

Alumnos:

Christian Andre Ayala Pichilingue

Leonardo Nicolás Zuñiga Dominguez

Italo Alexander Acuña Salas

Pregunta 1:

La base de datos se compartió como una forma de entrenar modelos de regresión, en este caso la cantidad de observaciones en total es de 442, ya que se tomaron datos de 442 pacientes. Como se mencionó, al ser un dataset que sirva como modelo para entrenar modelos de regresión, el target que debemos tener debe ser numérico, y en este caso la última columna, que vendría a ser el target o variable dependiente es un valor numérico, que en este caso se nos dice que vendría a ser una medida cuantitativa de la progresión de la enfermedad desde su primera medida.

Pregunta 2

Analizando medidas como media, mediana, desviación estándar y rangos para cada variable. También visualizando gráficas como histogramas, diagramas de caja y dispersión para identificar patrones, correlaciones y posibles valores atípicos.

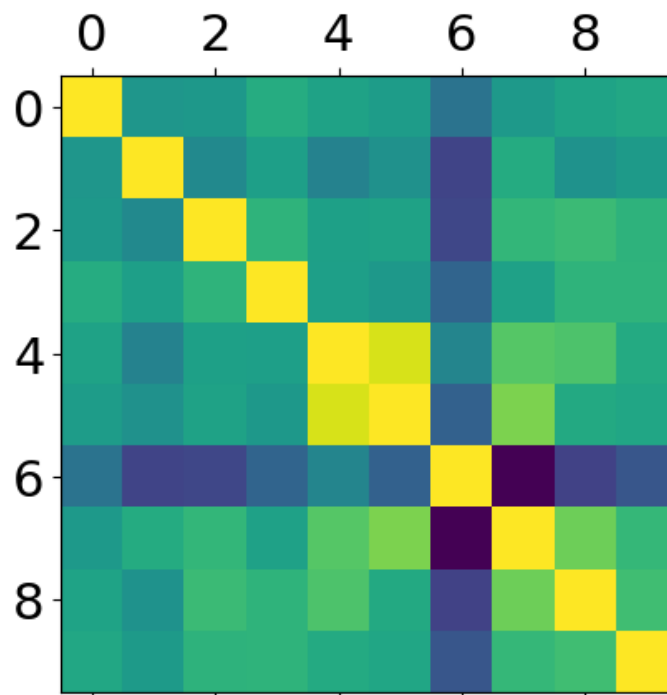


Figura 1. Matriz de correlaciones entre las variables del dataset

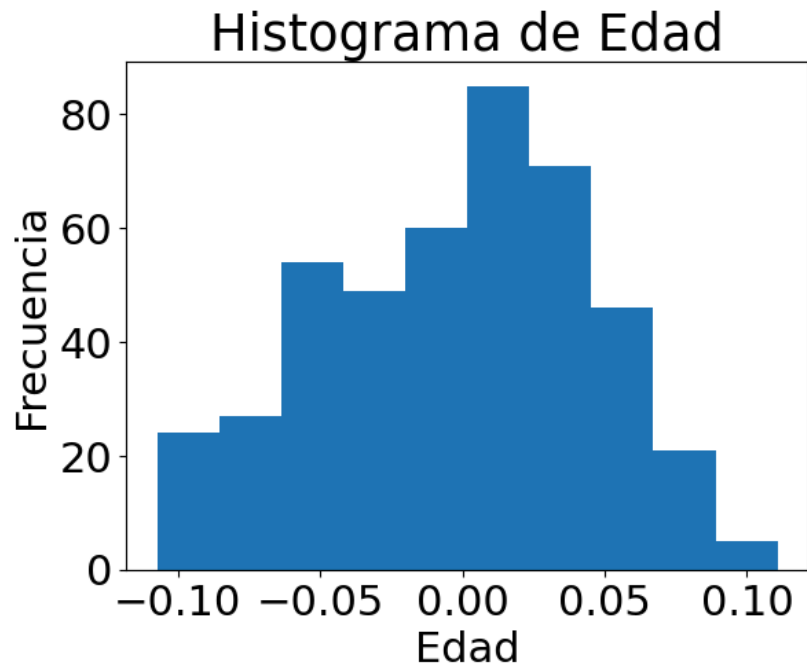
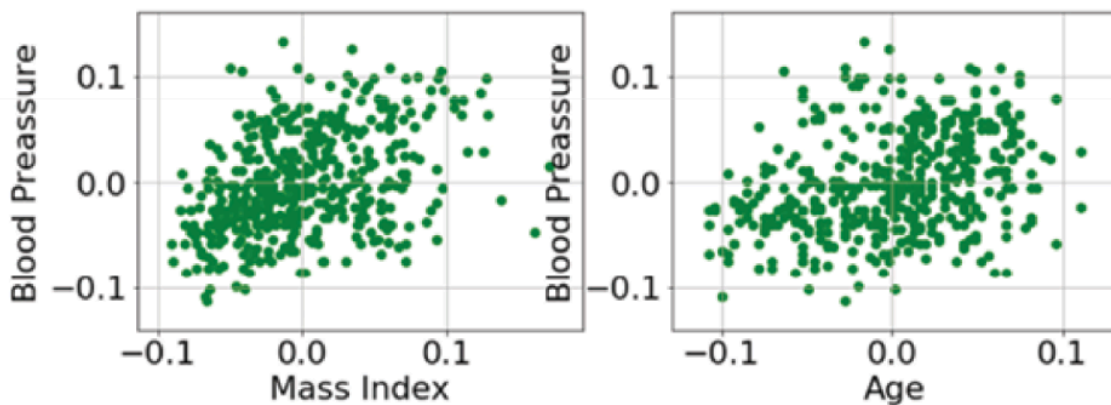


