Predicting bird classifcation by bone lenghts

Jouke Profijt

October 8, 2018

Classification of differing ecological bird populations

Introduction

Materials & Methods

During this research we used data from an external source, used weka 3.8.0 for classification and R version 3.5.1 in combination with Rstudio for data analysis.

Data

Birds' Bones and Living Habits, Kaggle dataset

Bone measurements were measured from a skeleton collection of Natural History Museum of Los Angeles County, provided by Dr. D. Liu of beijing Museaum of Natural History

The birds are separated in 6 diffrent groups:

- Swimming Birds, SW
- Wading Birds, W
- Terrestrial Birds, T
- Raptors, R
- Scansorial Birds, P
- Singing Birds, SO

Most samples have data for:

- Length and Diameter of the Humerus
- Length and Diameter of the Ulna
- Length and Diameter of the Femur
- Length and Diameter of the Tibiotarsus
- Length and Diameter of the Taesometatarsus

weka 3.8.0

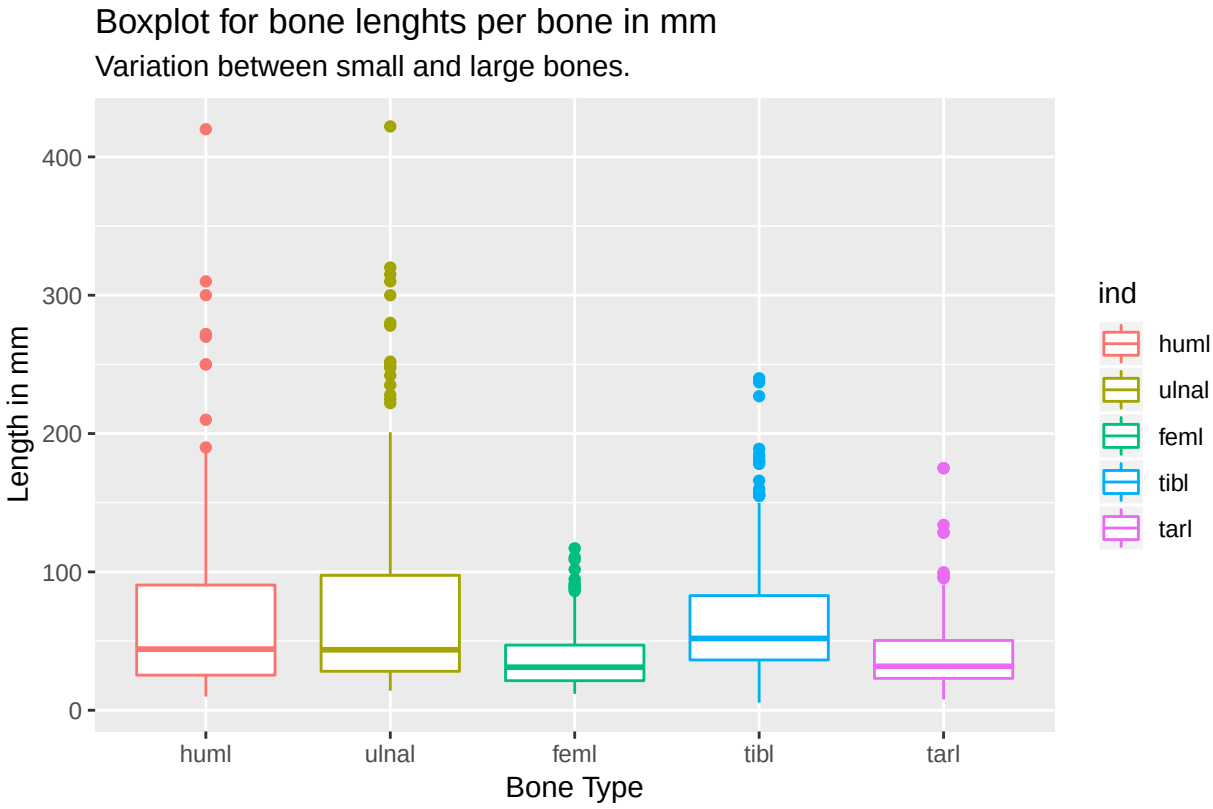
Weka (Waikato Environment for Knowledge Analysis) is a free to use datamining software written in Java. It is an java application that is capable of doing lots of things apliccable to datamining. We will be using it to determine what is the best classification algorithm for our usecase. As we will be creating a java application of our own that can be used to classify the different bird groups.

