

# Report di analisi

Sono stati creati tre dataset (ciascuno dei quali diviso in 80% per il training e 20% per il testing):

1. Dataset con le feature della frase
2. Dataset con l'embedding delle frasi
3. Dataset con le feature e l'embedding delle frasi

Per ognuno di questi sono state analizzate le performance di alcuni modelli di regressione (Linear, Random Forest e Gradient Boosting) utilizzati per predire i valori V, A, D.

## 1. Dataset con le feature della frase

Di seguito, le metriche ottenute dalle predizioni dei valori V, A, D dei modelli di regressione Linear, Random Forest e Gradient Boosting addestrati sul dataset contenente l'encoding dell'emozione, le dimensioni A, D e le feature delle frasi:

- **Valence**

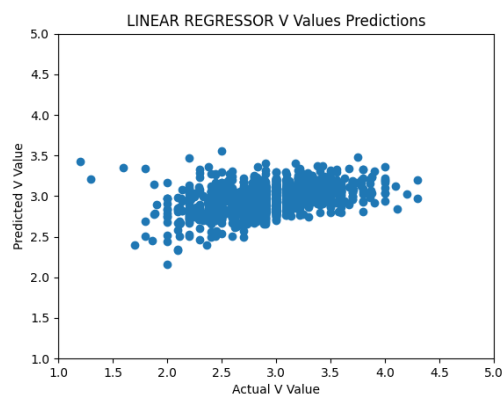
### LINEAR REGRESSOR

Mean Absolute Error (MAE): 0.23206861067831516

Mean Squared Error (MSE): 0.1046109991568542

**Root Mean Squared Error (RMSE): 0.3234362366168241**

R<sup>2</sup>: 0.14913207580493537



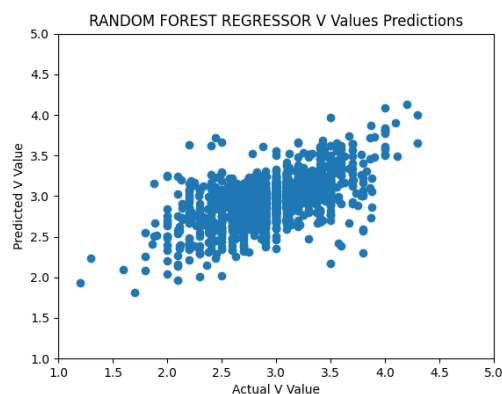
### RANDOM FOREST REGRESSOR

Mean Absolute Error (MAE): 0.20246897195819738

