



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

**FACTORES DE PREDICCIÓN PARA LA DETECCIÓN TEMPRANA DE TRASTORNOS
NEUROCOGNITIVOS EN PERSONAS INSTITUCIONALIZADAS EN CENTROS DE PROTECCIÓN
SOCIAL PERTENECIENTES A LA BENEFICENCIA DE CUNDINAMARCA**

PREDICTION FACTORS FOR THE EARLY DETECTION OF NEUROCOGNITIVE DISORDERS IN
INSTITUTIONALIZED PEOPLE IN SOCIAL PROTECTION CENTERS BELONGING TO LA
BENEFICENCIA DE CUNDINAMARCA

Andrés Parada Hernández: aparadah@libertadores.edu.co

John González Veloza: jjgonzalezv02@libertadores.edu.co

RESUMEN

Los trastornos neurocognitivos vienen aumentando de manera significativa en la población de personas mayores. A partir de esto, resulta importante comprender este fenómeno como el resultado de una multiplicidad de variables donde el contexto social y la afectación intelectual establecen elementos de estudio acordes con la identificación de factores de riesgo y/o contributivos. Las personas que presentan trastorno del desarrollo intelectual con afectación leve están más expuestas a adquirir un tipo específico de demencia la cual, generalmente, comienza a presentar signos tempranos. El presente estudio pretendió identificar casos potenciales en personas institucionalizadas en Centros de Protección Social de la Beneficencia de Cundinamarca con el fin de establecer acciones terapéuticas tempranas para controlar el deterioro mental y los costos asociados con los tratamientos. Para el estudio se tuvieron en cuenta 32 variables predictoras obtenidas por medio de la Escala Wechsler de Inteligencia para Adultos y Mini-Mental State Examination. A partir del análisis de datos existentes en los Centros de Protección Social de la Beneficencia de Cundinamarca en 2018, 2020 y 2022 (n = 294) se entrenaron modelos estadísticos que dividieron la variable objetivo: diagnóstico de demencia en 2022 en componentes independientes orientados a la predicción de casos potenciales de acuerdo con la prevalencia de algún tipo de afectación. Una vez procesados los datos, el modelo de clasificación que presentó mejor desempeño fue el Random Forest Classifier, arrojando un



AUC de 0.97, 0.01 por encima del Light Gradient Boosting Machine, y 0.1 de la Logistic Regression planteada inicialmente. Las variables que mejor contribuyeron a la predicción se relacionaron con la memoria evocativa, organización temporal, memoria inmediata, atención y cálculo, diagnóstico base, coeficiente ejecutivo y edad.

Palabras claves: Persona con discapacidad, demencia, CI, envejecimiento.

ABSTRACT

Neurocognitive disorders have been increasing significantly in the elderly population. Based on this, it is important to understand this phenomenon as the result of a multiplicity of variables where the social context and the intellectual affectation base study elements consistent with the identification of risk factors and/or adjuvants. People with Mildly Impaired Intellectual Development Disorder are more at risk of acquiring a specific type of dementia, which usually begins to show early signs. The objective of this study is to identify potential cases in people institutionalized in the Social Protection Centers of Cundinamarca to establish early therapeutic actions to control mental deterioration and the costs associated with the treatments. For the study, thirty-two predictor variables obtained through the Wechsler Intelligence Scale for Adults and the Mini-Mental State Examination were considered. Based on the analysis of the existing data in the Cundinamarca Social Protection Centers for Charities in the years 2018, 2020 and 2022 (n=295), statistical models were trained that divided the objective variable: diagnosis of dementia in 2022 into components independent oriented to the prediction of potential cases according to the prevalence of some type of affectation. Once the data was processed, the classification model that presented the best performance was the Random Forest Classifier, yielding precision metrics of 0.95, 0.01 above the Light Gradient Boosting Machine, and 0.02 above Logistic Regression. initially proposed. The variables that best contributed to the prediction were those related to age, intelligence quotient, executive quotient, and verbal quotient.

Keywords: Person With, Disability, Dementia, IC, Aging.



INTRODUCCIÓN

El aumento de la expectativa media de vida en Colombia ha alcanzado los 74 años (Departamento Administrativo Nacional de Estadística, 2023). Para el 2020 el DANE proyectó la población de Colombia en 51 millones de personas, de las cuales 4 millones y medio serán mayores de 65 años. Algunos autores, por ejemplo, consideran una prevalencia de la demencia en el 6% de la población mayor lo que arroja 260.000 afectados. Continuando con esto, y tomando las personas mayores con trastornos del desarrollo intelectual, se podría esperar cerca de 1.700.000 (Cabezas, Marulanda, & Martínez, 2013) personas afectadas lo que plantea enormes retos para las políticas de salud pública.

El informe mundial acerca de demencia tipo Alzheimer (DTA) estipuló que para el 2010 la demencia afectaba a 36.5 millones de personas, previendo un aumento progresivo del doble de personas cada 20 años. A partir de esto, para el 2040 se espera que 81.1 millones de personas estén afectadas por esta enfermedad, siendo los países en vía de desarrollo los más propensos con un 70% de los casos (Cabezas et al., 2013).

El diagnóstico temprano de demencia se ha sustentado en guías de práctica clínica como las desarrolladas por el Sistema General de Seguridad Social en Salud de Colombia, siendo la Guía N° 61 de 2017 un claro ejemplo. No obstante, la demencia constituye una enfermedad infradiagnosticada por razones que van desde el poco entrenamiento del personal de atención primaria, los síntomas precoces, sutiles y heterogéneos, y la mala utilización de los test de cribado (Contador, Fernández, Ramos, Tapias, & Bermejo, 2010). “El cribado cognitivo generalizado de demencia AP no es coste-efectivo y, al no existir un tratamiento curativo en la mayoría de las demencias, el beneficio del diagnóstico inicial no es bien conocido por el paciente, por su familia e incluso por el estamento sanitario” (Contador et al., 2010)

La aplicación de instrumentos de cribado cognitivo es el primero paso que debe llevarse para una evaluación detallada de casos sospechosos de demencia. La estrategia de detección de casos de riesgo en la atención primaria, contrario al cribado generalizado en individuos asintomáticos, típicos de los estudios epidemiológicos; podría iniciar la



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

intervención farmacológica y psicosocial orientada a contrarrestar la evolución de la demencia y la institucionalización. Por otra parte, facilitaría la toma de decisiones en la familia acerca de aspectos sociosanitarios relacionados con el enfermo, además de reducir los costos asociados al uso de servicios a domicilio y especializados (Contador et al., 2010)

A partir de lo anterior, los Centros de Protección Social pertenecientes a la Beneficencia de Cundinamarca han implementado el proceso de valoración a las personas institucionalizadas a través de instrumentos de tamizaje y cribado, con el fin de identificar síntomas tempranos de demencia. No obstante, la elección de instrumentos no resulta una tarea fácil y confiable debido a que muchas de las características psicométricas (fiabilidad, validez, etc.) son heterogéneas. Comúnmente la población institucionalizada presenta características sociodemográficas ancladas en altos niveles de pobreza, marginación, poca educación y afectaciones psicosociales por lo que los resultados, no pocas veces, pueden verse afectados por falsos positivos. “La etiología del deterioro cognitivo y las características sociodemográficas y culturales de la población repercuten en la validez discriminativa del test, esto es, en la capacidad del instrumento para diferenciar entre individuos o grupos” (Contador et al., 2010)

