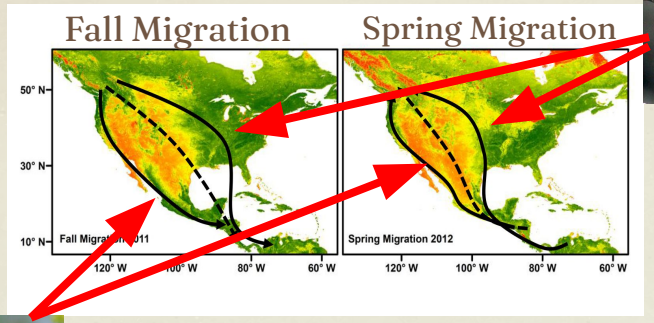
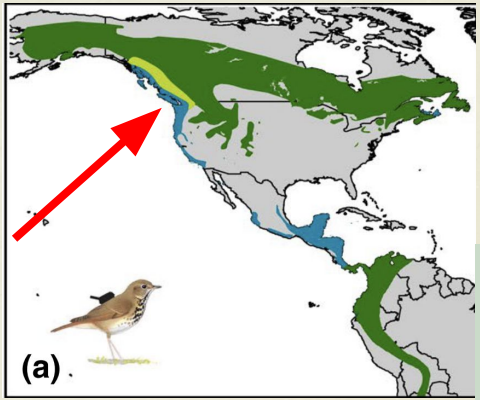
Survival of Migration

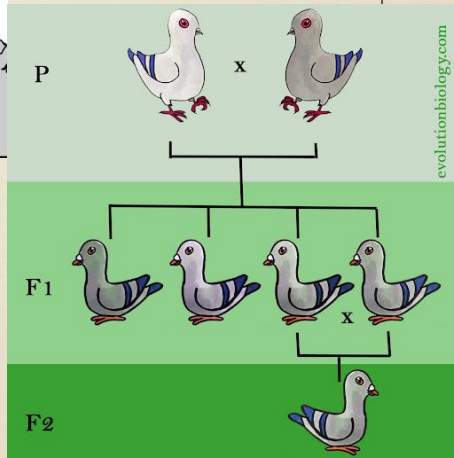
Background Information

Olive-backed Thrush



Russet-backed Thrush

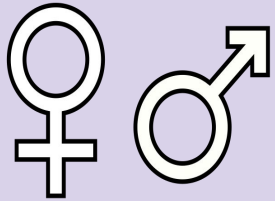




Which traits contribute to survival of migration for hybrids?

