

Proyecto Machine Learning – Movimiento ocular

Andrés Darío Higuera Pérez *
Jasmin Juliana Jaramillo Tapias *
Juan Guillermo Preciado Zapata *

* Universidad de Antioquia

ResumenEl presente trabajo aborda el estudio y análisis de distintos modelos de aprendizaje automático aplicados a problemas de predicción y clasificación, destacando la importancia de estos enfoques en la transformación de datos en conocimiento útil y su creciente relevancia en la toma de decisiones basadas en información. A través de las etapas de exploración, entrenamiento y validación de modelos, se demuestra cómo las técnicas de machine learning permiten identificar patrones complejos y generar respuestas precisas en diversos contextos de la vida cotidiana, científica e industrial. En este marco, se desarrolla un caso de aplicación sobre la base de datos *Eye Movements*, orientado al análisis de patrones visuales y su relación con la relevancia de la información leída. Asimismo, se realiza un análisis comparativo entre los métodos empleados, reflexionando sobre su capacidad para generalizar y adaptarse a distintos tipos de datos y escenarios, resaltando el papel del aprendizaje automático como herramienta fundamental para la innovación tecnológica y la comprensión de fenómenos complejos.

Índice de términos: Machine Learning, Predicción, Clasificación, Validación, Inteligencia Artificial, Eye Movements.

1. INTRODUCCIÓN

En los últimos años, el análisis de movimientos oculares (eye tracking) ha adquirido gran relevancia en el estudio del comportamiento humano, la atención visual y los procesos cognitivos durante la lectura o navegación en entornos digitales. La información derivada de los movimientos oculares permite inferir qué partes de un texto o imagen captaron la atención del observador, así como medir el nivel de comprensión, interés o relevancia percibida. El presente trabajo aborda el problema de clasificar la relevancia de oraciones leídas a partir de los movimientos oculares registrados durante una tarea de lectura, utilizando la base de datos Eye Movements disponible en OpenML. Este problema se enmarca en la categoría de aprendizaje supervisado, donde el objetivo es predecir una etiqueta de clase asociada a cada palabra leída: irrelevante, relevante o correcta. El uso de técnicas de Machine Learning en este contexto resulta especialmente valioso, ya que permite automatizar el análisis de señales complejas y multidimensionales, capturadas mediante dispositivos de eye-tracking, con el fin de construir modelos capaces de inferir patrones cognitivos y conductuales. El desarrollo de este tipo de soluciones tiene aplicaciones en campos como la evaluación de interfaces, la neurociencia cognitiva, la educación y la publicidad digital.

2. DESCRIPCIÓN DEL PROBLEMA

El problema planteado consiste en clasificar palabras leídas según su nivel de **relevancia** para una pregunta dada, basándose en las características obtenidas de los *movimientos oculares* de distintos sujetos durante la lectura.

Una solución basada en **aprendizaje automático** es adecuada porque permite modelar relaciones complejas entre las variables (características oculares) y los niveles de relevancia de las palabras leídas, automatizando la clasificación con mayor precisión y eficiencia que los métodos analíticos comunes.

El conjunto de datos **Eye Movements** contiene instancias que representan palabras individuales observadas en múltiples tareas de lectura. Cada tarea o “asignación” consiste en una pregunta y diez oraciones (títulos de noticias). De estas oraciones, una corresponde a la respuesta correcta (*Correcta*), cuatro son *Relevantes*, pero no responden directamente a la pregunta, y cinco son *Irrelevantes*.

El conjunto de datos está identificado en OpenML con el ID = 1044, está conformado por 10936 registros y 28 variables que describen el comportamiento ocular de distintos participantes durante tareas de lectura. La primera columna (**lineNo**) identifica la línea del texto, la segunda (**assgNo**) corresponde al número de asignación o tarea experimental, y las columnas 3 a 24 contienen las características principales de los movimientos oculares. Dentro de este grupo se incluyen variables numéricas –que cuantifican aspectos temporales, espaciales y fisiológicos, como duraciones de fijaciones, longitudes de sacadas y variaciones pupilares– y variables categóricas, que representan estados o posiciones específicas durante el proceso de lectura, como las ubicaciones de fijación o la presencia de regresiones.

