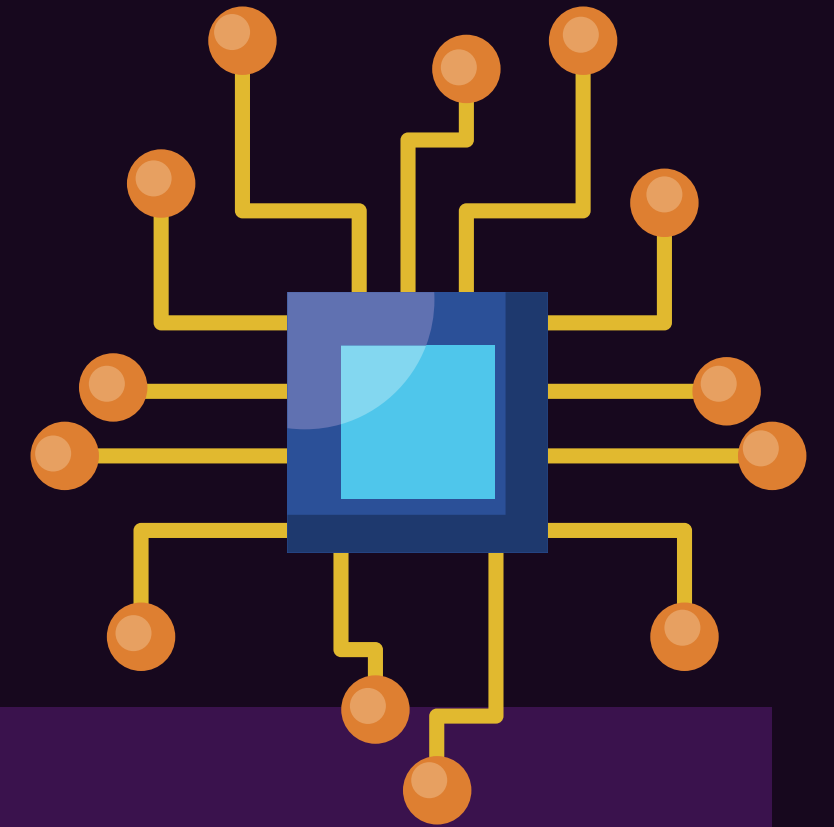


DETECCIÓN DE NIVELES DE ALZHEIMER SEGÚN CARACTERÍSTICAS EXTRAÍDAS DE IMÁGENES MÉDICAS

Grupo 4:

- Javier Urbina
- Diego Segura

Contenidos generales



1 Problemática

2 Metodología

3 Resultados

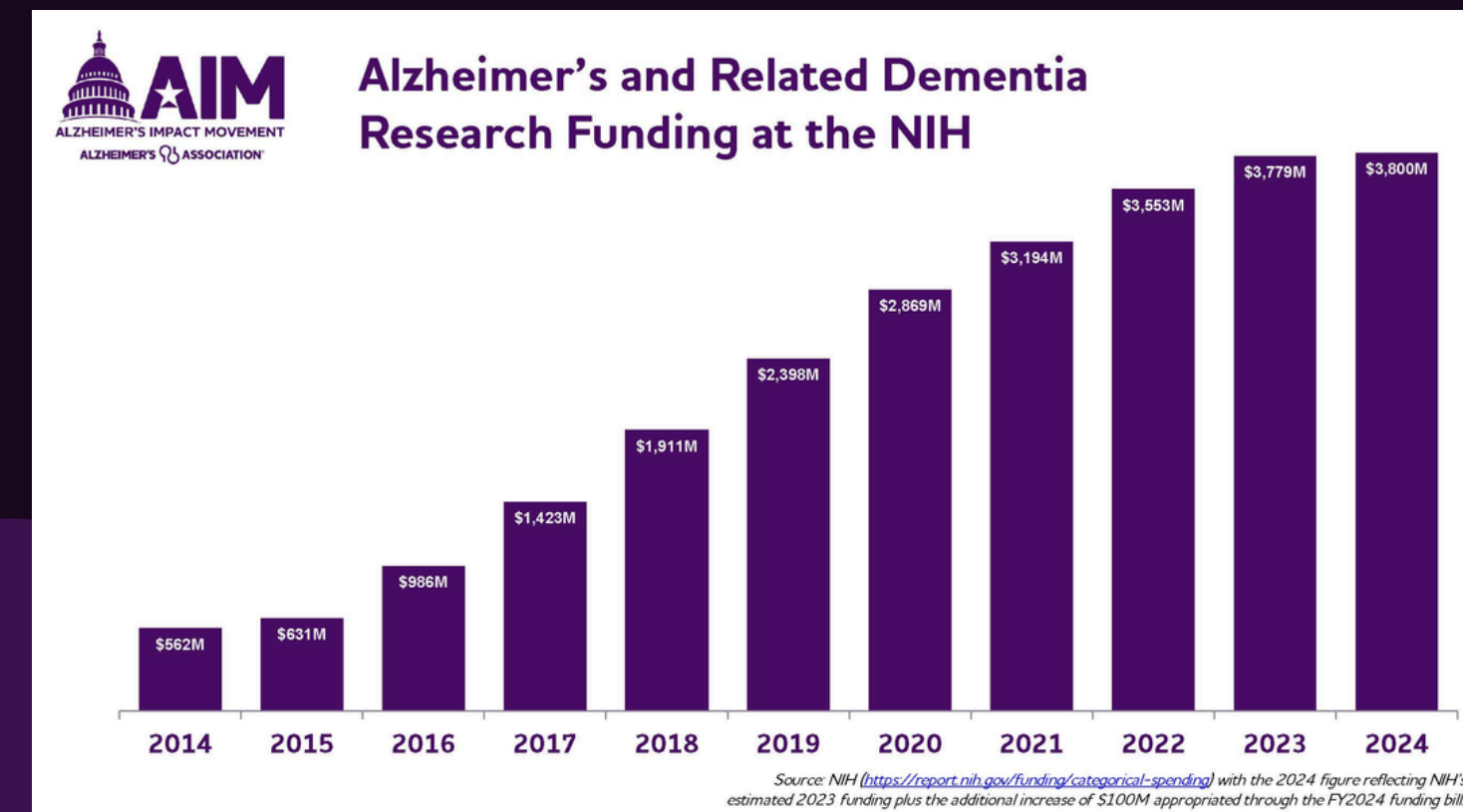
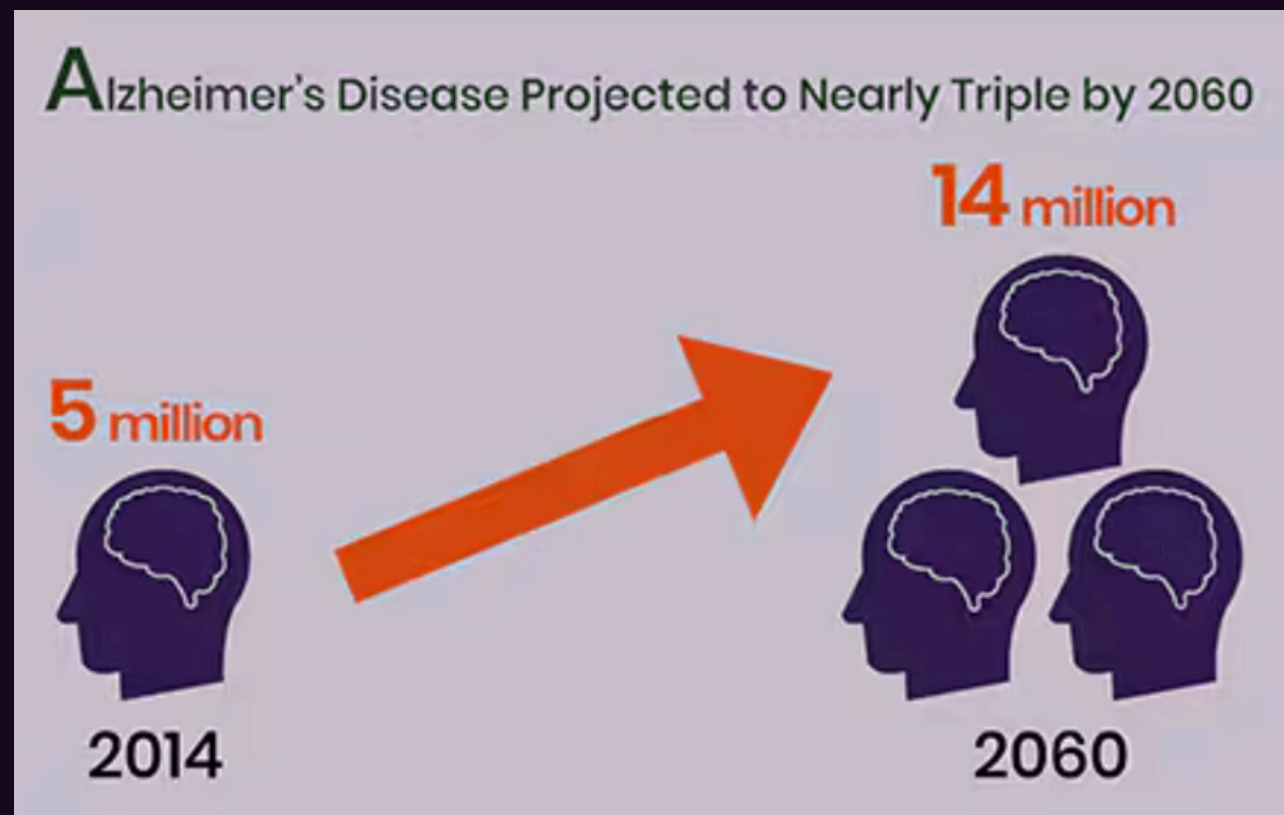
4 Discusión y Conclusiones

Problemática: Enfermedad de Alzheimer (EA)

Prevalencia Global: Aproximadamente 50 millones de personas padecen demencia en el mundo, con un aumento rápido en países de ingresos bajos y medios.

Diagnóstico: Imágenes de resonancia magnética (RM) son fundamentales para detectar cambios estructurales en el cerebro

Desafíos: La diferenciación precisa entre EA y el envejecimiento normal es complicada debido a síntomas superpuestos.



Metodología



1

Adquisición de datos

2

**Extracción de
características**

3

EDA

4

**Implementación del
modelo de ML**

5

**Entrenamiento y
evaluación del modelo**

6

Modelo Final

1. Adquisición de datos

The Open Access Series of Imaging Studies (OASIS)

Name	Description	Dimensions	Vox. size	Orient
OAS1_xxxx_MRy_mpr-z_anon	Individual scan (z=repetition)	256x256x128	1x1x1.25	Sag
OAS1_xxxx_MRy_mpr_ni_anon_sbj_111	Image averaged across scans (i=# of scans)	256x256x160	1x1x1	Sag
OAS1_xxxx_MRy_mpr_ni_anon_111_t88_gfc	Gain-field corrected atlas registered average	176x208x176	1x1x1	Trans
OAS1_xxxx_MRy_mpr_ni_anon_111_t88_masked_gfc	Brain-masked version of atlas registered image	176x208x176	1x1x1	Trans
OAS1_xxxx_MRy_mpr_ni__anon 111_t88_masked_gfc_fseg	Brain tissue segmentation	176x208x176	1x1x1	Trans

1. Adquisición de datos

The Open Access Series of Imaging Studies (OASIS)