La psiquiatría, neurología y psicología contemporáneas han centrado parte de su interés en la comprensión e identificación de los cambios cognitivos, conductuales y biológicos que ocurren dentro de la fase presintomática de la demencia. El deterioro cognitivo que ocurre sin cambios significativos en el funcionamiento de la vida diaria recibe el nombre de deterioro cognitivo leve (DCL), comprendiéndose como la fase intermedia entre la cognición normal y demencia. Para algunos autores no existe un evento fijo que determine el punto de inicio de la fase asintomática a la fase sintomática (Servicio de Neurología, Hospital Militar Central, Lima, Perú.). Una persona con DCL tiene 10 a 15% de probabilidad de desarrollar demencia en comparación con personas sanas donde la probabilidad está en 1 a 2% anual (Petersen, y otros, 2001, pág. 58 a 92). Un estudio de personas con DCL, seguidos durante 5 años, encontró que aquellos con presencia de factores de riesgo vascular tuvieron un aumento en la conversión a demencia tipo Alzheimer. Los resultados de este estudio mostraron que las personas que fueron tratadas a partir del control de



todos los factores de riesgo tuvieron menor evolución a demencia, mientras las personas que fueron intervenidas a partir de control de algunos factores de riesgo tuvieron mayores casos positivos (Cabezas et al., 2013).

De acuerdo con lo anterior, el objetivo del presente estudio se fundamenta en pronosticar la probabilidad de que una persona institucionalizada adquiera demencia a partir del análisis y clasificación de factores de riesgo por medio del uso de modelos de Machine Learning supervisados. En este trabajo, se buscará la detección temprana y prevención de demencia con el fin de ralentizar su avance y, sea el caso, reducir los efectos sociales, familiares y económicos que conlleva la enfermedad. No obstante, los resultados alcanzados no serán del todo eficientes por sí solos ya que deben actuar con pruebas neuropsicológicas, bioquímicas e imágenes diagnósticas.

Un diagnóstico de demencia debe hacerse sólo después de una evaluación integral, incluyendo aspectos del historial clínico, instrumentos de validación, revisión farmacológica y detección de trastornos del estado de ánimo, entre otros. Después de confirmado el caso, se deben evaluar las comorbilidades médicas y las características psiquiátricas claves asociadas con la enfermedad, incluyendo depresión y psicosis, para asegurar el tratamiento óptimo de condiciones coexistentes (Guía de Práctica Clínica para el diagnóstico y tratamiento del trastorno neurocognoscitivo mayor. Sistema General de Seguridad Social en Salud – Colombia, 2017)

En términos generales, este estudio busca identificar los factores individuales y generales que pueden pronosticar la aparición de demencia. Para tal fin, se realizará el análisis de 294 datos clasificados en 32 variables predictoras y 1 objetivo, obtenidas por medio de la Escala Wechsler de Inteligencia para Adultos (WAIS III) y Mini-Mental State Examination. Los datos analizados pertenecen a los Centros de Protección Social de la Beneficencia de Cundinamarca en 2018, 2020 y 2022. Las variables estudiadas toman como referencia las puntuaciones obtenidas para el coeficiente intelectual general, categorizando valores de la escala ejecutiva y verbal. Por otra parte, se toman variables que relacionan la velocidad de procesamiento con la organización perceptual, la ubicación temporal y la espacialidad.



METODOLOGÍA

Datos.

El presente estudio se fundamentó en los datos existentes en los Centro de Protección Social José Joaquín Vargas y Colonia ubicados en el municipio de Sibaté, Cundinamarca. La versión final de la base de datos contó con 33 variables relacionadas con 155 hombres y 139 mujeres para un total de 294 personas institucionalizadas desde 2018 a 2022. Las fases desarrolladas para el tratamiento de los datos comprendieron tres pasos: Limpieza de la base y selección de variables relevantes, ejecución de análisis descriptivos y, finalmente, modelación y evaluación de resultados.

Los datos que componen este estudio se obtienen a partir del proceso de valoración y actualización de información de las personas institucionalizadas. De acuerdo con los criterios establecidos en pliego de condiciones que establece la Beneficencia de Cundinamarca, debe evaluarse la evolución clínica por medio de la aplicación de instrumentos cuantitativos debidamente estandarizados. La actualización de información debe hacerse cada 2 años con el fin de detectar deterioro o fluctuaciones significativas en la población.

Limpieza de la base de datos y selección de variables relevantes.

Se incluyeron 33 variables con información de 2018 a 2022 sustentadas en la actualización de la evolución clínica de las personas institucionalizadas. Esta decisión se sustentó en la importancia de cuantificar diferencias en las puntuaciones obtenidas a lo largo de tres periodos. Se crearon, pues, variables a partir de las diferencias de puntuaciones producto de la actualización de información 2 años después. Se decidió conservar la variable lugar de procedencia para identificar posibles tendencias de tipo sociodemográfico (esta variable fue la única externa a la persona que se tuvo en cuenta). Se eliminó la variable escolaridad por la gran cantidad de valores nulos y la poca precisión de registros caracterizados.



Tabla 1. Transformación de variables

Variables 2018	Variables 2020	Nuevas Variables
coef_int_verbal_2018	coef_int_verbal_2020	dif_coeficiente_verbal_2a_atrás
coef_ejecutivo_2018	coef_ejecutivo_2020	dif_coeficiente_ejecutivo_2a_atrás
coef_verbal_2018	coef_verbal_2020	dif_comprension_verbal_2a_atras
org_perceptual_2018	org_perceptual_2020	dif_org_perceptual_2a_atras
mem_operativa_2018	mem_operativa_2020	dif_memoria_operativa_2a_atras
vel_procesamiento_2018	vel_procesamiento_2020	dif_vel_procesamiento_2a_atras
org_temporal_2018	org_temporal_2020	dif_org_temporal_2a_atras
org_espacial_2018	org_espacial_2020	dif_org_especial_2a_atras
mem_inmediata_2018	mem_inmediata_2020	dif_mem_inmediata_2a_atras
aten_calculo_2018	aten_calculo_2020	dif_aten_calculo_2a_atras
mem_evocativa_2018	mem_evocativa_2020	dif_mem_evocativa_2a_atras
com_lectora_2018	com_lectora_2020	dif_comp_lectora_2a_atras
coef_intelectual_2018	coef_intelectual_2020	dif_coef_intelectual_2a_atras
diag_demencia_2018	diag_demencia_2020	dif_diag_demencia_2a_atras

En la tabla 1 puede verse que las variables de 2018 y las variables de 2020 se fusionaron en una sola variable. Esta fusión se consiguió restando las puntuaciones de 2018 a las puntuación de 2020; esto se hizo con el fin de evaluar la disminución de capacidades y visualizar más fácilmente el deterioro general de las personas.

Procesamiento y Modelación.

Para el procesamiento de la modelación se tomó una base de datos conformada por 294 personas de género femenino y masculino institucionalizadas en Centros de Protección Social de la Beneficencia de Cundinamarca. Se tuvo en cuenta 1 variable objetivo y 32 explicativas. Una vez cargada la base de datos, se procedió a realizar el análisis de valores nulos. Se revisó la categoría de las variables encontrando 31 numéricas y 2 categóricas. para observar algunas características de las variables categóricas, estableciendo un conteo de 294 individuos clasificados por lugar de procedencia (17 provincias de Cundinamarca) y género (femenino y masculino). La procedencia que más se repitió fue Bogotá con una frecuencia de 60 individuos, mientras el género que más se repitió fue el masculino con 155 personas.

