**Project 4: Predicting Brand of Small Arms Propellants (SAP) Particles from Recovered IEDs**

**Problem Background:**

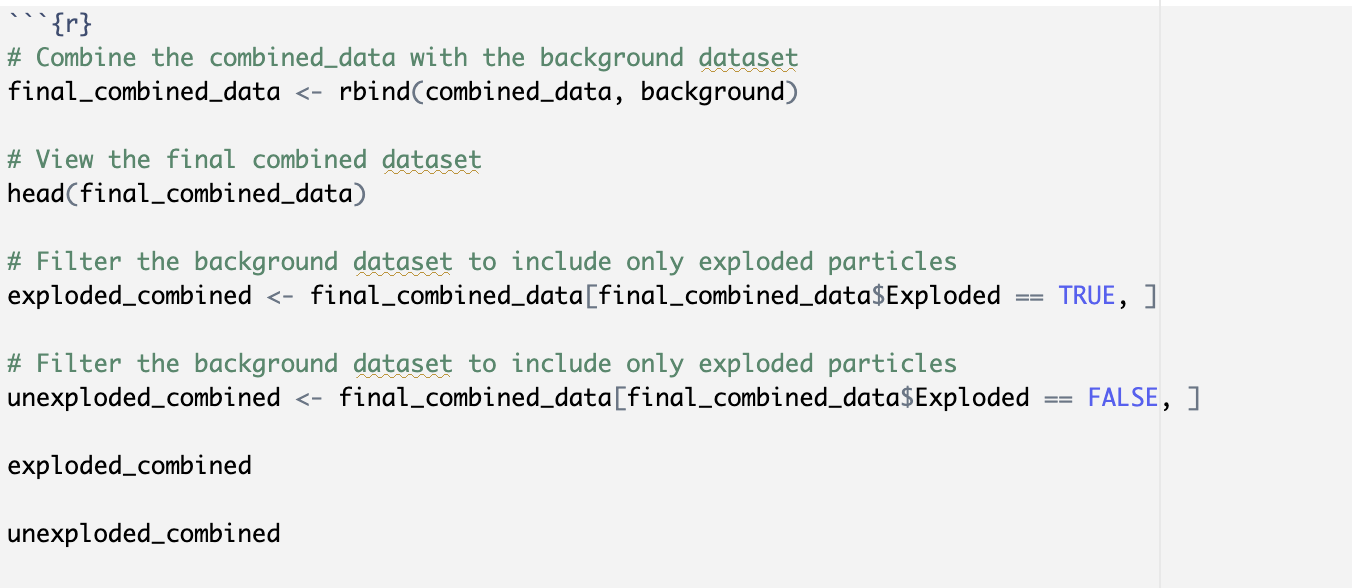
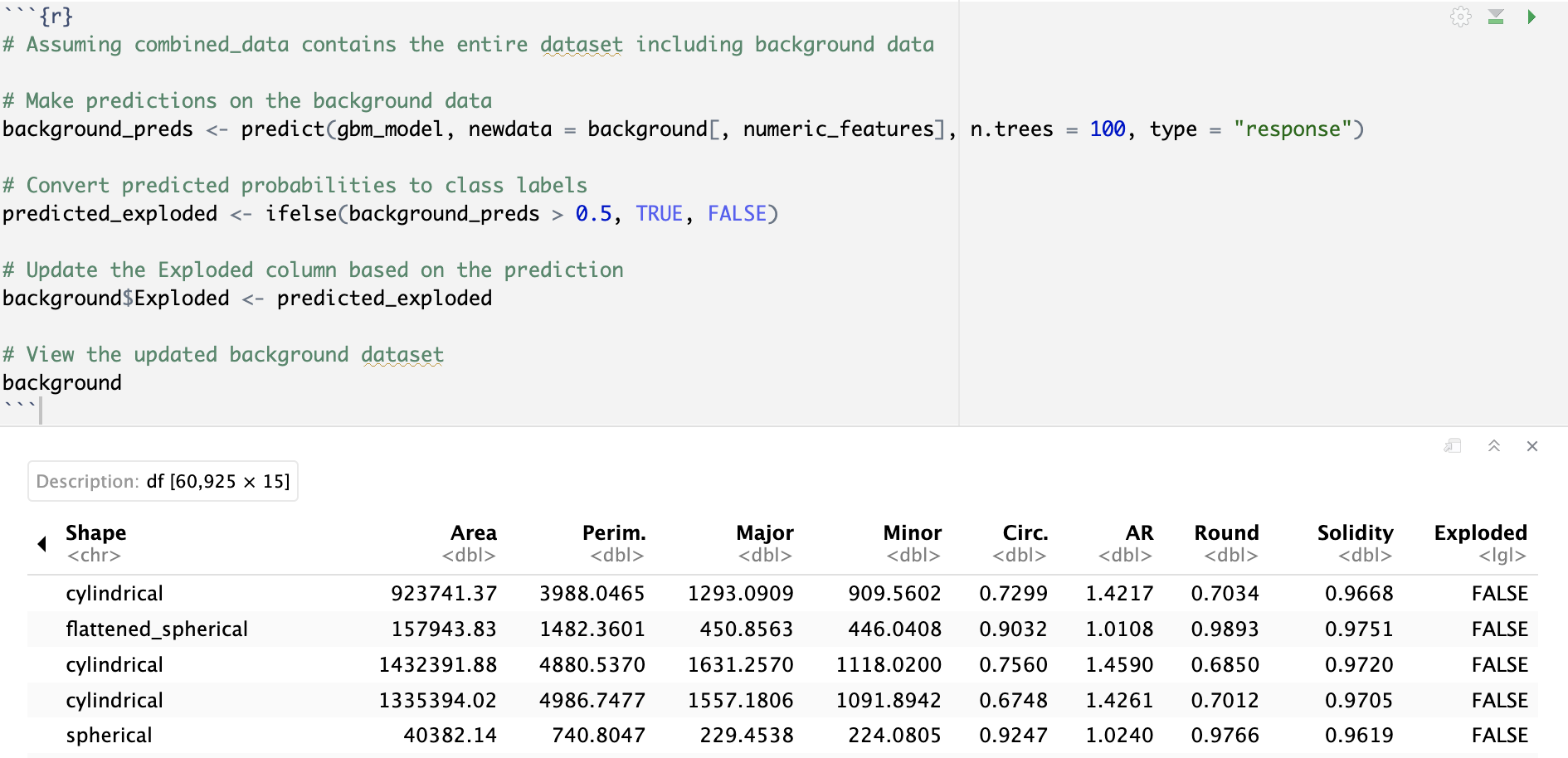
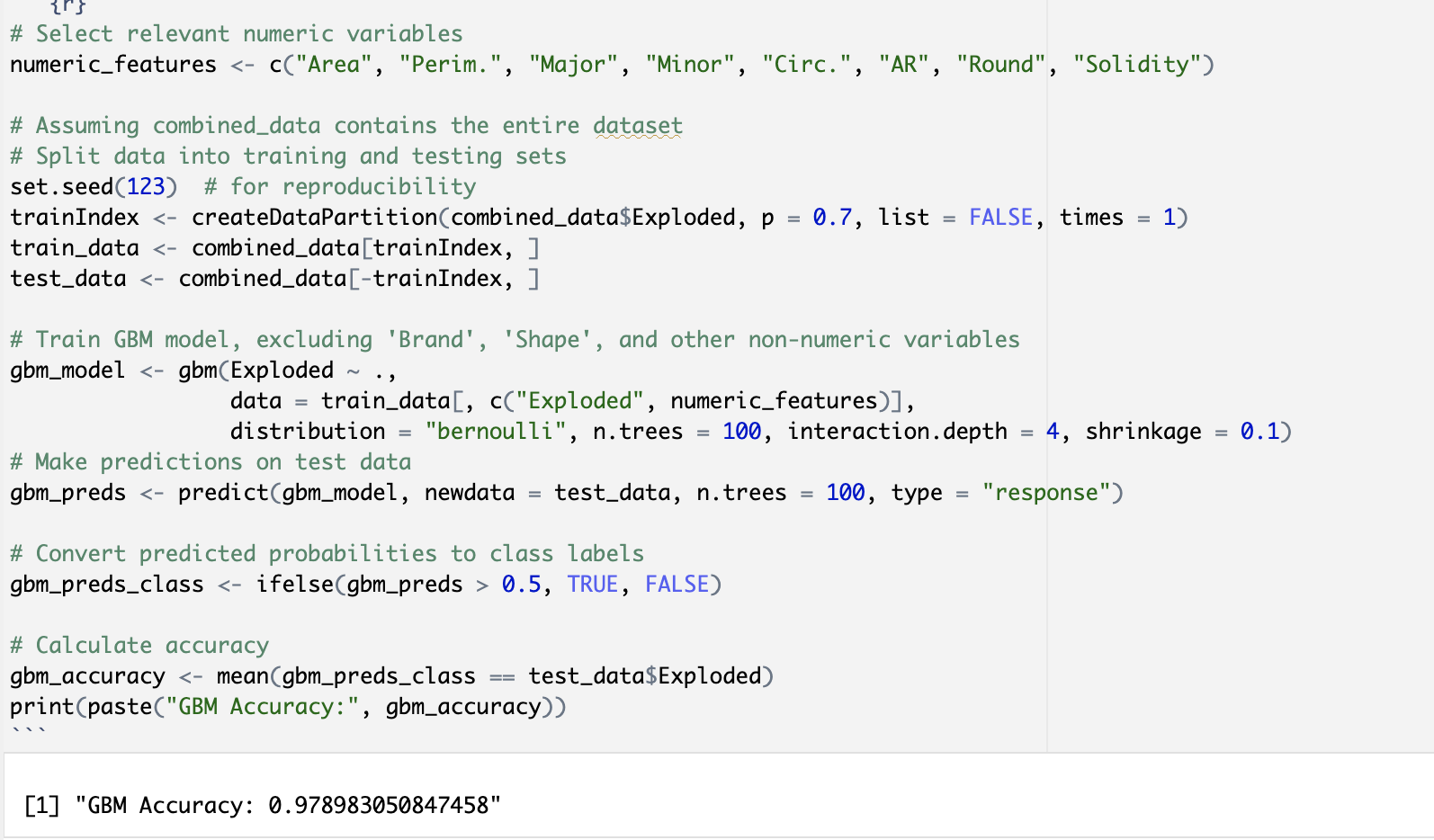
Small arms propellants (SAP), commonly known as canister powders, are materials that can be utilized for both legitimate ammunition assembly and constructing improvised explosive devices (IEDs). In this project, we were provided with 204 one-pound canisters of smokeless propellant, representing nine manufacturers and 154 unique brands. Each brand contributed a single sample of particles, which were analyzed using an automated image analysis routine to obtain size and shape measurements.