Table 2. Age and Diagnosis Characteristics of the Data Set										
Age Group	Total n	Without Dementia				With Dementia				CDR 0.5/1/2
		n	Mean	Male	Female	n	Mean	Male	Female	
<20	19	19	18.53	10	9	0		0	0	0/0/0
20s	119	119	22.82	51	68	0		0	0	0/0/0
30s	16	16	33.38	11	5	0		0	0	0/0/0
40s	31	31	45.58	10	21	0		0	0	0/0/0
50s	33	33	54.36	11	22	0		0	0	0/0/0
60s	40	25	64.88	7	18	15	66.13	6	9	12/3/0
70s	83	35	73.37	10	25	48	74.42	20	28	32/15/1
80s	62	30	84.07	8	22	32	82.88	13	19	22/9/1
≥90	13	8	91.00	1	7	5	92.00	2	3	4/1/0
Total	416	316		119	197	100		41	59	

Extracción de características

Selección de Cortes Axiales

Se enfocaron en los cortes axiales del cerebro, promediando los slices de las secciones 70 a 100

Coeficientes de Wavelet

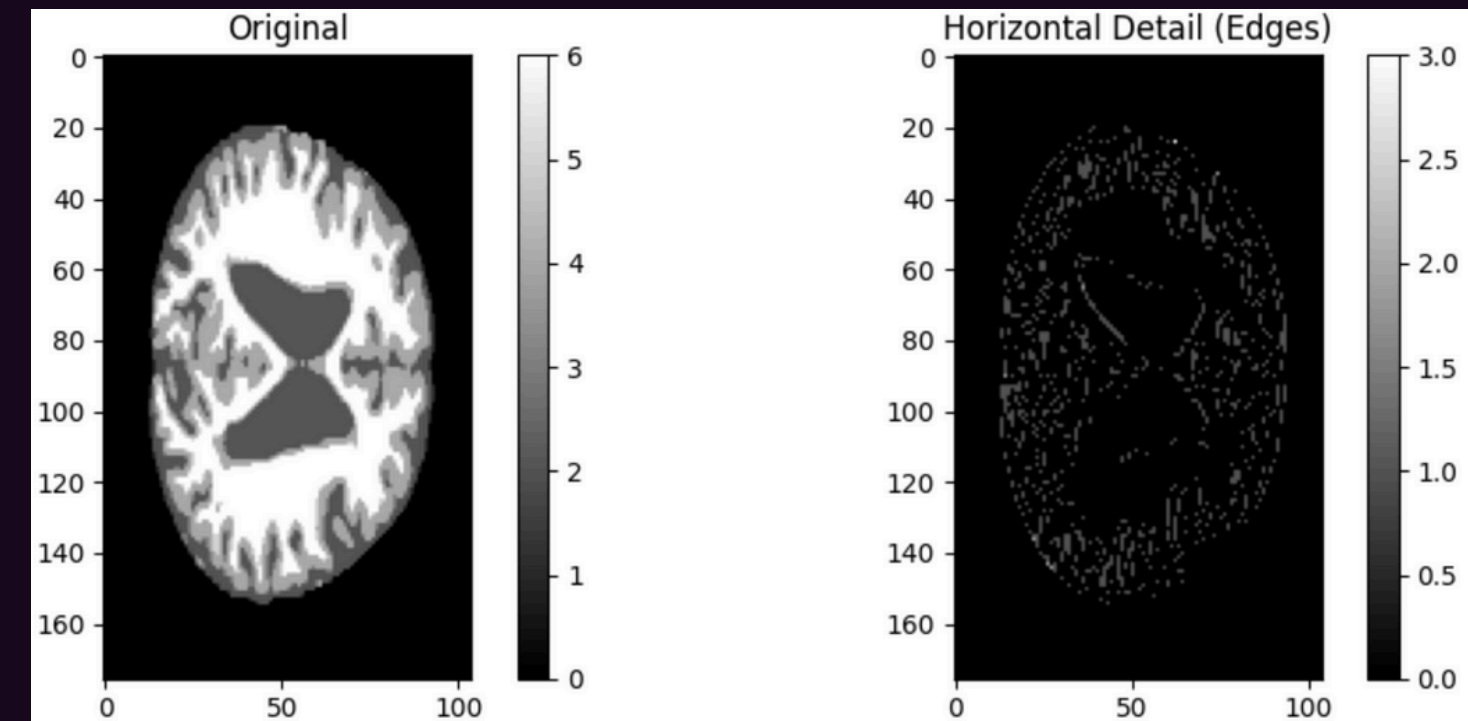
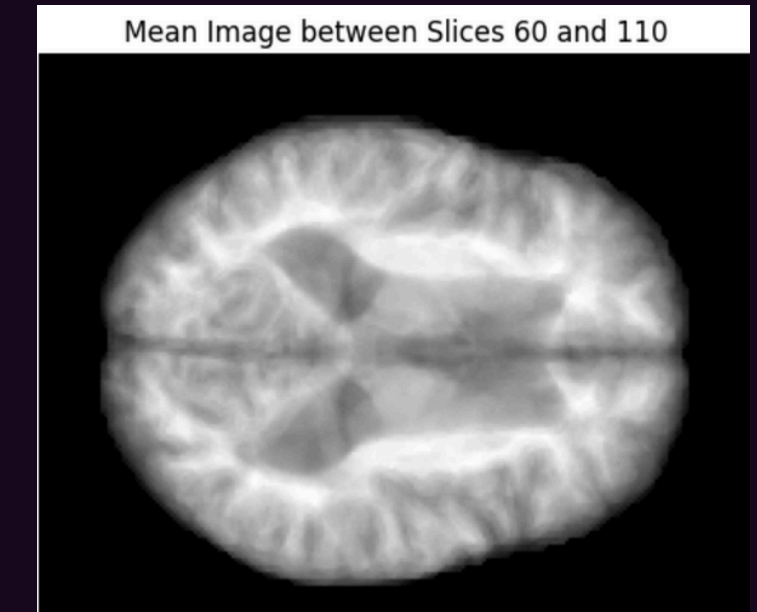
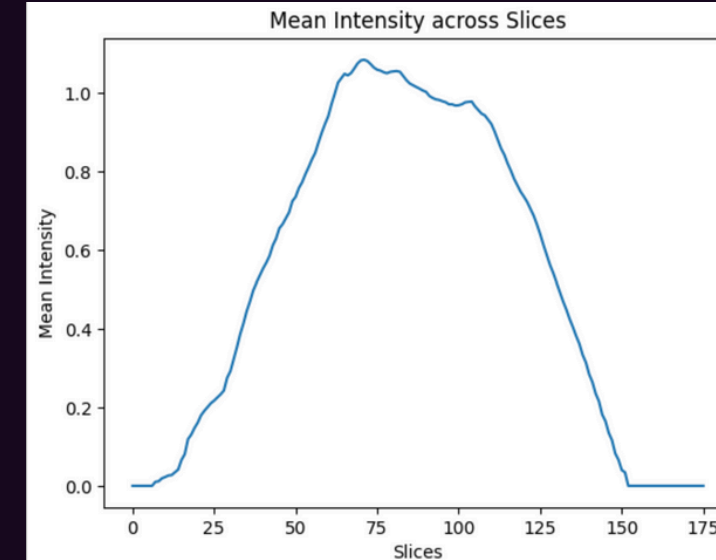
Descomponer la señal de imagen en diferentes frecuencias y escalas, proporcionando detalles sobre textura y bordes.

