que, aunque los modelos supervisados pueden explicar más del 80 % de la varianza en condiciones de test controlado, su generalización global ronda el 70–75 %, lo que exige más datos o estrategias de transferencia. Se confirma, no obstante, que variables como EMG medio, GSR y movimiento de manos son predictores robustos de la autopercepción del rendimiento, y que entrenar modelos por rol funcional optimize la precisión. Estos hallazgos sientan las bases para sistemas de entrenamiento personalizados y de retroalimentación psicofisiológica en ESports, con aplicación potencial en modalidades de voz y expresiones faciales y validación en entornos competitivos reales.

**Palabras clave:** ESports, ciencia de datos, inteligencia artificial