

**Aprenentatge Computacional**

**MD3: Kaggle**

**GPU Kernel  
Performance Dataset**

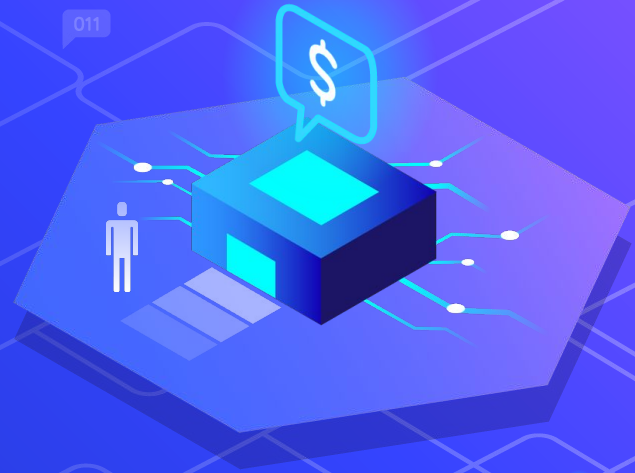
**Mario González 1566235**



# Índice de contenidos

- ⬡ Explicación de la BBDD
- ⬡ Análisis de los datos
- ⬡ Preprocesado de los datos
- ⬡ Modelos y búsqueda del número óptimo de variables
- ⬡ Conclusiones

# Explicación de la BBDD



# Explicación de la BBDD

- ⬡ Mesures de temps d'execució de multiplicacions de matrius  $2048 \times 2048$  en un kernel parametrizable
- ⬡ 241600 posibles combinaciones de parámetros



# Atributos de la BBDD

Tiempo de ejecución	Atributos parametrizables del kernel
Run (ms) (integer)	<ul style="list-style-type: none"><li>- MWG, NWG: {16, 32, 64, 128} (integer)</li><li>- KWG: {16, 32} (integer)</li><li>- MDIMC, NDIMC: {8, 16, 32} (integer)</li><li>- MDIMA, NDIMB: {8, 16, 32} (integer)</li><li>- KWI: {2, 8} (integer)</li><li>- VWM, VWN: {1, 2, 4, 8} (integer)</li><li>- STRM, STRN: {0, 1} (categorical)</li><li>- SA, SB: {0, 1} (categorical)</li></ul>

# Análisis de los datos



# Busqueda de valores nulos

⬡ No disponemos de valores NaN o null:

	MWG	NWG	KWG	MDIMC	NDIMC	MDIMA	NDIMB	KWI	VWM	VWN	STRM	STRN	SA	SB	Run1 (ms)	Run2 (ms)	Run3 (ms)	Run4 (ms)
0	16	16	16	8	8	8	8	2	1	1	0	0	0	0	115.26	115.87	118.55	115.80
1	16	16	16	8	8	8	8	2	1	1	0	0	0	1	78.13	78.25	79.25	79.19
2	16	16	16	8	8	8	8	2	1	1	0	0	1	0	79.84	80.69	80.76	80.97
3	16	16	16	8	8	8	8	2	1	1	0	0	1	1	84.32	89.90	86.75	85.58
4	16	16	16	8	8	8	8	2	1	1	0	1	0	0	115.13	121.98	122.73	114.81

```
001
MWG      0
NWG      0
KWG      0
MDIMC    0
NDIMC    0
MDIMA    0
NDIMB    0
KWI      0
VWM      0
VWN      0
STRM     0
STRN     0
SA       0
SB       0
Run1 (ms) 0
Run2 (ms) 0
Run3 (ms) 0
Run4 (ms) 0
dtype: int64
```



# Preprocesado de los datos





# Transformación del target

4 ejecuciones independientes → 1 sola (media)