Tabla 2. Descripción de variables categóricas.

	Procedencia	Género
conteo	294	294
valor	17	17
máximo	bogota	M
frecuencia	60	155



El set de datos de validación tomó 205 personas a partir del análisis de 49 variables; mientras el set de datos de testeo se estructuró en 89 personas a partir del análisis de 49 variables (Anexo. Figura S1)

Para la realización de la modelación, se procedió a efectuar el análisis de modelos de clasificación orientados a la variable objetivo en cuanto a la predicción de demencia por medio de su diagnóstico (0 = sin diagnóstico de demencia, 1 = con diagnóstico de demencia). A partir de esto, se compararon las métricas con los resultados más altos, siendo el Accuracy, AUC, Recall, Precisión y F1 las consideradas; además de contrastar los resultados de las matrices de confusión a través de la identificación de casos positivos verdaderos. La preparación de los datos se realizó por medio del software estadístico R, y la modelación a partir del lenguaje de programación Python desde los paquetes PyCaret versión 3, Scikit Learn y Sweetviz.

RESULTADOS

Las variables que mejor contribuyeron a la predicción se relacionaron con la memoria evocativa, organización temporal, memoria inmediata, atención y cálculo, diagnóstico base, coeficiente ejecutivo, edad y coeficiente intelectual general.

De acuerdo con el Feature Importance, se observó que la variable Memoria Evocativa presentó puntaje 1 equivalente al 17% de los casos de demencia, y 35% para los casos no diagnosticados. Este hecho resulta significativo debido a que esta competencia suele verse muy afectada en las personas con diagnósticos positivo (Figura 1). La evaluación de la función memoria constituye un requisito básico para el inicio del análisis exploratorio de la enfermedad. En este aspecto, el modelo Machine Learning es coherente con la evidencia clínica.

En cuanto a la variable Organización temporal, se evidencia que el puntaje 1 fue el más característico en las personas con diagnóstico positivo constituyendo el 17% de los casos (Figura 2). La organización temporal, comúnmente, suele estar afectada en casos de demencia



ya que la relación entre la codificación de información sensorial y el sistema perceptivo presenta alteraciones que distorsionan el principio de realidad. Una vez más el modelo descriptivo resulta coherente con la evidencia clínica.

Frente a la Memoria Inmediata, puede verse que el puntaje 1 fue el más característico ya que incluye al 20% de los caso positivos (Figura 3). A diferencia de la memoria evocativa, la inmediata hace relación a eventos ocurridos apenas unos segundo antes. La afectación marcada de esta función implica la presencia de estados avanzados de la enfermedad y se relaciona con Alzheimer. La modelación descriptiva supo categorizar adecuadamente la relevancia de esta variable debido a que, generalmente, es el mayor indiciador de afectaciones mayores.

La variable de atención y cálculo presentó una puntuación de 0 en el 14% de personas con diagnóstico de la enfermedad lo que muestra la relación existente entre el procesamiento de información numérica y la afectación que produce la demencia en los procesos mentales (Figura 4).

En cuanto a la variable Edad, la media estuvo en 57 años lo que establece un marco de referencia y acción importante para este estudio: la detección de casos potenciales de demencia (Figura 5).

Con lo que respecta al Diagnóstico Base, puede evidenciarse que 66% de la población presenta afectación intelectual, mientras el 34% presenta afectación mental. De los 23 casos de demencia confirmados en 2022, el 10% se asoció con afectación intelectual, mientras el 3% con afectación mental lo que establece una correlación importante entre el nivel cognitivo y la disposición a la demencia (Figura 6a y 6b)

La variable Coeficiente Ejecutivo, presentó una media de 44 puntos. De acuerdo con esto, puede verse que el grueso de los casos de demencia se ubica por debajo de la media lo que muestra la relevancia de la afectación ejecutiva en las características diferenciales de la enfermedad. La capacidad de las personas con demencia para razonar a partir de estímulos visoespaciales y para resolver problemas nuevos se encuentra muy afectada (Figura 7)



MEMORIA EVOCATIVA

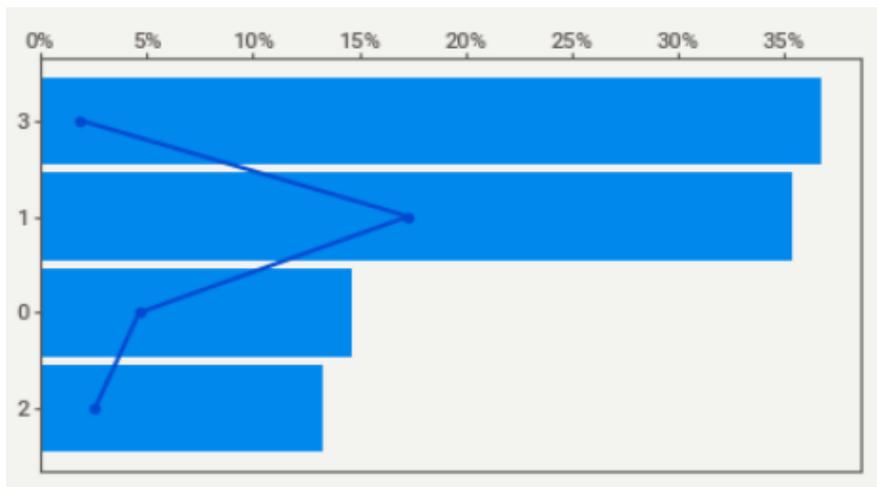


Figura 1. Histograma de la variable Memoria Evocativa. Puede observarse como la puntuación 1 es la que más se relaciona con los casos de demencia. La línea azul continua representa la distribución empírica acumulada de observaciones para casos positivos de demencia. El puntaje 3 implica sin afectación, el puntaje 2 afectación leve, el puntaje 1 afectación grave, y el puntaje 0 afectación profunda. La puntuación 1 se relacionó con el 17% de los casos positivos.

ORGANIZACIÓN TEMPORAL

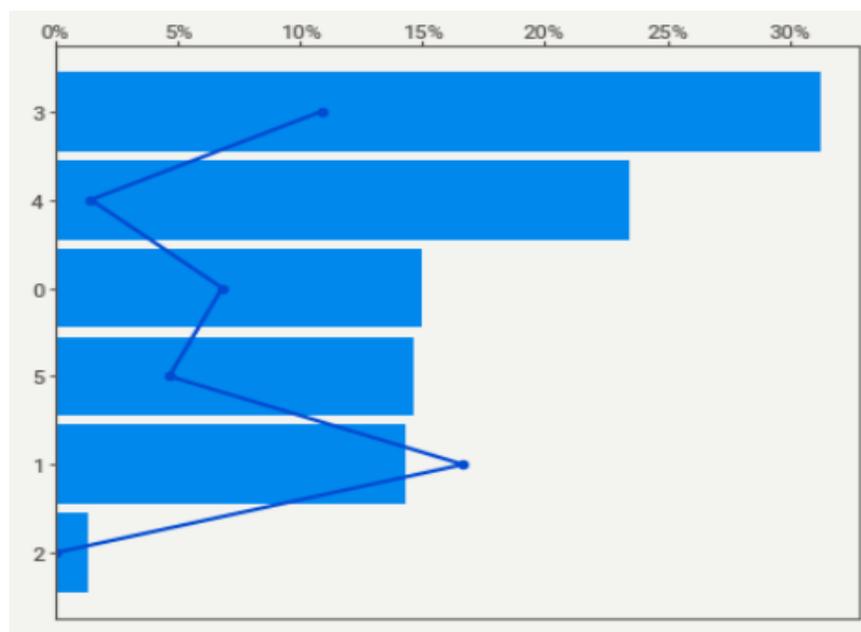


Figura 2. De acuerdo con la figura, puede verse como la puntuación 1 se relaciona con la mayor cantidad de casos positivos de demencia, representando el 17%.