Mean Squared Error (MSE): 0.08305877467893116

**Root Mean Squared Error (RMSE): 0.2881991927104085**

R<sup>2</sup>: 0.32443005260582813



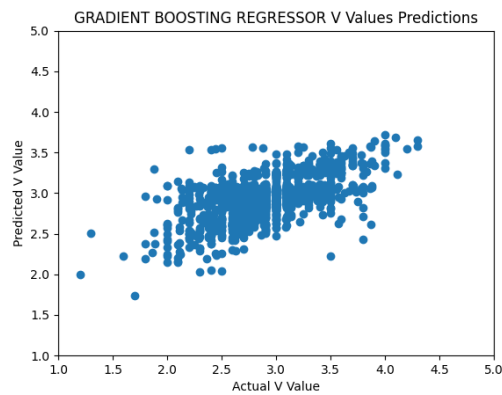
### GRADIENT BOOSTING REGRESSOR

Mean Absolute Error (MAE): 0.19997520705668495

Mean Squared Error (MSE): 0.07865752411190896

**Root Mean Squared Error (RMSE): 0.2804594874699534**

R<sup>2</sup>: 0.3602282283617966



- **Arousal**

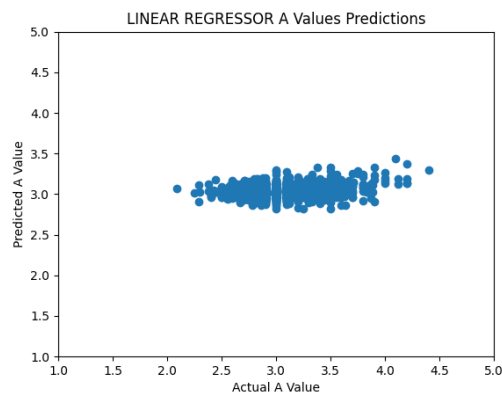
### LINEAR REGRESSOR

Mean Absolute Error (MAE): 0.18961689446946448

Mean Squared Error (MSE): 0.06356948704806066

**Root Mean Squared Error (RMSE): 0.25212990113840256**

R<sup>2</sup>: 0.07478418045086366



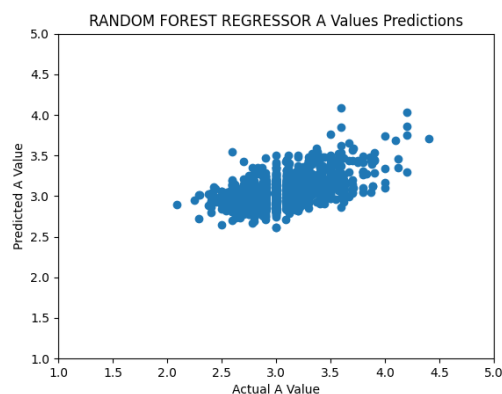
### RANDOM FOREST REGRESSOR

Mean Absolute Error (MAE): 0.16986956859523025

Mean Squared Error (MSE): 0.04903015439903892

**Root Mean Squared Error (RMSE): 0.22142753758066977**

R<sup>2</sup>: 0.2863954612276298



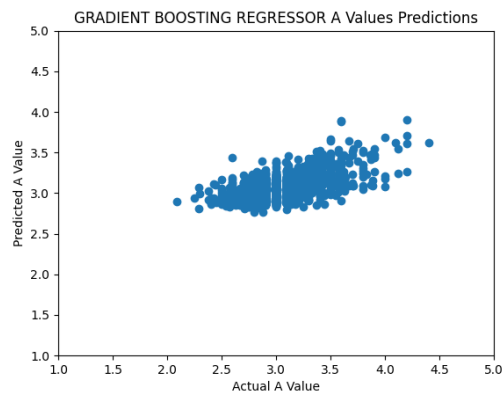
### GRADIENT BOOSTING REGRESSOR

Mean Absolute Error (MAE): 0.1615443504264046

Mean Squared Error (MSE): 0.04533628408885673

**Root Mean Squared Error (RMSE): 0.21292318823664258**

R<sup>2</sup>: 0.34015753175935515