**Task:**

Our task was to develop algorithms to predict the brand of each particle in a query object recovered from IEDs. Specifically, we needed to build separate models for exploded and unexploded query objects due to the differing distribution of particles. Additionally, we had to consider brands beyond the five lots of interest and account for the potential misclassification of smaller particles as unexploded due to Dr. Hietpas' conjecture.

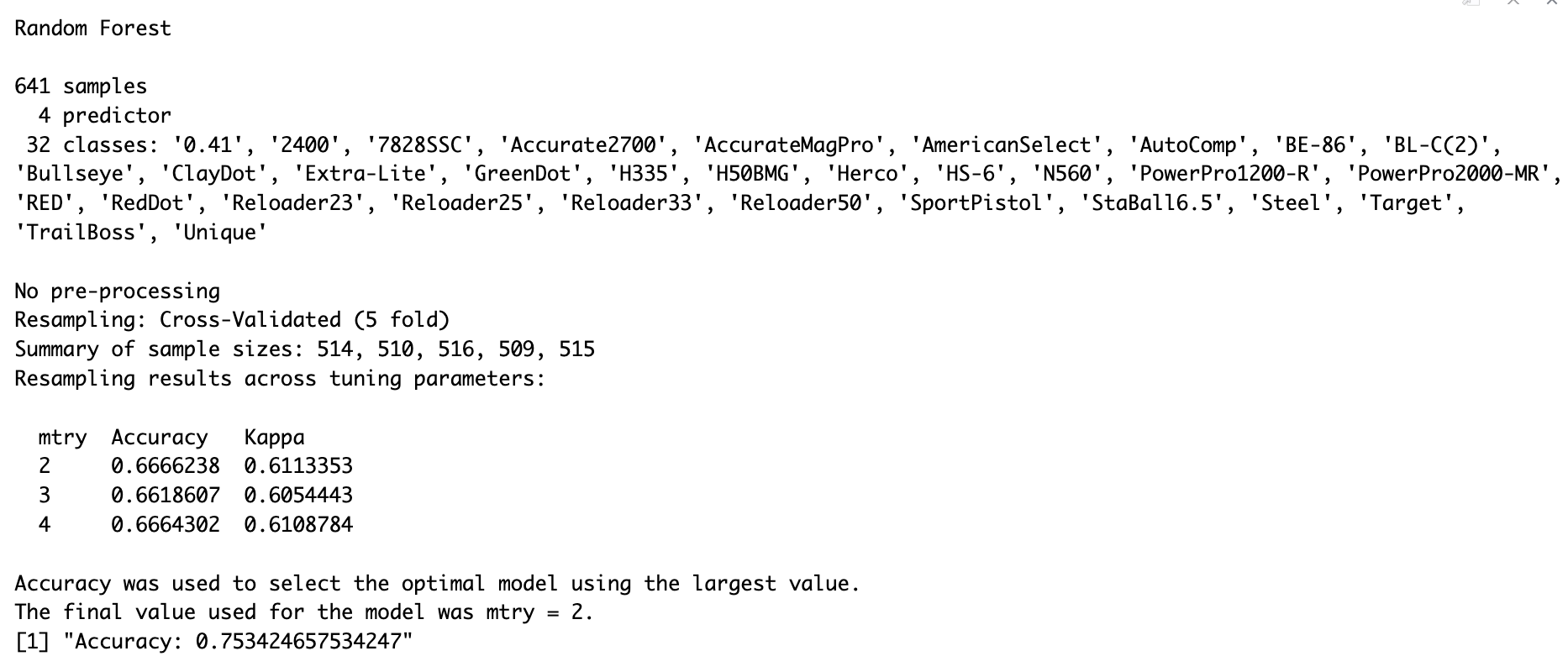
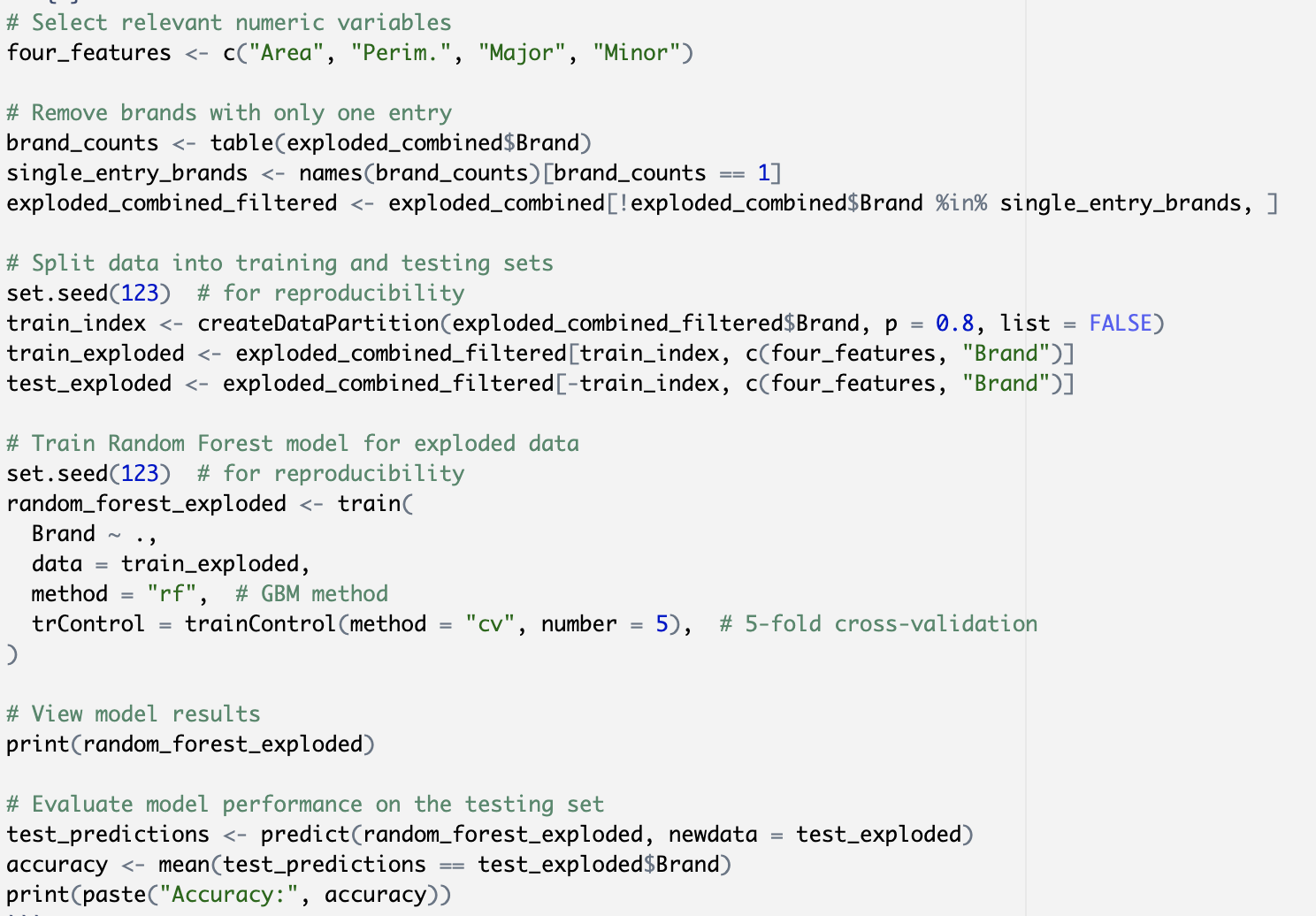
**Approach:**