MEMORIA INMEDIATA

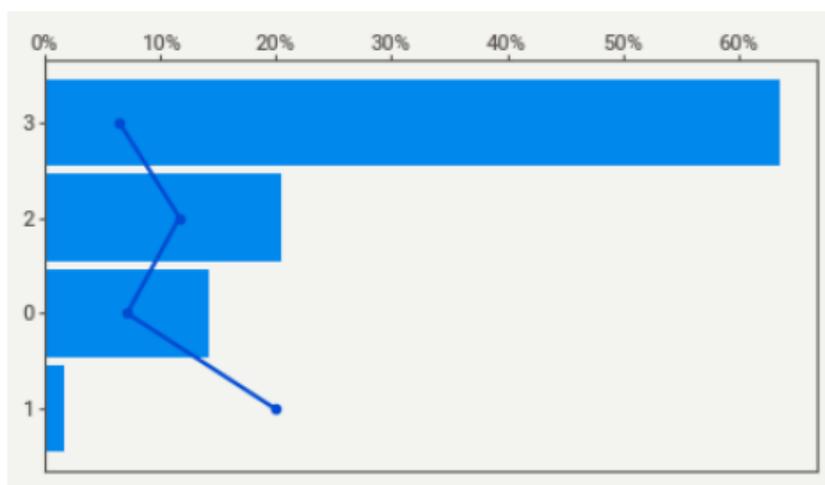


Figura 3. Puede observarse como la puntuación 1 es la que más se relaciona con los casos de demencia, constituyendo el 20%. La línea azul continua representa la distribución empírica acumulada de observaciones para casos positivos. La afectación de la memoria inmediata es un importante indicador para casos graves y, probablemente, Alzheimer.

ATENCIÓN Y CÁLCULO

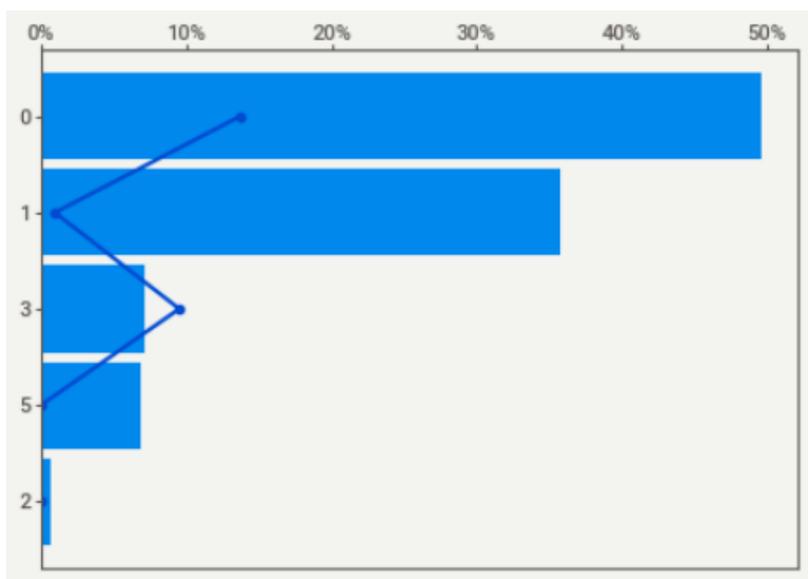


Figura 4. Puede observarse como la puntuación 0 es la que más se relaciona con los casos de demencia, constituyendo el 48%. La afectación de esta competencia es un indicador de la relación que existe entre el cálculo y los procesos intelectuales avanzados.



EDAD

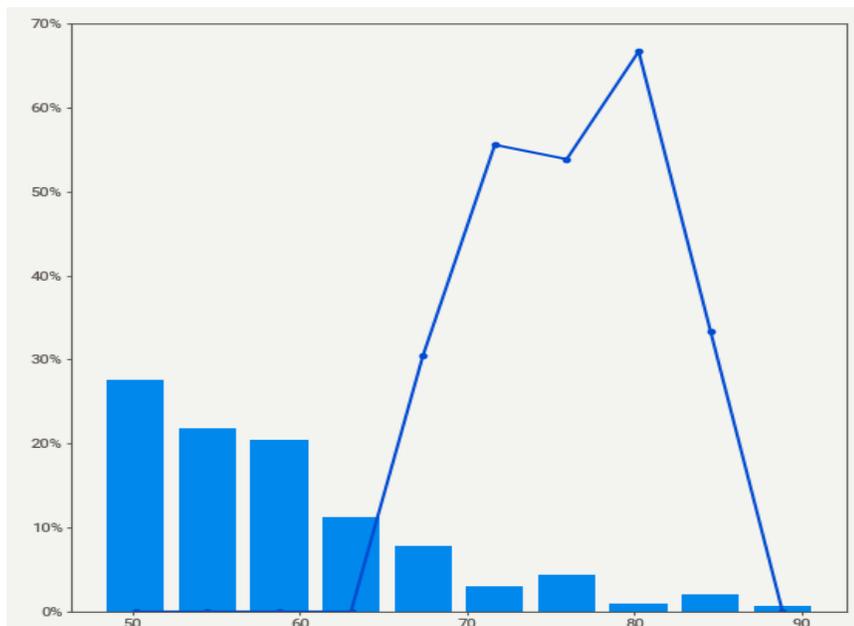


Figura 5. Histograma de la edad por estatus de distribución porcentual de personas institucionalizadas. De acuerdo con la gráfica, puede verse que la edad presenta importancia significativa para la explicación del modelo. Entre más edad tenga la persona más propensa es para adquirir demencia. La totalidad de casos están dentro del rango de 60 a 90 años, evidenciándose un importante incremento después de los 65 años (periodo crítico) y un descenso después de los 80 años.

DIAGNÓSTICO BASE VS COEFICIENTE INTELECTUAL

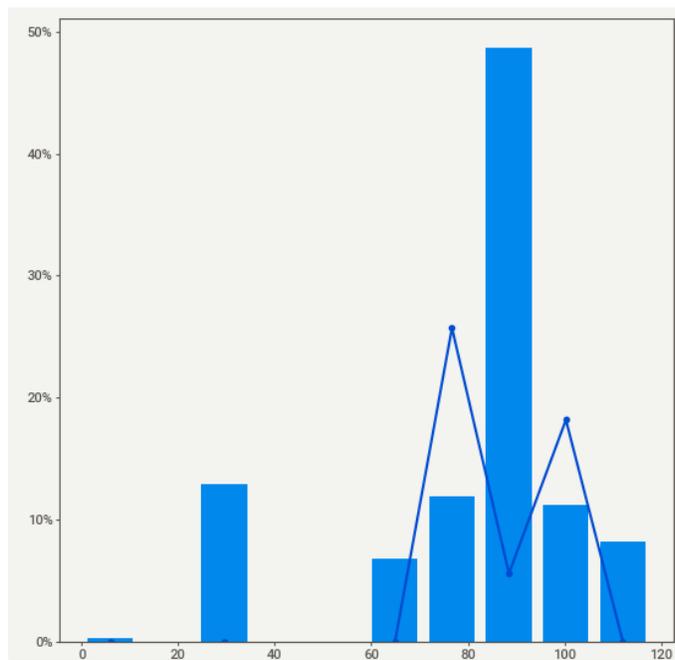
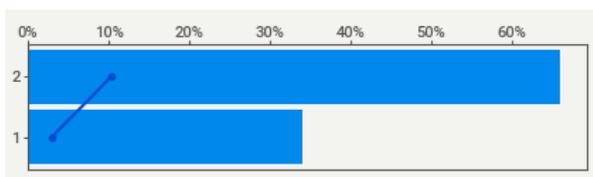