Figura 2. Histograma de Edad



Pregunta 3

Para enseñar técnicas de regresión lineal, regularización, validación cruzada, etc.

Para proyectos de estudiantes en análisis de datos y predicción médica.

En startups de salud digital para estimar riesgos de progresión de enfermedades crónicas. Como modelo base para desarrollar sistemas predictivos para seguros médicos o clínicas privadas.

Pregunta 4

Es un caso de regresión y los modelos más utilizados en estos casos son regresión lineal, Lasso, Random Forest, Regressor, etc. Respecto a las métricas serían el R2 score, Mean squared error, Root Mean Squared error.

Pregunta 5

El mejor modelo ha sido HuberRegressor

Model	Adjusted R-Squared	R-Squared	RMSE
HuberRegressor	0.52	0.58	51.65
OrthogonalMatchingPursuitCV	0.52	0.58	51.69
LassoCV	0.51	0.57	52.19
LassoLarsCV	0.51	0.57	52.19
TransformedTargetRegressor	0.51	0.57	52.19
LinearRegression	0.51	0.57	52.19
Ridge	0.51	0.57	52.22
LassoLarsIC	0.51	0.57	52.22
SGDRegressor	0.51	0.57	52.25
Lasso	0.51	0.57	52.35
LassoLars	0.51	0.57	52.35
RidgeCV	0.51	0.56	52.47
PoissonRegressor	0.51	0.56	52.50
ElasticNetCV	0.50	0.56	52.60
BayesianRidge	0.50	0.56	52.62
GradientBoostingRegressor	0.50	0.55	52.95
BaggingRegressor	0.48	0.54	53.63
PassiveAggressiveRegressor	0.48	0.54	53.70
ExtraTreesRegressor	0.47	0.53	54.16
RandomForestRegressor	0.47	0.53	54.56
LGBMRegressor	0.46	0.52	54.82
HistGradientBoostingRegressor	0.45	0.52	55.20
ElasticNet	0.44	0.50	56.06
AdaBoostRegressor	0.43	0.49	56.54
TweedieRegressor	0.39	0.46	58.57
KNeighborsRegressor	0.38	0.45	58.66
GammaRegressor	0.37	0.45	59.10
LarsCV	0.33	0.40	61.30
XGBRegressor	0.31	0.39	61.96
LinearSVR	0.31	0.39	61.97
OrthogonalMatchingPursuit	0.31	0.38	62.27
RANSACRegressor	0.30	0.38	62.59
SVR	0.08	0.18	71.83
NuSVR	0.08	0.18	71.85
DecisionTreeRegressor	-0.01	0.10	75.25
Lars	-0.07	0.05	77.48
ExtraTreeRegressor	-0.10	0.02	78.46
DummyRegressor	-0.13	-0.00	79.39
QuantileRegressor	-0.15	-0.02	80.00
GaussianProcessRegressor	-0.30	-0.15	85.12
MLPRegressor	-1.12	-0.88	108.84
KernelRidge	-3.67	-3.14	161.55

Model	Time Taken
HuberRegressor	0.02
OrthogonalMatchingPursuitCV	0.02
LassoCV	0.12
LassoLarsCV	0.03
TransformedTargetRegressor	0.01
LinearRegression	0.01
Ridge	0.01
LassoLarsIC	0.01
SGDRegressor	0.01
Lasso	0.02
LassoLars	0.01
RidgeCV	0.01
PoissonRegressor	0.02
ElasticNetCV	0.19
BayesianRidge	0.02
GradientBoostingRegressor	0.23
BaggingRegressor	0.08
PassiveAggressiveRegressor	0.01
ExtraTreesRegressor	0.29
RandomForestRegressor	0.34
LGBMRegressor	0.12
HistGradientBoostingRegressor	0.12
ElasticNet	0.02
AdaBoostRegressor	0.27
TweedieRegressor	0.01
KNeighborsRegressor	0.02
GammaRegressor	0.02
LarsCV	0.04
XGBRegressor	0.25
LinearSVR	0.01
OrthogonalMatchingPursuit	0.01
RANSACRegressor	0.10
SVR	0.02
NuSVR	0.03
DecisionTreeRegressor	0.02
Lars	0.02
ExtraTreeRegressor	0.03
DummyRegressor	0.02
QuantileRegressor	0.03
GaussianProcessRegressor	0.04
MLPRegressor	0.43
KernelRidge	0.02