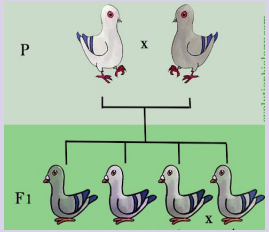
Traits

Sex Binary

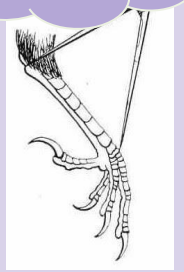


Ancestry

Heterozygosity

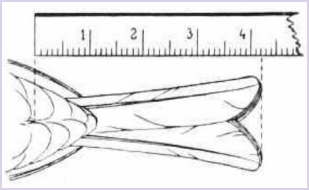


Tarsus Length

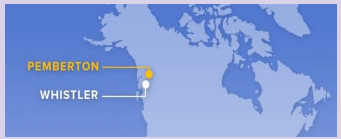


Weight

Tail length

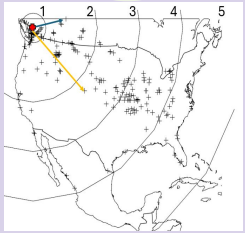


Release day

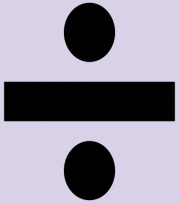


Fall detect day 1

Fall bearing 1

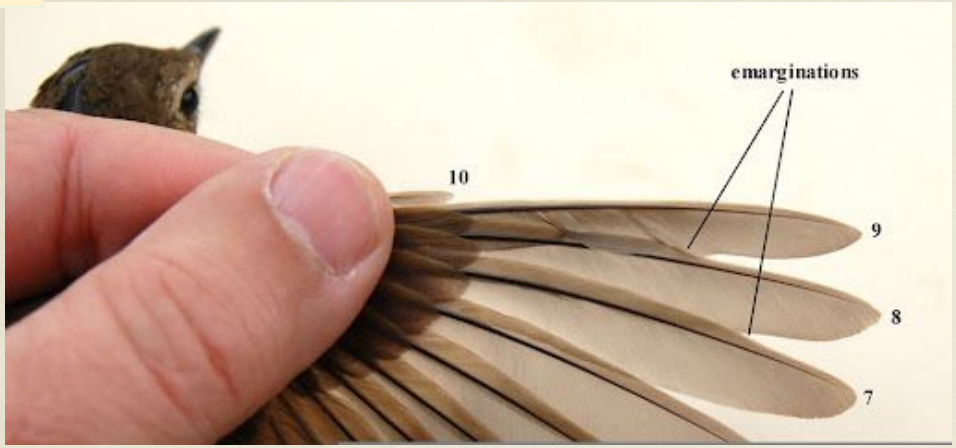


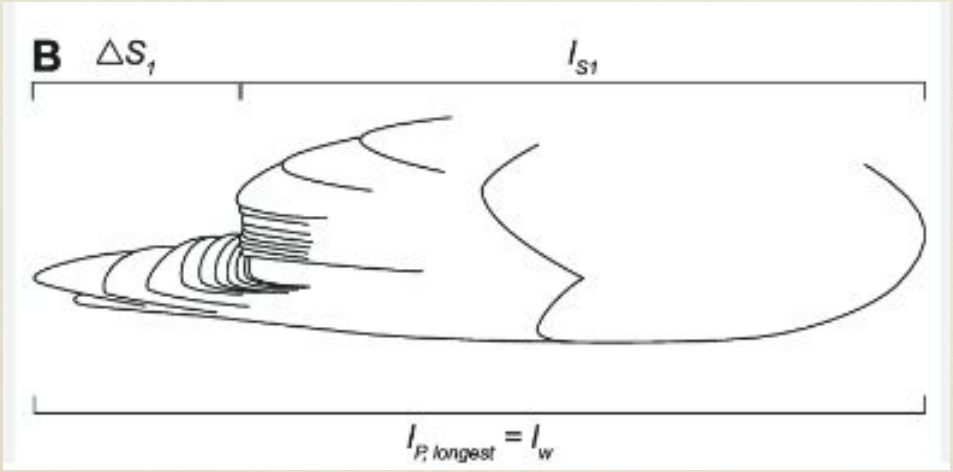
Body condition



Traits

Wing Measurements





p7

p8

distal

p9

p10

kipps

carpal

Wing cord

Goal:

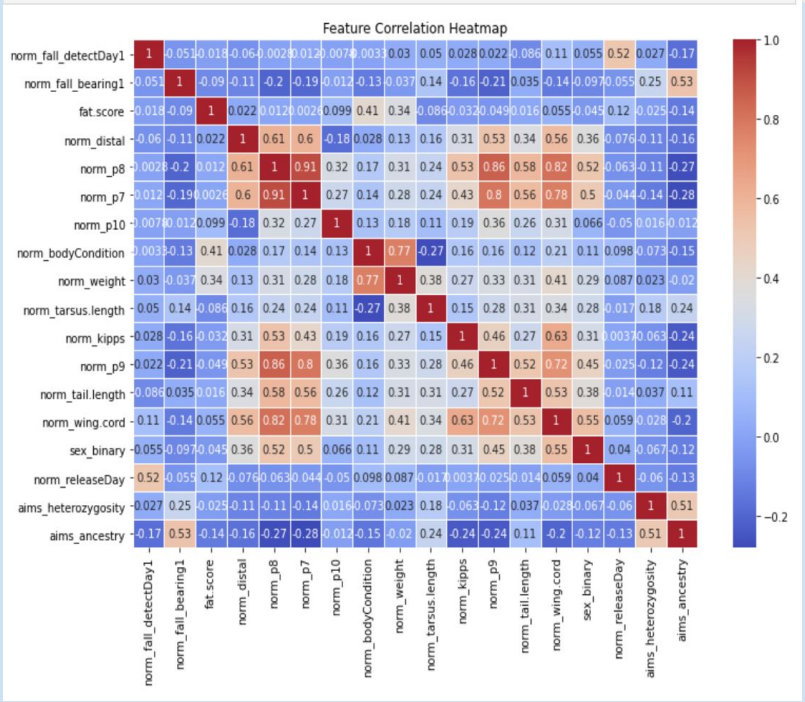