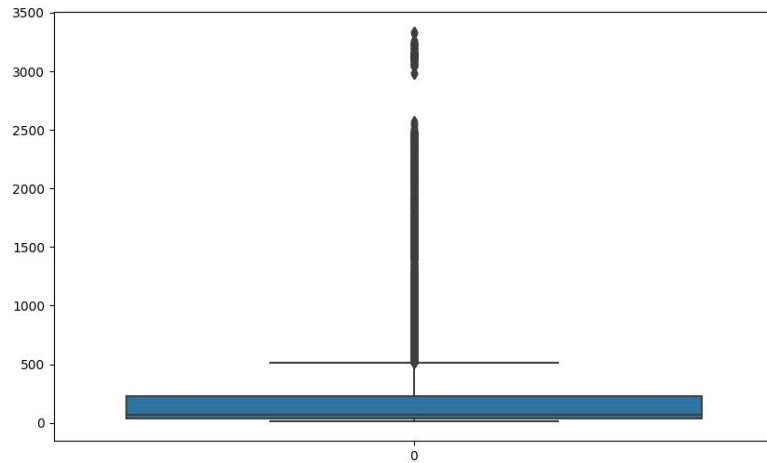
	MWG	NWG	KWG	MDIMC	NDIMC	MDIMA	NDIMB	KWI	VWM	VWN	STRM	STRN	SA	SB	Run (ms)
0	16	16	16	8	8	8	8	2	1	1	0	0	0	0	116.3700
1	16	16	16	8	8	8	8	2	1	1	0	0	0	1	78.7050
2	16	16	16	8	8	8	8	2	1	1	0	0	1	0	80.5650
3	16	16	16	8	8	8	8	2	1	1	0	0	1	1	86.6375
4	16	16	16	8	8	8	8	2	1	1	0	1	0	0	118.6625

# Eliminacion de outliers

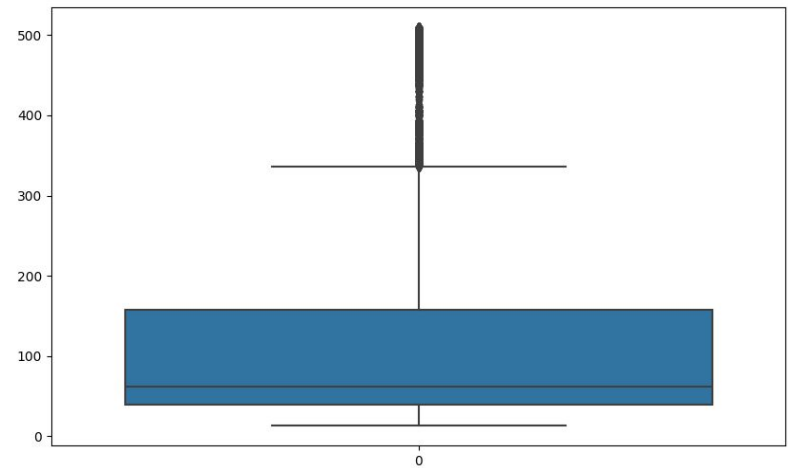
- ⬡ Objetivo: reducir la dispersion de valores
- ⬡ Descartamos aquellos valores que mas se alejan de los más repetidos