Figura 6a y 6b. Histograma de distribución porcentual por estatus de diagnóstico base contrastado con las puntuaciones de CI. La gráfica 6a muestra que la mayoría de los casos de demencia se relacionan con personas diagnosticadas con discapacidad intelectual. Una persona con CI dentro del rango de 80 a 100 (límite de normalidad) puede presentar mayor riesgo de contraer demencia, mientras una persona con CI alto (gráfica 6b) está menos expuesta. Frente a los CI muy bajos resulta difícil establecer criterios diferenciales por motivos asociados con múltiples afectaciones.



COEFICIENTE EJECUTIVO

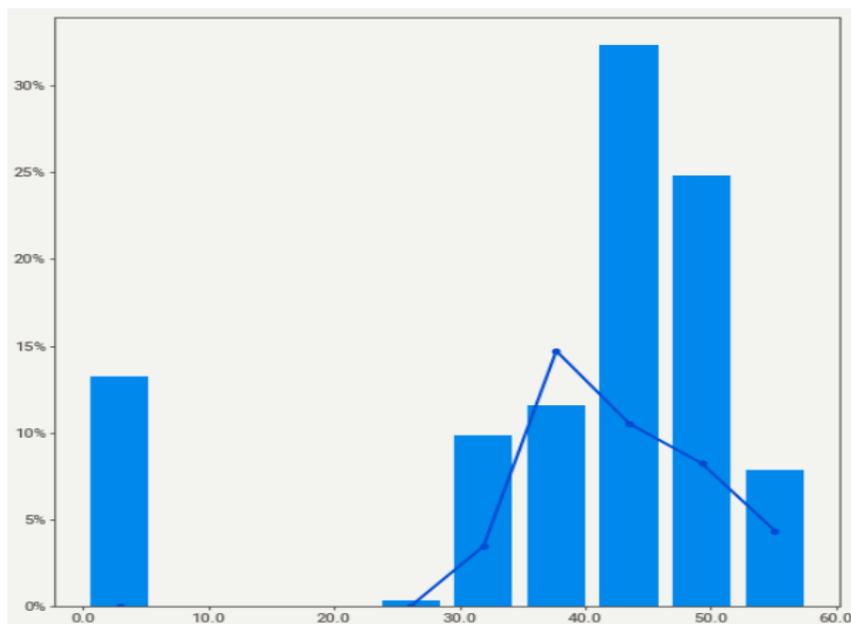


Figura 7. Puede observarse como la mayoría de los casos de demencia se relacionan con un coeficiente ejecutivo cercano a los 40 puntos relacionándose, a su vez, con el 15% de los casos.

En términos generales, de las 32 variables explicativas seleccionadas, 7 fueron las que tuvieron mayor relevancia. Un hecho que resulta importante para este estudio tiene que ver con el acierto del modelo descriptivo con respecto a la evidencia clínica. En sus términos más generales, la demencia constituye una pérdida de la función cerebral que afecta a competencias como la memoria, el pensamiento, el juicio y raciocinio, entre otros. Se ve como el modelo selecciona las variables que se asocian con estas competencias, dándoles relevancia y significancia de acuerdo con investigaciones (ver introducción).

La elección de variables de acuerdo con la relevancia descriptiva mostró una acertada ejecución del proceso de medición, además de la importancia que tienen los instrumentos de valoración para la recolección de información necesaria para el diseño de mecanismos contingentes. En resumidas cuentas, el análisis descriptivo permitió observar la coherencia de los datos con la práctica clínica, además de la relación existente entre los resultados con los esperado.



Una vez establecidos los criterios descriptivos, se procedió a encontrar el mejor modelo explicativo a través de la comparación de métricas.

Tabla 3. Desempeño promedio de los mejores modelos relacionados con los datos.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
rf	Random Forest Classifier	0.9464	0.9734	0.3000	0.4000	0.3333	0.3288	0.3378	0.5520
lightgbm	Light Gradient Boosting Machine	0.9367	0.9602	0.4500	0.5167	0.4633	0.4386	0.4491	0.4230
xgboost	Extreme Gradient Boosting	0.9221	0.9510	0.5000	0.4500	0.4533	0.4154	0.4269	0.2790
gbc	Gradient Boosting Classifier	0.9414	0.9444	0.5000	0.6000	0.5333	0.5059	0.5144	0.4280
ada	Ada Boost Classifier	0.9424	0.9235	0.5500	0.5833	0.5567	0.5345	0.5394	0.4340
et	Extra Trees Classifier	0.9419	0.8997	0.2500	0.4000	0.3000	0.2931	0.3067	0.6770
lda	Linear Discriminant Analysis	0.9464	0.8757	0.5000	0.6000	0.5333	0.5127	0.5218	0.2710
lr	Logistic Regression	0.9269	0.8681	0.4500	0.5500	0.4833	0.4454	0.4539	1.1090
knn	K Neighbors Classifier	0.9319	0.8389	0.2500	0.2667	0.2467	0.2366	0.2431	0.1770
dt	Decision Tree Classifier	0.8979	0.7183	0.5000	0.3667	0.4067	0.3556	0.3682	0.3480
nb	Naive Bayes	0.5952	0.5893	0.6000	0.0912	0.1569	0.0342	0.0879	0.2120
qda	Quadratic Discriminant Analysis	0.9319	0.5500	0.1000	0.2000	0.1333	0.1288	0.1378	0.3580
dummy	Dummy Classifier	0.9224	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2850
svm	SVM - Linear Kernel	0.9026	0.0000	0.0000	0.0000	0.0000	-0.0211	-0.0225	0.1360
ridge	Ridge Classifier	0.9464	0.0000	0.5000	0.6000	0.5333	0.5127	0.5218	0.1340

En la Tabla 3 puede observarse como el modelo Random Forest Classifier presenta una exactitud del 0.9464 y un AUC del 0.9734. Con base en esto, puede decirse que la medición del acierto del modelo en la predicción de eventos binarios es muy alta. Al revisar el AUC se evidencia que la elección del indicador parte de la decisión de predecir correctamente los casos de demencia (0 = sin diagnóstico para demencia, 1 = con diagnóstico para demencia). El resultado de 0.9734, establece que la predicción de ocurrencia para un caso positivo es mucho más alta que la probabilidad de fallar, esto es, la relación entre todos los eventos que ocurren es mayor que la relación de los eventos que no ocurren en este modelo (Anexos: Figura S2)



LOS LIBERTADORES FUNDACIÓN UNIVERSITARIA

Al evaluar los aciertos y errores que presenta el modelo de detección de demencia, se establece que la relación de la estimación con respecto al sesgo es baja debido al alto número de verdaderos positivos y verdaderos negativos (Figura 5)