Create a machine learning model in which, when we input the traits of a specific bird, it predicts if that bird will survive the whole migration

Feature Reduction

Nulls

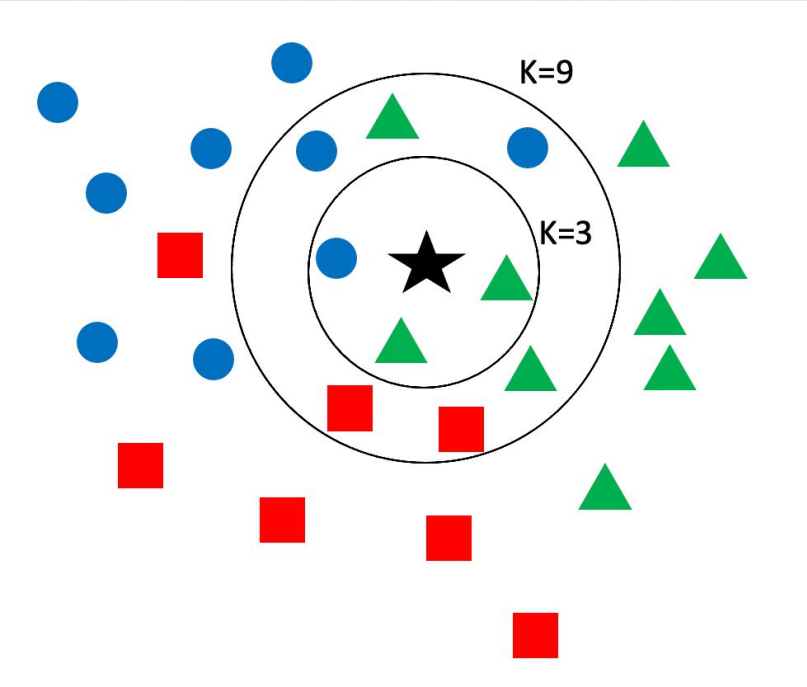
Carpal	37%	p10	13%	Tail length	1%
Fall detect day 1	28%	Body condition	3%	Wing cord	0%
Fall bearing 1	28%	Weight	2%	Sex binary	0%
Distal	13%	Tarsus length	1%	Release day	0%
p8	13%	Kipps	1%	Heterozygosity	0%
p7	13%	p9	1%	Ancestry	0%

Correlations



K-Nearest Neighbors (KNN) Imputation

By finding the “k” closest neighbors to a given missing data point and then imputing the missing value based on the average of the values of the “k” neighbors.

