|  |  |  |
| --- | --- | --- |
| Original Dataset (n=28,499) | | |
| Class | Serious | Slight |
| Number of samples | 4,479 | 24,020 |
| Percentage | 15.72% | 84.28% |

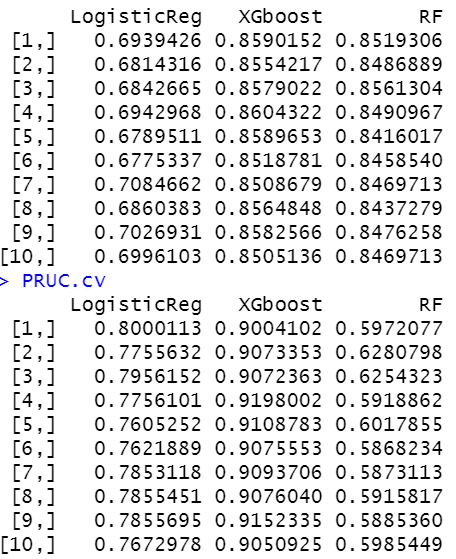
|  |  |  |
| --- | --- | --- |
| 10% of Original Dataset (n=2,849) | | |
| Class | Serious | Slight |
| Number of samples | 447 | 2,402 |
| Percentage | 15.69% | 84.31% |

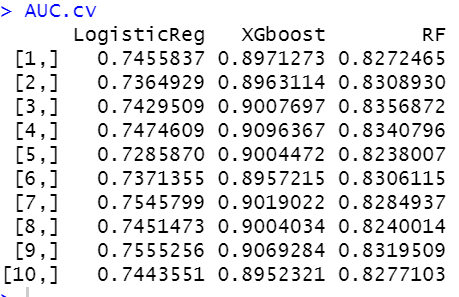
|  |  |  |
| --- | --- | --- |
| 90% of Original Dataset (n=25,650) | | |
| Class | Serious | Slight |
| Number of samples | 4,032 | 21,618 |
| Percentage | 15.72% | 84.28% |

