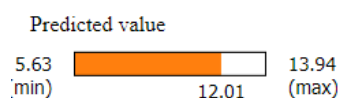
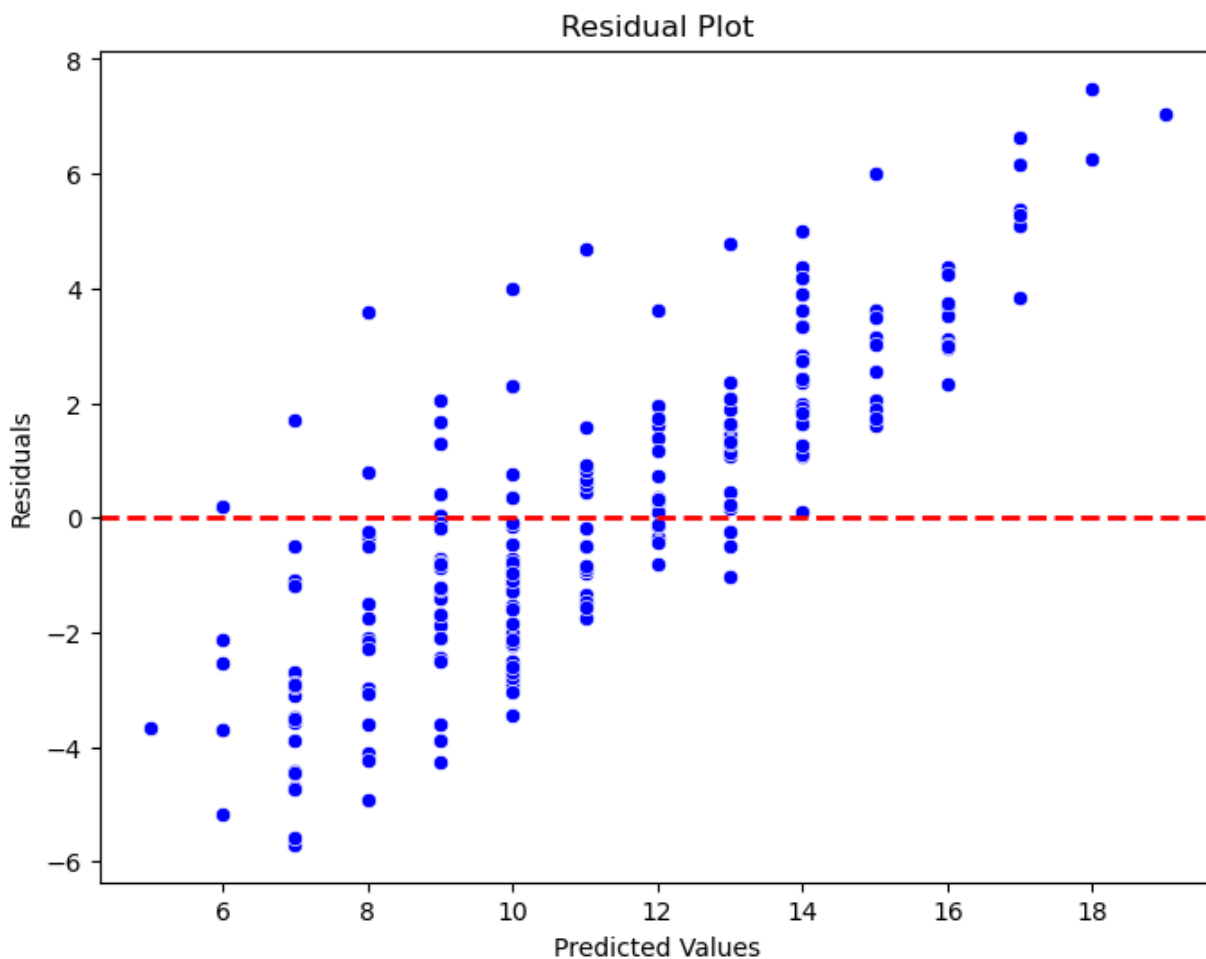


# Our Approach

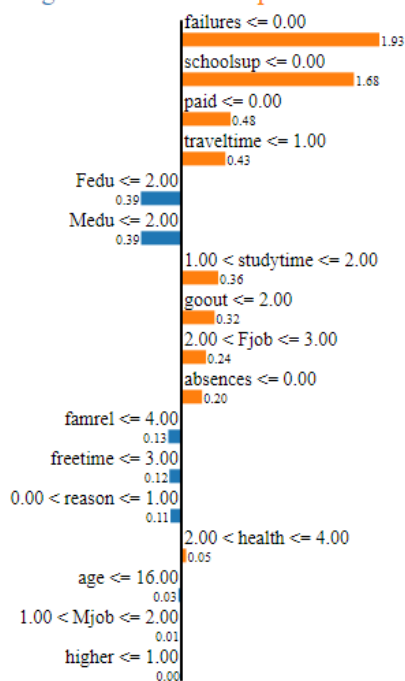
We started by removing the “school” from the features. This was in consideration to the fact that we had no information about the “school” column and a user has no way to enter this information. Since this was a Regression Problem we used R-squared (R2) , Mean Absolute Error (MAE), Mean Squared Error (MSE) as the metrics. Next we used the following models to predict G1.