- First we use the explorer to make our datafiles usable in weka by removing duplicate id collums
- when the data is ready we can use the classify module in weka to test diffrent classification algorithm's that have reasonable accuracy's.
- at last when we found some interesting algorithm's we can use the experimenter to try diffrent settings to find optimal ones.

R and Rstudio

Resutls

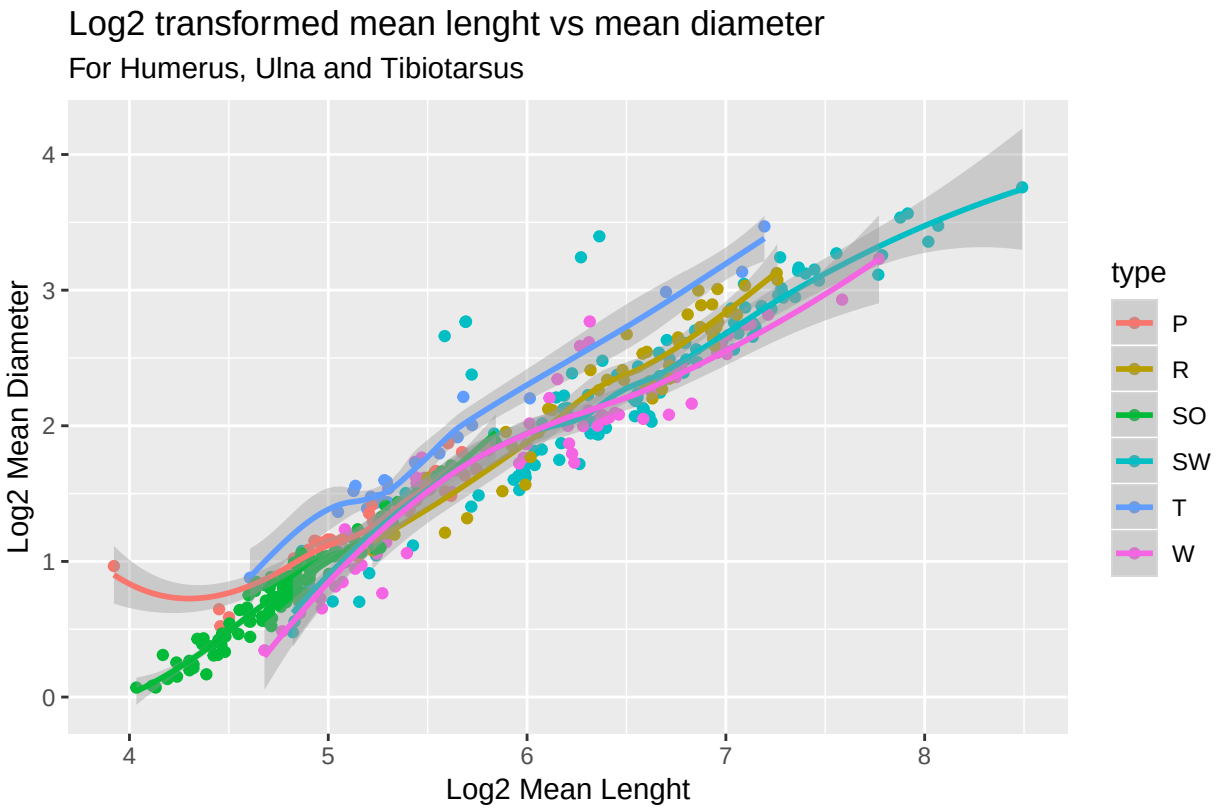
EDA



Graph 1: Bone lengths

Graph 1 shows the average lenght per bone for all ecological groups. As we can see there are 2 bone types which are quite a bit smaller than the other 3 bones. The Femuer(green) is the smallest of these 2, this stands out because in us humans this is our biggest bone. after the Femeur the next smallest bone is the Taesometatarsus, this is expected as this bone connects the feet of the birds to their legs.

