



80

81 **## Modeling**

82

83 We will end up building 20 models! Here is a quick synopsis of our methodology:

84

85 baseline model

86 baseline model addressing class imbalance

87 Baseline model addressing class imbalance with GridSearchCV performed to optimize hyperparameters

88 model with unimportant features dropped

89 model with unimportant features dropped addressing class imbalance

90 model with unimportant features dropped addressing class imbalance with GridSearchCV performed to optimize hyperparameters

91

92 This methodology will be applied to all 3 types of models we are building:

93

94 Decision Tree

95 Random Forest

96 K-Nearest Neighbors

97

98 Then we will read in all results and pick the 4 best models. Those 4 best models will be used to create a Stacking Ensemble and then a Stacking Ensemble with GridSearchCV to finetune hyperparameters. Once our best model is created, we can discuss results.

99

100 **## Conclusion**

101

116 The features with some of the biggest impact on our model are 'Overtime', 'Age',
'TotalWorkingYears', 'WorkLifeBalance', and 'MonthlyIncome'. Based on our EDA earlier
we know that the distribution of employee age is skewed to the right, meaning most
employees are going to be earlier in their career and some of them are going to be older
executives. More holistically, the feature importances tell us that a shift in company