Model	R-squared (R2)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Linear Regression	0.201084	2.122629	7.009377
Decision Tree Regressor	-0.315521	2.672249	11.541866
Support Vector Machine Regressor	0.164724	2.18967	7.328388
XGBoost Regressor	0.216802	2.074647	6.871473



negative

positive

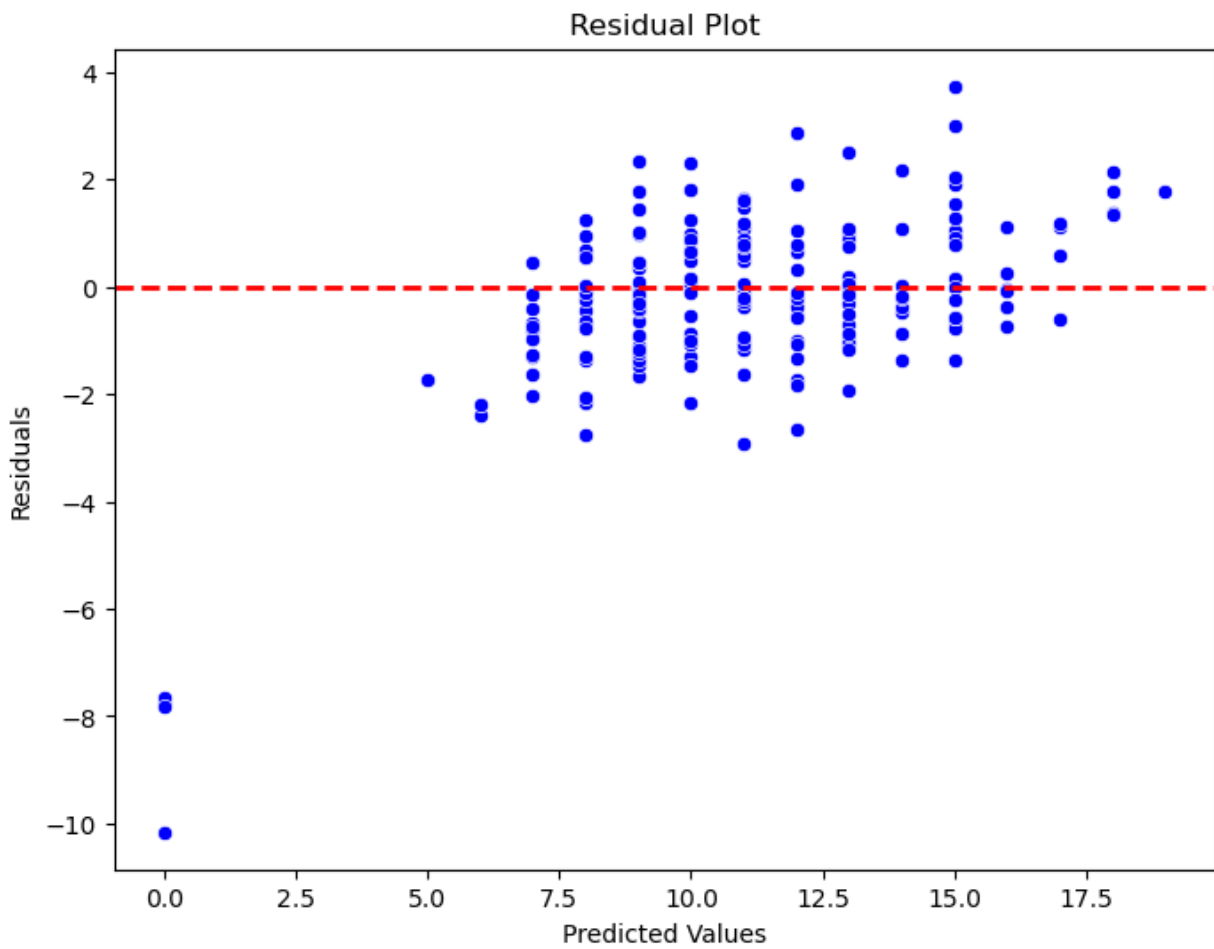


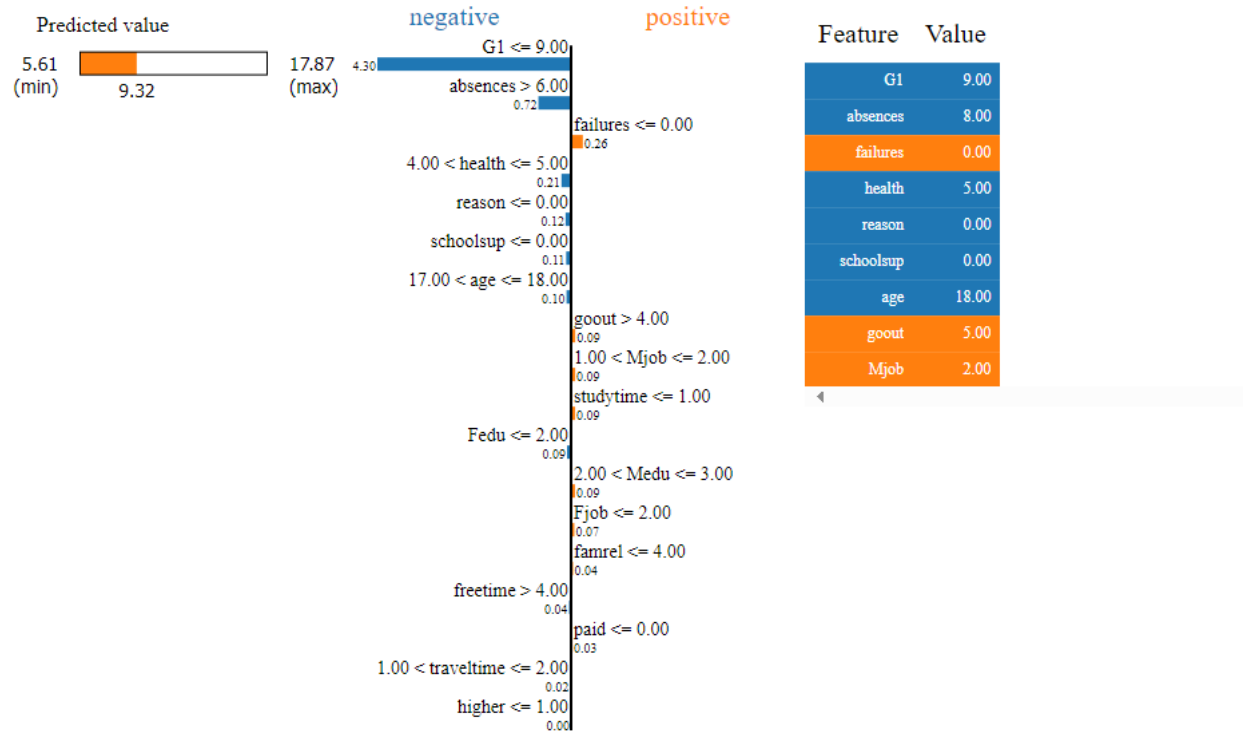
Feature Value

failures	0.00
schoolsup	0.00
paid	0.00
traveltime	1.00
Fedu	1.00
Medu	2.00
studytime	2.00
goout	2.00
Fjob	3.00

When we G1 feature. Next we used the following models to predict G2.

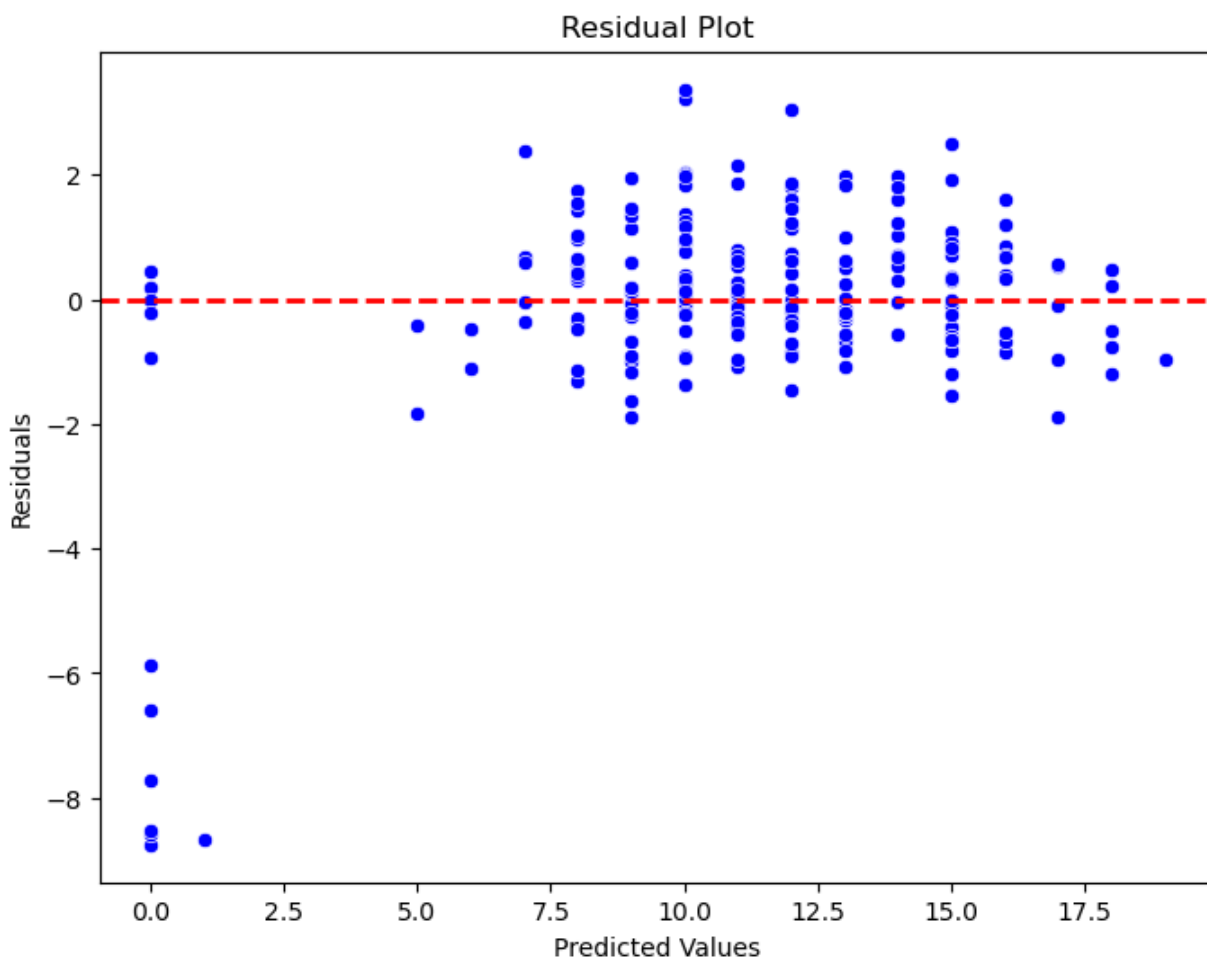
Model	R-squared (R2)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Linear Regression	0.721009	1.153713	3.012782
Decision Tree Regressor	0.457677	1.617225	5.856459
Support Vector Machine Regressor	0.728015	1.097734	2.937119
XGBoost Regressor	0.638752	1.339506	3.901061