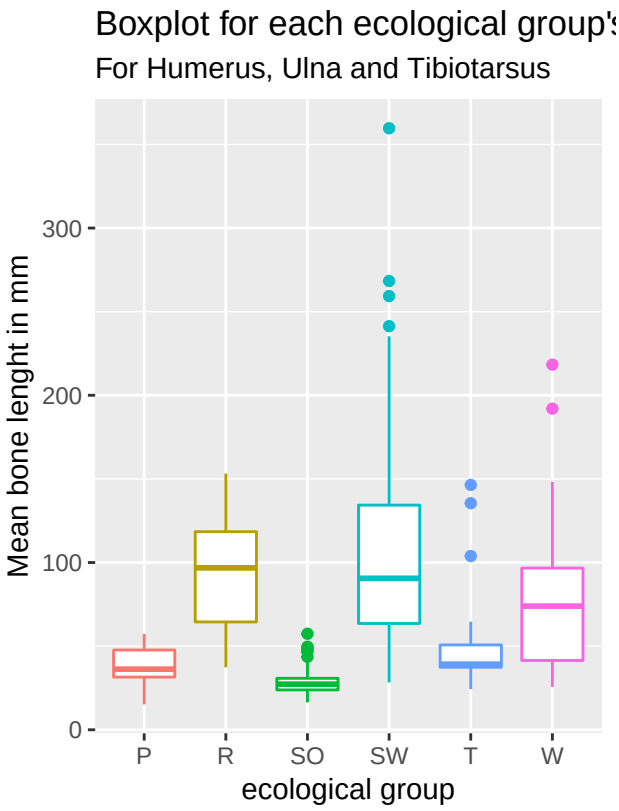
These 2 bones are quite small and show less variation than the other 3 bones. As we are trying to classify these bones they might be less important in our final classification algorithm.



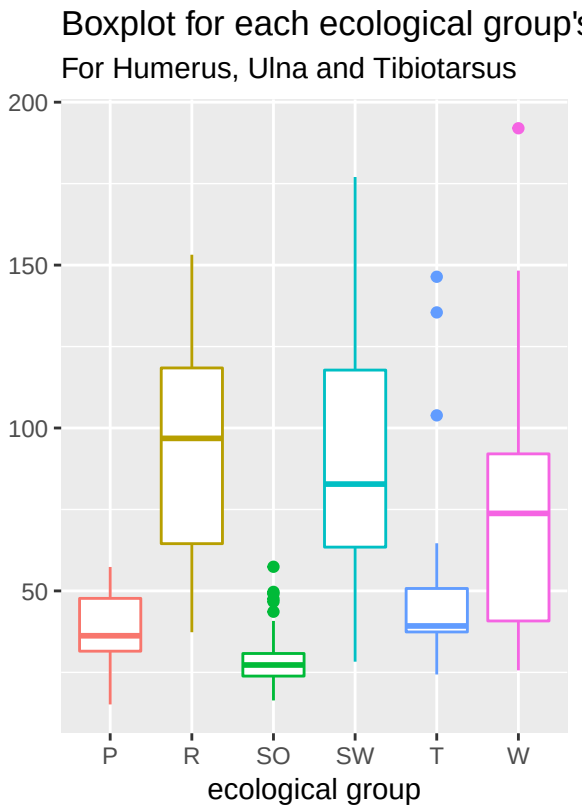
graph 2: lenght vs diameter

graph 2 shows us that the total bone lenght for the Swimming Birds is quite a bit larger than the other birds. Also we can see that the Terrestrial birds have thicker bones than the rest of the birds but are stull quite small. the singing birds are overall very small ans should be easy to classify. The raptors and Wading birds are a bit average and could become diffcult to classify but the raptors are a bit thicker and the Wading birds are a bit smaller.