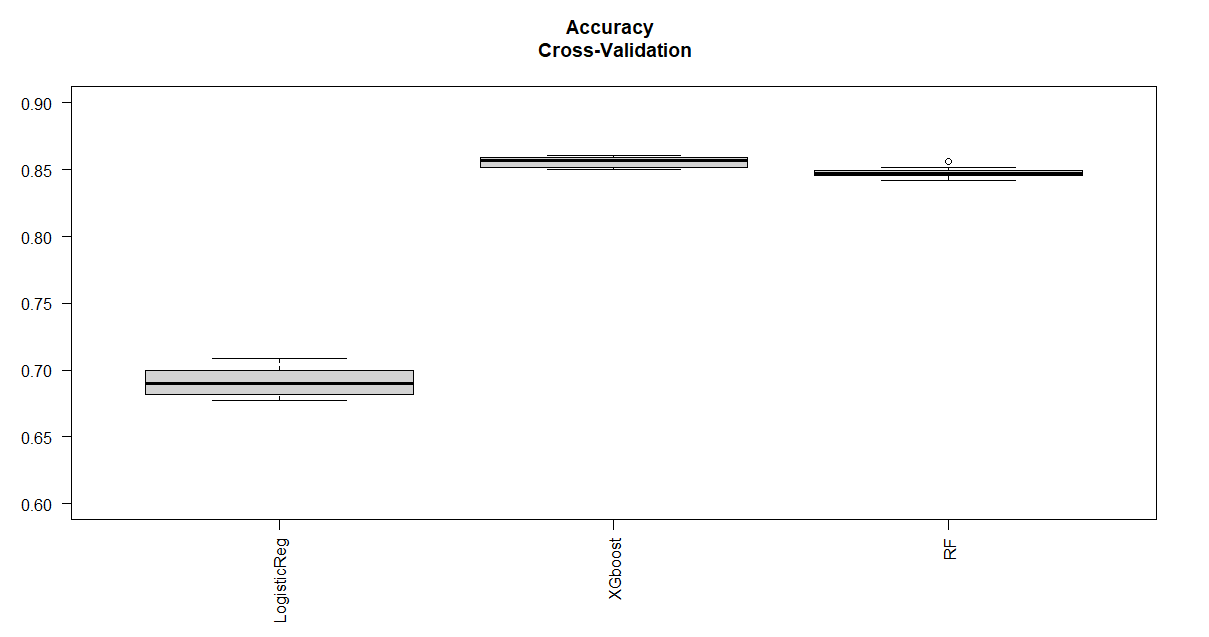
**Approach 1: SMOTE**

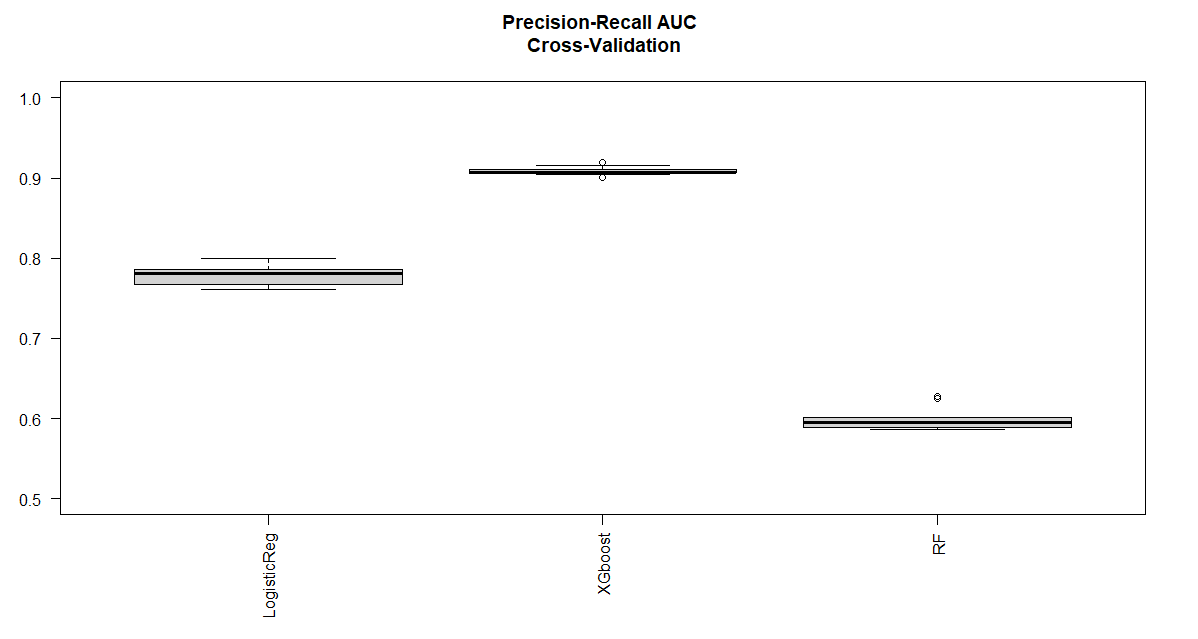
|  |  |  |
| --- | --- | --- |
| SMOTE Dataset (n=28,224) | | |
| Class | Serious | Slight |
| Number of samples | 12,096 | 16,128 |
| Percentage | 42.86% | 57.14% |