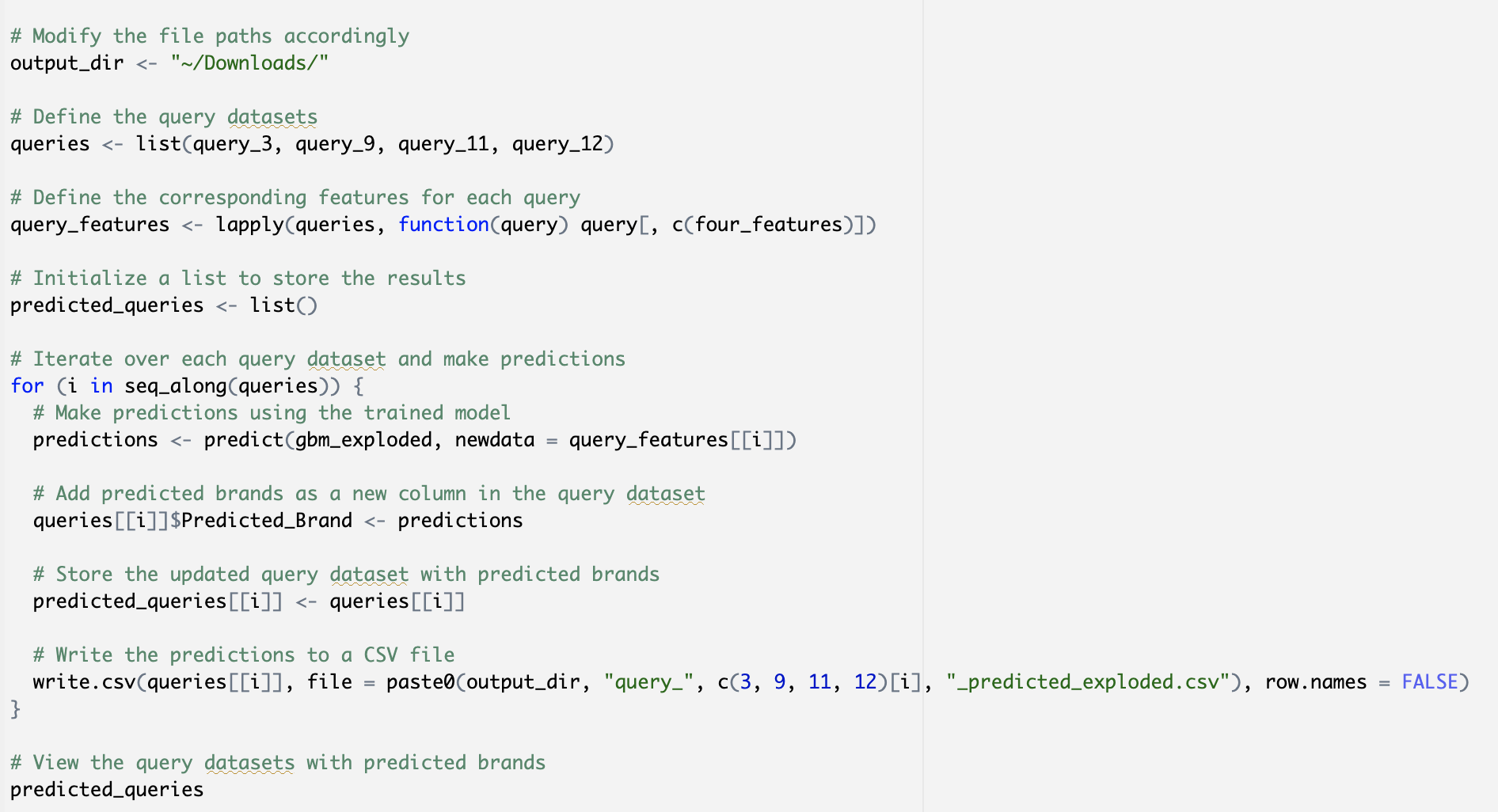
We began by using the model developed in Project 3 to predict whether particles were exploded or unexploded. Considering Dr. Hietpas' conjecture, we predicted the exploded or unexploded status of the background data, which initially contained unexploded particles. After combining exploded, unexploded, and background data, we split the dataset into exploded and unexploded datasets.