# Eliminacion de outliers

Antes



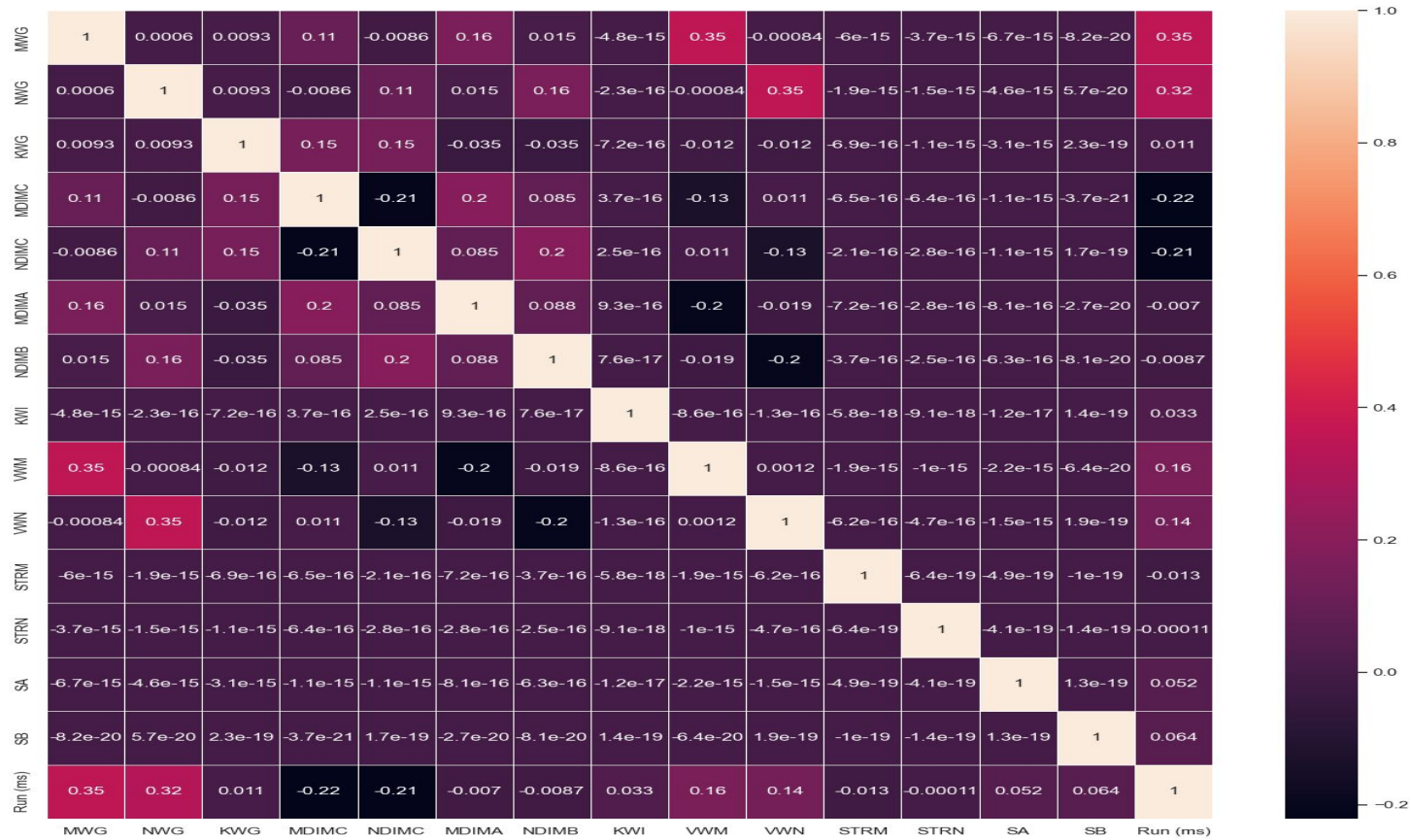
Después



# Matriz de correlación

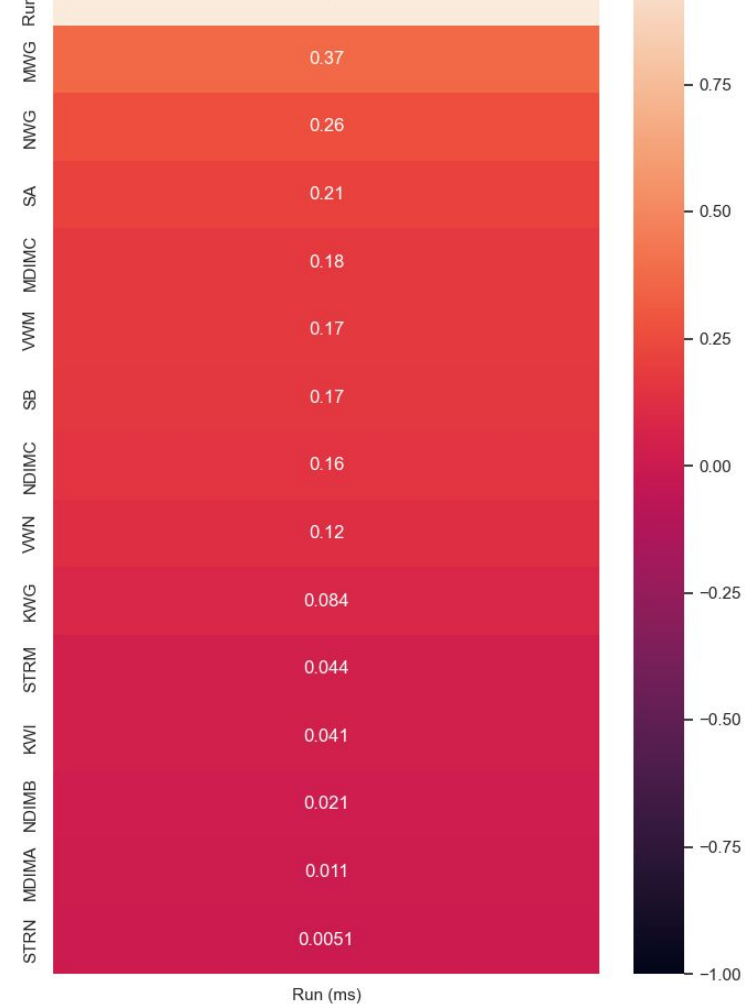
Observación de la matriz de correlación

Atributos con una mayor  
correlación: **MWG, NWG, SA, VWN**



# Matriz de correlación

Representación de la correlación de cada atributo respecto al target



# Escalado de los datos

- Creemos un dataset reducido con los atributos con una mayor correlación: **MWG, NWG, SA, VWN**

Valores:

	MWG	NWG	SA	VWN	Run (ms)
<b>count</b>	214833.000000	214833.000000	214833.000000	214833.000000	214833.000000
<b>mean</b>	75.688037	75.761619	0.486038	2.345156	114.554350
<b>std</b>	41.968313	41.997377	0.499806	1.862122	113.825481
<b>min</b>	16.000000	16.000000	0.000000	1.000000	13.317500
<b>25%</b>	32.000000	32.000000	0.000000	1.000000	39.095000
<b>50%</b>	64.000000	64.000000	0.000000	2.000000	61.790000
<b>75%</b>	128.000000	128.000000	1.000000	4.000000	157.892500
<b>max</b>	128.000000	128.000000	1.000000	8.000000	509.962500



# Escalado de los datos

- Creemos un dataset reducido con los atributos con una mayor correlación: **MWG, NWG, SA, VWN**

Valores:

	MWG	NWG	SA	VWN	Run (ms)
<b>count</b>	214833.000000	214833.000000	214833.000000	214833.000000	214833.000000
