

Evaluación de modelos de Machine Learning para explicar la concentración a partir de estados emocionales

Alfredo Emir Puente Medrano, Danaé Sánchez Gutiérrez, Oscar Arreola Alvarado, Esteban Sierra, Baccio, Edsel De Jesus Cisneros Bautista, Alejandra Coeto Sánchez, Rodrigo Garcia Torres

Introducción

La creciente adopción de plataformas digitales en educación ha impulsado nuevas formas de evaluar la experiencia de aprendizaje más allá de los métodos tradicionales de autoreporte. En este contexto, la Multimodal Learning Analytics (MMLA) integra señales fisiológicas, neuronales y expresiones faciales para analizar estados cognitivos durante el aprendizaje. Utilizando el dataset NPFC-Test 23A, que recopila datos sincronizados de 41 participantes en interacciones humano-computadora, este trabajo estudia la relación entre emociones y concentración. En particular, esta investigación busca responder: ¿hasta qué punto las emociones detectadas automáticamente y autodefinidas explican la variación en la concentración del estudiante durante un test?

Metodología

Para el análisis de ciencia de datos se siguió la siguiente pipeline, mediante la cual se limpiaron los datos, se calculó la concentración y realizó diversos modelos predictivos para la concentración.

01 EDA + Limpieza 47,905 filas 43 columnas con NA 97% HeadBandOn (Muse) 97.3% Face_Detection	02 Índice EEG Índices de concentración con Muse band Agrupación por persona y task	03 KMeans 9 perfiles demograficos k via codo + silhouette PCA para visualizar	04 Regresión Gradient Boosting Random Forest Regresion Lineal Validacion GroupKFold
---	---	--	--

Índice de Concentración

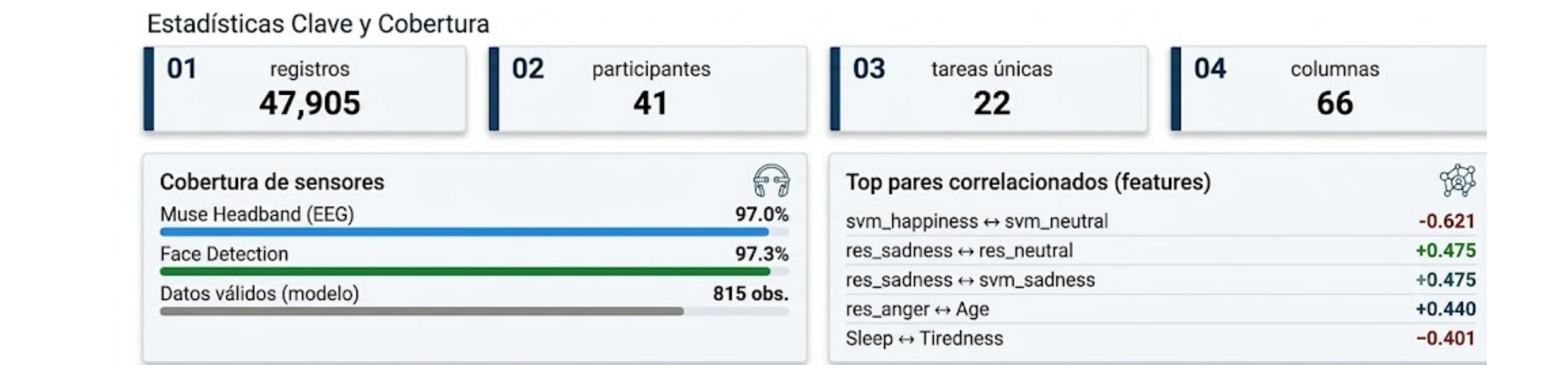
Para determinar la concentración, se calculó un índice de concentración basado en un estudio de la NASA ($\beta / (\alpha + \theta)$) que prioriza el enfoque activo y penaliza la relajación. Estos datos se agruparon por persona y tarea, comparando cada medición con la media basal de cada participante, clasificando como "concentrado" todo valor superior al promedio personal y obteniendo un porcentaje por cada tarea.