Data Cleaning



graph 3: mean bone lenghts



graph 4: Cleaned data

For our cleaning of data we removed all rows with NA's which cost us 7 datapoints. After that we calculated the 1st and 3rd quartile outliers from the humurus which can be seen in graph 3 & 4, and removed 9 rows.

After all cleaning we left with 404 datapoints to use for Weka Analyses and classification.

Weka

In weka the goal was to find a classification algorithem that has the highest possible accuracy because classifying a birds herritage it does not matter that much if there is a False positive, the goal was to keep false negatives low and keep true positives high.

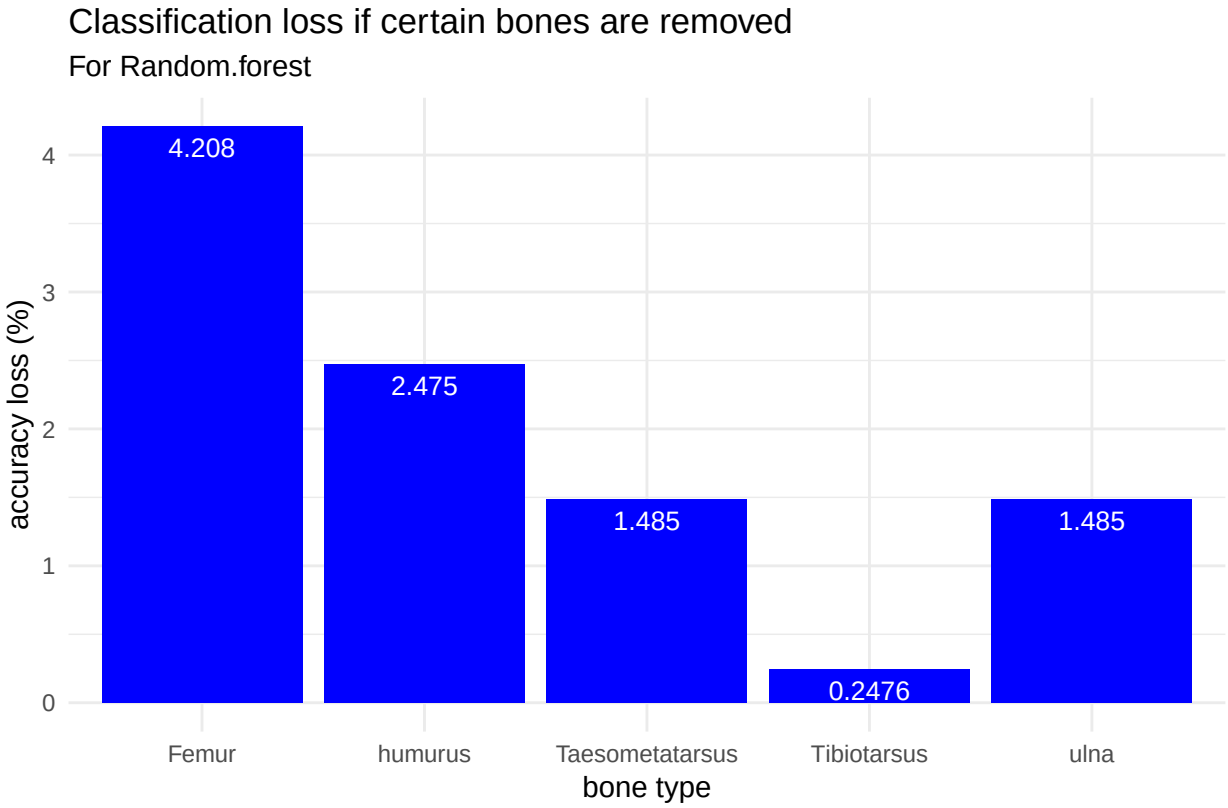
Chosen classifier: Random.Forest

Table 1: Random.Forest Confusion matrix as chosen classifier

	SW	W	T	R	P	SO
SW	87	11	0	4	0	6
W	17	34	0	2	4	6
T	2	0	12	2	4	3
R	7	1	0	37	3	0
P	0	0	1	1	30	6
SO	0	0	1	1	1	121

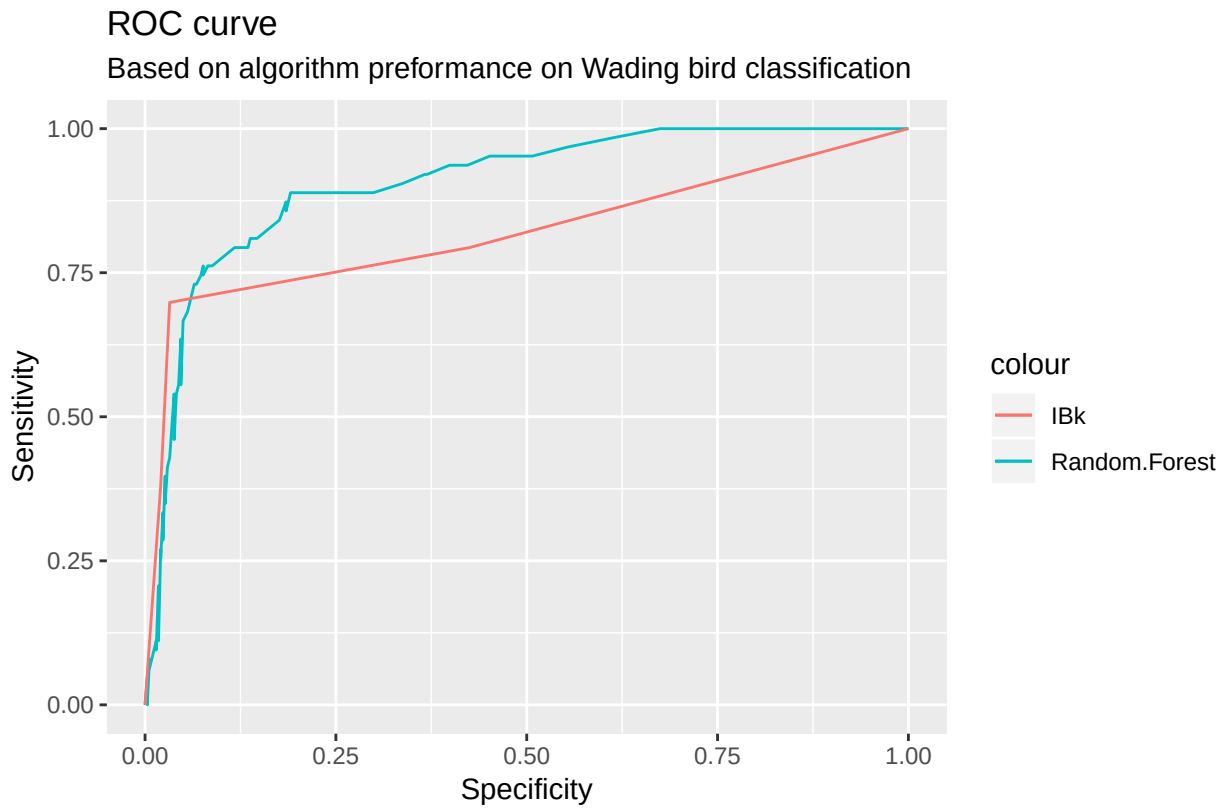
Table 2: Random.Forest as chosen classifier