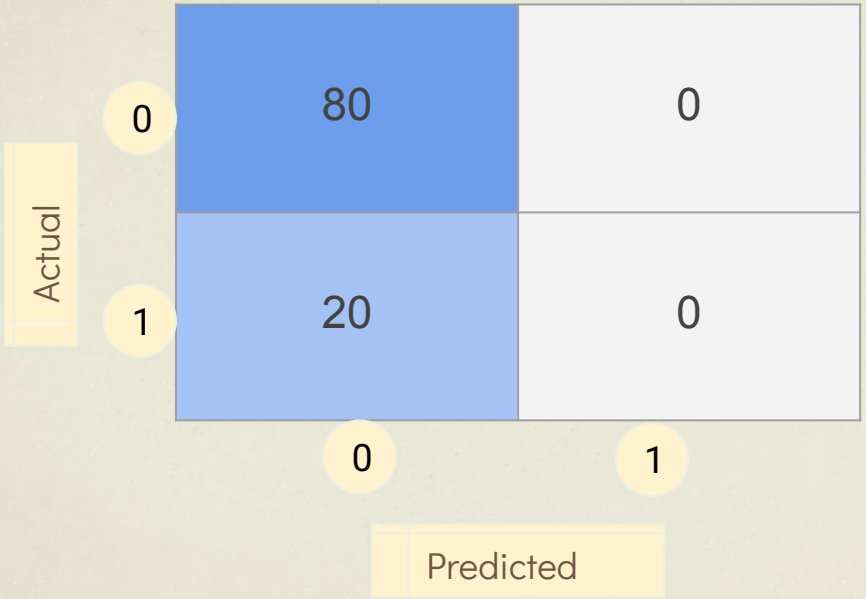


Example for k=2

Wing length	Tail length	Ancestry
NA	21.3	0.64
32.1	21.2	0.63
32.3	21.4	0.65

(32.1 + 32.3)/2 = 32.2

Imbalanced data: why it's a problem

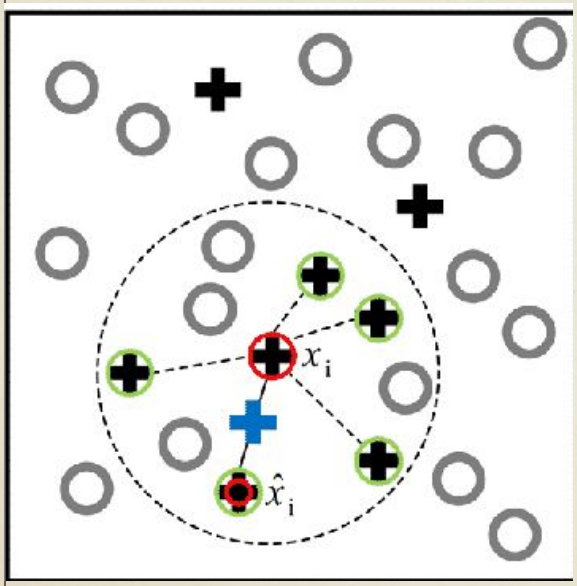


80/(80+20) = 0.8

80% accuracy

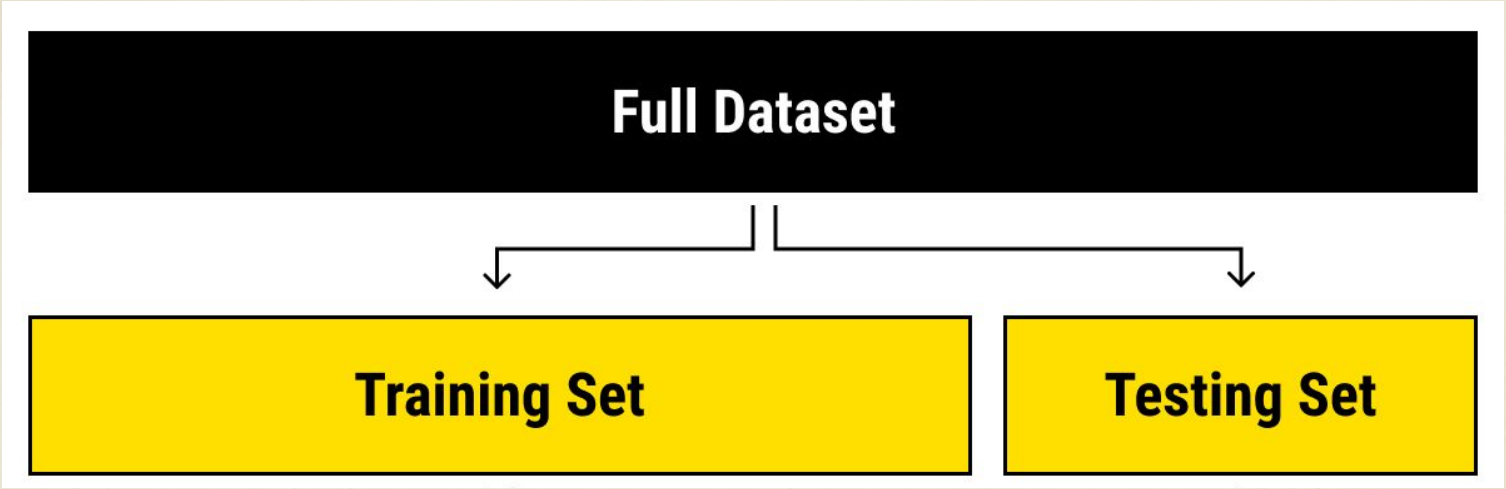
SMOTE (Synthetic Minority Over-sampling Technique)

By finding the “k” closest neighbors to the minority class and then creating a new sample based on a random number between the values of the “k” neighbors.