Muse Headband

Delta: sueño profundo. Theta: somnolencia. Alpha: relajación, calma. Beta: Enfoque, pensamiento activo. Gamma: procesamiento complejo

EDA y Correlaciones

Las emociones faciales de resmasknet superan al SVM en señal predictiva, con resmasknet_neutral como la más correlacionada ($r = -0.131$). La alta multicolinealidad entre emociones limita el poder predictivo individual de cada feature.



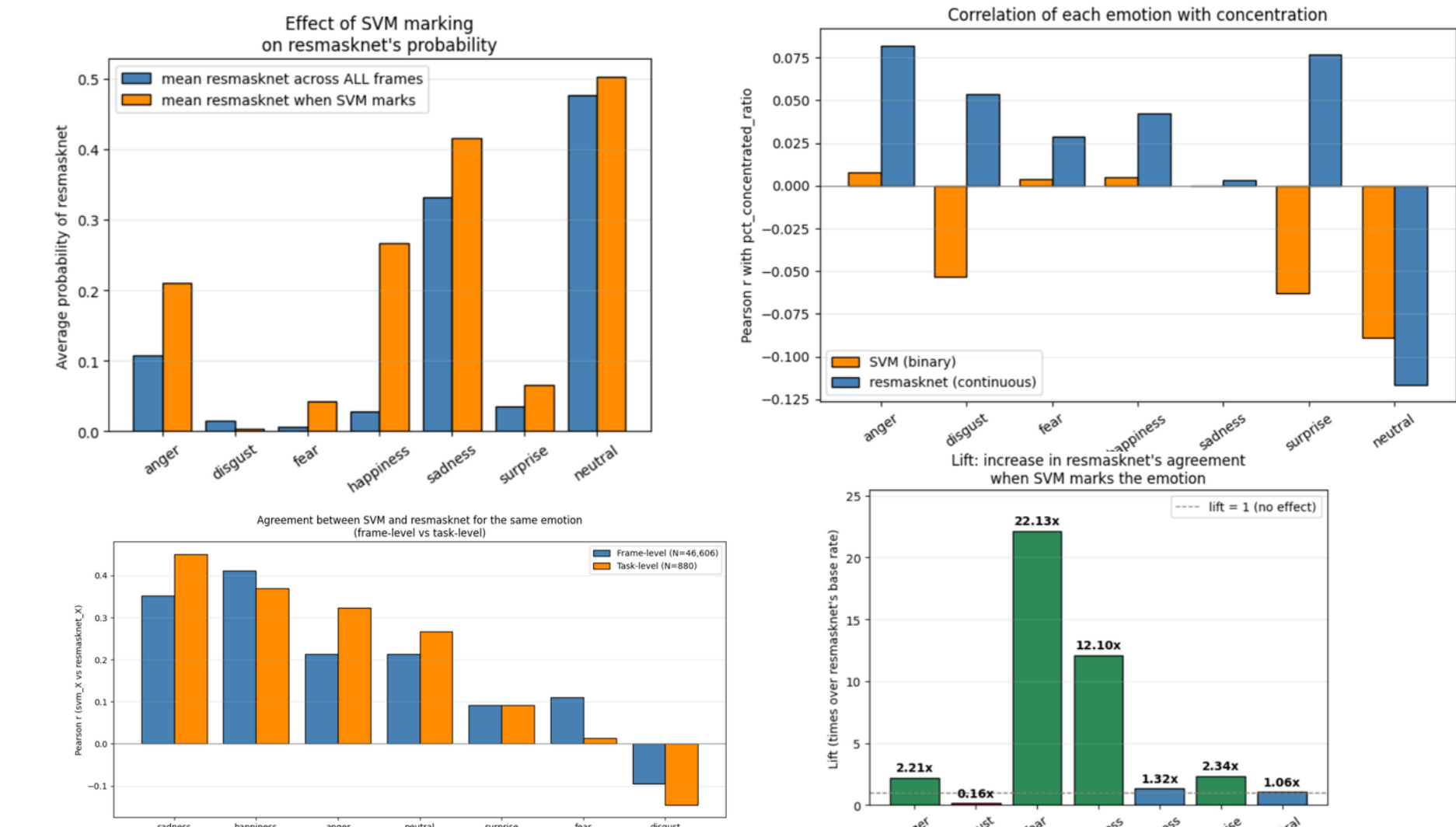
KMeans

Se aplicó K-Means para identificar perfiles demográficos dentro de la muestra. El método del codo sugirió inicialmente una solución de 9 clusters; sin embargo, la visualización mediante PCA mostró un solapamiento considerable entre varios grupos, dificultando su interpretación. Tras evaluar distintas configuraciones, se seleccionó $k = 4$, ya que permitió obtener perfiles más claros y representativos. La solución final obtuvo un silhouette de 0.318 y permitió identificar grupos diferenciados principalmente por edad y promedio de sueño.

