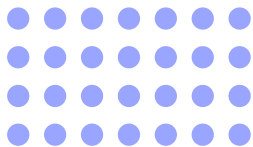


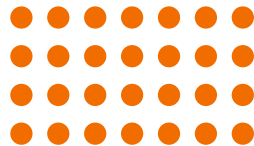
# EXPRESIÓN DE PROTEÍNAS EN EL PROGRESO DE LA ENFERMEDAD DE PARKINSON

Fiorella Ojeda & Carlos Pariona





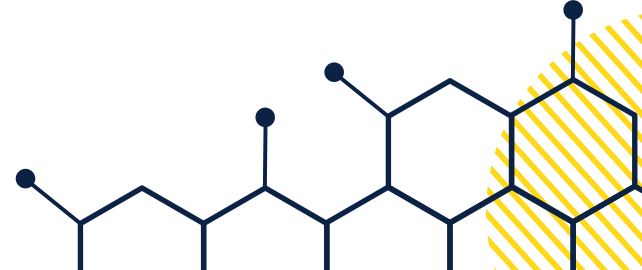
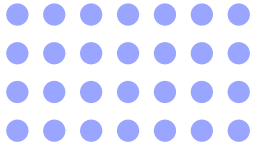
# INTRODUCCIÓN



# ENFERMEDAD DE PARKINSON



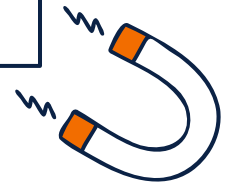
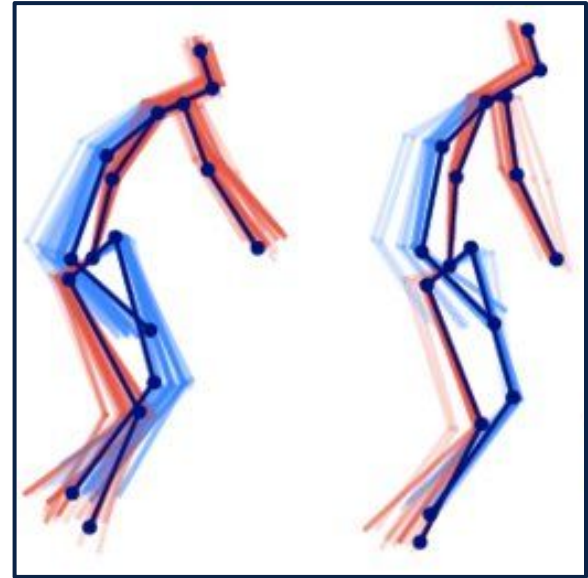
1. Condición crónica neurodegenerativa que afecta a millones de personas en el mundo y es la segunda enfermedad neurodegenerativa más común.
2. Se caracteriza por la degeneración progresiva de las células nerviosas en una región del cerebro, resultando en deficiencia de dopamina, crucial para el control del movimiento y la coordinación muscular.
3. Ciertas proteínas, como la alfa-sinucleína, desempeñan un papel importante en su desarrollo y progresión.



# UNIFIED PARKINSON'S DISEASE RATING SCALE

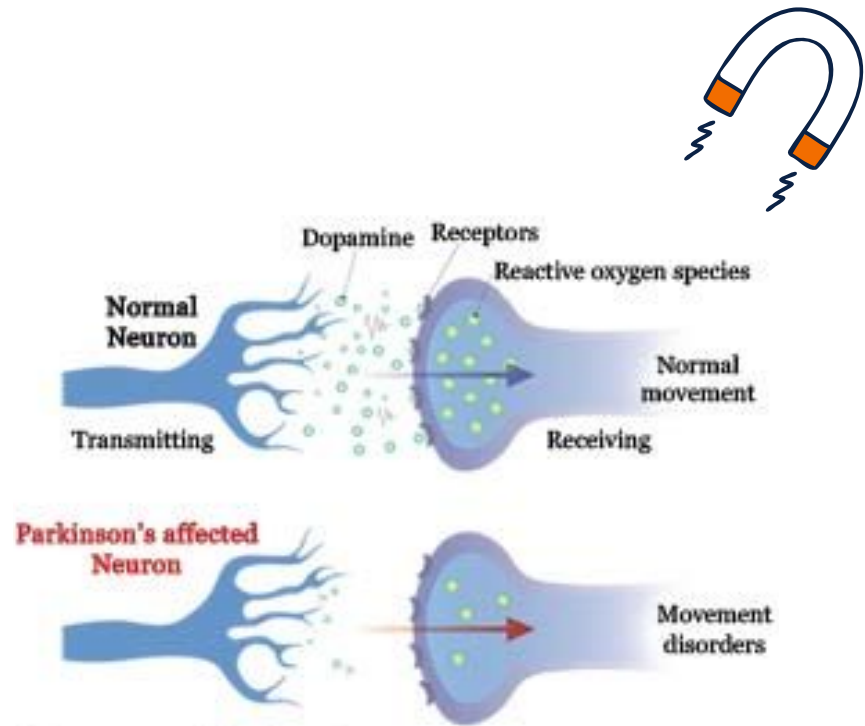
La Escala Unificada de Calificación de la Enfermedad de Parkinson (UPDRS) es una herramienta de evaluación estandarizada para evaluar y cuantificar con precisión los síntomas motores y no motores en pacientes con Parkinson. Evalúa varios dominios:

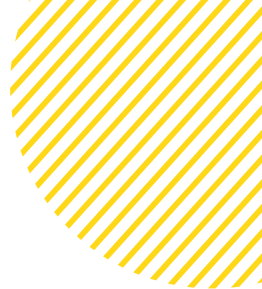
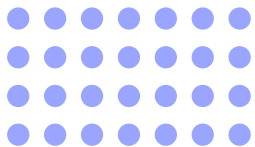
- Función motora
- Actividades de la vida diaria
- Complicaciones motoras
- Síntomas no motores: depresión, ansiedad y calidad de vida.



# OBJETIVOS DEL PROYECTO

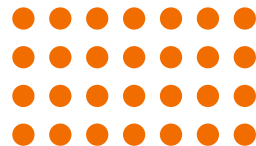
Explorar el potencial predictivo de las proteínas en la enfermedad de Parkinson para la detección temprana y el desarrollo de tratamientos personalizados.





02

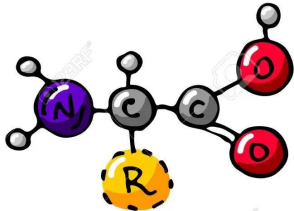
# METODOLOGÍA



# PROCESAMIENTO Y MODELOS

## PASO I

Características de  
la data de  
proteínas

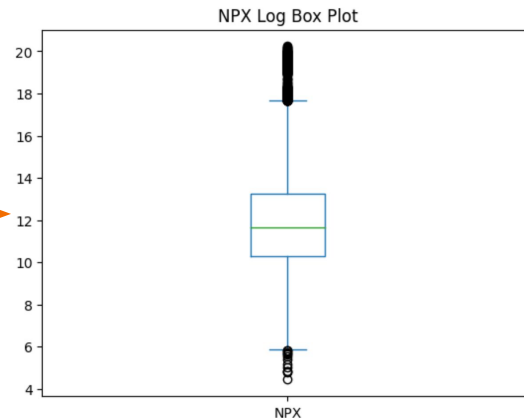


# CARACTERÍSTICAS DE LA DATA DE PROTEÍNAS

Proteins dataframe has # Rows: 232741, # Columns: 5

	visit_id	visit_month	patient_id	UniProt	NPX
0	55_0	0	55	O00391	11254.3
1	55_0	0	55	O00533	732430.0
2	55_0	0	55	O00584	39585.8
3	55_0	0	55	O14498	41526.9
4	55_0	0	55	O14773	31238.0

np.log



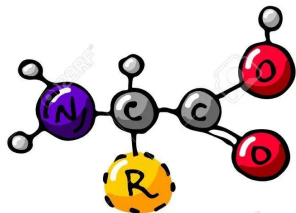
	visit_id	000391	000533	000584	014498	014773	014791	015240	015394	043505	...
0	55_0	9.328506	13.504123	10.586226	10.634097	10.349391	8.343485	12.088274	11.049273	12.717026	...
1	1517_0	9.362967	12.945662	10.464639	0.000000	9.768029	7.931788	11.684793	10.603484	12.021681	...
2	1923_0	9.969360	13.672777	10.374587	0.000000	9.501001	0.000000	12.502861	11.309342	12.662674	...
3	2660_0	0.000000	13.270489	10.249196	10.305052	9.859227	7.541476	11.808643	11.165703	12.270431	...
4	3636_0	9.295325	12.496508	9.340526	9.661180	8.801795	7.993907	11.027691	10.163071	11.193212	...



