|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Linear outlier | Polynomial Outlier | Linear without outlier | Polynomial without outlier |
| ScreenSize | *Goodness of Fit of Model Train Dataset*  *Explained Variance (R^2) : 0.31933165189620694*  *Mean Squared Error (MSE) : 0.03834475833749644*  *Goodness of Fit of Model Test Dataset*  *Explained Variance (R^2) : 0.10179997642308691*  *Mean Squared Error (MSE) : 0.048913316476672734* | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.3203876561071426  Mean Squared Error (MSE) : 0.03828526941549131  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.09930844678620487  Mean Squared Error (MSE) : 0.04904899781094225 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.26094451094874993  Mean Squared Error (MSE) : 0.0368336400994604  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.4140030805077928  Mean Squared Error (MSE) : 0.03147045940312778 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.2610244304052446  Mean Squared Error (MSE) : 0.036829657009501554  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.4131736150000693  Mean Squared Error (MSE) : 0.03151500513317995 |
|  | R^2 and MSE remains relatively same after using polynomial. Significant improvement in MSE and R^2 after removing outliers. | | | |
| Rom | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.2610244304052446  Mean Squared Error (MSE) : 0.036829657009501554  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.4131736150000693  Mean Squared Error (MSE) : 0.03151500513317995 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.25901081763666134  Mean Squared Error (MSE) : 0.04281069966960534  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.2736626346070964  Mean Squared Error (MSE) : 0.0355346337732239 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.18393089062698265  Mean Squared Error (MSE) : 0.04182425305417913  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.2364044652802635  Mean Squared Error (MSE) : 0.03654925602971778 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.2662798288174759  Mean Squared Error (MSE) : 0.03760379820536333  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.32589458634093516  Mean Squared Error (MSE) : 0.0322658400089871 |
|  | R^2 did not improve after using polynomial. No improvement after removing outliers. | | | |
| Ram | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.18521298645868323  Mean Squared Error (MSE) : 0.045403154876042864  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.23498304181337692  Mean Squared Error (MSE) : 0.04367285018541827 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.2779274105636822  Mean Squared Error (MSE) : 0.04023674047949218  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.3533109179776871  Mean Squared Error (MSE) : 0.036917816126131445 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.21040470677943157  Mean Squared Error (MSE) : 0.0391259254719013  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.150172853655415  Mean Squared Error (MSE) : 0.04603390366178158 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.33316721091445045  Mean Squared Error (MSE) : 0.033042813491915145  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.2340539475211404  Mean Squared Error (MSE) : 0.04149018649450959 |
|  | R^2 is and MSE improved after using polynomial. No improvement after removing outliers. | | | |
| Battery | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.23687455690815495  Mean Squared Error (MSE) : 0.04475041661658099  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.16199229923355962  Mean Squared Error (MSE) : 0.037820959515904515 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.24340716700603815  Mean Squared Error (MSE) : 0.04436733802036811  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.16583519851199235  Mean Squared Error (MSE) : 0.03764752180417421 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.20534475525199025  Mean Squared Error (MSE) : 0.039504482126895384  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.1723423444557033  Mean Squared Error (MSE) : 0.04430189729459535 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.3185099934911013  Mean Squared Error (MSE) : 0.033878729121489294  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.2843203553364382  Mean Squared Error (MSE) : 0.03830806844028604 |
|  | Not much change when using poynomial. R^2 increased significantly when using cleaned polynomial however best MSE is still from linear uncleaned. | | | |
| Price | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.1859703602265501  Mean Squared Error (MSE) : 0.04371243346180819  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.22424965181327294  Mean Squared Error (MSE) : 0.050746377582231446 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.25843717920913944  Mean Squared Error (MSE) : 0.03982105058311236  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.29977686622230726  Mean Squared Error (MSE) : 0.045805699760953236 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.30233033574082324  Mean Squared Error (MSE) : 0.03172138346295044  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.34738129210957946  Mean Squared Error (MSE) : 0.03112778089816519 | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.3297948023038453  Mean Squared Error (MSE) : 0.03047263936516012  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.3744187353074736  Mean Squared Error (MSE) : 0.029838183162557604 |
|  | R^2 and MSE improved after using polynomial. R^2 and MSE improved significantly after removing outliers. | | | |

|  |  |
| --- | --- |
| Polynomial regression, 2 predictor : Price,Ram) | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.3635922059663619  Mean Squared Error (MSE) : 0.029080058056459208  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.28517474218624594  Mean Squared Error (MSE) : 0.03326075945196376 |
| (Linear regression, 2 predictor : Price,Ram) | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.33934698811698616  Mean Squared Error (MSE) : 0.030187920576782275  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.2685715817535982  Mean Squared Error (MSE) : 0.03403330311806424 |
|  | MSE did not improve. |
| (Polynomial regression, 3 predictor : Price,Ram,Battery) | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.33607611135716264  Mean Squared Error (MSE) : 0.03012410055825894  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.38413601835643596  Mean Squared Error (MSE) : 0.029486920372112126 |
|  | MSE improved. |
| (Linear regression, 3 predictor : Price,Ram,Battery) | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.3242127102704859  Mean Squared Error (MSE) : 0.03066237654653304  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.33032901776552326  Mean Squared Error (MSE) : 0.03206314302707604 |
| (Polynomial regression, 4 predictor : Price,Ram,Battery,Scrensize) | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.4291521675331773  Mean Squared Error (MSE) : 0.02635377155704498  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.16377432648079282  Mean Squared Error (MSE) : 0.036859156039343616 |
|  | MSE did not improve. |
| (Linear regression, 4 predictor : Price,Ram,Battery,Scrensize) | Goodness of Fit of Model Train Dataset  Explained Variance (R^2) : 0.36385184866691345  Mean Squared Error (MSE) : 0.029368427281613634  Goodness of Fit of Model Test Dataset  Explained Variance (R^2) : 0.23498889062371742  Mean Squared Error (MSE) : 0.033720160412755 |

There are several reasons why the R-squared (R^2) value might decrease after removing outliers from the data:

Reduction in sample size: Outliers are data points that are significantly different from the rest of the data. When outliers are removed from the data, the sample size is reduced, which can lead to a decrease in the R^2 value. This is because with a smaller sample size, the model has less data to work with, and therefore might not be able to capture the variability in the dependent variable as well as before.

Non-linear relationships: Outliers can sometimes represent non-linear relationships in the data. When these outliers are removed, the model might lose some of its ability to capture the non-linear relationships, leading to a decrease in the R^2 value.

Influence on regression line: Outliers can have a strong influence on the regression line. When outliers are removed, the regression line might change significantly, which can lead to a decrease in the R^2 value.

Other outliers: Outliers can sometimes be indicative of other outliers in the data. When one outlier is removed, it might reveal other outliers that were previously hidden, leading to a decrease in the R^2 value.

In summary, the R^2 value might decrease after removing outliers due to a reduction in sample size, loss of ability to capture non-linear relationships, influence on the regression line, or revealing other outliers in the data. However, it's important to carefully evaluate the reasons for the decrease and consider the impact of outliers on the model's performance.