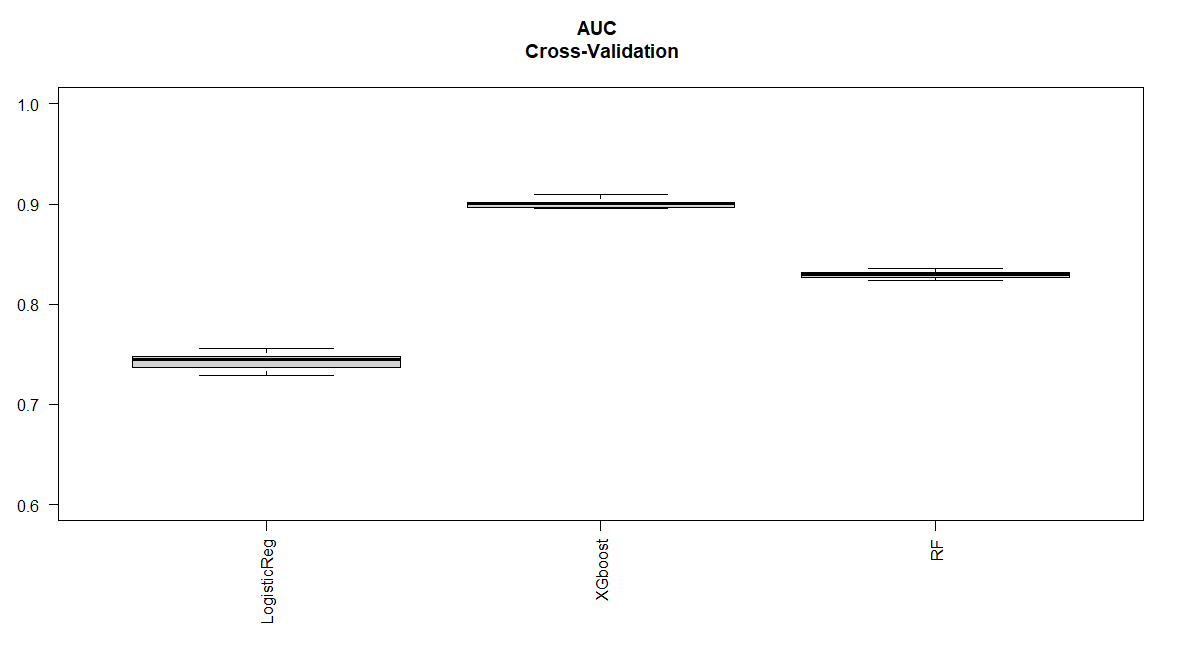
-10-fold CV



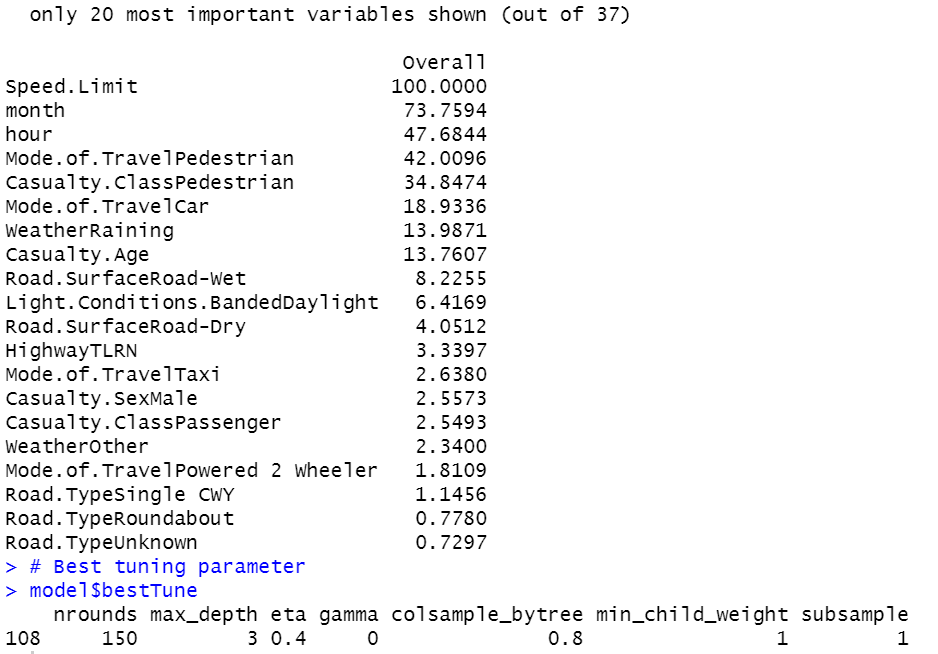








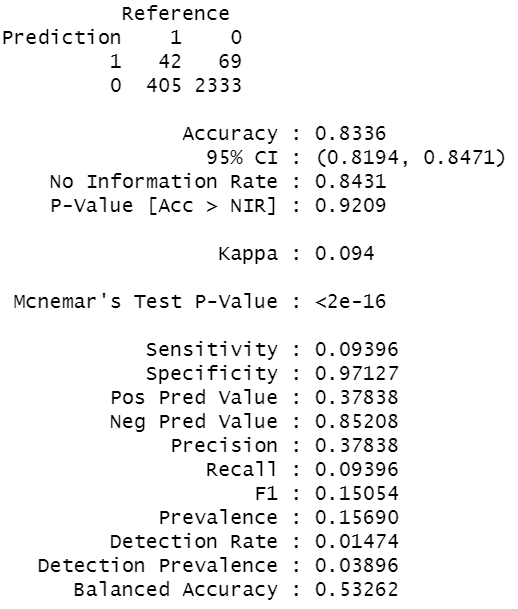
-pick XGboost -> tuning

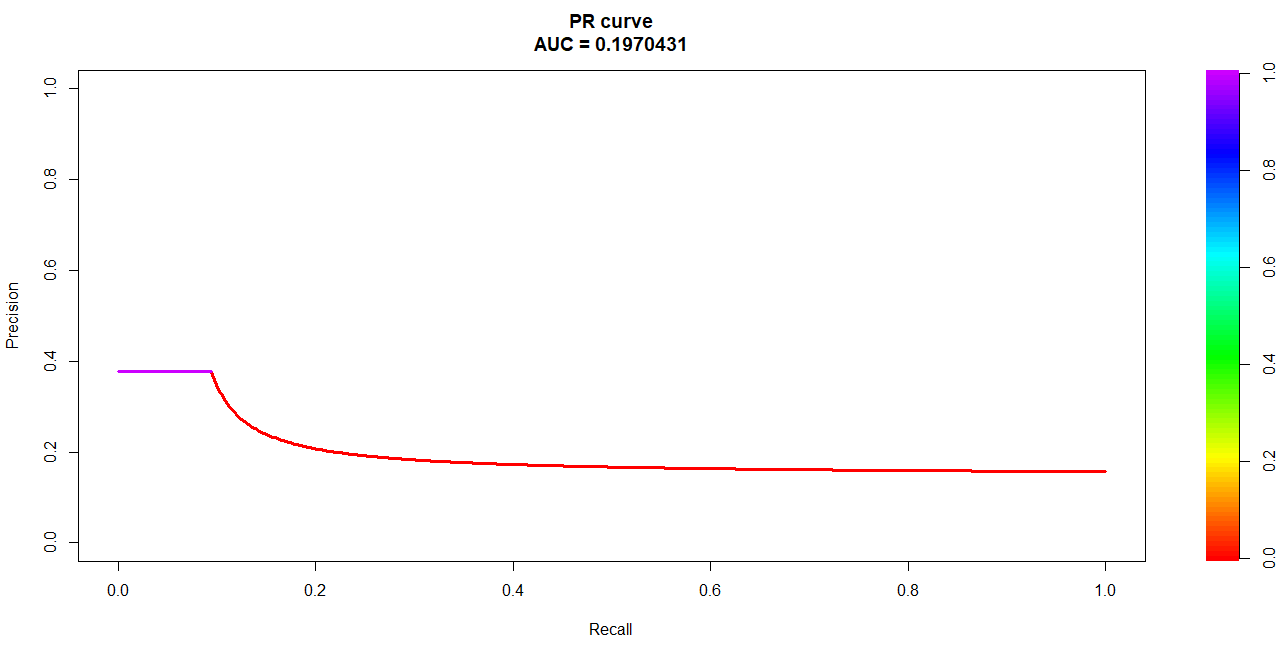