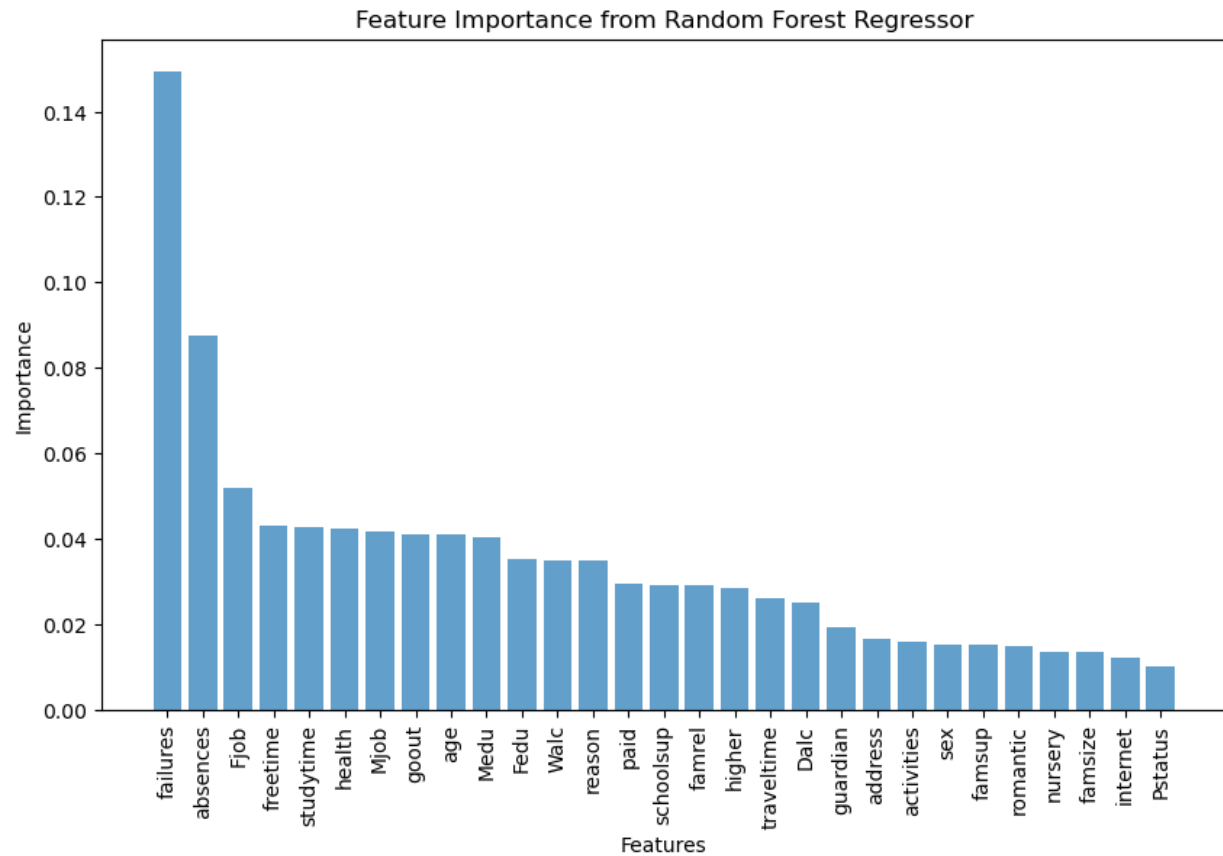
When we G1 and G2 feature. Next we used the following models to predict G3.

Model	R-squared (R2)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Linear Regression	0.798866	1.028168	3.109693
Decision Tree Regressor	0.662985	1.162679	5.210526
Support Vector Machine Regressor	0.785694	0.990189	3.313344
XGBoost Regressor	0.638752	1.339506	3.901061