# PROCESAMIENTO Y MODELOS

## PASO I

Características de  
la data de  
proteínas



## PASO II

Características de  
la data clínica

# CARACTERÍSTICAS DE LA DATA CLÍNICA

Clinical dataframe has # Rows: 2615, # Columns: 8

	visit_id	patient_id	visit_month	updrs_1	updrs_2	updrs_3	updrs_4	upd23b_clinical_state_on_medication
0	55_0	55	0	10.0	6.0	15.0	NaN	NaN
1	55_3	55	3	10.0	7.0	25.0	NaN	NaN
2	55_6	55	6	8.0	10.0	34.0	NaN	NaN
3	55_9	55	9	8.0	9.0	30.0	0.0	On
4	55_12	55	12	10.0	10.0	41.0	0.0	On

```
visit_id      0
patient_id    0
visit_month    0
updrs_1        1
updrs_2        2
updrs_3       25
updrs_4     1038
upd23b_clinical_state_on_medication 1327
```

$\text{updrs\_1} + \text{updrs\_2} + \text{updrs\_3} + \text{updrs\_4}$

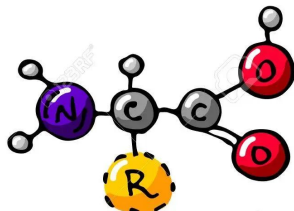
$\text{updrs\_1} + \text{updrs\_2}$

	visit_id	updrs_1	updrs_2	updrs_3	updrs_4	upd23b_clinical_state_on_medication	updrs_non_motor	updrs
0	55_0	10	6	15	<NA>	NaN	16	<NA>
1	55_3	10	7	25	<NA>	NaN	17	<NA>
2	55_6	8	10	34	<NA>	NaN	18	<NA>
3	55_9	8	9	30	0	On	17	47
4	55_12	10	10	41	0	On	20	61

# PROCESAMIENTO Y MODELOS

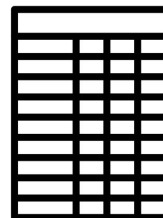
## PASO I

Características de  
la data de  
proteínas



## PASO III

Generación del  
conjunto de datos  
de entrenamiento



## PASO II

Características de  
la data clínica

# GENERACIÓN DE BD DE ENTRENAMIENTO

## NPX & MDS-UPDRS

	000391	000533	000584	014498	014773	014791	015240	015394	043505	060888	...
0	0.013127	0.670496	0.763220	0.980540	1.045347	0.723823	0.493669	0.411514	1.117953	0.392114	...
1	0.016781	-0.230696	0.533761	-0.001925	0.058720	0.099265	0.018511	-0.314723	-0.208763	-0.637695	...
2	0.081074	0.942653	0.363814	-0.001925	-0.394453	0.000186	0.981905	0.835193	1.014250	1.429801	...
3	-0.975931	0.293478	0.127174	0.311463	0.213493	-0.492851	0.164362	0.601191	0.265852	0.966835	...
4	0.009609	-0.955498	-1.587675	-0.997780	-1.581073	0.193502	-0.755321	-1.032201	-1.789479	-0.344913	...

5 rows × 228 columns

```
# Import KNNImputer
from sklearn.impute import KNNImputer

# Impute dataframe
imputer = KNNImputer()
imputed = imputer.fit_transform(df)
df_imputed = pd.DataFrame(imputed, columns=df.columns)
df_imputed.head()
```

...	Q9NQ79	Q9NYU2	Q9UBR2	Q9UBX5	Q9UH62	Q9UKV8	Q9UNU6	Q9Y646	Q9Y6R7	updrs
...	0.076783	-0.231475	0.911468	1.140102	1.095780	-0.000278	0.065964	0.540623	0.066043	28.6
...	-0.334073	-0.005700	-0.225343	0.278439	-0.442750	-0.142632	0.071873	0.094211	-0.281903	47.0
...	0.663979	0.250438	-0.023872	1.327897	1.436155	0.052761	0.050811	0.780066	0.366023	17.2
...	-0.036434	0.488983	-0.009597	-0.575823	0.538580	-0.150511	0.014832	0.714222	-0.005308	7.6
...	-0.003852	-1.002169	-1.137671	-1.516837	-0.698239	-0.000278	-0.967985	-0.002969	-1.017786	22.6