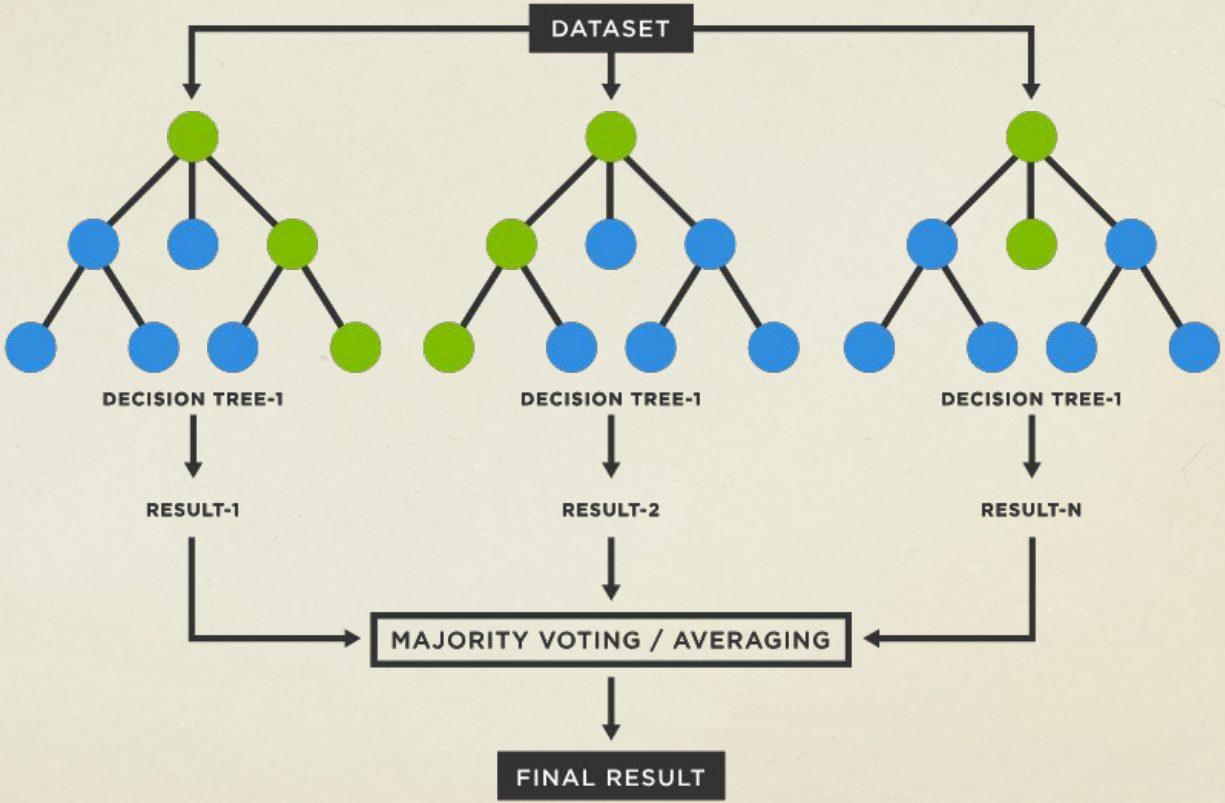


Example for k=1

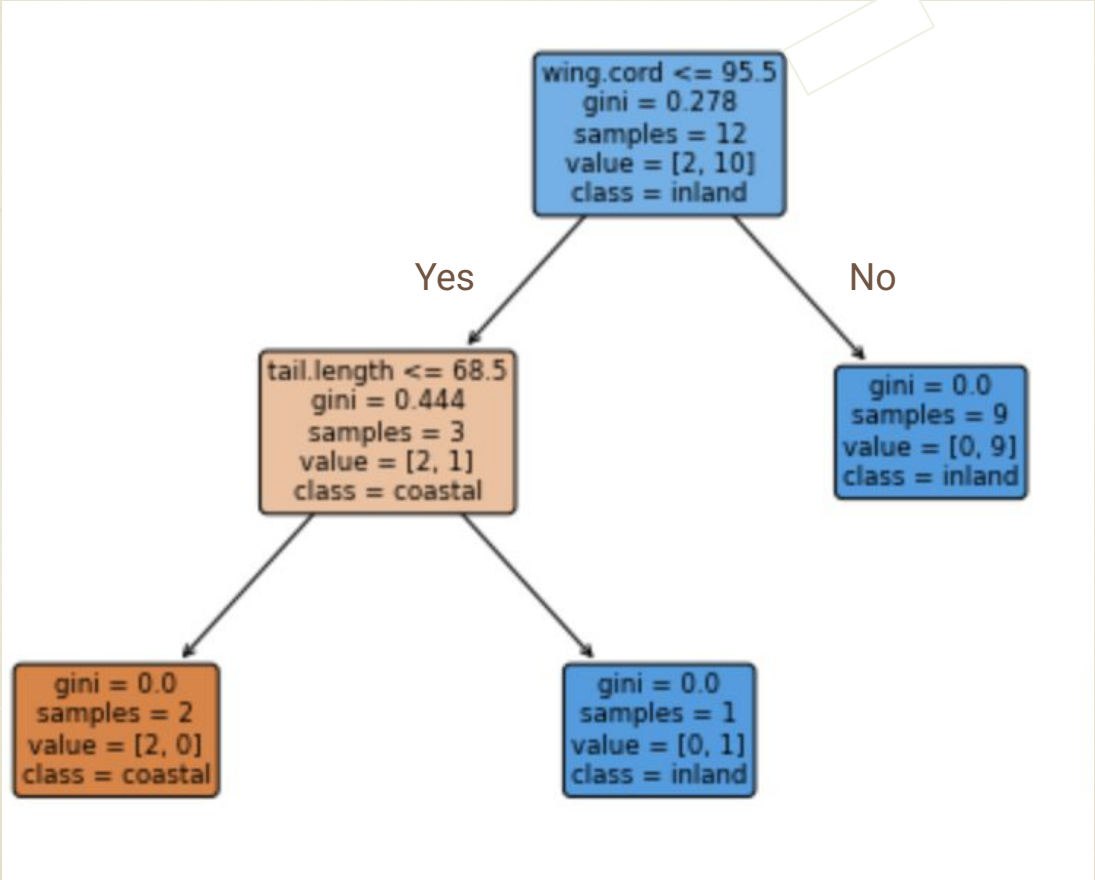
	Wing length	Tail length	Ancestry	Survival
original	32.2	21.3	0.64	1
Closest neighbor	32.1	21.2	0.63	1
New synthetic sample	**32.15**	**21.25**	**0.635**	**1**