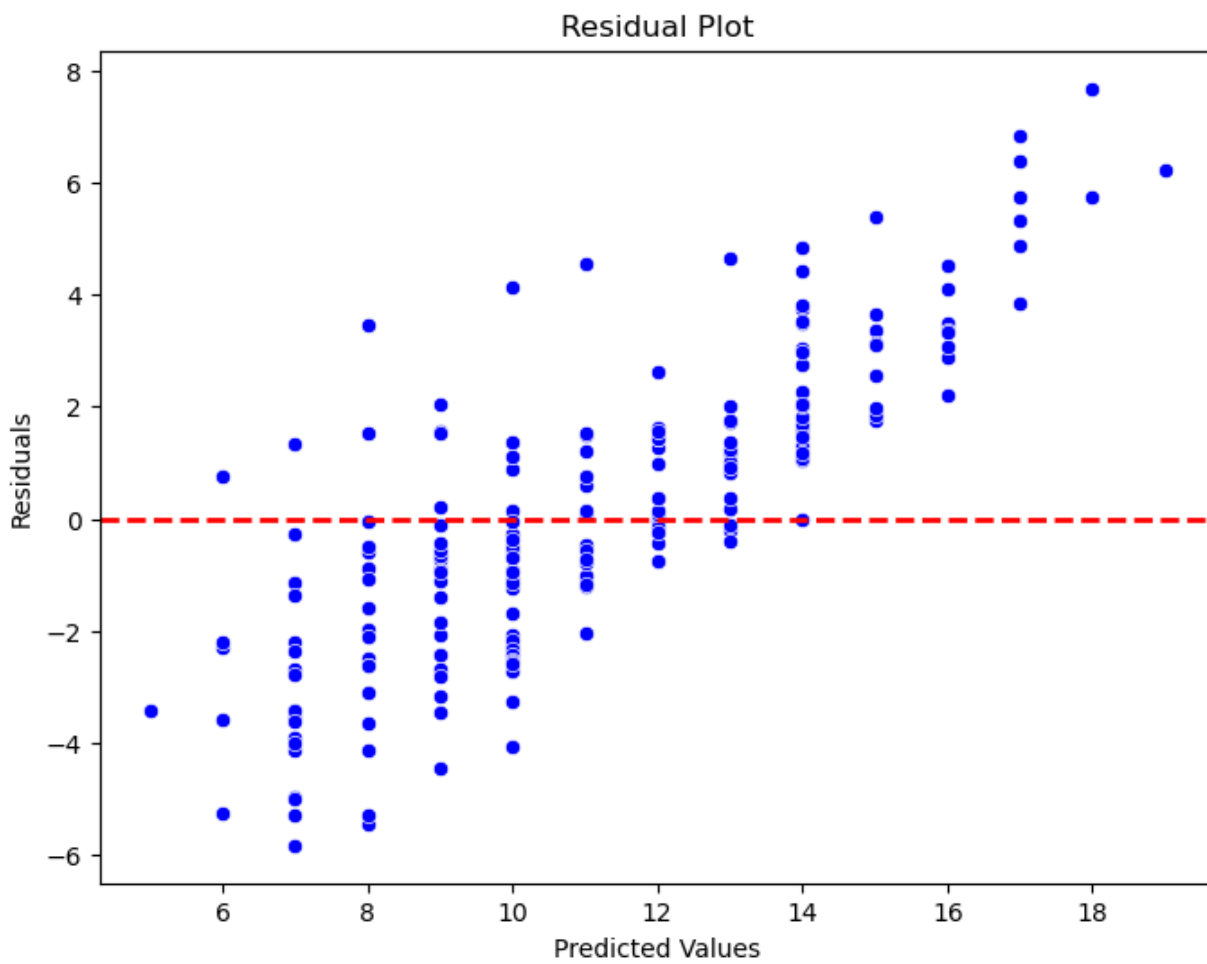


Next we used feature selection to and using the feature importance selected the following: ['failures', 'absences', 'Fjob', 'freetime', 'health', 'studytime', 'Medu', 'Mjob', 'goout', 'age', 'Fedu', 'reason', 'famrel', 'higher', 'schoolsup', 'paid', 'traveltime'].