<b>mean</b>	75.688037	75.761619	0.486038	2.345156	114.554350
<b>std</b>	41.968313	41.997377	0.499806	1.862122	113.825481
<b>min</b>	16.000000	16.000000	0.000000	1.000000	13.317500
<b>25%</b>	32.000000	32.000000	0.000000	1.000000	39.095000
<b>50%</b>	64.000000	64.000000	0.000000	2.000000	61.790000
<b>75%</b>	128.000000	128.000000	1.000000	4.000000	157.892500
<b>max</b>	128.000000	128.000000	1.000000	8.000000	509.962500

# Escalado de los datos

- Mediante el escalado de los datos, los dejamos todos en una escala parecida.

Valores:

```
[ [-1.42222005 -1.42298787 -0.97245519 -0.7223801  0.01595121]
  [-1.42222005 -1.42298787 -0.97245519 -0.7223801 -0.31495086]
  [-1.42222005 -1.42298787  1.02832501 -0.7223801 -0.29861002]
  ...
  [ 1.24646624  1.24385153 -0.97245519  0.8886892 -0.69751108]
  [ 1.24646624  1.24385153  1.02832501  0.8886892 -0.75643915]
  [ 1.24646624  1.24385153  1.02832501  0.8886892 -0.84987187]]
```

# Modelos y búsqueda de hiperparámetros



# Selección de modelos

- ⬡ Para abordar la regresion haremos uso de los modelos:
  - Linear Regression
  - Decision Tree
  - SGD Regressor

# Entrenamiento de modelos

⬡ Rendimiento con las 4 mejores variables (correlación):