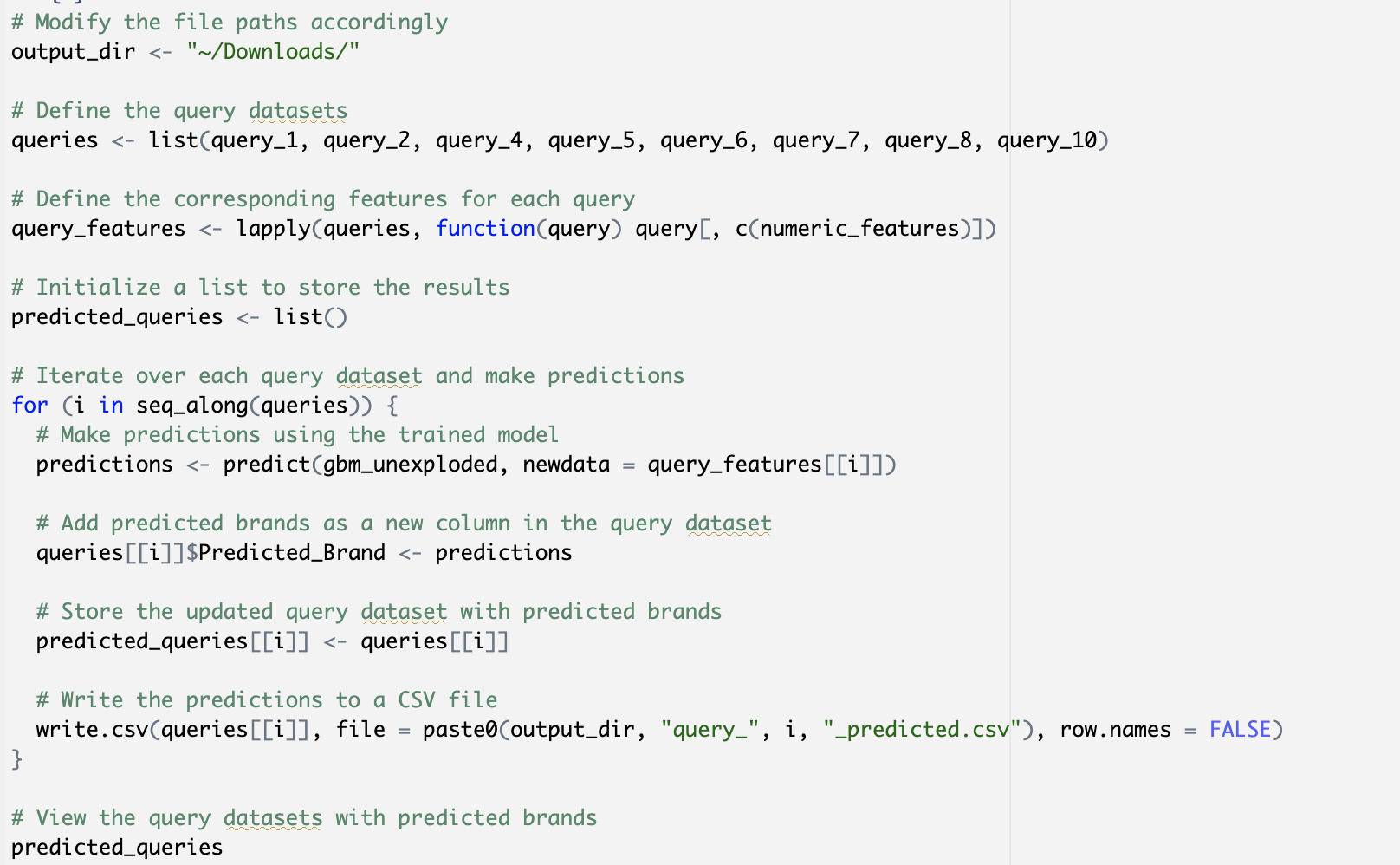
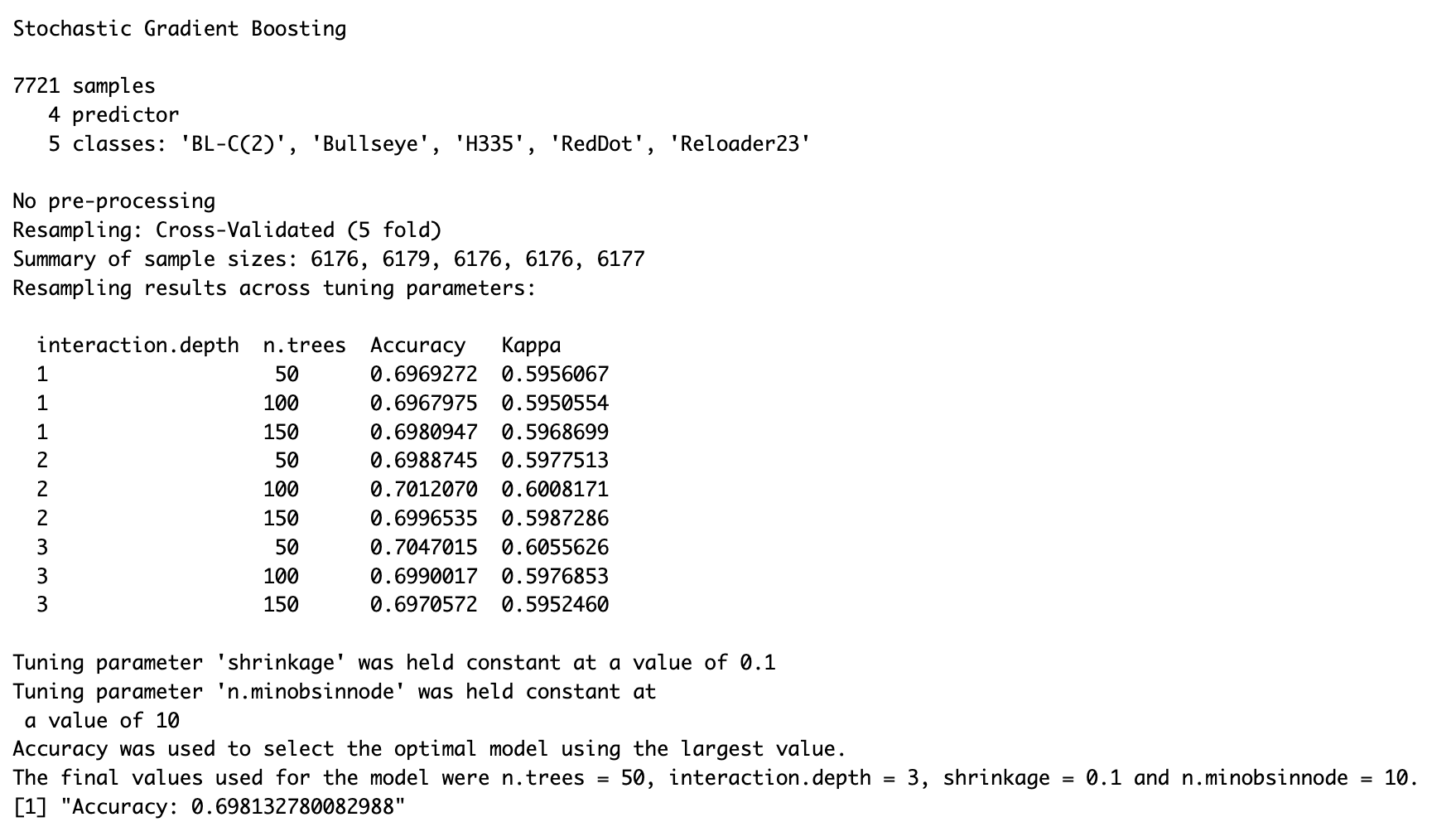