Full Dataset
Training Set
Testing Set

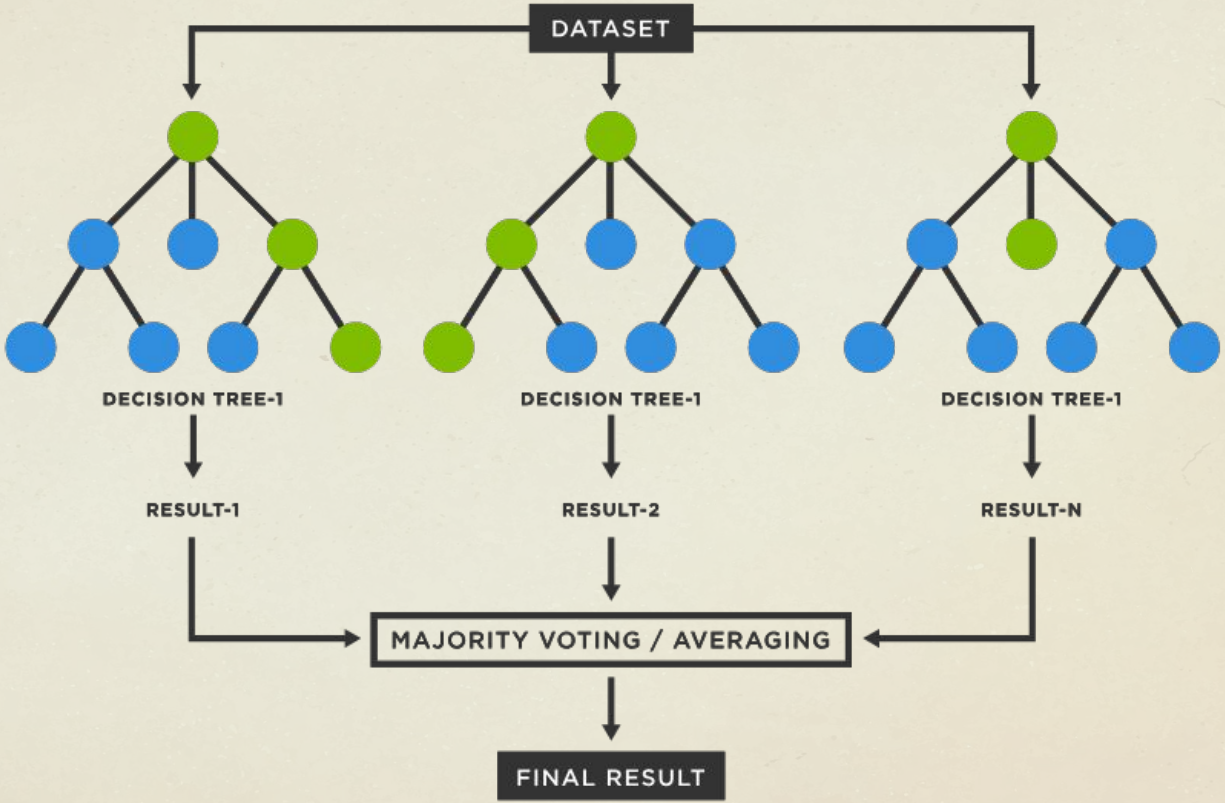
Random Forest



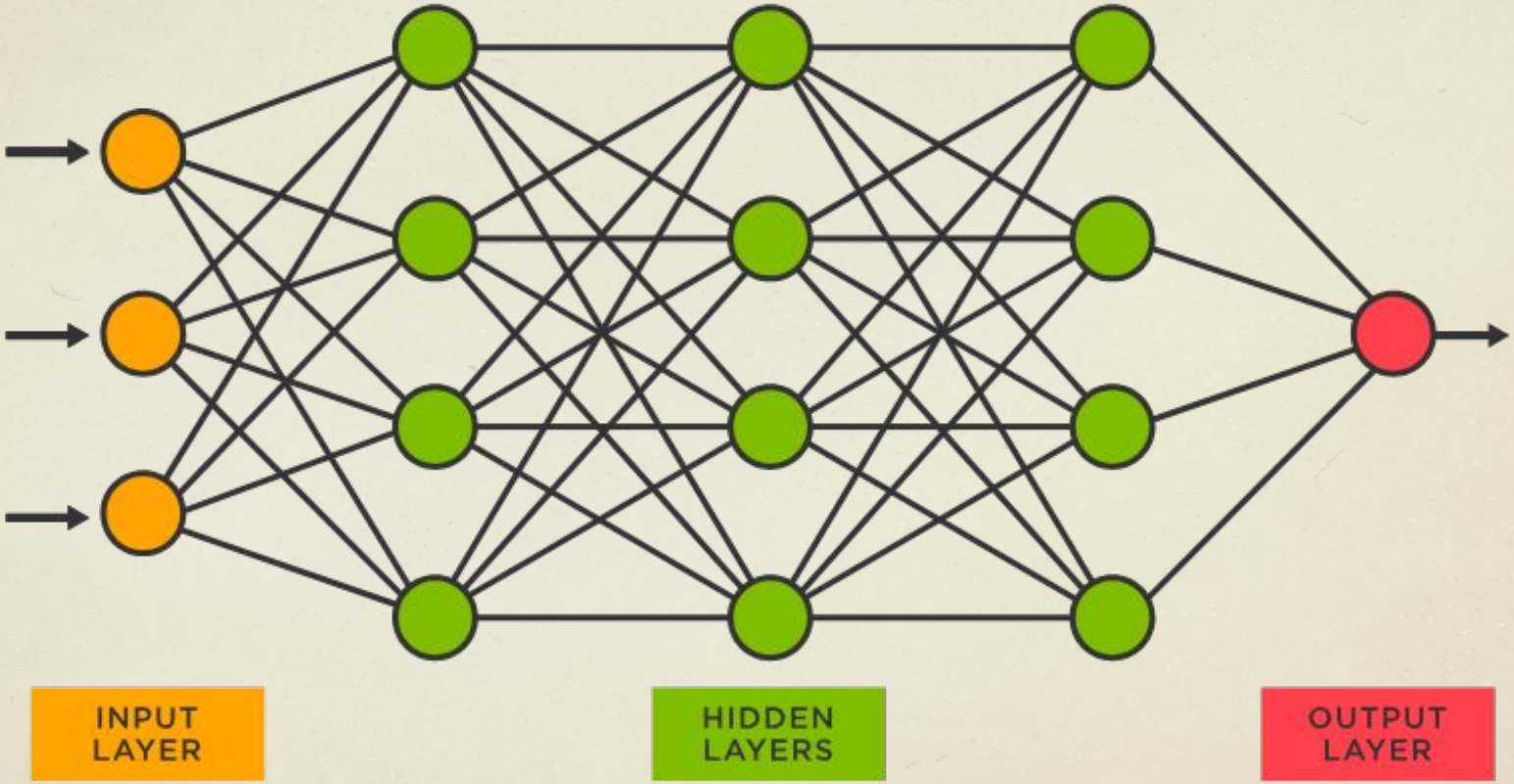
Decision Tree



Random Forest

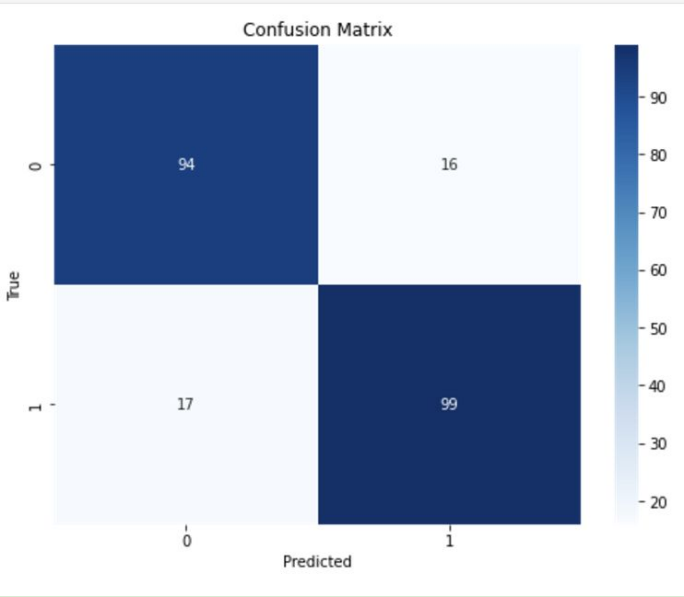


Neural Network