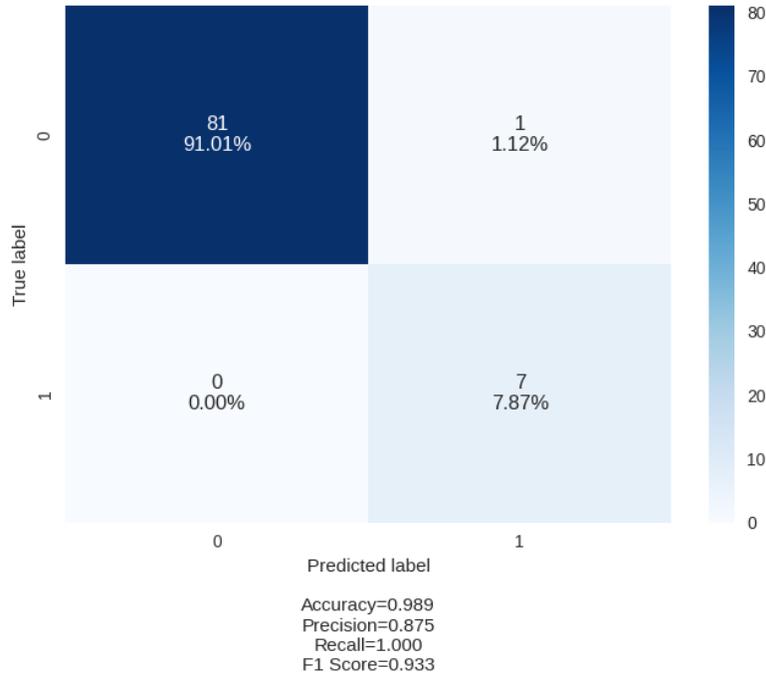


Figura 5. Matrix de Confusión del modelo Random Forest Classifier. Se evidencia que el Accuracy es de 0.989 lo que establece que el 99% de las predicciones positivas son correctas. En cuanto a la Precisión, el valor alcanzado por el modelo de 0.875 indica que la dispersión del conjunto de valores no fue significativa, esto es, se detectó un alto porcentaje de casos positivos. El modelo presentó un Recall de 1.0 lo que muestra la alta capacidad del estimador para discriminar los casos positivos de los negativos: tanto la sensibilidad como la especificidad alcanzaron el máximo valor. El F1 Score, más allá de promediar la Precisión y el Recall, muestra que la distribución de clases del modelo es desigual ya que el número de personas en condición de demencia es del 7.87% vs el 91% de personas sin esta enfermedad.

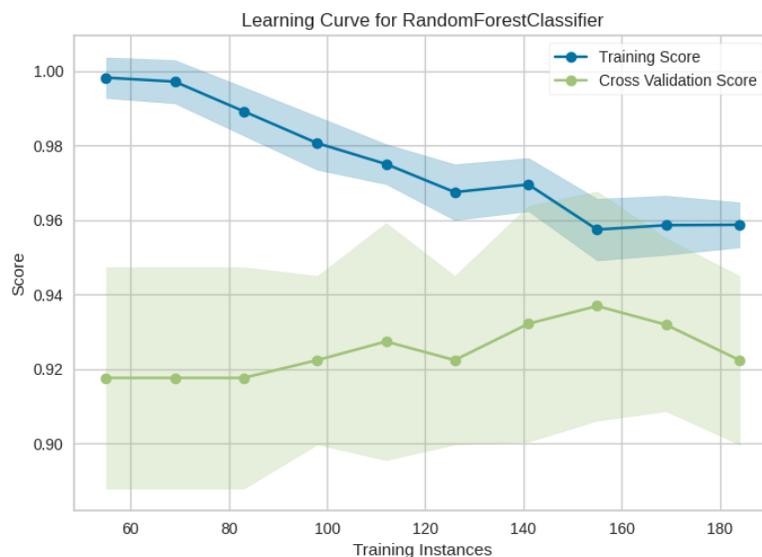


Figura 6. Al revisar la curva de aprendizaje del modelo, podemos observar el comportamiento de los datos de entrenamiento con respecto a la validación cruzada



DISCUSIÓN

El presente estudio pretendió predecir casos de demencia a partir de la implementación de modelos de Machine Learning. Los resultados obtenidos indicaron que los modelos con mejor desempeño fueron el *Random Forest Classifier*, *Light Gradient Boosting Machine* y *Extreme Gradient Boosting* (todos basado en árboles de decisión). De acuerdo con los resultados arrojados por la métricas, pudo verse la alta precisión del *modelo Random Forest Classifier* para predecir la relación de las variables explicativas con la variable objetivo. Por otra parte -y dada mi experiencia en la atención de personas con demencia- las variables que el modelo estableció como relevantes para la predicción, efectivamente son las más discriminativas dentro de los instrumentos diagnósticos, exactamente en los test de cribado y pruebas de inteligencia. Este hecho, indica que los modelos Machine Learning pueden ofrecer aportes valiosos para el diseño de estrategias de salud pública y control de riesgos dentro del proceso de atención individual de personas propensas a la enfermedad.

Los factores relevantes identificados por el modelo *Random Forest Classifier* para el pronóstico de casos de demencia se asociaron con fallas en los siguientes campos: memoria evocativa, organización temporal, memoria inmediata, atención y cálculo, diagnóstico de base, coeficiente ejecutivo y organización espacial (Anexo: Figura S3). Resulta interesante que el modelo haya identificado la memoria evocativa como la más característica: uno de los principales síntomas de la predemencia tiene que ver con la pérdida de memoria y con las dificultades para organizar información.

De acuerdo con estudios realizados (Fernández Cobas, 2021), los algoritmos con mejores resultados para la predicción (Anexo: Figura S4) temprana de demencia mediante el estudio de factores de riesgo se basan en k-NN, Navie Bayes, Red Neuronal Artificial, SVM, Árboles de decisión y Random Forest. En este estudio, precisamente, el modelo seleccionado identificó aspectos que se concluyeron en la investigación previamente citada. El modelo explicativo ofreció un método experimental para detectar las interacciones de las variables por medio de la comparación de la prevalencia de riesgos en personas ya diagnosticadas.



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Por otra parte, se logró identificar aspectos claves para la predicción de casos potenciales de demencia, no obstante, hubo factores que no tuvieron la relevancia esperada. El lugar de procedencia no contribuyó de manera significativa a la predicción de resultados lo que podría explicarse, en parte, en los pocos datos registrados en comparación con los índices poblacionales de los municipios que conforma el departamento. Otros factores poco relevantes fueron los resultados producto de la resta de las variables de 2018 a las de 2020. De acuerdo con el Feature Importance (Anexo: Figura S5), los resultados de las diferencias ocuparon los últimos lugares lo cual puede entenderse por el simple hecho de que otras variables arrojaron mejores resultados con menor esfuerzo.

Otro aspecto para tener en cuenta tiene que ver con la limitación misma de los datos: una población de 294 personas resulta limitada a la hora de establecer conclusiones amplias. Los datos de testeo y validación tomaron valores reducidos lo que puede afectar predicciones en poblaciones más extensas (y la validez misma del modelo). Un ejemplo de esto puede observarse en la procedencia de las personas que presentaron mayores casos de demencia. De acuerdo con el modelo, la provincia de Medina presentó la proporción más alta de casos, no obstante, el número de personas pertenecientes a este lugar es muy pequeño en comparación, por ejemplo, con Bogotá. En los centro de protección social hay tres personas que proceden de Medina, y uno de ellos presenta diagnóstico de demencia: esta proporción puede resultar alta sin ser necesariamente significativa.