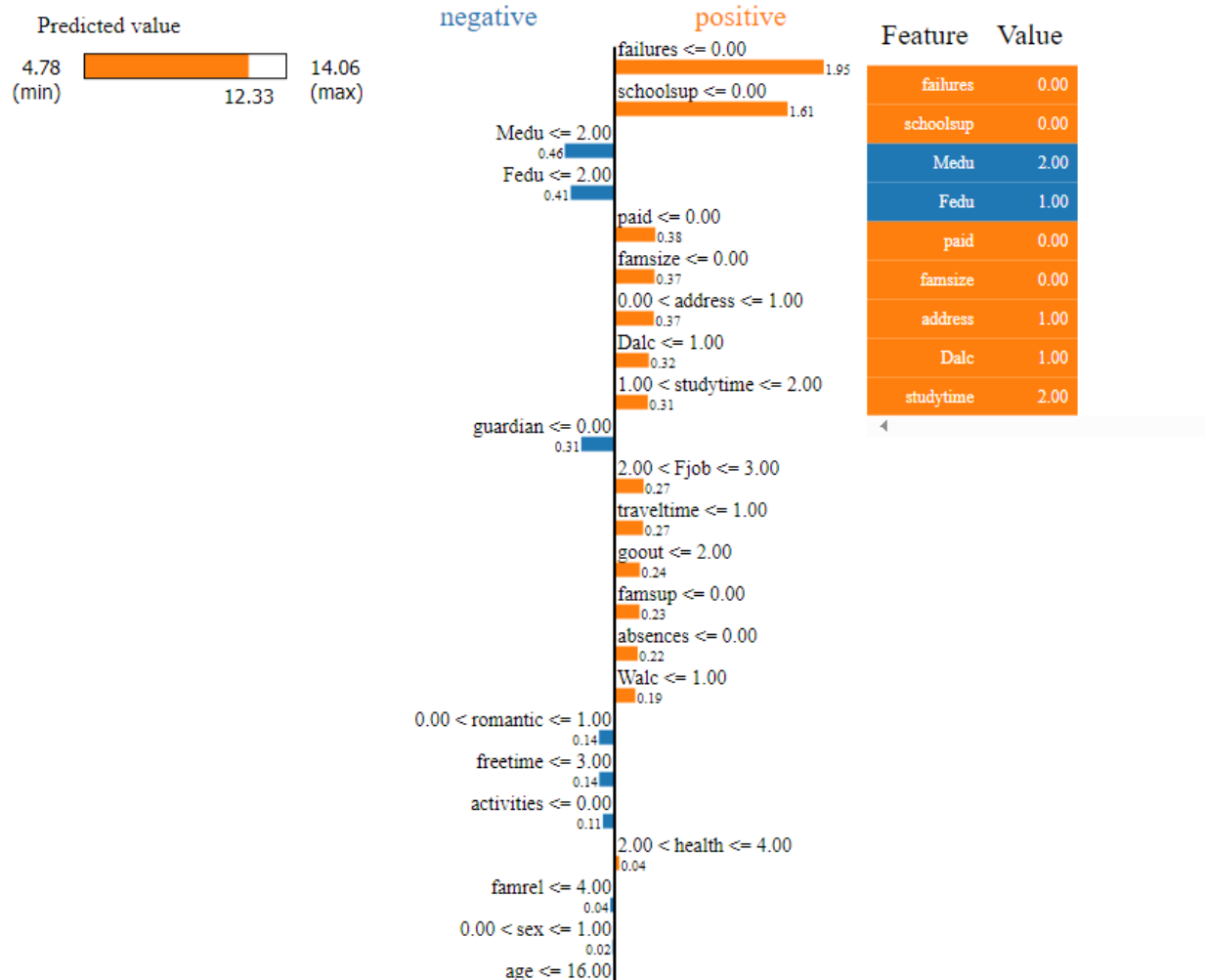


we used the following models to predict G1.

Model	R-squared (R2)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Linear Regression	0.180566	2.166261	7.189391
Decision Tree Regressor	-0.367329	2.72488	11.996411
Support Vector Machine Regressor	0.15346	2.217581	7.427212
XGBoost Regressor	0.179581	2.124667	7.198035