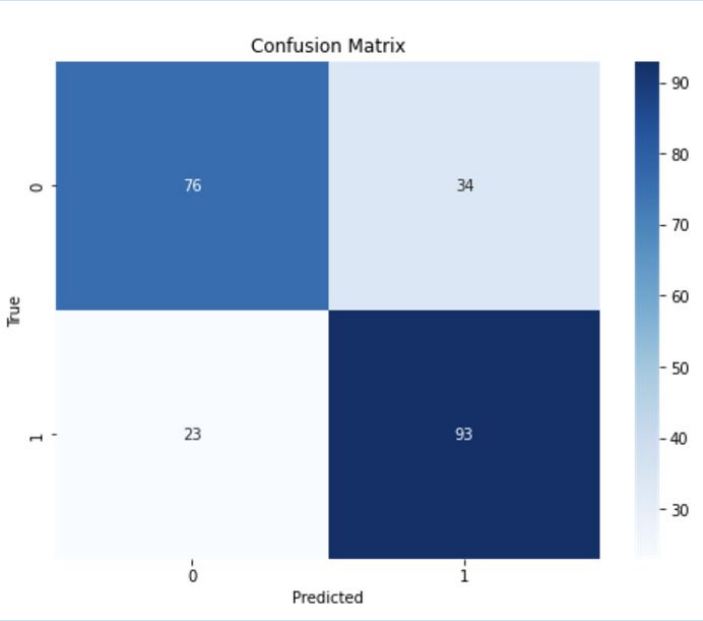
Performance

Random Forest



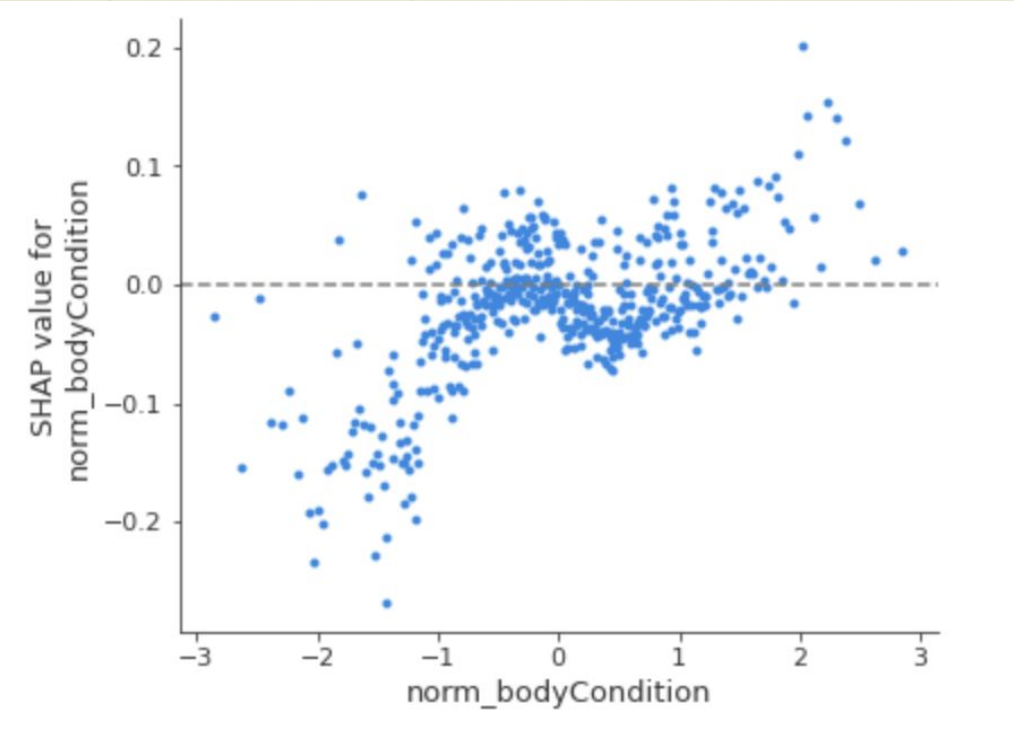
Accuracy: 0.85

Neural Network



Accuracy: 0.75

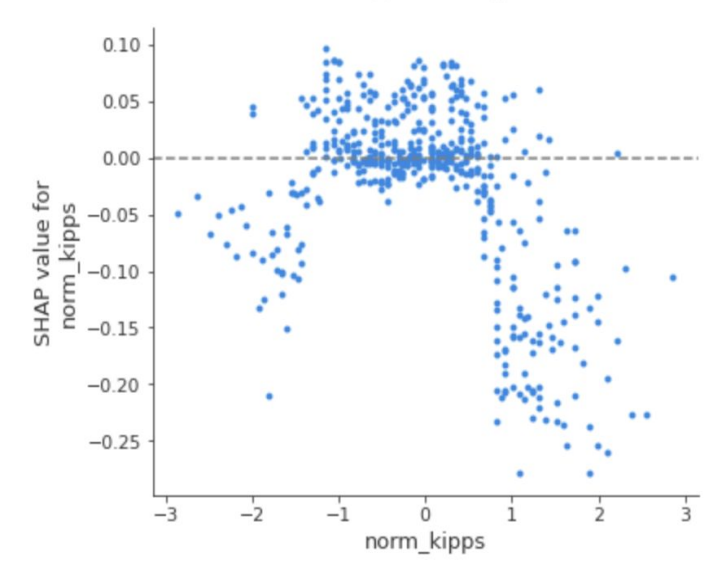
SHAP dependence plot (SHapley Additive exPlanations)