- **Dominance**

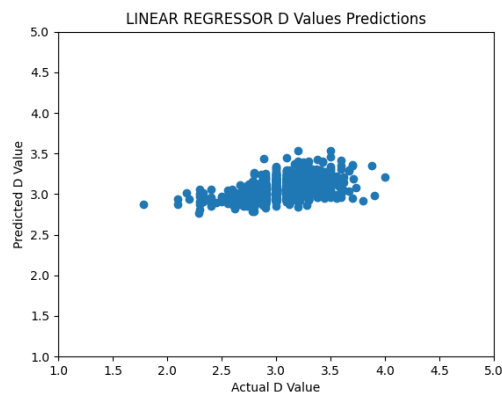
### LINEAR REGRESSOR

Mean Absolute Error (MAE): 0.1392052809295413

Mean Squared Error (MSE): 0.035283994983582534

**Root Mean Squared Error (RMSE): 0.18784034439806196**

R<sup>2</sup>: 0.19955215553728578



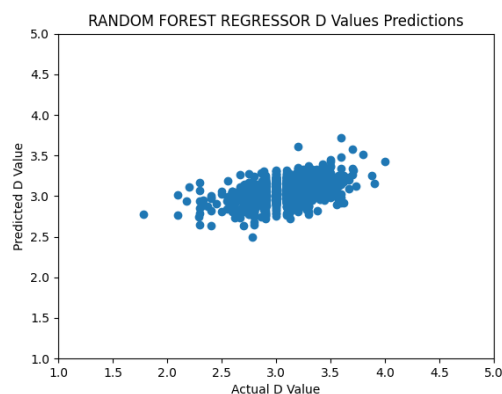
### RANDOM FOREST REGRESSOR

Mean Absolute Error (MAE): 0.1405024677259186

Mean Squared Error (MSE): 0.03535261926690389

**Root Mean Squared Error (RMSE): 0.18802292218478014**

R<sup>2</sup>: 0.19799535450928607



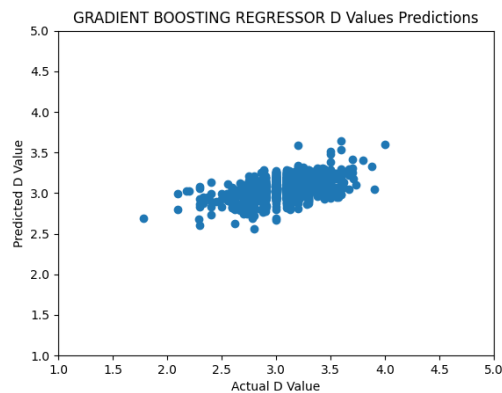
### GRADIENT BOOSTING REGRESSOR

Mean Absolute Error (MAE): 0.1615443504264046

Mean Squared Error (MSE): 0.04533628408885673

**Root Mean Squared Error (RMSE): 0.21292318823664258**

R<sup>2</sup>: 0.34015753175935515



## 2. Dataset con l'embedding delle frasi

Di seguito, le metriche ottenute dalle predizioni dei valori V, A, D dei modelli di regressione Linear, Random Forest e Gradient Boosting addestrati sul dataset contenente l'encoding dell'emozione, le dimensioni A, D e l'embedding delle frasi:

- **Valence**

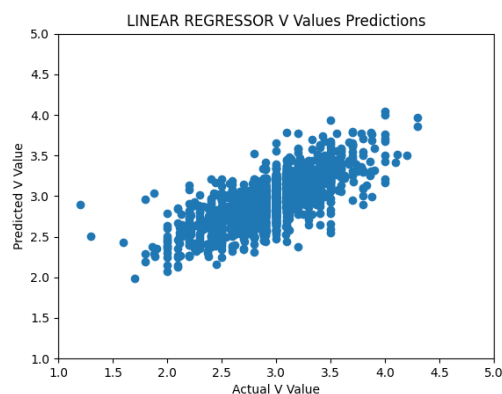
### LINEAR REGRESSOR

Mean Absolute Error (MAE): 0.17994883497224234

Mean Squared Error (MSE): 0.05911930602644773

**Root Mean Squared Error (RMSE): 0.24314461957124966**

R<sup>2</sup>: 0.5191450076568696



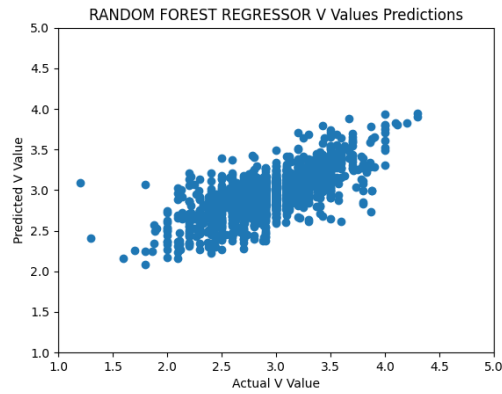
### RANDOM FOREST REGRESSOR

Mean Absolute Error (MAE): 0.1779534922211886

Mean Squared Error (MSE): 0.06234062845746442

**Root Mean Squared Error (RMSE): 0.24968105346113953**

R<sup>2</sup>: 0.4929439393931737



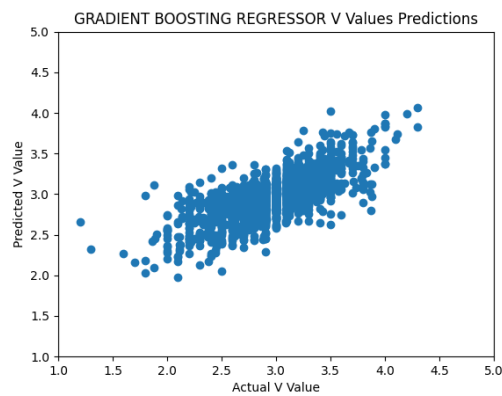
### GRADIENT BOOSTING REGRESSOR

Mean Absolute Error (MAE): 0.17665499559246653

Mean Squared Error (MSE): 0.05903367560889545

**Root Mean Squared Error (RMSE): 0.24296846628502114**

R<sup>2</sup>: 0.5198414944146474



- Arousal**

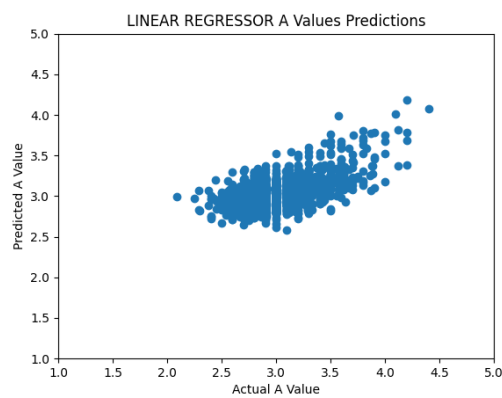
### LINEAR REGRESSOR

Mean Absolute Error (MAE): 0.17050541429271918

Mean Squared Error (MSE): 0.048371224917990556

**Root Mean Squared Error (RMSE): 0.21993459236325366**

R<sup>2</sup>: 0.29598578526740593



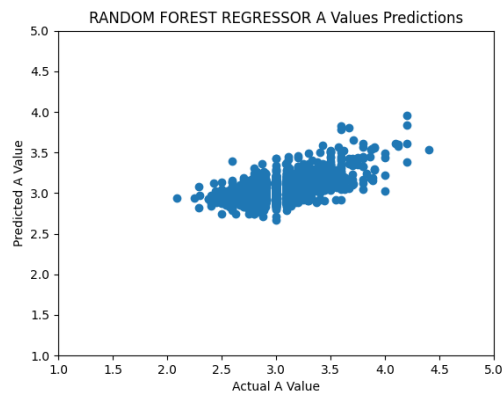
### RANDOM FOREST REGRESSOR

Mean Absolute Error (MAE): 0.1590289946564525

Mean Squared Error (MSE): 0.04357181385709752

**Root Mean Squared Error (RMSE): 0.2087386256951442**

R<sup>2</sup>: 0.3658383394448555



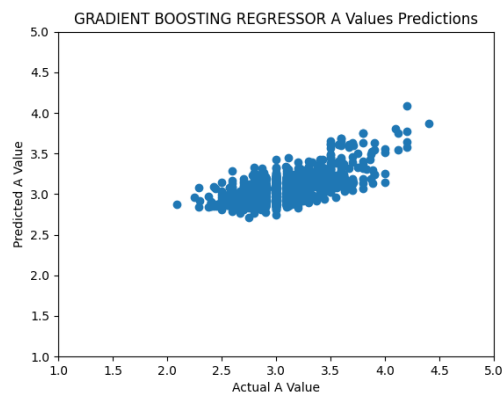
