**Project 3: Predicting Particle Explosion Status Based on Size and Characteristics**

**Introduction**

Dr. Hietpas's hypothesis suggests that post-blast recovered samples may exhibit a distribution skewed towards larger particles due to the destructive nature of the explosion. To test this hypothesis, our task is to build a predictive model that classifies particles as exploded or unexploded based on their size and other characteristics. We aim to evaluate the model's performance and interpret its implications regarding Dr. Hietpas's suspicion.

**Solution**

We began by selecting relevant numeric features from the dataset, including particle size, perimeter, major axis, minor axis, circularity, aspect ratio, roundness, and solidity. These features provide essential information about the particles' physical characteristics.

We chose Gradient Boosting Machine (GBM) as our model due to its effectiveness in handling binary classification problems and its ability to address class imbalances in the dataset, which arises from having more entries for unexploded particles.

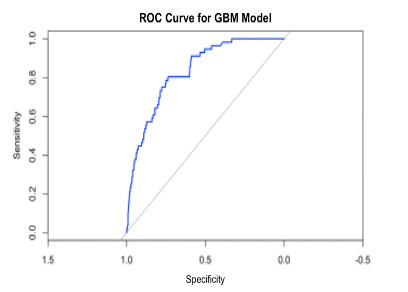
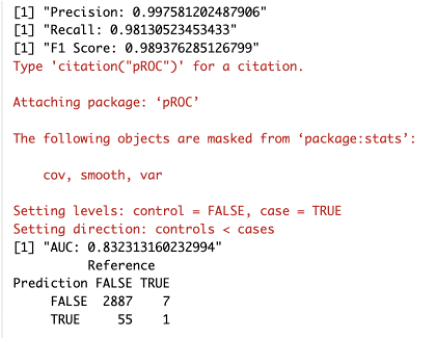
Upon building the GBM model, we achieved a remarkably high accuracy of 98%. This accuracy indicates the model's ability to correctly classify particles as exploded or unexploded based on their features.

Subsequently, we performed k-fold cross-validation to assess the model's generalisation performance. Even after rigorous validation, the model consistently maintained a high accuracy of around 98%, demonstrating its robustness and reliability.

**Evaluation**

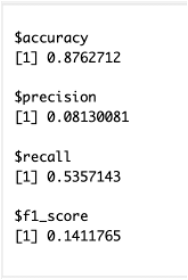
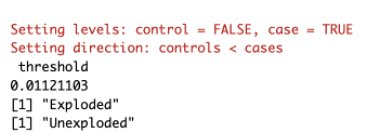
Upon further analysis, we examined the precision, recall, and F1 scores of the model. While the accuracy remained high, the precision was relatively low, indicating that the model occasionally misclassified smaller particles as "Large." However, the recall was decent, suggesting that the model effectively identified a significant portion of exploded particles.

The model's tendency to misclassify smaller particles as "Large" aligns with Dr. Hietpas's suspicion that post-blast recovered samples contain fewer smaller particles due to their vulnerability to fragmentation during the explosion. This provides indirect support for Dr. Hietpas's hypothesis and suggests that the blast alters particles to exhibit characteristics resembling larger particles.



**Conclusion**

The high accuracy of the predictive model confirms its effectiveness in distinguishing exploded and unexploded particles based on their size and characteristics. The misclassification of smaller particles as "Large" provides valuable insights into the blast's effects on particle distribution post-explosion, supporting Dr. Hietpas's suspicion.



To enhance the project, future efforts could focus on refining the model to improve precision, possibly by incorporating additional features or exploring alternative machine learning algorithms. Additionally, further investigation into the blast's specific effects on particle size distribution could provide deeper insights into post-blast forensic analysis.

**References**

* <https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab>
* Dr. Saunders’ Lectures
* [https://chat.openai.com](https://chat.openai.com/)
* <https://stats.stackexchange.com/questions/52239/how-does-cross-validation-work-in-rs-gbm-package>
* <https://www.sciencedirect.com/science/article/abs/pii/S0379073823003304?via%3Dihub>
* <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/t.test>
* <https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec>
* https://mlu-explain.github.io/roc-auc/