- Score
- MSE

⬡ Variables: **MWG, NWG, SA, VWN**

# Entrenamiento de modelos - 4 mejores atributos

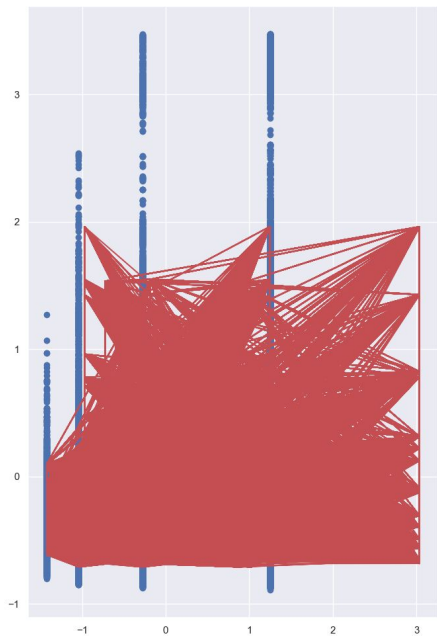
Resultados:

Modelo	Score	MSE
Linear Regression	0.272	0.724
Decision Tree	0.336	0.665
SGD Regressor	0.339	0.656

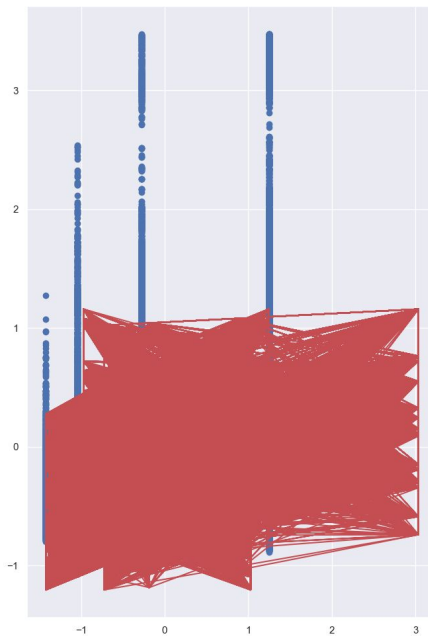


# Predicciones de los modelos - 4 mejores atributos

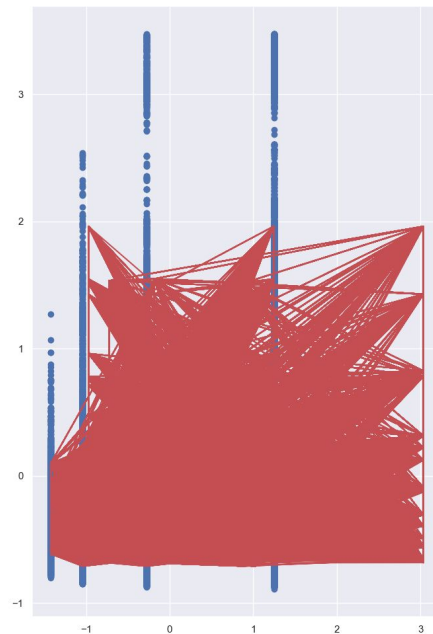
Decision Tree



Linear Regression



SGD Regressor





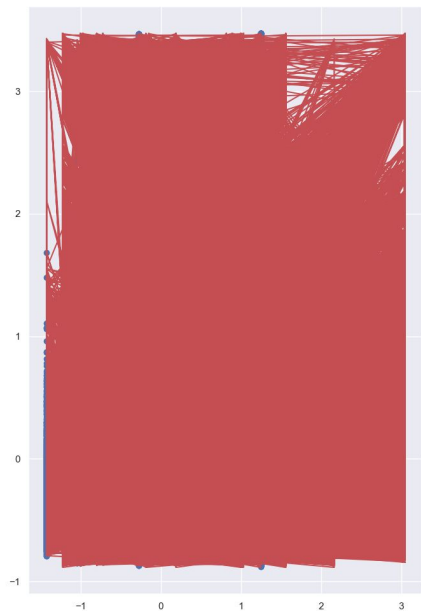
# Entrenamiento de modelos - todos los atributos

Resultados:

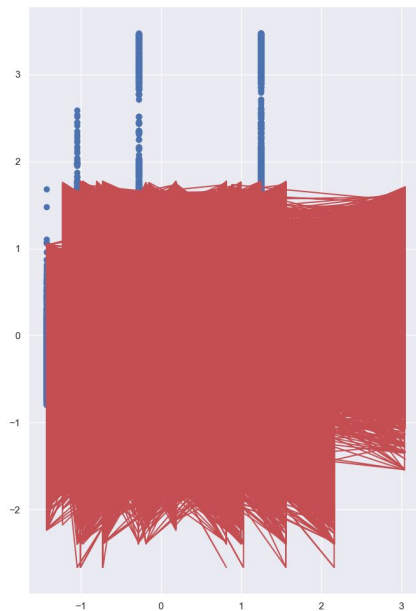
Modelo	Score	MSE
Linear Regression	0.469	0.530
Decision Tree	0.999	0.0006
SGD Regressor	0.467	0.531

# Predicciones de los modelos - todos los atributos

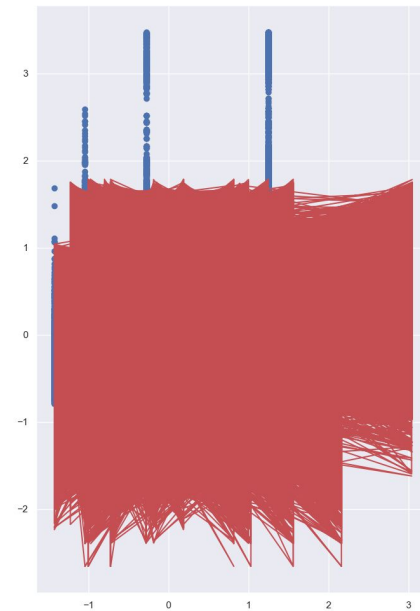
Decision Tree



Linear Regression



SGD Regressor

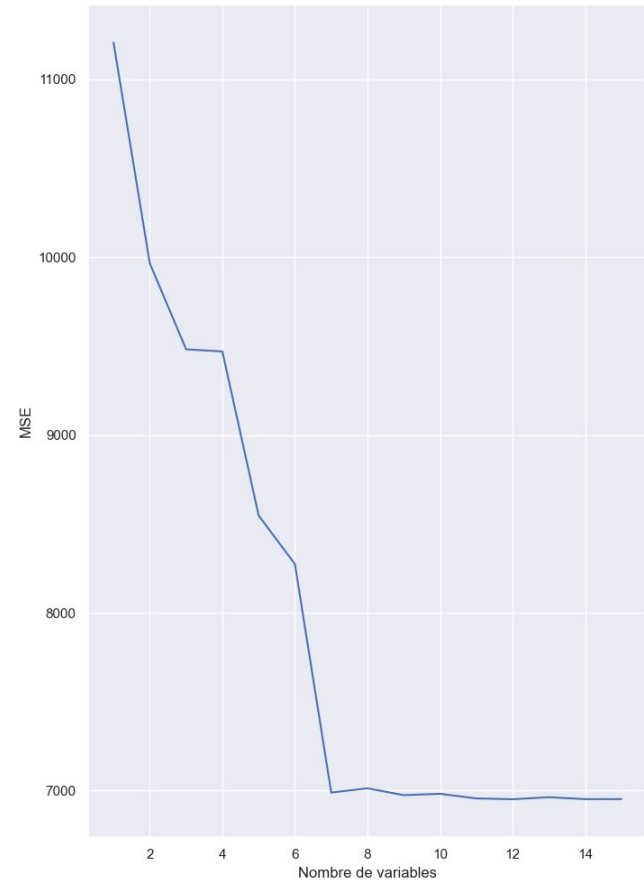


# Entrenamiento de modelos - búsqueda óptima de variables

- ⬡ Minimizar el nº de variables → overfitting
- ⬡ Minimizar MSE → mejores predicciones

# Entrenamiento de modelos - búsqueda óptima de variables

- Estancamiento MSE al utilizar mas de 7 variables



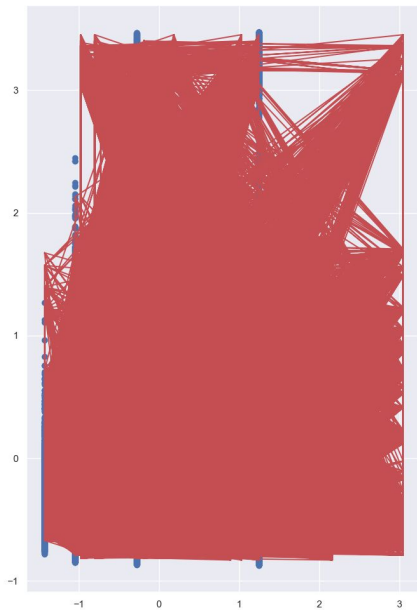
# Entrenamiento de modelos - numero óptimo de variables