## NPX & VISIT MONTH

...	Q9NQ79	Q9NYU2	Q9UBR2	Q9UBX5	Q9UHG2	Q9UKV8	Q9UNU6	Q9Y646	Q9Y6R7	visit_time
...	0.076783	-0.231475	0.911468	1.140102	1.095780	-14.428781	0.065964	0.540623	0.066043	0
...	-0.334073	-30.470617	-0.225343	0.278439	-0.442750	-0.142632	0.071873	0.094211	-0.281903	0
...	0.663979	0.250438	-0.023872	1.327897	1.436155	0.052761	0.050811	0.780066	0.366023	0
...	-0.036434	0.488983	-0.009597	-0.575823	0.538580	-0.150511	0.014832	0.714222	-0.005308	0
...	-12.077402	-1.002169	-1.137671	-1.516837	-0.698239	-14.428781	-0.967985	-2.071915	-1.017786	0

# PROCESAMIENTO Y MODELOS

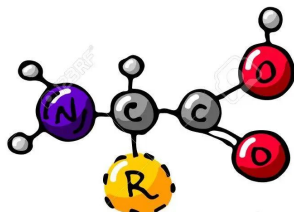
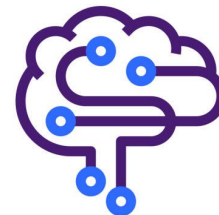
## PASO I

Características de  
la data de  
proteínas



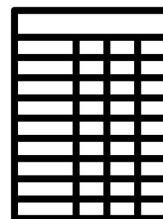
## PASO III

Generación del  
conjunto de datos  
de entrenamiento



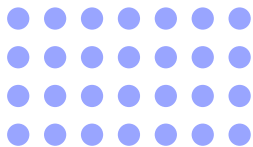
## PASO II

Características de  
la data clínica



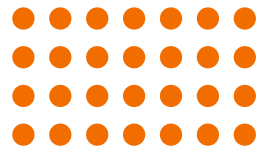
## PASO IV

Prueba de  
diferentes modelos



03

# RESULTADOS



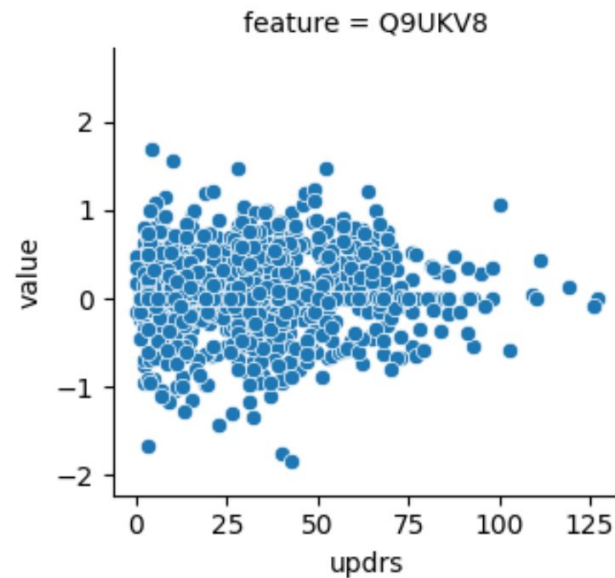
# NPX & MDS-UPDRS

Regression Model	R2 Score
<b>k-Nearest Neighbor</b>	<b>29.00 %</b>
Support Vector Machine	07.00 %
Decision Tree	00.00 %
Random Forest	24.00 %
Adaptive Boosting	14.00 %
Bootstrap Aggregating	13.00 %
Gradient Boosting	22.00 %
Histogram Gradient Boosting	26.00 %
Linear Regression	00.00 %
Multilayer Perceptron	11.00 %