SVM vs Resmasknet

Ambos detectan emociones sobre el mismo frame, pero el SVM marca varias emociones a la vez como binarios, mientras que resmasknet entrega siete probabilidades continuas que suman uno.

Al comparar los dos modelos, solo coinciden en happiness y sadness, disgust se contradice y el SVM satura neutral como categoría refugio. Aun así, resmasknet iguala o supera al SVM en las siete emociones en correlación con la concentración.



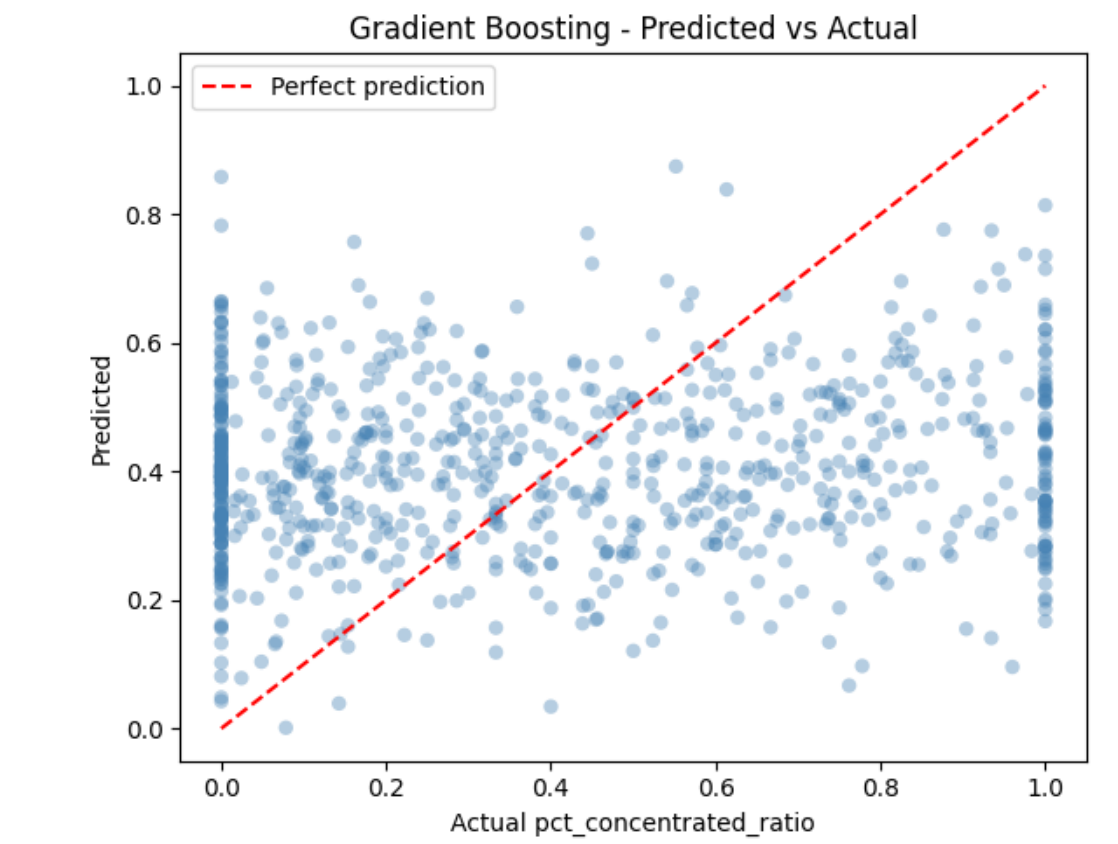
Conservar resmasknet y descartar el SVM: las decisiones binarias colapsan la intensidad de la expresión y saturan categorías refugio. La prob continua preserva los matices que dominan el ranking de importancia en GB y RF.