0.85

-making prediction on test dataset

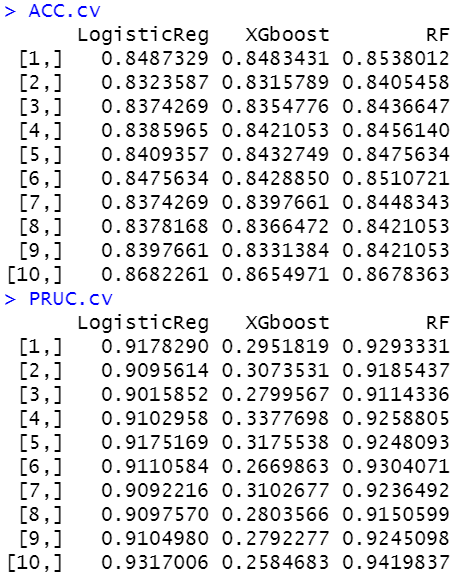
Accuracy= 0.8339

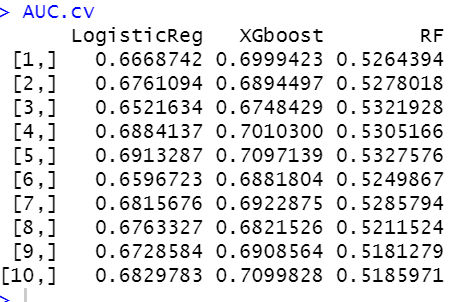


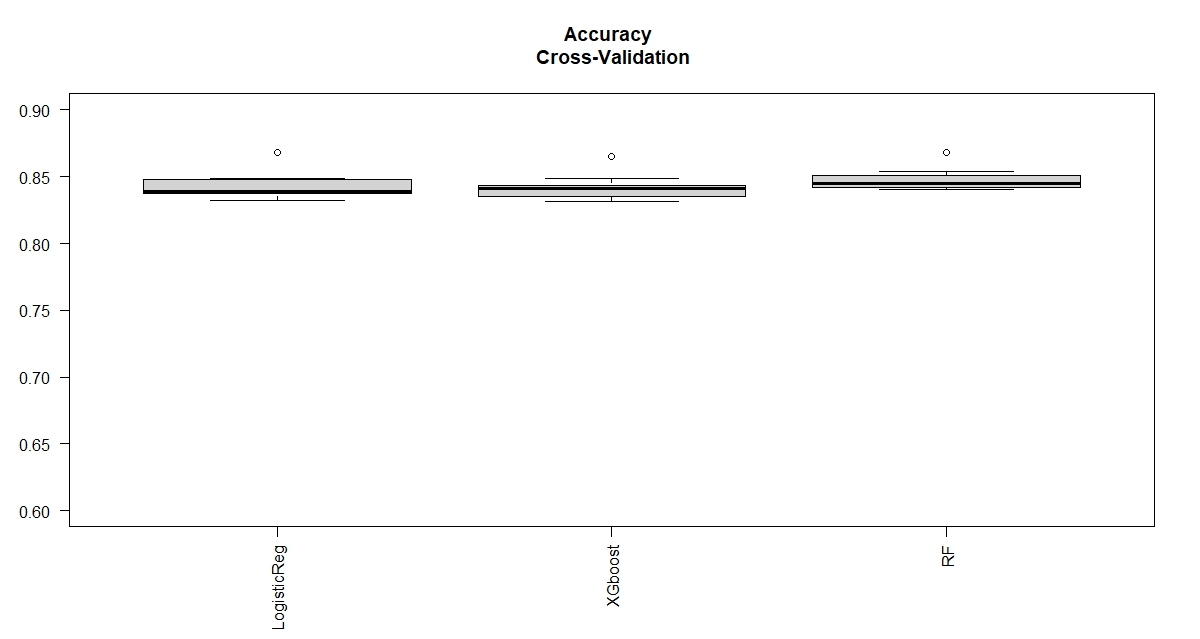


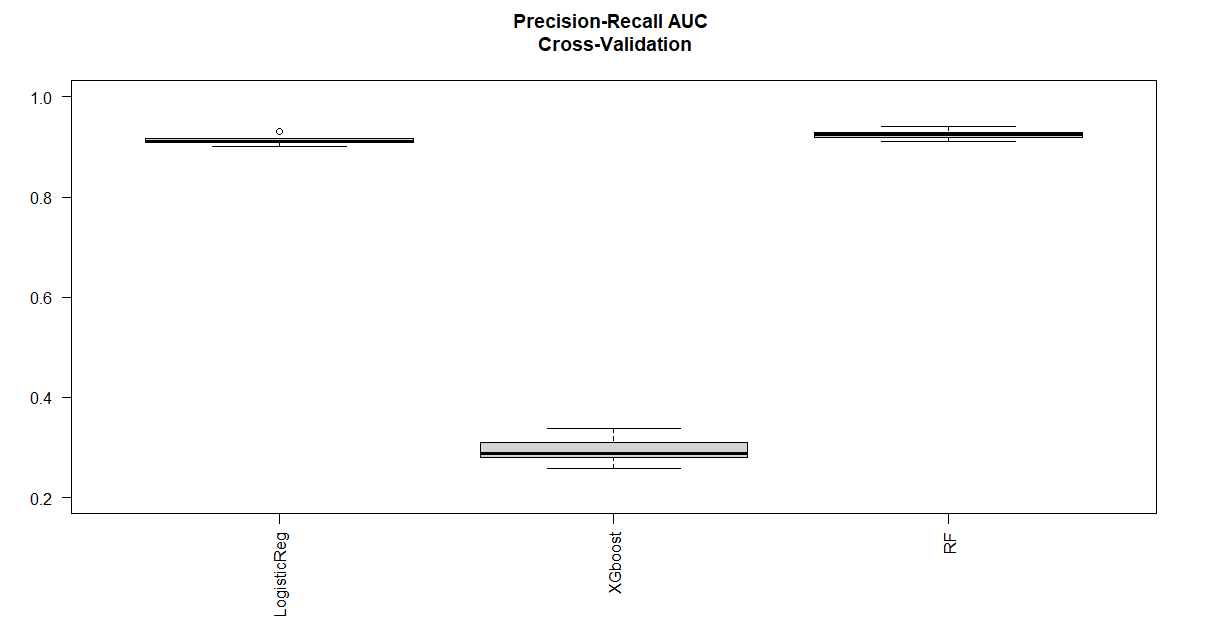