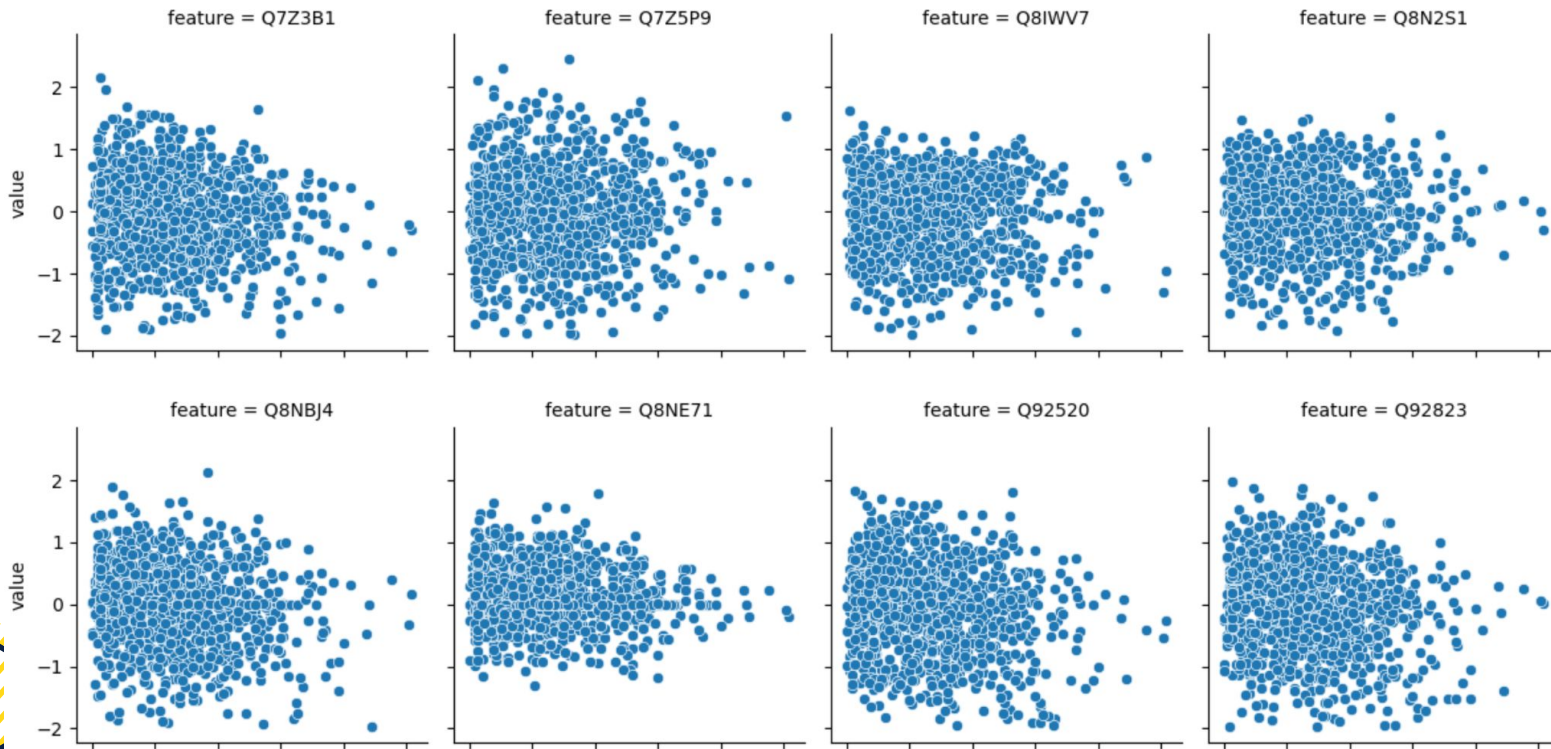


# Q9UKV8 PROTEIN DISTRIBUTION VS MDS-UPDRS

La distribución observada entre las proteínas y la escala MDS-UPDRS parece ser una nube dispersa sin ningún patrón.