### GRADIENT BOOSTING REGRESSOR

Mean Absolute Error (MAE): 0.15582117461767503

Mean Squared Error (MSE): 0.04106488014311282

**Root Mean Squared Error (RMSE): 0.20264471407641704**

R<sup>2</sup>: 0.4023252585383844



- **Dominance**

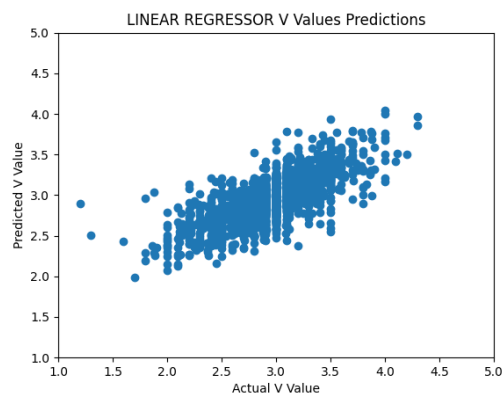
### LINEAR REGRESSOR

Mean Absolute Error (MAE): 0.17994883497224234

Mean Squared Error (MSE): 0.05911930602644773

**Root Mean Squared Error (RMSE): 0.24314461957124966**

R<sup>2</sup>: 0.5191450076568696



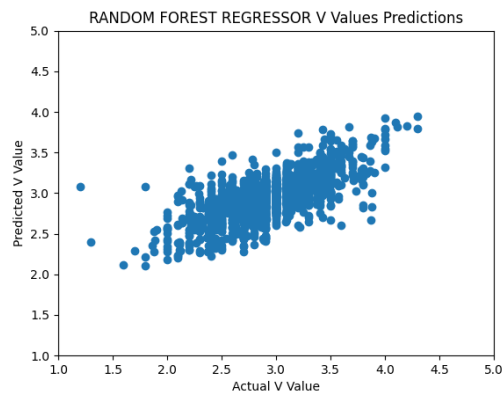
### RANDOM FOREST REGRESSOR

Mean Absolute Error (MAE): 0.1778880503144654

Mean Squared Error (MSE): 0.062465722922873224

**Root Mean Squared Error (RMSE): 0.24993143644382398**

R<sup>2</sup>: 0.49192646638394366



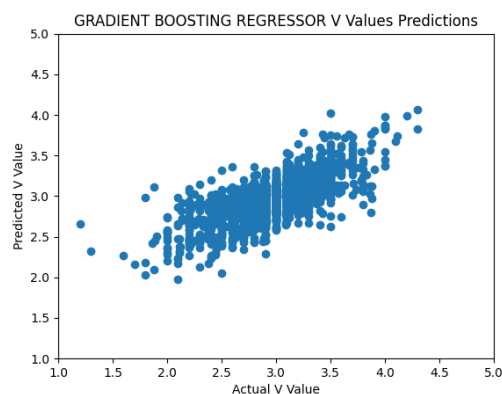
### GRADIENT BOOSTING REGRESSOR

Mean Absolute Error (MAE): 0.17665291411676748

Mean Squared Error (MSE): 0.059033988786268185

**Root Mean Squared Error (RMSE): 0.24296911076568597**

R<sup>2</sup>: 0.5198389471434888



## 2. Dataset con le feature e l'embedding delle frasi

Di seguito, le metriche ottenute dalle predizioni dei valori V, A, D dei modelli di regressione Linear, Random Forest e Gradient Boosting addestrati sul dataset contenente l'encoding dell'emozione, le dimensioni A, D, le feature e l'embedding delle frasi:

- **Valence**

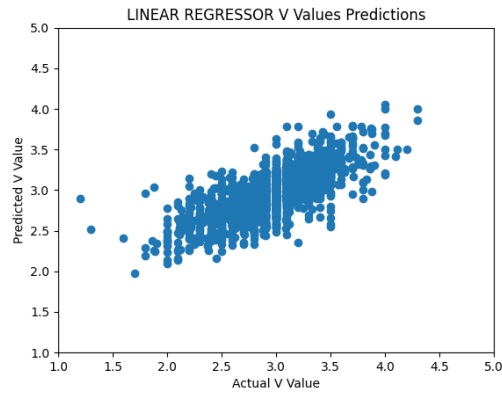
### LINEAR REGRESSOR

Mean Absolute Error (MAE): 0.18092736621519268

Mean Squared Error (MSE): 0.059611516263669465

**Root Mean Squared Error (RMSE): 0.2441546974024245**

R<sup>2</sup>: 0.5151415481144896



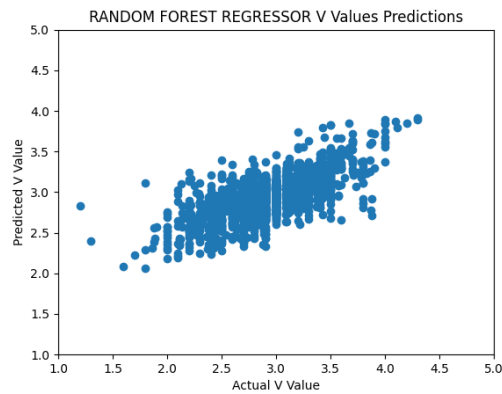
### RANDOM FOREST REGRESSOR

Mean Absolute Error (MAE): 0.1796016787251147

Mean Squared Error (MSE): 0.06327626307532082

**Root Mean Squared Error (RMSE): 0.2515477351822529**

R<sup>2</sup>: 0.4853338267710121



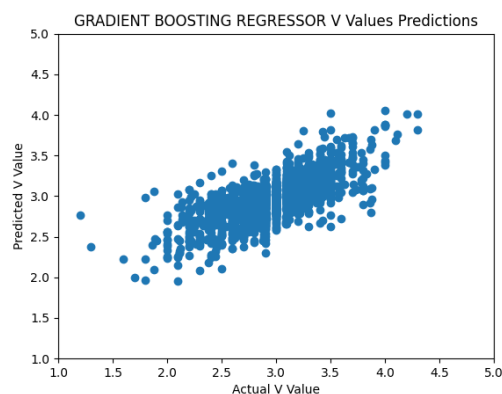
### GRADIENT BOOSTING REGRESSOR

Mean Absolute Error (MAE): 0.17615152586923183

Mean Squared Error (MSE): 0.05922424284673725

**Root Mean Squared Error (RMSE): 0.24336031485584755**

R<sup>2</sup>: 0.5182914896217584





- **Arousal**

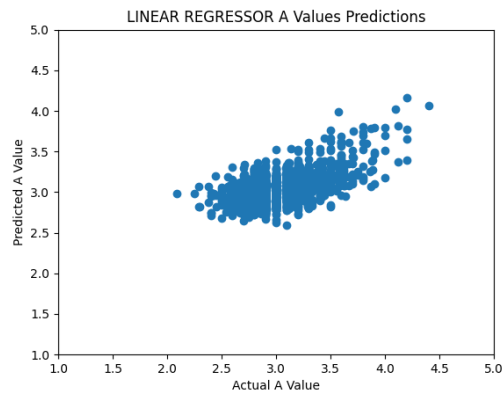
**LINEAR REGRESSOR**

Mean Absolute Error (MAE): 0.17054580914805398

Mean Squared Error (MSE): 0.04837929683090907

**Root Mean Squared Error (RMSE): 0.21995294231018842**

R<sup>2</sup>: 0.2958683034082148

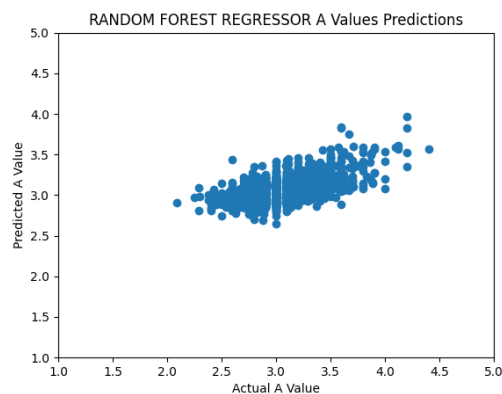
**RANDOM FOREST REGRESSOR**

Mean Absolute Error (MAE): 0.15961035905408175

Mean Squared Error (MSE): 0.04366539800472087

**Root Mean Squared Error (RMSE): 0.20896267131887664**

R<sup>2</sup>: 0.36447627821295225

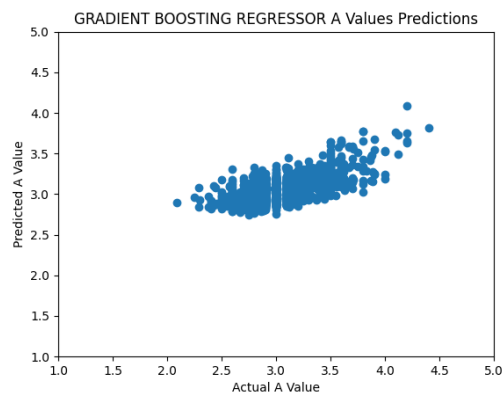
**GRADIENT BOOSTING REGRESSOR**

Mean Absolute Error (MAE): 0.15578277347216507

Mean Squared Error (MSE): 0.04110526754237497

**Root Mean Squared Error (RMSE): 0.20274434034609937**

R<sup>2</sup>: 0.40173744412547996



