# Feature selection June 15

XGBOOST 16-25 validated on 1-4 with different parameter selection

* All (295) 59.97
* All (200) that appeared in top 140 of any of the 16-40 models 59.98
* All (142) that appeared in top 100 of any of the 16-40 models 60.00
* All (80) that appeared in top 50 of any of the 16-40 models 59.70
* Consider 200 features in final blending but 142 in third major iteration submission (62%) – probably not losing much information at the benefit of reduced computation time!

# Hyperparameter selection June 15

Validation batches 5-8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Hyperset** | **#Rounds** | **ETA constant** | **Colsample** | **Tree depth** | **MAP@3** |
| Reference | 100 | 10 | 0.5 | 12 | 59.77 |
| 1 | 100 | 9 | 0.4 | 11 | 59.69 |
| 2 | **200** | 9 | 0.4 | 11 | 59.76 |
| 3 | **100** | **11** | 0.4 | 11 | 59.80 |
| 4 | **200** | 11 | 0.4 | 11 | 59.84 |
| 5 | **100** | **9** | **0.6** | 11 | 59.73 |
| 6 | **200** | 9 | 0.6 | 11 | 59.80 |
| *~~7~~* | *~~100~~* | *~~11~~* | *~~0.6~~* | *~~11~~* |  |
| *~~8~~* | *~~200~~* | *~~11~~* | *~~0.6~~* | *~~11~~* |  |
| Extra 1 | 100 (faster) | 12 (best dir) | 0.7 (best dir) | 11 (var next) | 59.83 |
| Extra 2 | 100 (faster) | 12 | 0.6 | 12 | 59.8 |
| Extra 3 | 100 (faster) | 12 | 0.6 | 10 | 59.83 |
| Final | **200** |  |  |  |  |

Conclusion: very little difference but better results for

* Increasing the number of rounds
* Higher ETA constants
* Colsample of 0.6