# PROTEIN DISTRIBUTION VS MDS-UPDRS

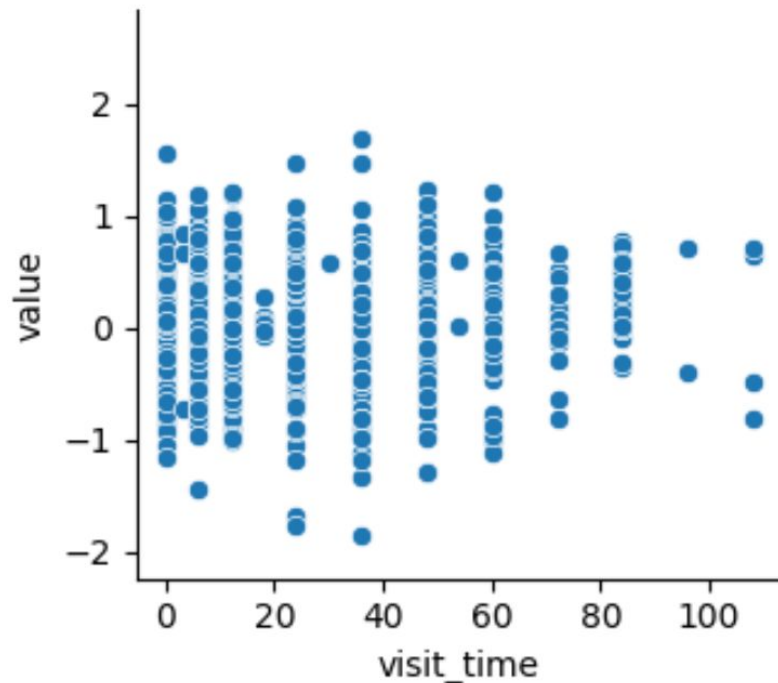


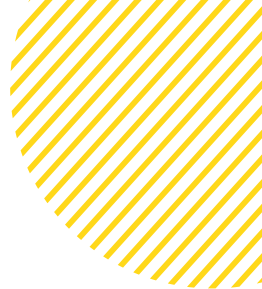
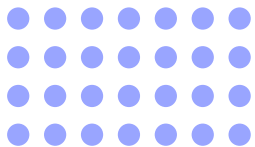
# NPX & VISIT MONTH

Classification Model	Accuracy
k-Nearest Neighbor	08.00 %
Support Vector Machine	12.00 %
Decision Tree	14.00 %
Random Forest	10.00 %
Adaptive Boosting	13.00 %
<b>Bootstrap Aggregating</b>	<b>16.00 %</b>
Gradient Boosting	11.00 %
Histogram Gradient Boosting	07.00 %
Multilayer Perceptron	07.00 %

# VISIT TIME DISTRIBUTION VS MDS-UPDRS

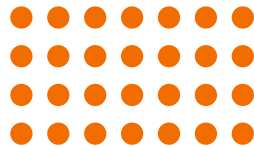
La distribución observada entre las proteínas y la escala MDS-Visit\_Time parece ser una nube dispersa sin ningún patrón.





04

# DISCUSIÓN



# ASPECTOS RELEVANTES

## ASOCIACIÓN PROTEÍNAS - EP

Los estudios revisados evidencian resultados inconclusos

01

02

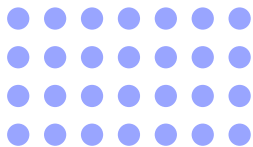
03

## RECOMENDACIONES A FUTURO

Abarcar enfermedades con evolución temporal y considerar el estado de la medicación con levodopa

## ENFOQUE METODOLÓGICO

Fortalezas y limitaciones de la metodología empleada

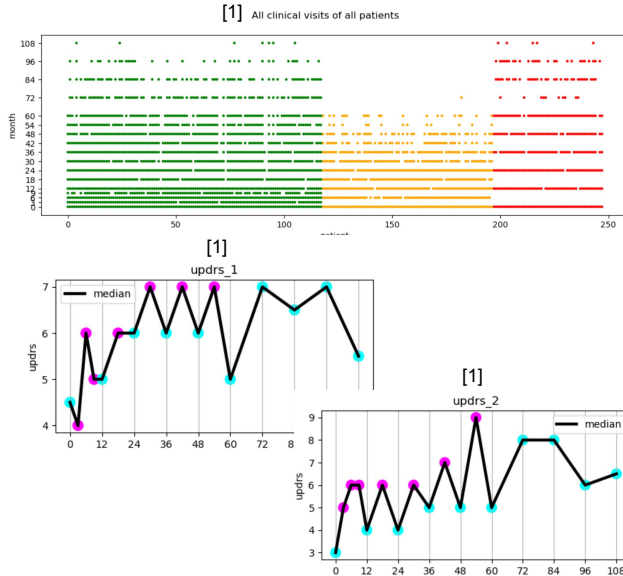


05

## CONCLUSIONES Y APRENDIZAJE



# CONCLUSIONES Y APRENDIZAJE



1. Para el approach escogido y con la data proporcionada, no hay evidencia suficiente para afirmar que exista una relación entre la expresión proteica y la evolución del parkinson.
2. Es importante realizar un análisis exhaustivo de los datos (EDA) y adaptarlo al conjunto de datos específico. Las herramientas automatizadas de análisis de datos no pueden descubrir la información oculta.
3. La disponibilidad y el costo de los datos médicos son limitados. Si no hemos logrado demostrar una conexión entre las proteínas y los síntomas de Parkinson, esto no significa que no exista.

[1] "AMP@Parkinson's Disease Progression Prediction | Kaggle". Kaggle: Your Machine Learning and Data Science Community.  
<https://www.kaggle.com/competitions/amp-parkinsons-disease-progression-prediction/discussion/411388>  
(accedido el 28 de junio de 2023).





# ¡GRACIAS!

¿TIENEN PREGUNTAS?

