**Data Science Career Track**

**Model Metrics Exercise**

1. Look at the table below. If the goal is to optimize the True Positives which model would you choose and why?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Recall** | **Precision** | **Accuracy** | **F1** |
| **Logistic** | 0.746 | 0.775 | 0.999 | 0.761 |
| **Logistic with auto threshold** | 0.891 | 0.061 | 0.976 | 0.114 |
| **Logistic with class weights** | 0.878 | 0.110 | 0.988 | 0.195 |
| **Hinge with auto threshold** | 0.905 | 0.014 | 0.890 | 0.028 |
| **Hinge with class weights** | 0.878 | 0.103 | 0.987 | 0.185 |

I would choose Logistic model. The reason I would choose it is the precision for standard error is smaller on this model. Other models have the precision rate that is very low. The F1 score is closest to 1 compared to other models.

1. Calculate the F-1 scores for each model and identify the best model based on the F1 score.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Recall** | **Precision** | **F1** | **Auc/Roc** |
| **Deep NN** | 0.79 | 0.82 | .40 | 0.92 |
| **Logistic Regression** | 0.75 | 0.79 | .38 | 0.90 |
| **Random Forest** | 0.80 | 0.66 | .36 | 0.90 |
| **LinearSVC** | 0.74 | 0.75 | .37 | 0.82 |

1. Identify the best parameter values for ‘alpha’ and ‘L1-ratio’ based on the above comparison.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Parameter** | **Parameter** | **Metric** | **Metric** | **Metric** |
|  | **Alpha** | **L1-ratio** | **MAE** | **R-squared** | **RMSE** |
| **Linear Regression** | 0.5 | 0.2 | 84.27 | 0.277 | 158.1 |
| **Linear Regression** | 0.2 | 0.5 | 84.08 | 0.264 | 159.6 |
| **Linear Regression** | 0.5 | 0.5 | 84.12 | 0.272 | 158.6 |
| **Linear Regression** | 0 | 0 | 84.49 | 0.249 | 161.2 |