**Model Building Process:**

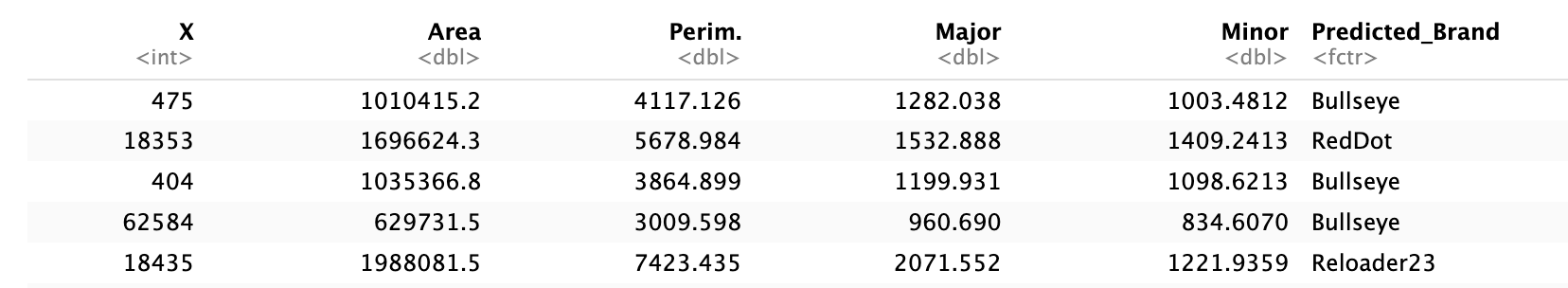
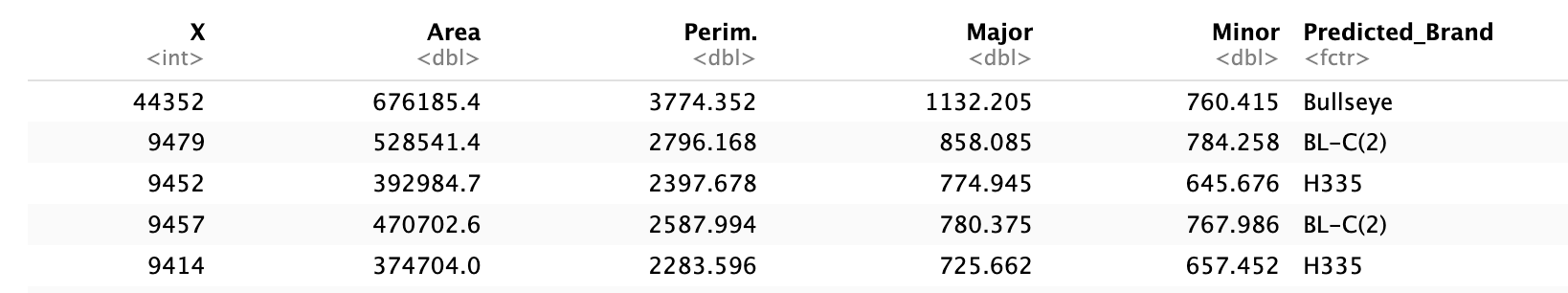
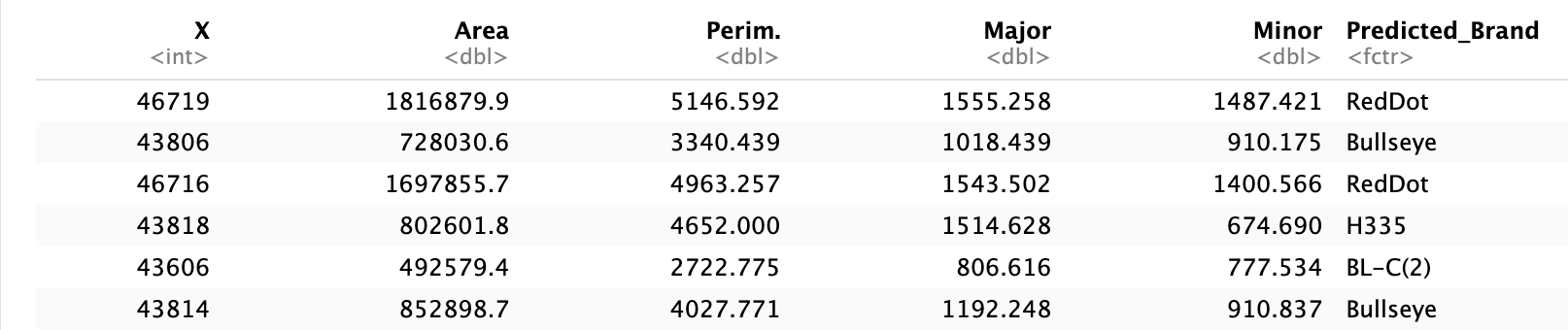
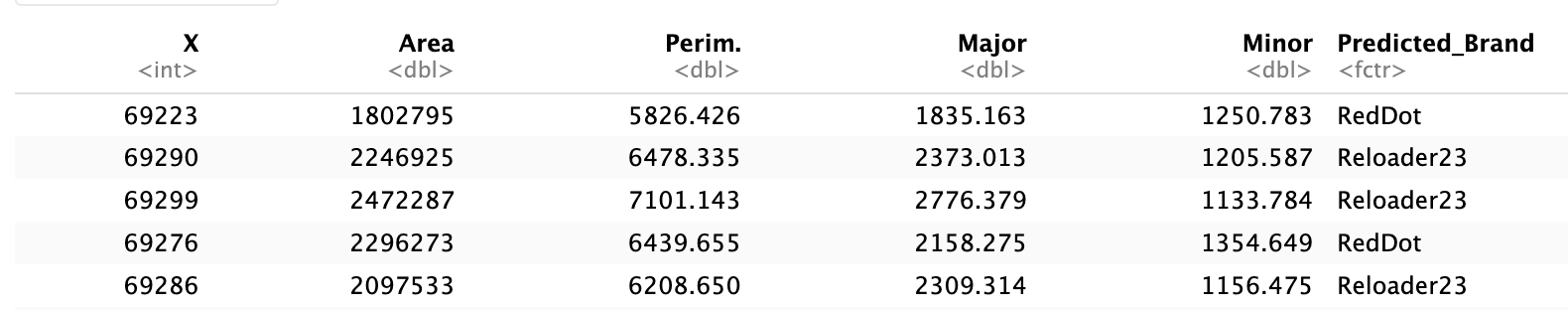
For both exploded and unexploded queries, we explored the performance of random forest and GBM algorithms. Random forest yielded higher accuracy for exploded queries, while GBM performed better for unexploded queries. Due to challenges with brands having only one entry, we omitted them from the analysis to ensure the focus remained on the five brands of interest.