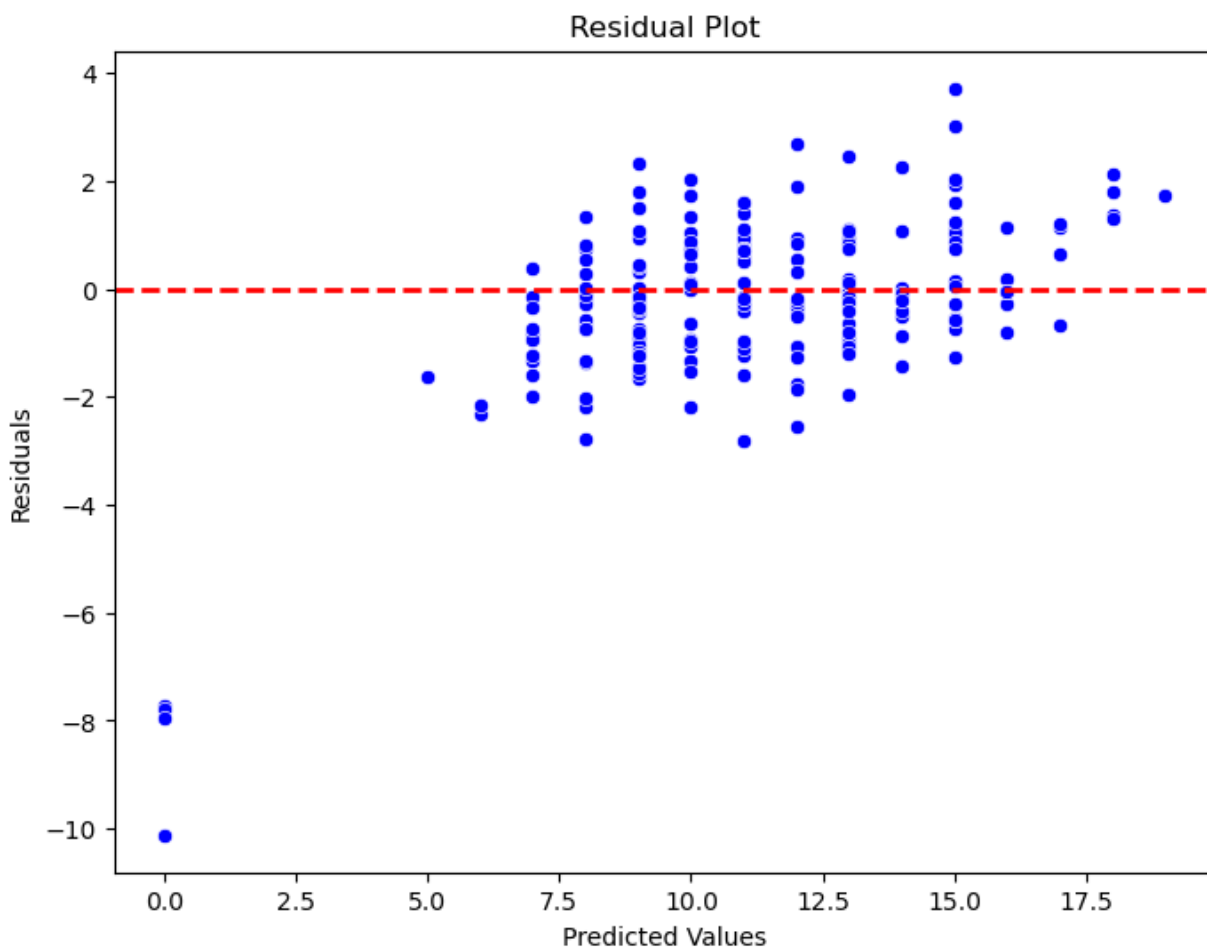


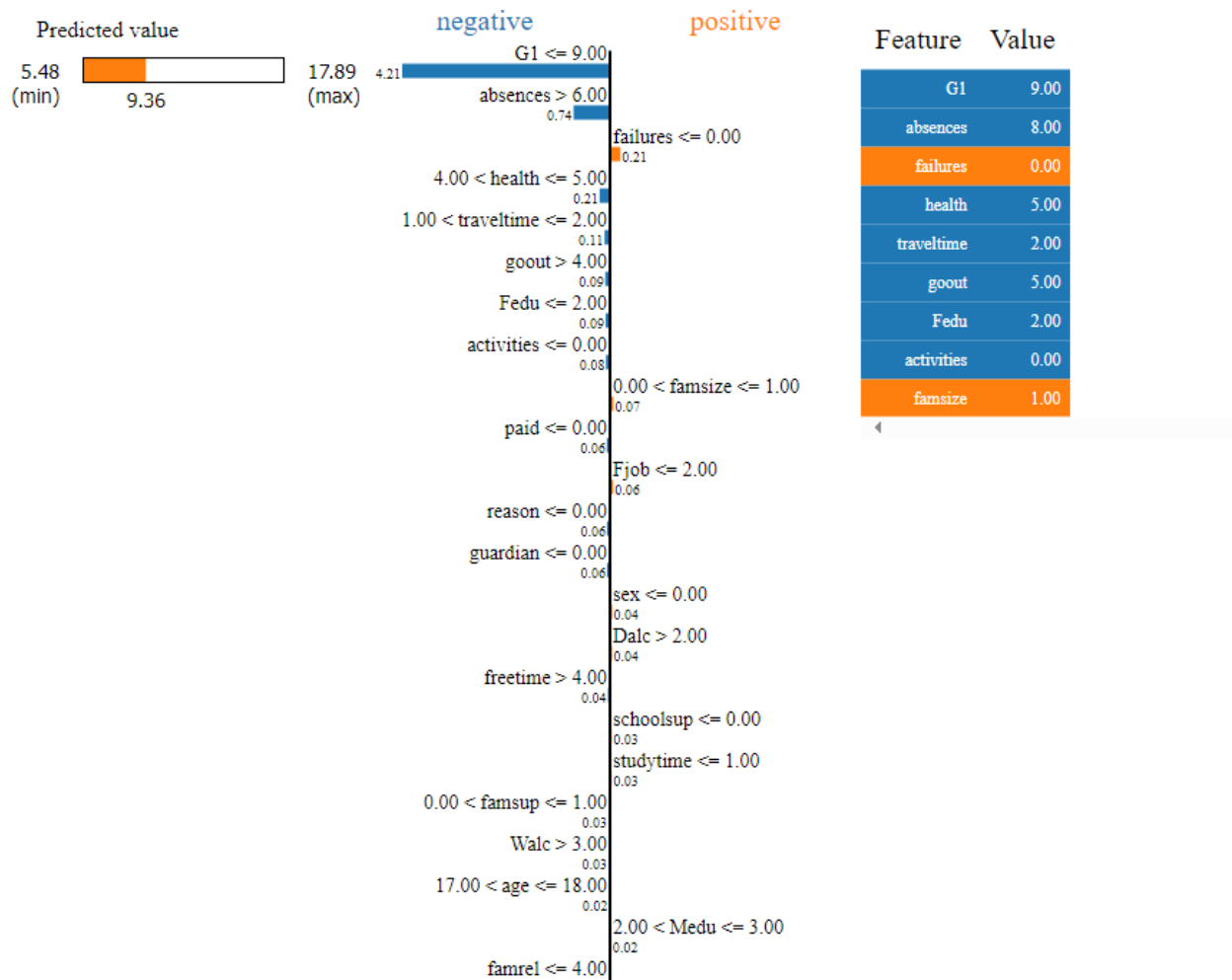




When we included G1 feature. Next we used the following models to predict G2.

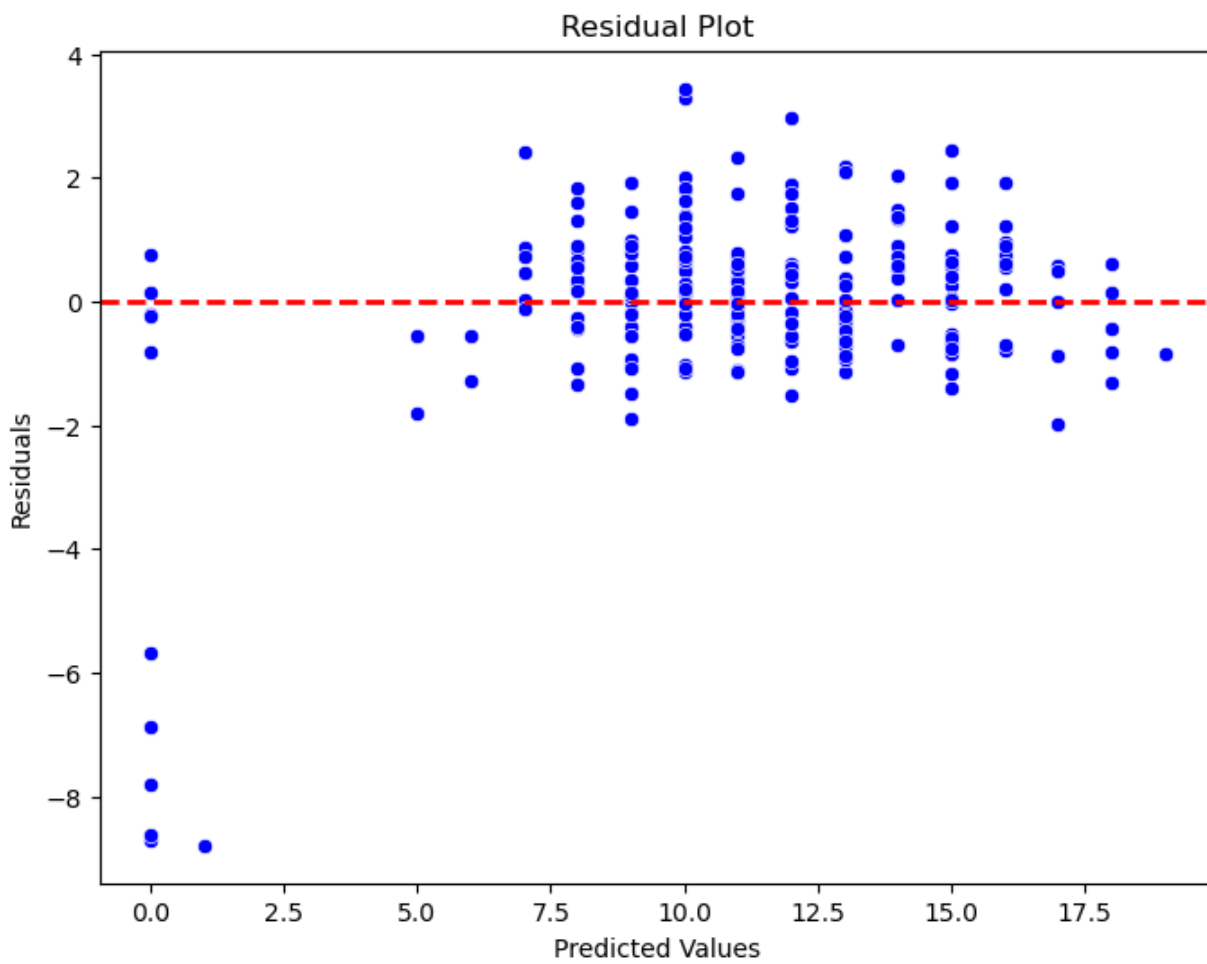
Model	R-squared (R2)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Linear Regression	0.726571	1.129493	2.952712
Decision Tree Regressor	0.410157	1.681818	6.369617
Support Vector Machine Regressor	0.72775	1.097762	2.939981
XGBoost Regressor	0.703859	1.257416	3.197976