Este estudio puede tener dificultades en el momento de extender los resultados de las predicciones a muestras más amplias y con variables que no se tuvieron en cuenta. Las características de las personas con discapacidad mental e intelectual no son suficientes para explicar la complejidad de los factores de riesgo, ni pueden establecer de manera confiable la relevancia del nivel educativo, las condiciones sociodemográficas y económicas de una población más diversa.

Finalmente, otra limitación puede estar en la estructura misma del modelo. De acuerdo con las revisiones del estado del arte, el Random Forest presenta una alta eficiencia en bases de



datos grandes, sin embargo, esta eficiencia podría verse reducida al momento de modelar bases de datos pequeñas.

CONCLUSIÓN

Este estudio pretendió identificar personas pertenecientes a Centros de Protección que por su condición están más expuesta a presentar demencia. La base de datos y las variables tenidas en cuenta se sustentaron en características particulares de un tipo específico de población: personas con discapacidad mental e intelectual. No fue objetivo de este estudio extender los resultados a poblaciones diferentes a la categorizada. Por esta razón, y dados los resultados, el modelo aportó de manera significativa información básica para el control y direccionamiento de técnicas terapéuticas de ejecución temprana, estableciendo factores relevantes para la identificación de síntomas iniciales. Otro aspecto de valor aportado por esta investigación tiene que ver con la importancia de monitorear las competencias intelectuales de la población por medio de instrumentos cuantitativos, no en vano los datos que sustentaron este estudio partieron de puntuaciones en escalas Weschler y test de cribado. En términos generales, el aporte del modelo a los Centros de Protección Social será significativo ya que permitirá reducir costos a partir de la identificación temprana de casos potenciales.



ANEXOS

	Description	Value
0	Session id	1234
1	Target	diag_demencia_2022
2	Target type	Binary
3	Original data shape	(294, 33)
4	Transformed data shape	(294, 49)
5	Transformed train set shape	(205, 49)
6	Transformed test set shape	(89, 49)
7	Ordinal features	1
8	Numeric features	30
9	Categorical features	2
10	Preprocess	True
11	Imputation type	simple
12	Numeric imputation	mean
13	Categorical imputation	mode
14	Maximum one-hot encoding	25
15	Encoding method	None
16	Fold Generator	StratifiedKfold
17	Fold Number	10
18	CPU Jobs	-1
19	Use GPU	False
20	Log Experiment	False
21	Experiment Name	clf-default-name
22	USI	9d5a

Figura S1. Tabla de descripción de nuevas variables.

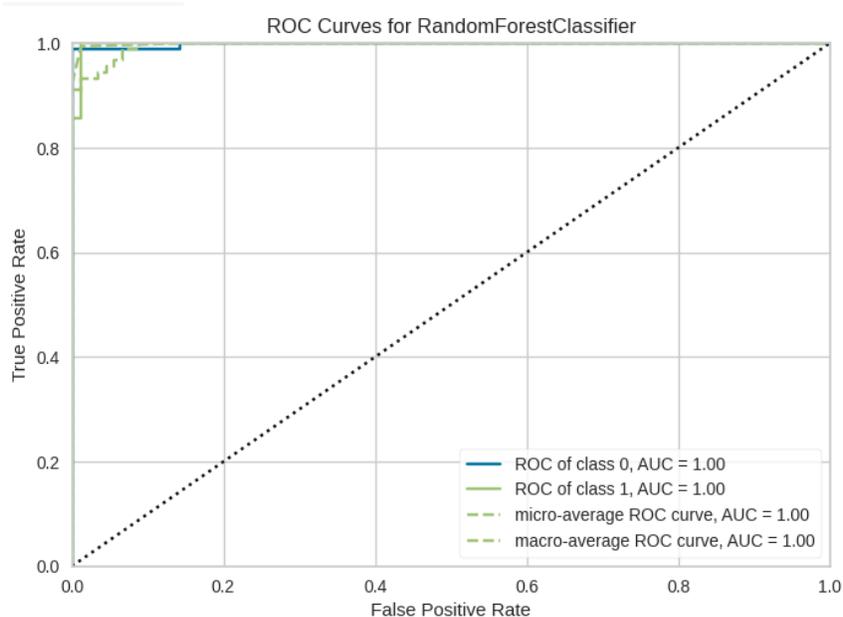


Figura S2. Curva ROC.

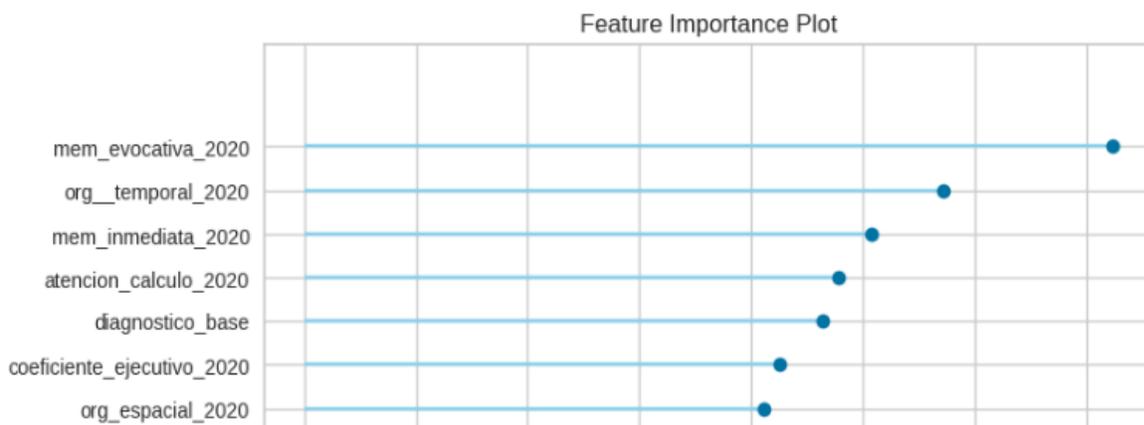


Figura S3. En esta gráfica puede verse la importancia de las variables en la explicación del modelo.

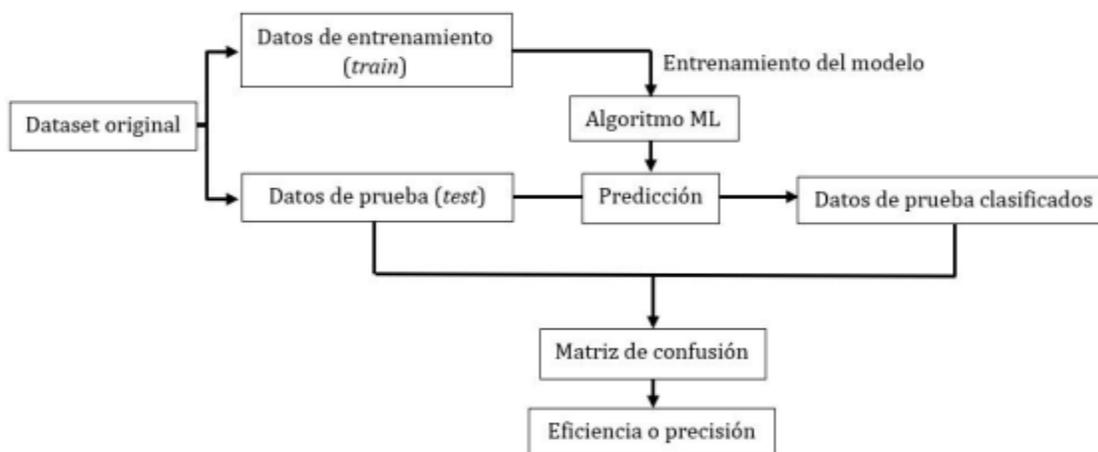


Figura S4. En esta gráfica puede verse un esquema general del proceso de modelación (Fernández Cobas, 2021)