- **Dominance**

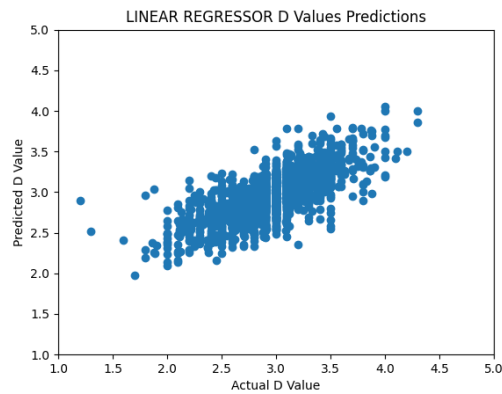
**LINEAR REGRESSOR**

Mean Absolute Error (MAE): 0.18092736621519268

Mean Squared Error (MSE): 0.059611516263669465

**Root Mean Squared Error (RMSE): 0.2441546974024245**

R<sup>2</sup>: 0.5151415481144896

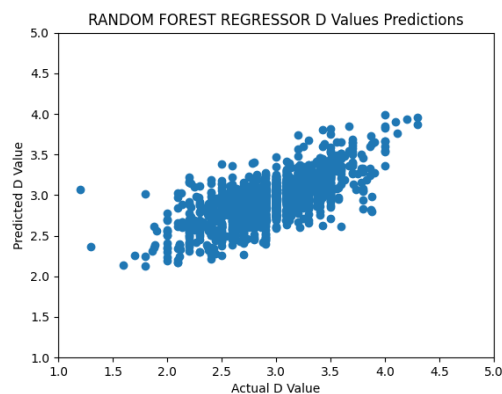
**RANDOM FOREST REGRESSOR**

Mean Absolute Error (MAE): 0.1777807372204095

Mean Squared Error (MSE): 0.0623745106194169

**Root Mean Squared Error (RMSE): 0.24974889513152385**

R<sup>2</sup>: 0.49266835417708643

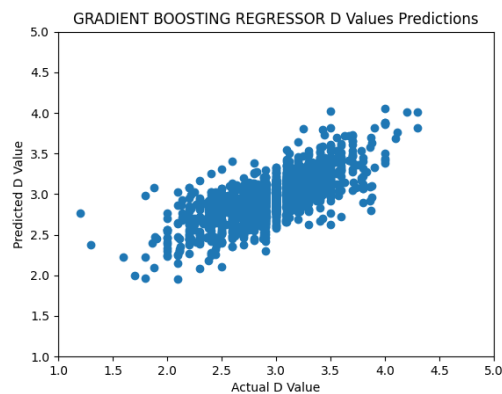
**GRADIENT BOOSTING REGRESSOR**

Mean Absolute Error (MAE): 0.17620046104121792

Mean Squared Error (MSE): 0.059272482855059164

**Root Mean Squared Error (RMSE): 0.24345940699644195**

R<sup>2</sup>: 0.5178991228909668



## - Summary dei risultati

Per la predizione dei valori di Valence e di Arousal, i risultati migliori (tenendo particolare attenzione all'RMSE) sono stati ottenuti dai dataset contenenti l'embedding delle frasi e le feature delle frasi (dataset 2 e 3). Per entrambi i dataset si sono ottenuti valori molto simili di RMSE con tutti e tre i modelli utilizzati. Per i valori di Dominance, i risultati migliori sono stati ottenuti utilizzando i modelli di linear regression e random forest sul dataset contenente solo le feature delle frasi. Di seguito una tabella ricapitolativa dove vengono mostrati solamente i valori della metrica RMSE per ogni dataset e modello utilizzato:

Dataset con le feature della frase	