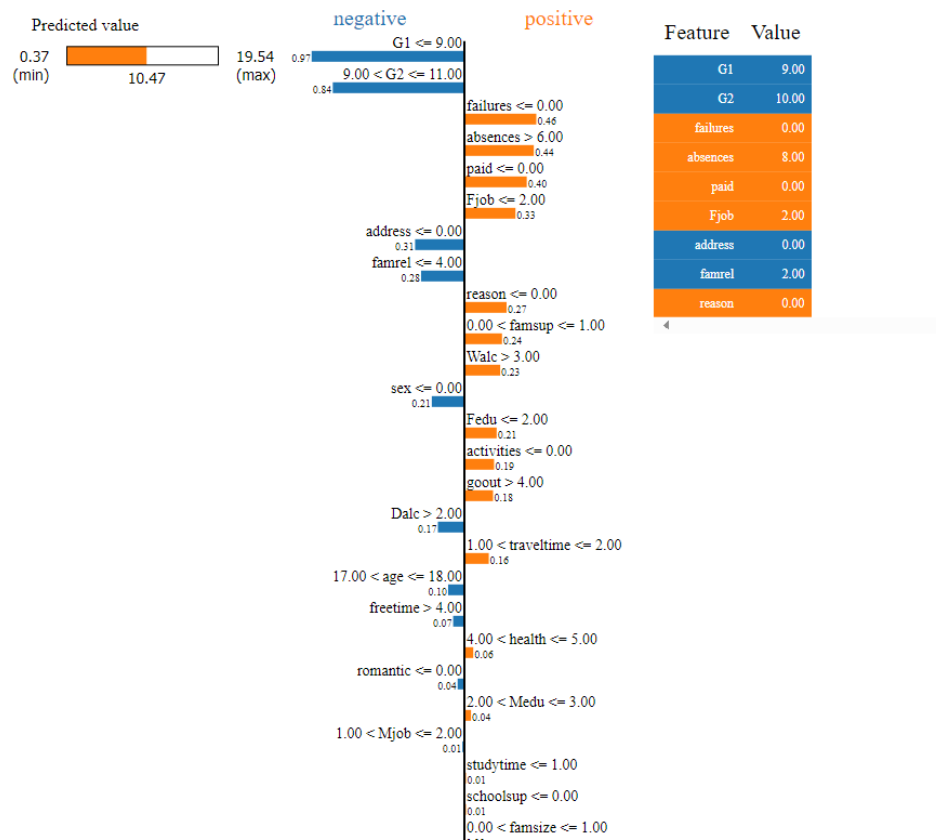




When we have the G1 and G2 feature. Next we used the following models to predict G3.

Model	R-squared (R2)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Linear Regression	0.802102	1.011723	3.059661
Decision Tree Regressor	0.612851	1.267943	5.985646
Support Vector Machine Regressor	0.789561	0.975559	3.253566
XGBoost Regressor	0.785801	1.077647	3.311697





We have also used the Hybrid model and target mean approach model to make predictions.

Hybrid Model Results:

model	R-squared (R2)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Linear Regression	0.599932	1.629285	6.18538
Decision Tree Regressor	0.327827	1.990431	10.392344
Support Vector Machine Regressor	0.588115	1.498565	6.368079
XGBoost Regressor	0.649568	1.578049	5.417968

Mean G Results:

model	R-squared (R2)	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Linear Regression	0.122859	2.474294	11.608247
Decision Tree Regressor	-0.248649	2.95614	16.524854
Support Vector Machine Regressor	0.1198	2.470506	11.648729
XGBoost Regressor	0.103347	2.56255	11.866478