Resultados

Regresiones

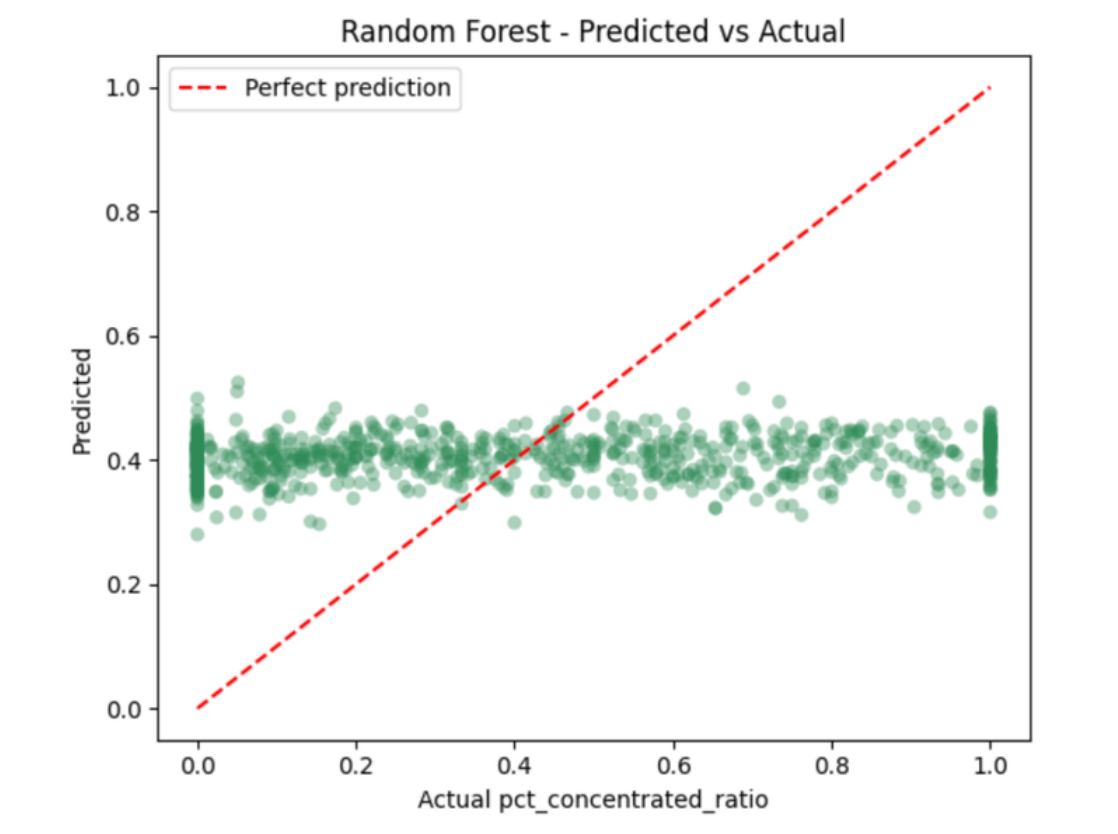
Gradient Boosting

Validación GroupKFold por sujeto (38 personas). **MAE = 0.31, RMSE = 0.36, Test $R^2 = -0.10 (\pm 0.14)$.** El R^2 negativo indica que el modelo no supera a predecir la media. Con **Train $R^2 = 0.79$ vs Test $R^2 = -0.10$** , hay sobreajuste severo (**gap = 0.90**): memoriza pero no generaliza a sujetos nuevos. Features más influyentes: **EDA, felicidad facial (resmasknet) y temperatura.**