**Approach 2: normal**

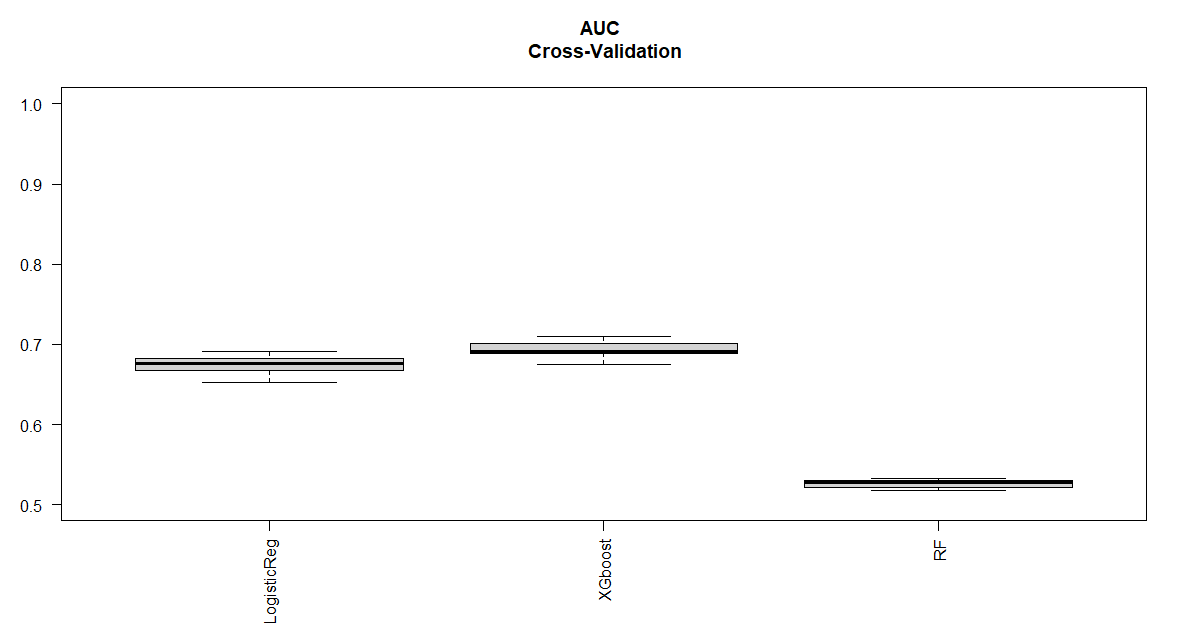
|  |  |  |
| --- | --- | --- |
| Normal Dataset (n=25,650) | | |
| Class | Serious | Slight |
| Number of samples | 4,032 | 21,618 |