GLCM (Matriz de co-ocurrencia de niveles de gris)

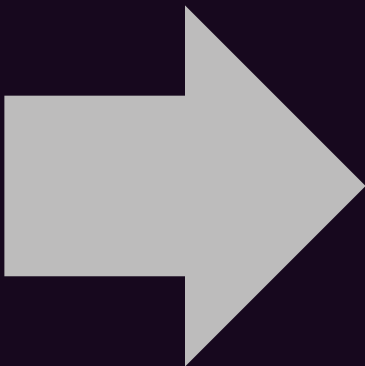
Analizar la distribución y relación de intensidades de píxeles adyacentes, caracterizando la textura y homogeneidad de las regiones cerebrales.

Volúmenes de regiones cerebrales

Se calculó la proporción de materia gris con respecto a la materia blanca en el cerebro (GWR)



Dataframe original (OASIS)



Dataframe con características añadidas

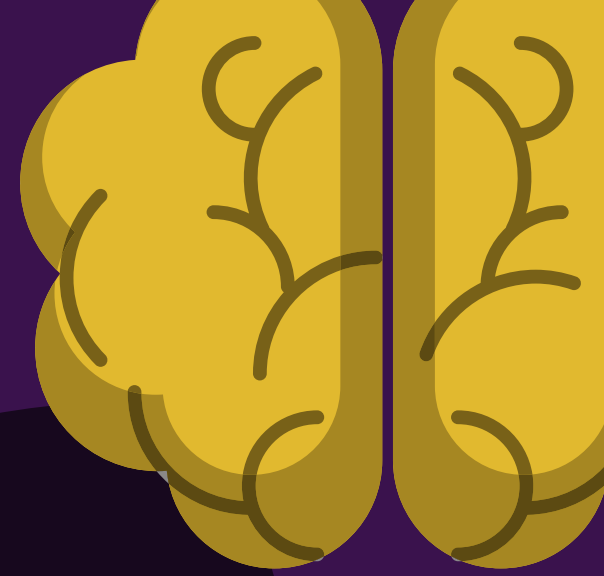


Coeficientes de Wavelet
GLCM

Volúmenes de regiones cerebrales
(15 características añadidas)

Table 4. Measures Included in the Data Set	
Age	Age at time of image acquisition (years)
Sex	Sex (male or female)
Education	Years of education
Socioeconomic status	Assessed by the Hollingshead Index of Social Position and classified into categories from 1 (<i>highest status</i>) to 5 (<i>lowest status</i>) (Hollingshead, 1957)
MMSE score	Ranges from 0 (<i>worst</i>) to 30 (<i>best</i>) (Folstein, Folstein, & McHugh, 1975)
CDR scale	0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2 = moderate AD (Morris, 1993)
Atlas scaling factor	Computed scaling factor (unitless) that transforms native-space brain and skull to the atlas target (i.e., the determinant of the transform matrix) (Buckner et al., 2004)
eTIV	Estimated total intracranial volume (cm ³) (Buckner et al., 2004)
nWBV	Expressed as the percent of all voxels in the atlas-masked image that are labeled as gray or white matter by the automated tissue segmentation process (Fotenos et al., 2005)