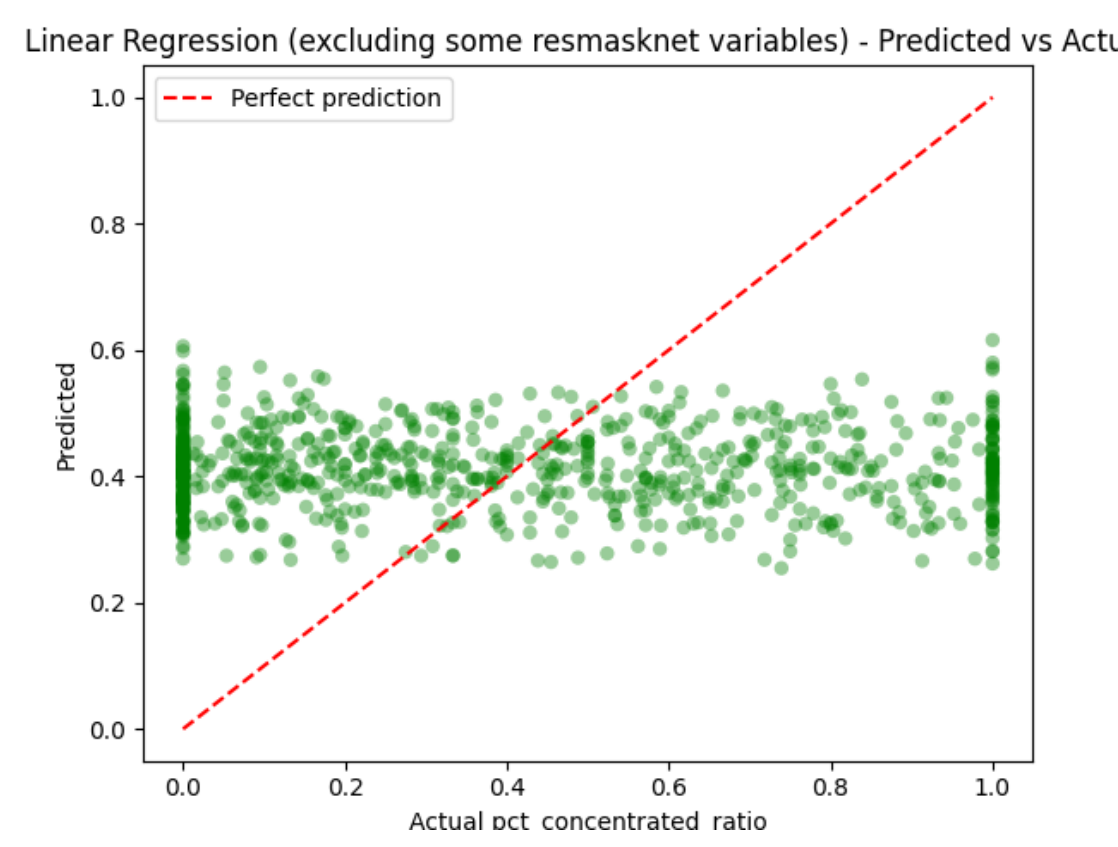
Random Forest

Validación GroupKFold por sujeto (38 personas). **MAE = 0.30, RMSE = 0.34, Test $R^2 = -0.005 (\pm 0.04)$.** El R^2 es prácticamente cero, equivalente a predecir la media global. Con **Train $R^2 = 0.11$ vs Test $R^2 = -0.005$** , el gap es bajo (**0.11**). El modelo está bien regularizado y no sobreajusta, pero el techo predictivo se queda en cero. Confirmando que los features no tienen suficiente señal para generalizar entre sujetos. Features más influyentes: **neutral y sorpresa faciales (resmasknet), EDA y temperatura.**