Valence	LINEAR REGRESSOR: (RMSE): 0.3234362366168241
	RANDOM FOREST REGRESSOR (RMSE): 0.2881991927104085
	GRADIENT BOOSTING REGRESSOR (RMSE): 0.2804594874699534
Arousal	LINEAR REGRESSOR: (RMSE): 0.25212990113840256
	RANDOM FOREST REGRESSOR (RMSE): 0.22142753758066977
	GRADIENT BOOSTING REGRESSOR (RMSE): 0.21292318823664258
Dominance	LINEAR REGRESSOR: (RMSE): 0.18784034439806196
	RANDOM FOREST REGRESSOR (RMSE): 0.18802292218478014
	GRADIENT BOOSTING REGRESSOR (RMSE): 0.21292318823664258
Dataset con l'embedding delle frasi	
Valence	LINEAR REGRESSOR: (RMSE): 0.24314461957124966
	RANDOM FOREST REGRESSOR (RMSE): 0.24968105346113953
	GRADIENT BOOSTING REGRESSOR (RMSE): 0.24296846628502114
Arousal	LINEAR REGRESSOR: (RMSE): 0.21993459236325366
	RANDOM FOREST REGRESSOR (RMSE): 0.2087386256951442
	GRADIENT BOOSTING REGRESSOR (RMSE): 0.20264471407641704
Dominance	LINEAR REGRESSOR: (RMSE): 0.24314461957124966
	RANDOM FOREST REGRESSOR (RMSE): 0.24993143644382398
	GRADIENT BOOSTING REGRESSOR (RMSE): 0.24296911076568597
Dataset con le feature della frase e l'embedding delle frasi	
Valence	LINEAR REGRESSOR: (RMSE): 0.2441546974024245
	RANDOM FOREST REGRESSOR (RMSE): 0.2515477351822529
	GRADIENT BOOSTING REGRESSOR (RMSE): 0.24336031485584755
Arousal	LINEAR REGRESSOR: (RMSE): 0.21995294231018842
	RANDOM FOREST REGRESSOR (RMSE): 0.20896267131887664
	GRADIENT BOOSTING REGRESSOR (RMSE): 0.20274434034609937
Dominance	LINEAR REGRESSOR: (RMSE): 0.2441546974024245
	RANDOM FOREST REGRESSOR (RMSE): 0.24974889513152385
	GRADIENT BOOSTING REGRESSOR (RMSE): 0.24345940699644195