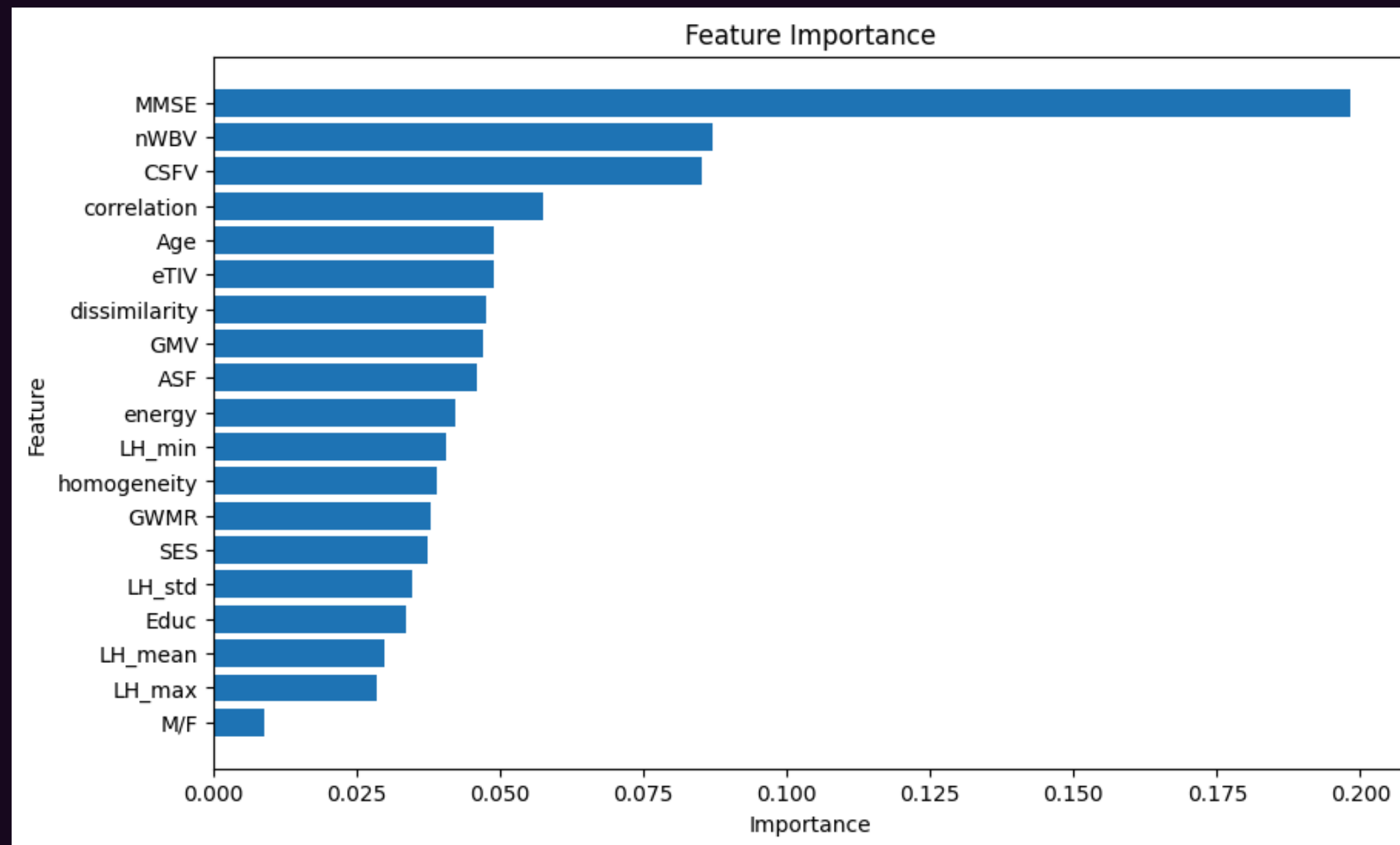
EDA



Implementación del Modelo

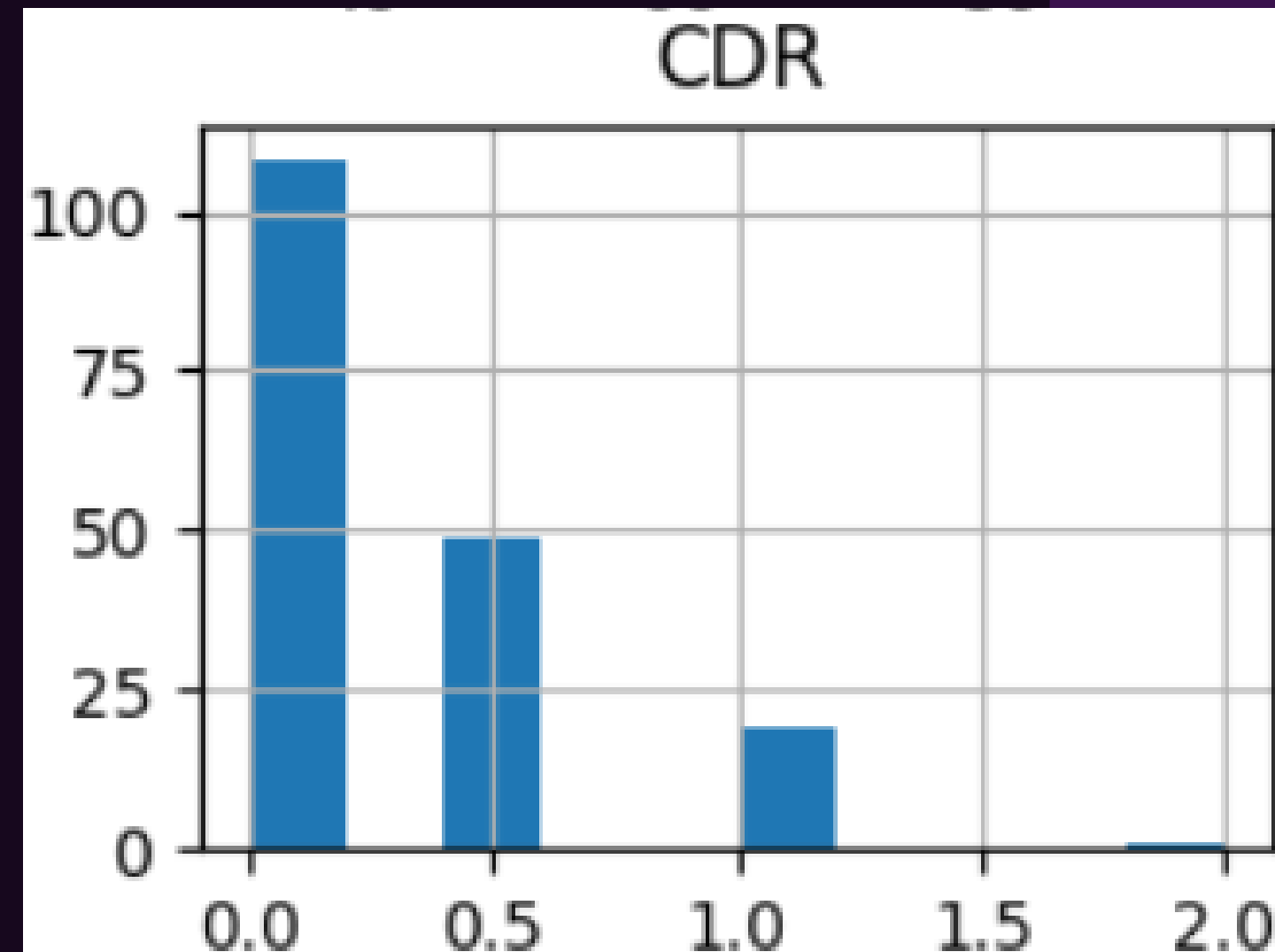
EXTRACCIÓN DE CARACTERÍSTICAS

FEATURE IMPORTANCE



Implementación del Modelo

CATEGORIZAR EL TARGET: CDR



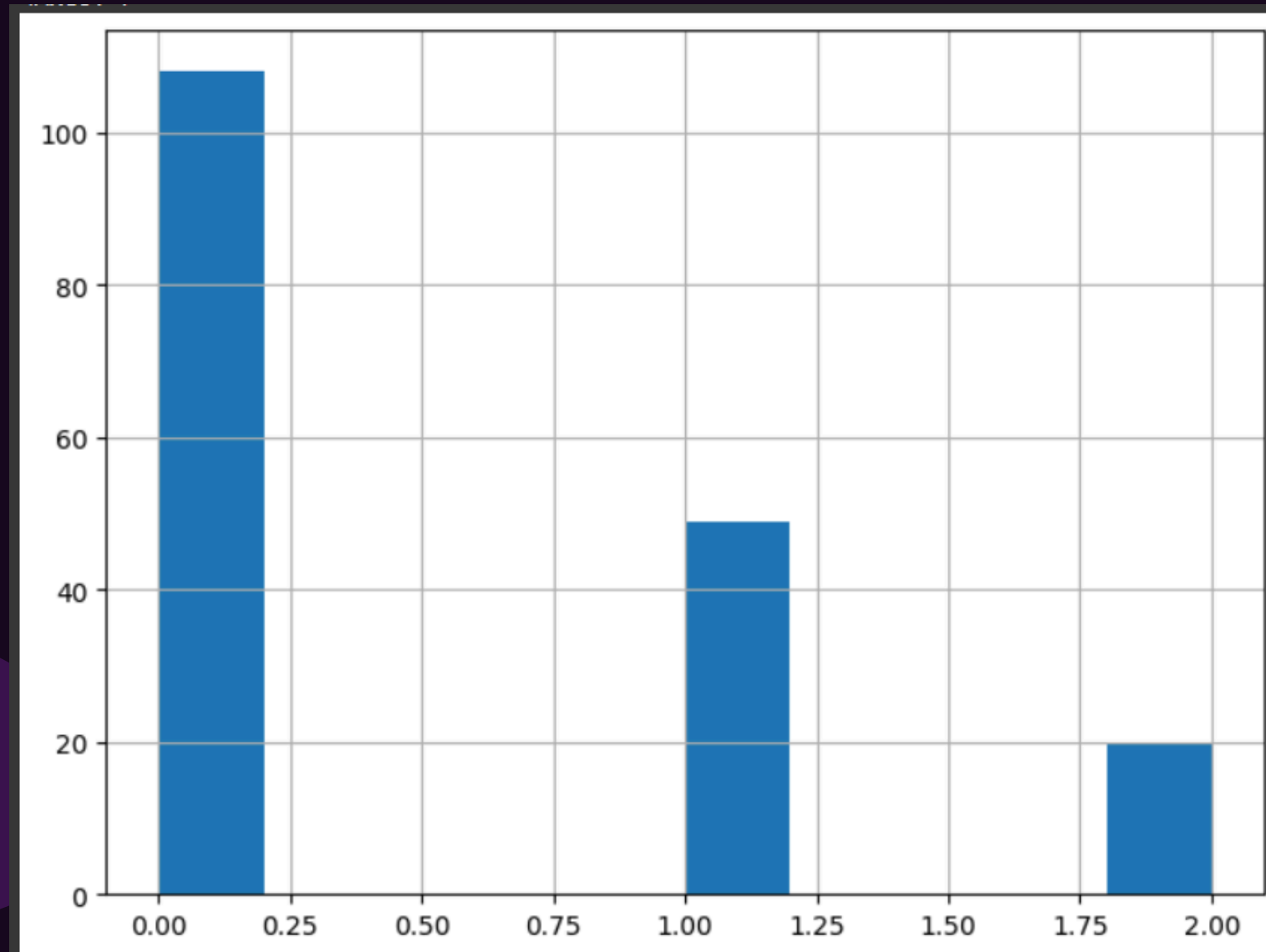
Eliminar categoria

```
expanded_features_df0['CDR'] = expanded_features_df0['CDR'].replace(2, 1.)
```

Rescalar

```
cdr_mapping = {0: 0, 0.5: 1, 1: 2}# {0: 'no dementia', 0.5: 'very mild AD', 1: 'mild AD'}  
  
# Replace values in the target column "CDR"  
expanded_features_df0['CDR'] = expanded_features_df0['CDR'].map(cdr_mapping )
```