Resultados:

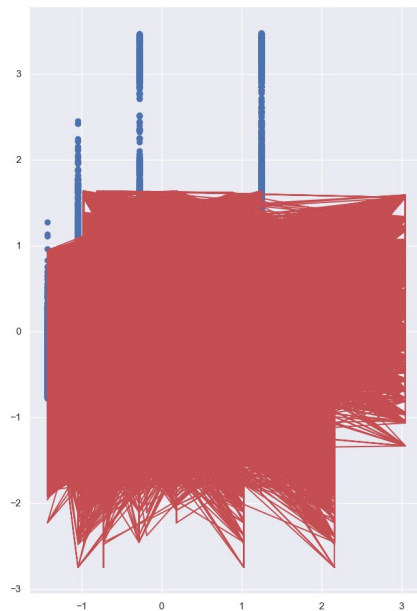
Modelo	Score	MSE
Linear Regression	0.457	0.537
Decision Tree	0.935	0.063
SGD Regressor	0.457	0.538

# Predicciones de los modelos - n° optimo de atributos

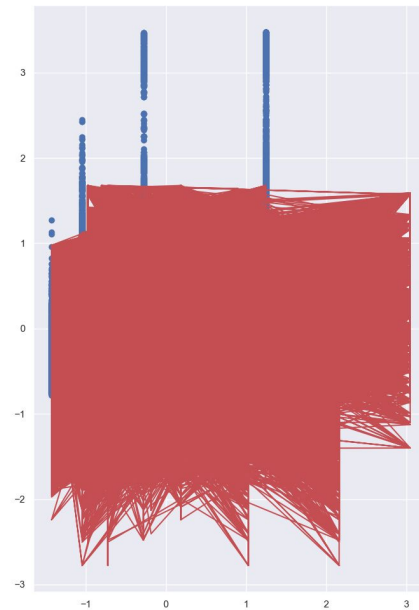
Decision Tree



Linear Regression

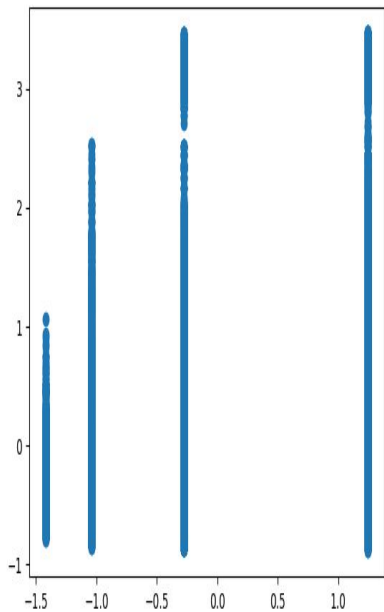


SGD Regressor



# Modelos - Decision tree

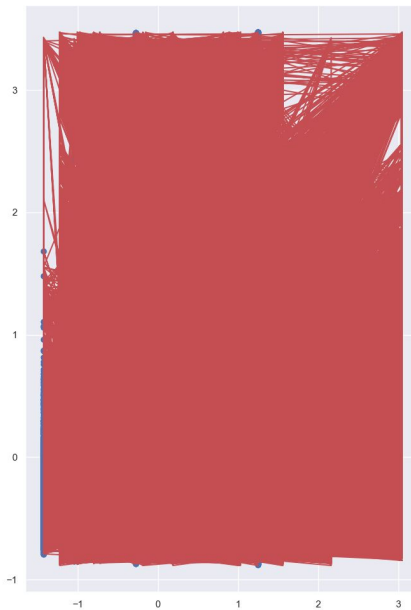
Valores test



Predicción nº óptimo de variables



Predicción con todas las variables





# Conclusiones

