My topic is about the classification of dog outcome using three different methods: MLR, RF and XGBoost.

I collect the data of over 15,000 instances covering the period of 3 years of AAC. My model comprises 10 independent variables: size, age, group… and 1 dependent variable: the outcome with 5 different levels: Adoption, Died, Euthanasia, Return to owner and Transfer.

Because of limited time, I will go directly into some noticeable results of each model.

First model is MLR:

* It shows that the younger dogs have more chance to get adopted, however, they tend to be in death before reaching their new home.
* Obviously, the older dogs and their human friends have more sentimental long-term relationship. Therefore, they are 4 times more likely to return to their previous owner or they’re euthanized.
* Regarding to the breed, people seems to prefer keeping their purebred rather than multiple and hybrid breed. This might come from the fact that purebred is healthier.
* I’m pretty sure that all of you know Pit bull and may have some opinions about them. My study shows that they hardly get adopted and also more likely to involve in mercy killing at least 4 times than the other groups. This caused by the negative stereotype of human being towards them.
* And last but not least, the possibility of successful adoption in neutered or spayed dog is 20 times compared to the intact one. This suggest you to neuter your dogs if you want to keep one.

Top 3 significant variables in this model is: Age, Name and Intactness

Second model is RF: RF ranks Age, Size and Intactness as most important features. This points out the size is under evaluated in the former model but it plays an important role in this model.

The last model is XGboost. It yields very similar results with regression, in which Age, Name and Intactness are important.

RF and XGBoost can predict nearly 60% new intances. And XGboost seems to be more reliable and stable compared to random forest.

When I train the latter 2 models. I was trying to observe the ability of each model in every instance of the outcome. What I observe is that these two model can predict quite well the instances related to adoption, return to owner and transfer. However, it predicts poorly the instances related to the death of the animal because of inadequate of information. This result suggests that the death of animal is unexpected and caused by some other features such as the personality which beyond the scope of my study.

In conclusion, I think it will be more interesting to perform more model when doing classification since each model has its own advantages and disadvantages. When it comes to the accuracy, I recommend using the Boosting to classify. I hope that my topic is interesting, especially for the one who is dog-lover.