| Percentage | 15.72% | 84.28% |



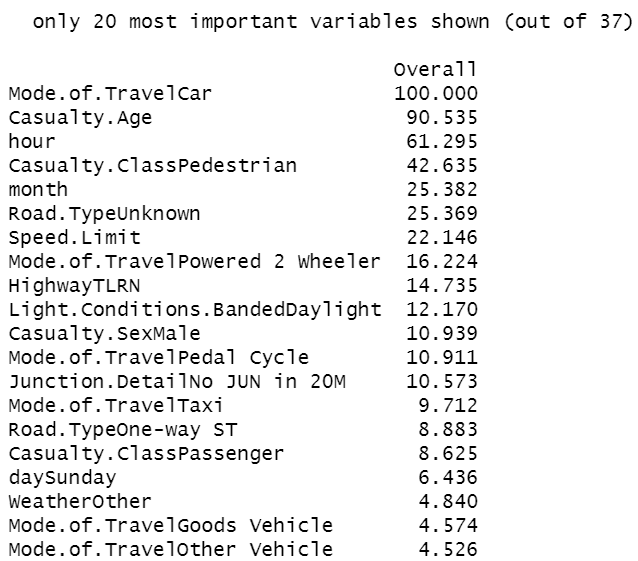


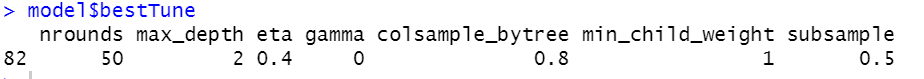






-pick XGboost -> tuning





-making prediction on test dataset

0.8431 (predict all 0 class Slight)