Las columnas 25 a 27 recogen información contextual y estructural del texto, como el número total de palabras en la oración, la posición del título y la ubicación ordinal de cada

palabra. Finalmente, la columna 28 (**label**) corresponde a la variable objetivo del modelo supervisado, con tres categorías de clasificación: 0 = Irrelevante, 1 = Relevante y 2 = Correcta. Las características y su descripción se resumen en la **Tabla 1**.

El conjunto de datos se encuentra preprocesado y estructurado en formato **ARFF** (*Attribute-Relation File Format*), compatible con la herramienta Weka y ampliamente utilizado en entornos de aprendizaje automático. Este formato permite representar de manera explícita los nombres, tipos y valores de cada atributo, así como la variable objetivo de clasificación, lo que garantiza una correcta interpretación por parte de diferentes librerías y entornos de análisis, para nuestro caso Python.

El conjunto de datos no presenta valores faltantes en ninguna de sus variables, lo que facilita su utilización directa en los procesos de análisis y modelado sin requerir estrategias de imputación o limpieza adicional. Dado que el objetivo es predecir una categoría discreta, el problema se enmarca en una clasificación supervisada multiclase, donde se busca representar el nivel de relevancia de las palabras leídas. El modelo debe aprender a relacionar las características de los movimientos oculares con las clases objetivo (irrelevante, relevante y correcta).

Este paradigma resulta apropiado porque permite capturar relaciones no lineales y de alta dimensionalidad entre las variables, ofreciendo mayor capacidad predictiva y de generalización frente a los métodos tradicionales. Entre los algoritmos adecuados para abordar este tipo de problema se incluyen **Support Vector Machines (SVM)**, **Random Forests**, **Redes Neuronales** y **Gradient Boosting**.

Nota: Los valores en las columnas marcadas con un asterisco (*) son los mismos para todas las apariciones de la palabra.

Tabla 1. Descripción de las variables del conjunto de datos *Eye Movements*

Nº	Variable	Tipo de dato	Descripción
1	lineNo	Numérica (int64)	Número de línea del texto dentro de la tarea de lectura
2	assgNo	Numérica (int64)	Identificador de la asignación o tarea experimental (pregunta con 10 oraciones)
3	fixcount	Numérica (uint8)	Número total de fijaciones realizadas en la palabra
4*	firstPassCnt	Numérica (uint8)	Número de fijaciones durante la primera pasada de lectura (antes de cualquier regresión)
5*	P1stFixation	Categórica	Posición de la primera fijación en la palabra (inicio, medio o final)
6*	P2stFixation	Categórica	Posición de la segunda fijación en la palabra (si existió)
7*	prevFixDur	Numérica (int64)	Duración de la fijación anterior a la palabra actual (en milisegundos)
8*	firstfixDur	Numérica (int64)	Duración de la primera fijación en la palabra actual (ms)
9*	firstPassFixDur	Numérica (int64)	Duración total de las fijaciones durante la primera pasada de lectura (ms)
10*	nextFixDur	Numérica (int64)	Duración de la fijación inmediatamente posterior (en la siguiente palabra)
11	firstSaccLen	Numérica (float64)	Longitud de la primera sacada (movimiento ocular rápido) hacia la palabra actual (en grados visuales o píxeles)
12	lastSaccLen	Numérica (float64)	Longitud de la última sacada desde la palabra actual hacia otra palabra
13	prevFixPos	Numérica (float64)	Posición horizontal de la fijación anterior (en píxeles o proporción del ancho del texto)
14	landingPos	Numérica (float64)	Posición de aterrizaje de la mirada en la palabra actual (proporción respecto a su longitud)
15	leavingPos	Numérica (float64)	Posición desde la que el ojo abandona la palabra (píxeles o porcentaje)
16	totalFixDur	Numérica (int64)	Duración total de todas las fijaciones realizadas en la palabra (ms)
17	meanFixDur	Numérica (float64)	Duración promedio de las fijaciones sobre la palabra (ms)
18*	nRegressFrom	Numérica (uint8)	Número de regresiones iniciadas desde la palabra (cuando la mirada vuelve a una palabra previa)
19*	regressLen	Numérica (int64)	Longitud promedio de las regresiones (en píxeles o grados visuales)
20*	nextWordRegress	Categórica	Indica si la siguiente palabra fue objeto de una regresión (sí/no)
21*	regressDur	Numérica (int64)	Duración total de las fijaciones durante una regresión desde esta palabra (ms)
22	pupilDiamMax	Numérica (float64)	Diámetro máximo de la pupila durante la lectura de la palabra (indicador de carga cognitiva)
23	pupilDiamLag	Numérica (float64)	Diferencia de diámetro pupilar con respecto a la fijación anterior (cambio en la dilatación)
24	timePrctctg	Numérica (float64)	Proporción de tiempo dedicada a esta palabra respecto al total de la oración
25	nWordsInTitle	Numérica (uint8)	Número total de palabras en la oración o título leído
26	titleNo	Numérica (uint8)	Identificador de la oración o título dentro de la tarea (1 a 10)
27	wordNo	Numérica (uint8)	Posición ordinal de la palabra dentro de la oración
28	label	Categórica	Variable objetivo o de clase: 0 = Irrelevante, 1 = Relevante, 2 = Correcta

