

# Random Forest Review

## 1. Random Forest:

Categorical Features: *member\_casual* , *season*

Continuous Features: *TMAX*

Target Variable: *rideable\_type*

NaN Values: dropped all the rows with null values

Random Forest Accuracy: 0.64

Random Forest Confusion Matrix:  $\begin{bmatrix} 1650972 & 10486 \\ 937518 & 11376 \end{bmatrix}$

The Random Forest model achieves an accuracy of 64%, indicating moderate effectiveness in predicting the correct *rideable\_type*. However, the confusion matrix reveals a significant imbalance in predictive performance: while true negatives are high at 1,650,972, suggesting good specificity, the model suffers from a high number of false negatives (937,518), indicating poor sensitivity or a strong bias toward the more prevalent class. This results in only 11,376 true positives, which is concerning for applications needing reliable identification of the positive class. The low false positives (10,486) suggest that the model is conservative, potentially at the cost of missing many true positive cases.

Categorical Features: *member\_casual* , *season*

Continuous Features: *TMAX*

Target Variable: *rideable\_type*

NaN Values: Handled NaN Values.

Random Forest Accuracy: 0.53

Random Forest Confusion Matrix:  $\begin{bmatrix} 749533 & 916209 \\ 674585 & 1075940 \end{bmatrix}$

The Random Forest model shows an accuracy of 53%, which is lower than might be desirable for robust classification tasks. The confusion matrix indicates a nearly balanced but still challenging scenario with considerable false positives

(916,209) and false negatives (674,585), alongside true positives (1,075,940) and true negatives (749,533). This performance suggests that handling null values by imputation rather than dropping them has not necessarily improved the model's predictive accuracy or the balance between sensitivity and specificity. The model's ability to correctly classify the positive class has improved in terms of true positives but at the cost of increasing false positives, highlighting a trade-off between detecting more positives and making more errors in classifying negatives.

Categorical Features: *member\_casual* , *day\_of\_week*

Continuous Features: *TMAX*

Target Variable: *rideable\_type*

Random Forest Accuracy: 0.54

Random Forest Confusion Matrix:  $\begin{bmatrix} 843729 & 822013 \\ 746061 & 1004464 \end{bmatrix}$

The Random Forest model's performance with a new feature set achieves an accuracy of 54%, a slight improvement from previous models. The confusion matrix reveals a more balanced result between true positives (1,004,464) and true negatives (843,729), but also high numbers of false positives (822,013) and false negatives (746,061). This indicates that while the model has become slightly better at identifying both classes, it still struggles with a significant error rate in both predicting false outcomes. The introduction of new features appears to have moderately enhanced the model's ability to detect positive cases

Categorical Features: *member\_casual* , *day\_of\_week*, *season*

Continuous Features: *TMAX*, *Elevation\_Change*

Target Variable: *rideable\_type*

Random Forest Accuracy: 0.56

Random Forest Confusion Matrix:  $\begin{bmatrix} 925987 & 739755 \\ 774323 & 976202 \end{bmatrix}$

The Random Forest model with a new feature set yields an accuracy of 56%, showing a modest improvement over previous versions. The confusion matrix presents a mixed scenario with 925,987 true positives and 976,202 true negatives, balanced against high false positives (739,755) and false negatives

(774,323). This result indicates a slight improvement in model performance, capturing more true cases but still suffering from a considerable number of incorrect predictions.