CDR actualizado



NECESIDAD DE BALANCEAR LOS DATOS

```
# Evaluating each sampling strategy
sampling_strategies = {
    'SMOTEENN': SMOTEENN(random_state=42),
    'ADASYN': ADASYN(random_state=42),
    'RandomOverSampler': RandomOverSampler(random_state=42),
    'SMOTE': SMOTE(random_state=42),
    'BorderlineSMOTE': BorderlineSMOTE(random_state=42),
    'SVMSMOTE': SVMSMOTE(random_state=42),
    'KMeansSMOTE': KMeansSMOTE(random_state=42)
}
```

Evaluación de un modelo de clasificación utilizando un clasificador Random Forest

- Equilibrar clases
- División de 70/30
- Entrenar el modelo con las características seleccionadas
- Evalua el modelo
- Validación cruzada



Busqueda de hiperparámetros

```
grid_search = GridSearchCV(estimator=rf_classifier, param_grid=param_grid, cv=5, scoring='accuracy', n_jobs=-1)
```

Mejores hiperparámetros: {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 300}

Resultados RF con mejores hiperparámetros

Accuracy: 0.6979166666666666

Classification Report:

	precision	recall	f1-score	support
0	0.61	0.66	0.63	29
1	0.77	0.56	0.65	41
2	0.71	0.96	0.82	26
accuracy			0.70	96
macro avg	0.70	0.73	0.70	96
weighted avg	0.71	0.70	0.69	96

for SMOTEENN

Results for SMOTEENN

Accuracy: 0.975609756097561

Classification Report:

	precision	recall	f1-score	support
0	1.00	0.93	0.97	15
1	0.90	1.00	0.95	9
2	1.00	1.00	1.00	17
accuracy			0.98	41
macro avg	0.97	0.98	0.97	41
weighted avg	0.98	0.98	0.98	41

for RandomOverSampler

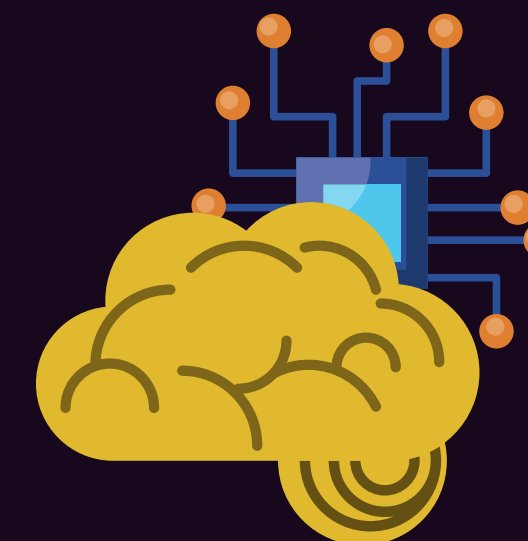
Results for RandomOverSampler

Accuracy: 0.8367346938775511

Classification Report:

	precision	recall	f1-score	support
0	0.81	0.70	0.75	30
1	0.81	0.81	0.81	37
2	0.89	1.00	0.94	31
accuracy			0.84	98
macro avg	0.83	0.84	0.83	98
weighted avg	0.83	0.84	0.83	98

Resultados



for ADASYN

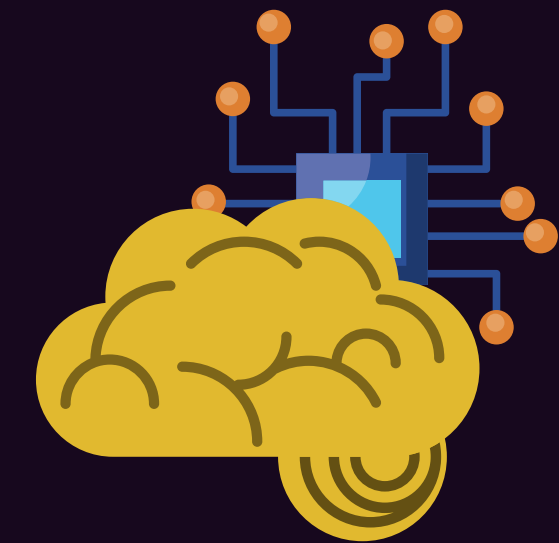
Results for ADASYN

Accuracy: 0.7835051546391752

Classification Report:

	precision	recall	f1-score	support
0	0.70	0.72	0.71	29
1	0.81	0.66	0.72	38
2	0.83	1.00	0.91	30
accuracy			0.78	97
macro avg	0.78	0.79	0.78	97
weighted avg	0.78	0.78	0.78	97

Resultados



for SMOTE

Results for SMOTE

Accuracy: 0.6836734693877551

Classification Report:

	precision	recall	f1-score	support
0	0.65	0.57	0.61	30
1	0.62	0.62	0.62	37
2	0.77	0.87	0.82	31
accuracy			0.68	98
macro avg	0.68	0.69	0.68	98
weighted avg	0.68	0.68	0.68	98

for SVM SMOTE

Results for SVM SMOTE

Accuracy: 0.7555555555555555

Classification Report:

	precision	recall	f1-score	support
0	0.75	0.77	0.76	31
1	0.68	0.52	0.59	25
2	0.79	0.91	0.85	34
accuracy			0.76	90
macro avg	0.74	0.74	0.73	90
weighted avg	0.75	0.76	0.75	90