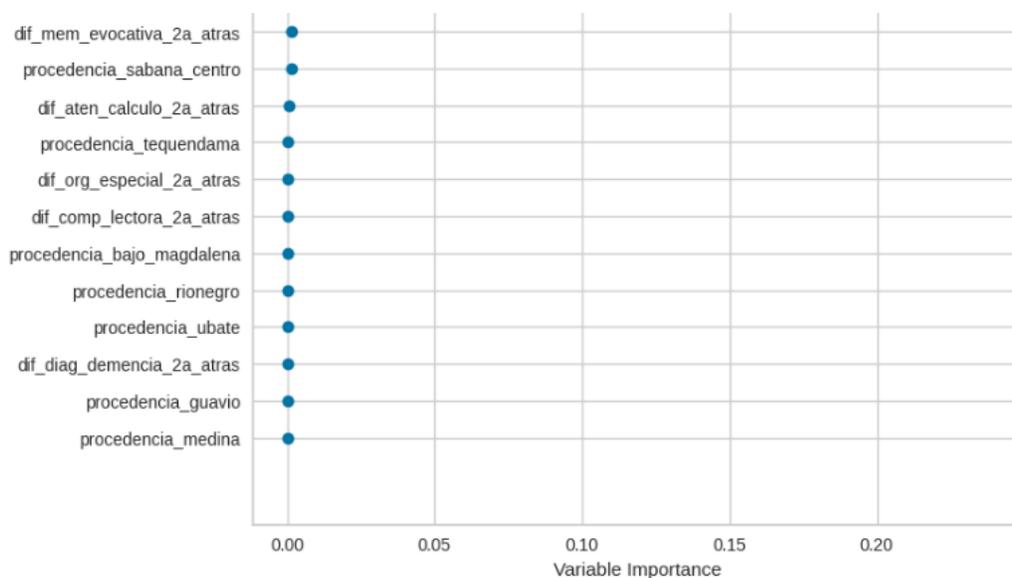


Figura S5. Gráfica de variables con poca relevancia explicativa



Figura S6. En esta gráfica puede verse la distribución de los datos que componen las métricas del modelo.

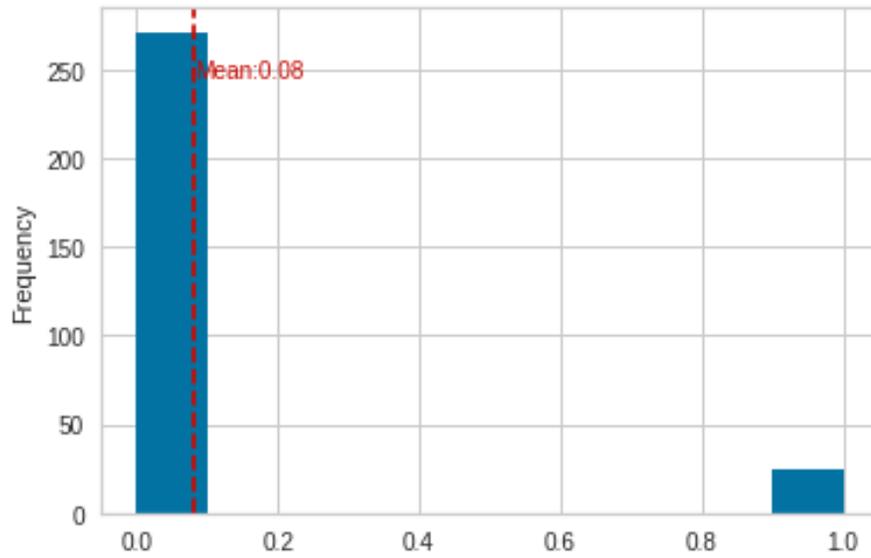


Figura S7. En esta gráfica puede verse que a partir de la media, en dirección a la derecha, comenzarán los valores de predicción.



REFERENCIAS BIBLIOGRÁFICA

Abad, V. M., Santaefemia, F. C., & Yáñez, S. R. (2020). Aplicación de Técnicas de Machine Learning para la Identificación de Patrones de Planificación de la FCR.

Bargalló, E. V., Domènech, J. V., Pallàs, J. A., Boquet, J. E., Pueyo, T. A., & Ramírez, E. L. (2002). Concordancia entre el Mini-Examen Cognoscitivo y el Mini-Mental State Examination en el cribado del déficit cognitivo. *Atención primaria*, 30(1), 5-13.

Contador, I., Fernández-Calvo, B., Ramos, F., Tapias-Merino, E., & Bermejo-Pareja, F. (2010). El cribado de la demencia en atención primaria. Revisión crítica. *Rev Neurol*, 51(11), 677-686.

Custodio, N., Becerra-Becerra, Y., Cruzado, L., Castro-Suárez, S., Montesinos, R., Bardales, Y., ... & Lira, D. (2018). Nivel de conocimientos sobre demencia frontotemporal en una muestra de médicos que evalúan regularmente a pacientes con demencia en Lima-Perú. *Revista chilena de neuro-psiquiatría*, 56(2), 77-88.

Custodio, N., Herrera, E., Lira, D., Montesinos, R., Linares, J., & Bendejú, L. (2012, October). Deterioro cognitivo leve:¿ dónde termina el envejecimiento normal y empieza la demencia?. In *Anales de la Facultad de Medicina* (Vol. 73, No. 4, pp. 321-330). UNMSM. Facultad de Medicina.

del Carmen, C. V. M. (2018). *La atención de la locura en la Beneficencia de Cundinamarca, durante el periodo 1950-1970. Una lectura desde las historias clínicas* (No. 016340). Universidad del Valle-CIDSE.

Díaz Cabezas, R., Marulanda Mejía, F., & Martínez Arias, M. H. (2013). Prevalencia de deterioro cognitivo y demencia en mayores de 65 años en una población urbana colombiana. *Acta Neurologica Colombiana*, 29(3), 141-151.

Fernández Cobas, H. (2021). Análisis de los factores de riesgo de la enfermedad del Alzheimer y su detección temprana mediante machine learning.



LOS LIBERTADORES
FUNDACIÓN UNIVERSITARIA

González Veloza, J. J. F. (2022). Análisis estadístico para la predicción de la probabilidad de encontrar adultos mayores desaparecidos mediante la aplicación de modelos de aprendizaje automático.

Gualdrón, D. F. R., Martínez, M. R., Márquez, M. C., Vallejo, C., Llano, A. M. S., Martínez, J., & Bernal, D. R. (2022). Valoración crítica de la guía colombiana de práctica clínica para la demencia con los instrumentos AGREE-II y el AGREE-REX. *Revista Colombiana de Psiquiatría*.

Subirana, J., Bruna, O., Puyuelo, M., & Virgili, C. (2009). Lenguaje y funciones ejecutivas en la valoración inicial del deterioro cognitivo leve y la demencia tipo Alzheimer. *Revista de logopedia, foniatría y audiología*, 29(1), 13-20.

Wechsler, D. (2002). WAIS-III: test de inteligencia para adultos de Wechsler.