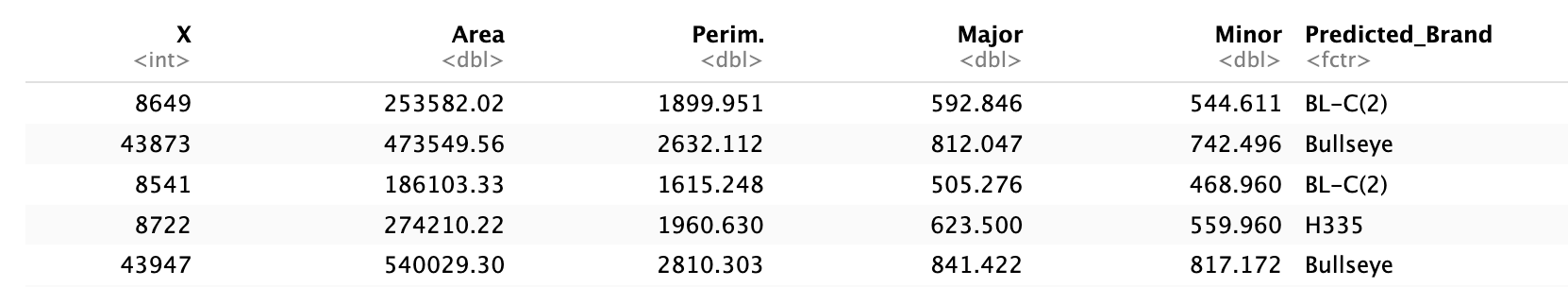
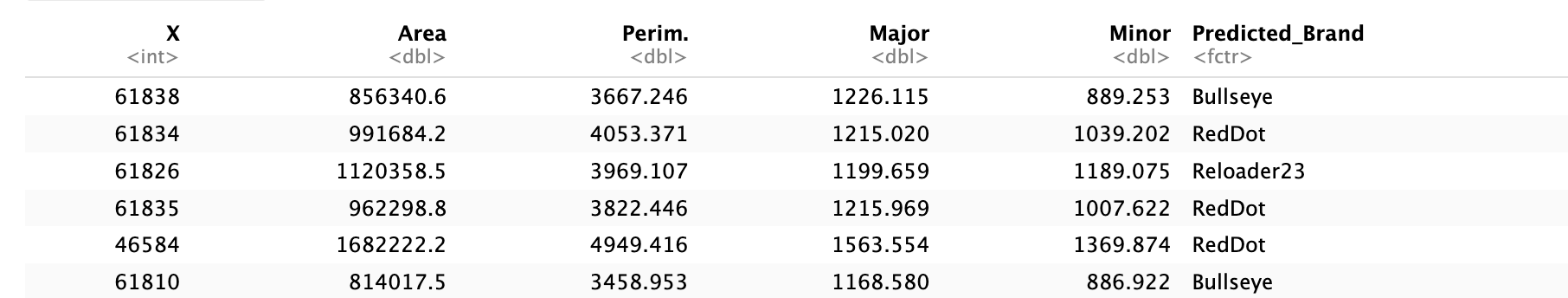
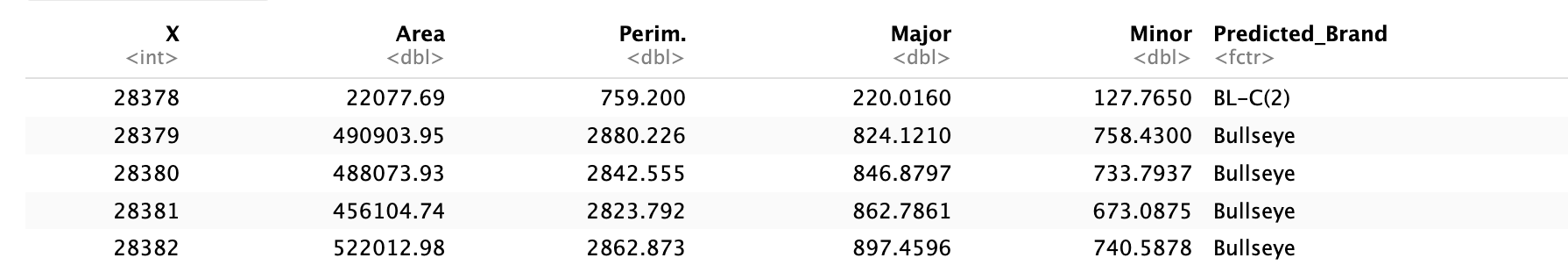
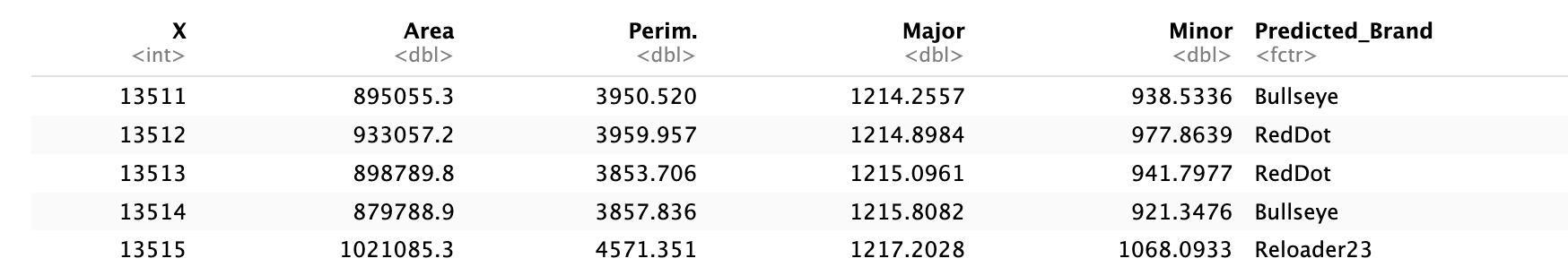
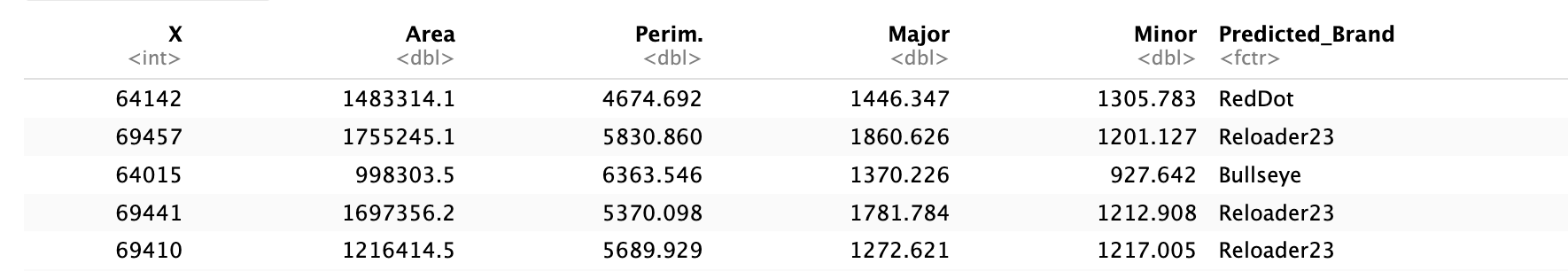
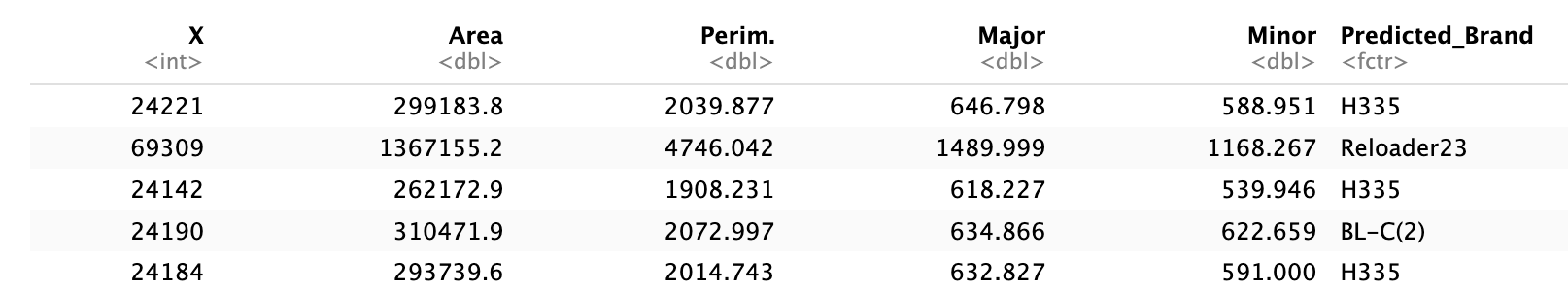
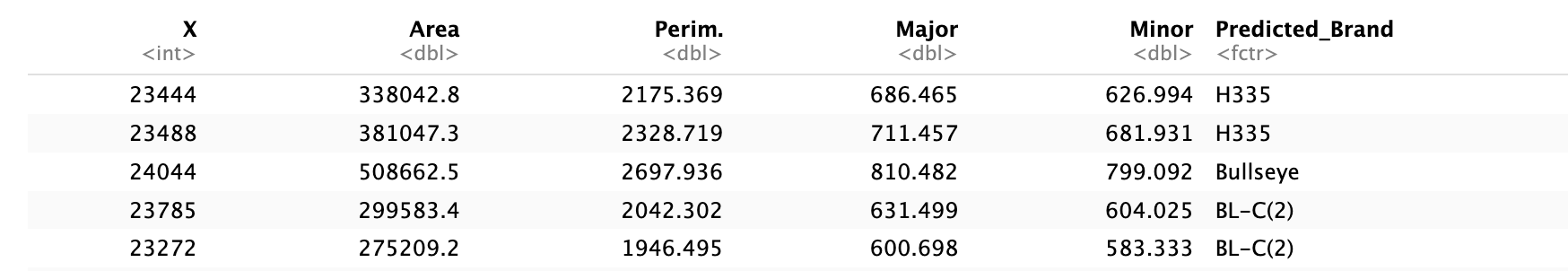
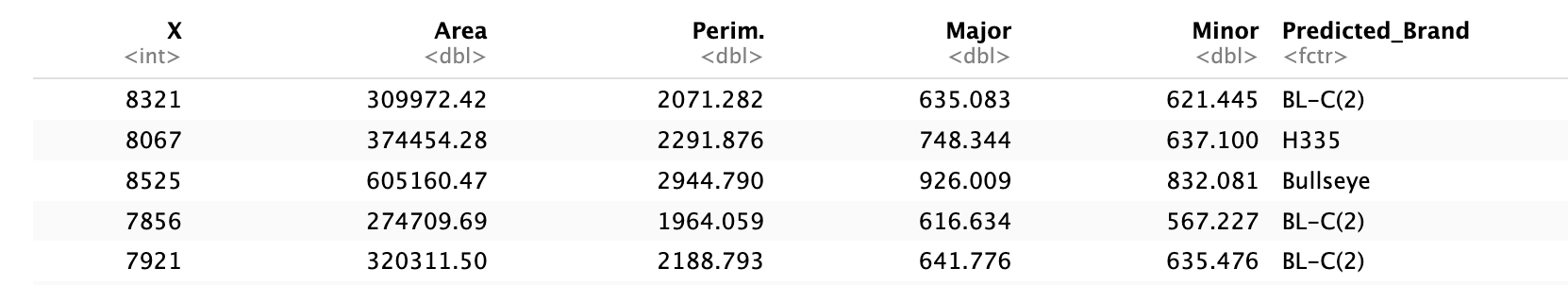




**Query Predictions Exploded Examples:**

****

**Query Predictions Unexploded Examples:**

****

**Evaluation Challenges:**

Despite successful model development, evaluating the models for brand prediction posed challenges. We encountered errors when calculating confusion matrices using caret, preventing us from computing precision, recall, and F1 scores accurately. Manual calculation of confusion matrices also had limitations.

**Conclusion and Future Directions:**

In conclusion, we developed models to predict the brand of SAP particles from recovered IEDs. However, further exploration of model evaluation techniques is needed to overcome challenges encountered during the project. Collaboration with domain experts and investigating methods to address issues with brands having minimal representation could enhance model performance and understanding.

**References**

* <https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab>
* Dr. Saunders’ Lectures
* <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://chat.openai.com/c/338382ac-7e47-4006-b7f5-1dd32dd28319>
* <https://www.r-bloggers.com/2021/04/random-forest-in-r/>