	Correct	Incorrect
Instances	321	83
Percentage	79.4554 %	20.5446 %



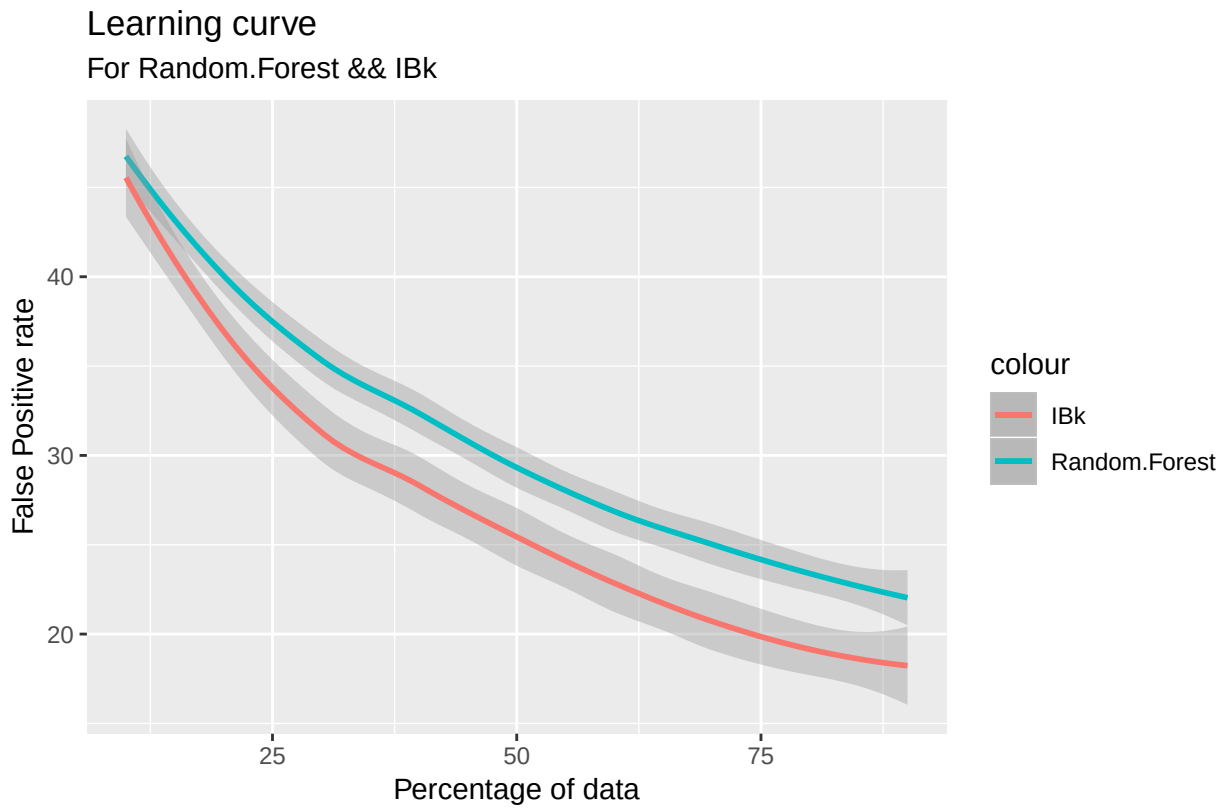
graph 5: accuracy loss per bone

in graph 5 we can see what the importance is from certain bones if we are talking about classification. the diffrences we can see are becouse of their importance for the functionallity of the bird groups.



Graph 6: ROC

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



graph 7: Learning curve inspected algorithm's

Conclusion & discussion

What is the most important bone for each ecological group their function? we can conclude that (At least for classification) the Femur is the most important as when we remove this from our classification algorithems the accuracy loss is great. we can see this in graph 5 where the loss per bone is displayed. while researching the subject we made the assumption that the Femeur and Taesometatarsus were the least important for classification. Becouse of this the first classification algorithems were done only using the longer bones but further in reverted this desicion. The classification algorithem that was chosen however wasn't changed. In the future we should first collect some data about the importance for some bones before making such rational desisions.

We do wan to think about removing some bones as when we have less data to be put in the more unknown fossils with missing bones could possibly be classified

Minor Proposal

References