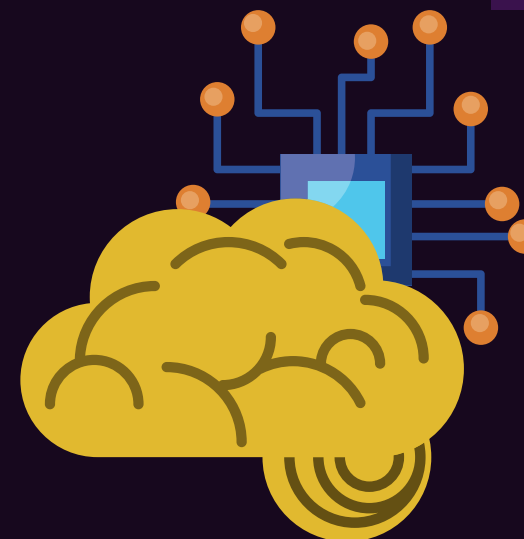
for Borderline SMOTE

Results for Borderline SMOTE

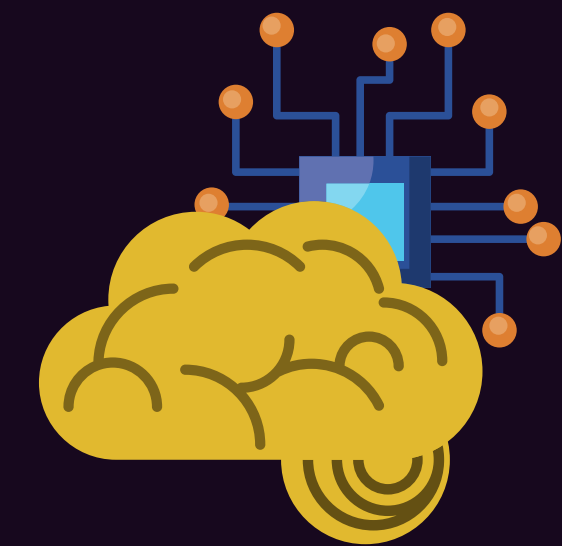
Accuracy: 0.6836734693877551

Classification Report:

	precision	recall	f1-score	support
0	0.67	0.60	0.63	30
1	0.64	0.62	0.63	37
2	0.74	0.84	0.79	31
accuracy			0.68	98
macro avg	0.68	0.69	0.68	98
weighted avg	0.68	0.68	0.68	98



Resultados



for KMeansSMOTE

Results for KMeansSMOTE

Accuracy: 0.7474747474747475

Classification Report:

	precision	recall	f1-score	support
0	0.68	0.68	0.68	31
1	0.69	0.69	0.69	36
2	0.88	0.88	0.88	32
accuracy			0.75	99
macro avg	0.75	0.75	0.75	99
weighted avg	0.75	0.75	0.75	99


Discusión

Clasificador	Características	Base de datos	Resultados
K-NN y ADABOOST (Kamathe et al.)	GLCM	OASIS	Acc: KNN 92.75%, ADABOOST 100%
Random Forest, SVM, Logistic Regression and XGBoost (Sarica et al.)	Datos demográficos de ADNI, biomarcadores, MMSE, volumen de hipocampo y ventrículos	ADNI	Acc: RF 93.6%, SVM 90%, Logistic Regression 89.2%, XGBoost 92% Precisión: RF 95.2%, SVM 90%, Logistic Regression 90.3%, XGBoost 93%
SVM, IVM and regularized extreme learning machine (RELM) (Lama et al.)	Grosor de la corteza cerebral, volúmenes, curvatura intrínseca, área de superficie	ADNI	Acc: SVM 80.32%, IVM 74.47%, RELM 77.88% Sen: SVM 83.37%, IVM 87.10%, RELM 68.85%
Random Forest (presente estudio)	Datos demográficos de OASIS, GLCM, coeficientes de Wavelet, volúmenes (materia gris, materia blanca y líquido cerebroespinal) y ratio de volúmenes de materia gris y materia blanca	OASIS	Acc: RF 97.56% Precisión: RF 97% Sen: RF 98%

Conclusiones

- Nuestro clasificador RF mostró un rendimiento similar o incluso superior en comparación con otros enfoques como SVM, XGBoost y Regresión Logística.
- Comparamos métodos de sobremuestreo para abordar el desequilibrio de clases en el conjunto de datos OASIS, lo cual es relevante en la investigación médica debido a que muchos conjuntos de datos médicos presentan un desequilibrio similar.

Limitaciones

- Base de datos con un desbalance de datos notorio
 - Evaluar el desempeño del modelo con otros datasets
 - Análisis con datos 2D (imágenes) y no con volúmenes
- 

Referencias

- Alzheimer's Disease International, World Alzheimer Report 2018—The State of the Art of Dementia Research: New Frontiers. London, UK: Alzheimer's Disease International, 2018, vol. 48.
- J. Huang, P. C. M. van Zijl, X. Han, C. M. Dong, G. W. Y. Cheng, K.-H. Tse, L. Knutsson, L. Chen, J. H. C. Lai, E. X. Wu, J. Xu, and K. W. Y. Chan, "Altered d-glucose in brain parenchyma and cerebrospinal fluid of early alzheimer's disease detected by dynamic glucose-enhanced MRI," *Sci. Adv.*, vol. 6, no. 20, p. eaba3884, May 2020
- R. K. Lama, J. Gwak, J.-S. Park, and S.-W. Lee, "Diagnosis of alzheimer's disease based on structural mri images using a regularized extreme learning machine and pca features," *Journal of Healthcare Engineering*, vol. 2017, p. 1–11, 2017. [Online]. Available: <http://dx.doi.org/10.1155/2017/5485080>
- R. S. Kamathe and K. R. Joshi, "A robust optimized feature set based automatic classification of alzheimer's disease from brain mr images using k-nn and adaboost," *ICTACT Journal on Image and Video Processing*, vol. 8, no. 3, p. 1665–1672, Feb. 2018. [Online]. Available: <http://dx.doi.org/10.21917/ijivp.2017.0234>
- A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in alzheimer's disease: A systematic review," *Frontiers in Aging Neuroscience*, vol. 9, Oct. 2017. [Online]. Available: <http://dx.doi.org/10.3389/fnagi.2017.00329>