|  |  |
| --- | --- |
| Slide | Text |
| 13 | Results After combining and analyzing the information from the learning curve and all metrics calculated, we classified our results into:   * Error * Generalization * Learning |
| 14 | The model shows great fit and low error. The reasons are:  \* The train and test RMSE suggest decent performance, considering the potential task complexity.  \* Similarly, the MAE and the MSE on the test set, further support this.  \* Notably, the high train and test scores indicate strong performance on both seen and unseen data. |
| 15 | The model also generalizes well. This is supported by:  \* The mean cross-validation score closely aligns with the test score, suggesting the model doesn't overfit significantly.  \* Also, (switch to graph) the relatively constant gap between training and cross-validation error in the learning curve further supports this.  \* (Back) And the OOB score being near the test score also reinforces the model's ability to generalize. |
| 16 | (Graph) |
| 17 | And as we can see,  \* The learning curve seems to plateau around 75 - 100,000 training samples, indicating the model might reach its optimal capacity efficiently without requiring excessive data. |
| 18 | Conclusions |
| 19 | Potential areas for further investigation:  \* While the model performs well overall, the significant difference between train and test RMSE warrants further investigation.  \* This could be due to factors like data imbalance, noise in the test data, or insufficient model complexity. |
| 20 | Feature importance and potential overfitting:  \* Analyzing feature importance could reveal if certain features contribute to overfitting or noise in the model.  \* Techniques like Partial Dependence Plots could be used to understand how individual features influence the predictions. |
| 21 | In Summary,  This random forest model shows promising performance with good accuracy, generalization, and potentially efficient learning.  However, the higher test\_rmse compared to train\_rmse and potential feature-related issues warrant further investigation. |