3. ESTADO DEL ARTE

Artículo 1:

Salojärvi, J., Puolamäki, K., Simola, J., Kovanen, L., Kojo, I., & Kaski, S. (2005). “Inferring Relevance from Eye Movements: Feature Extraction.” Helsinki University of Technology.

Este estudio constituye la fuente original de la base de datos *Eye Movements*. Los autores desarrollaron un experimento en el que los participantes respondieron preguntas mediante la lectura de títulos de noticias, mientras sus movimientos oculares eran registrados. Se propuso un enfoque de clasificación supervisada basado en la extracción de 22 características derivadas de los registros de fijación y sacadas.

Para la construcción del modelo, se evaluaron varios clasificadores, entre ellos *Naive Bayes*, *Decision Trees* y *k-Nearest Neighbors (k-NN)*. La validación se realizó mediante validación cruzada estratificada, y las métricas empleadas incluyeron exactitud (*accuracy*) y tasa de error promedio. Los mejores resultados se obtuvieron con *Naive Bayes*, con una precisión promedio del 73 %.

Inferring Relevance from Eye Movements: Feature Extraction — Salojärvi et al. (2005)
<https://research.ics.aalto.fi/events/eyechallenge2005/>

Artículo 2:

Papoutsaki, A., Gokaslan, A., Laskey, J., Huang, J., & Hays, J. (2016). “WebGazer: Scalable Webcam Eye Tracking Using User Interactions.” IJCAI.

Aunque no utiliza directamente la misma base de datos, este trabajo propone un enfoque de aprendizaje automático para inferir la atención visual a partir de movimientos oculares capturados por cámara web, demostrando la aplicabilidad del *eye tracking* para tareas de inferencia cognitiva.

El modelo utiliza un paradigma de aprendizaje supervisado con regresión lineal y redes neuronales para predecir la posición de la mirada. La validación se llevó a cabo con conjuntos de entrenamiento y prueba independientes, empleando métricas como el error medio cuadrático (*MSE*) y el error angular. Los autores lograron una precisión promedio de $1,06^\circ$ de error angular, mostrando que el análisis de patrones oculares puede ser abordado eficientemente mediante modelos de *machine learning* livianos y escalables.

WebGazer: Scalable Webcam Eye Tracking Using User Interactions — Papoutsaki, Sangkloy, Laskey, Daskalova, Huang & Hays (IJCAI 2016)

<https://cs.brown.edu/people/apapouts/>