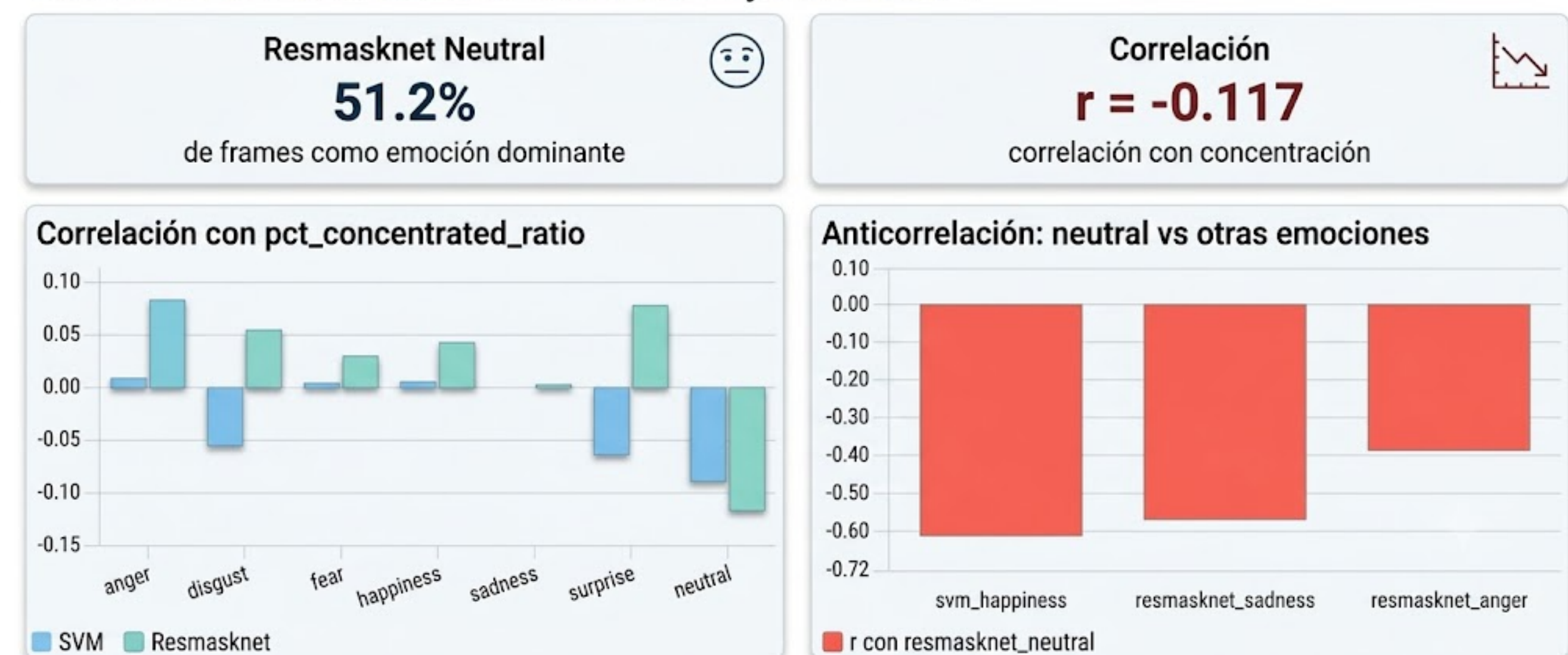
Regresión Lineal

El rendimiento del modelo es deficiente, con un R^2 promedio de **-0.0659 +/- 0.0857**, lo que indica que funciona peor que simplemente predecir la media de la variable objetivo. **MAE es 0.3073 +/- 0.0471** y el (**RMSE**) es **0.3483 +/- 0.0425**. El gráfico confirma el ajuste deficiente y los coeficientes de características, que aún muestran cierta inestabilidad.



Expresión Neutral y Concentración (EEG)

La expresión neutral fue la más frecuente en el dataset (51.2% de frames). Ambos detectores muestran una correlación negativa débil con el índice de concentración (resmasknet: $r = -0.117$; SVM: $r = -0.089$), sugiriendo que mayor neutralidad facial coincide con menor activación beta.



Discusión y Conclusiones

Las expresiones faciales detectadas por resmasknet resultaron los predictores más informativos de la concentración, por encima de las señales del clasificador SVM, las variables demográficas y los autorreportes estáticos (edad, sueño, estrés), que aportaron poca capacidad explicativa por su naturaleza estática; las emociones happiness y neutral destacaron como las más relevantes, posiblemente vinculadas a estados de flow o atención focalizada, consistente con la literatura previa (Bosch et al., 2015; Mitchell, 2023). No obstante, los valores de R^2 negativos bajo validación GroupKFold por sujeto indican que los modelos no generalizan a participantes no vistos, reflejo de la fuerte heterogeneidad inter-sujeto y del tamaño muestral modesto ($n=38$), patrón que coincide con Serrano-Mamolar et al. (2021), quienes favorecen enfoques intra-sujeto. En conjunto, las modalidades objetivas (faciales y fisiológicas) caracterizan la concentración mejor que las subjetivas, respaldando la hipótesis inicial, y el bajo R^2 inter-sujeto señala que los enfoques personalizados por participante son la vía más prometedora. Trabajo futuro podría ser orientado a modelos con normalización por baseline individual y arquitecturas secuenciales.