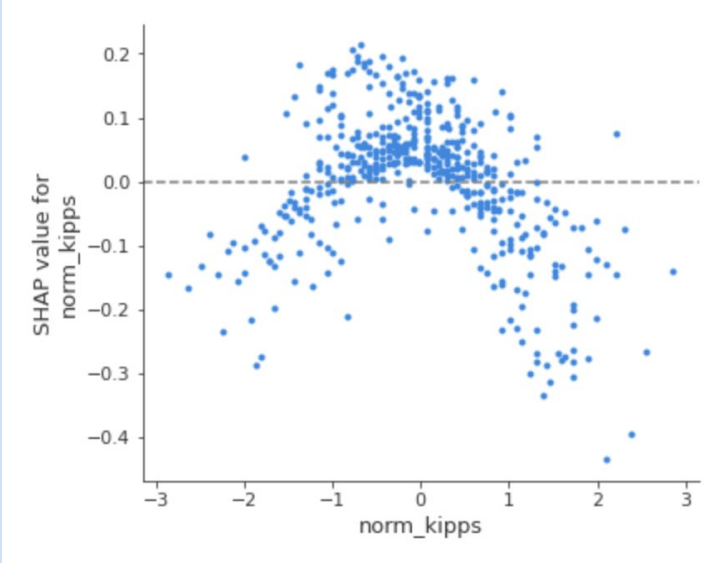


Random Forest

Kipps

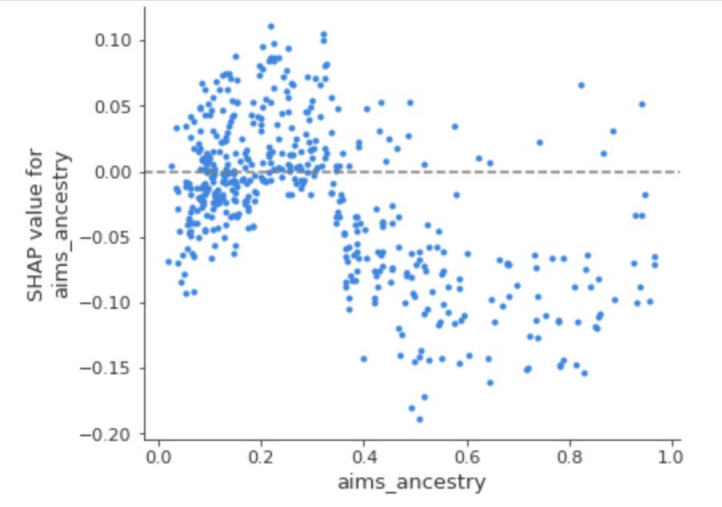
Neural Network



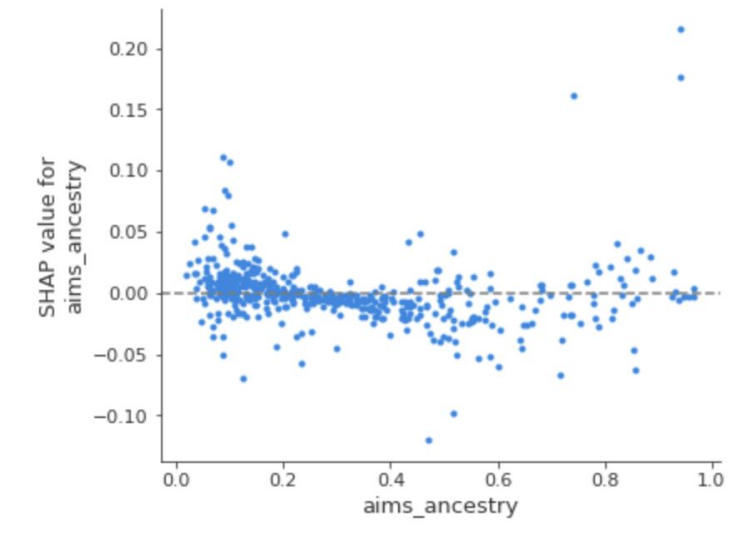


Ancestry

Random Forest



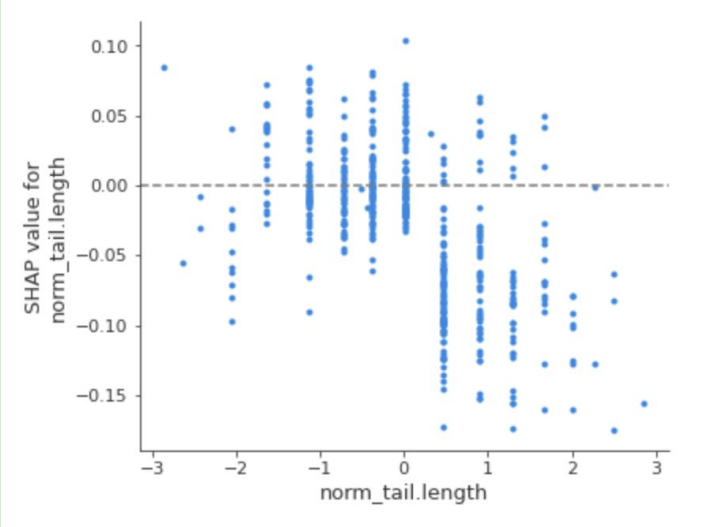
Neural Network

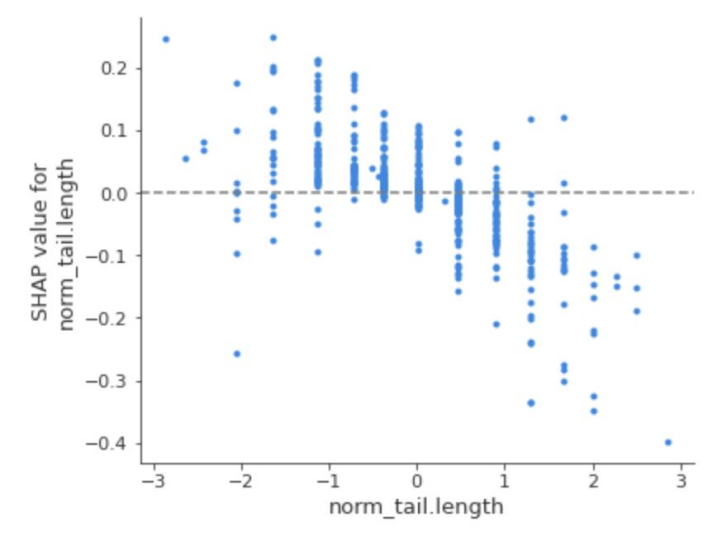


Random Forest

Tail length

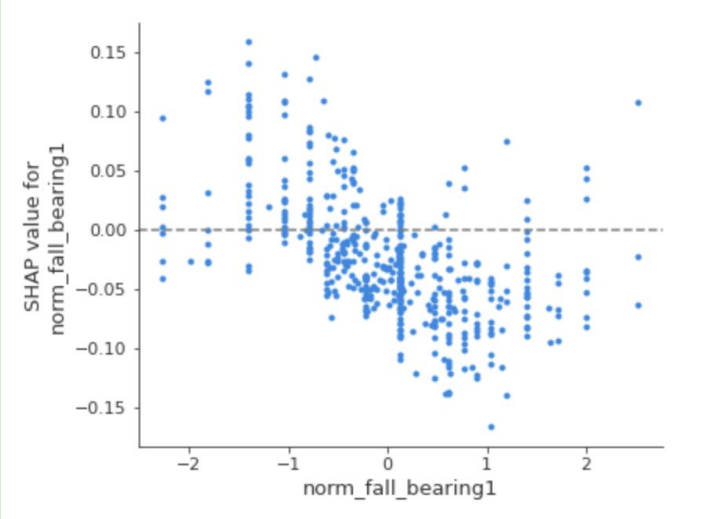
Neural Network



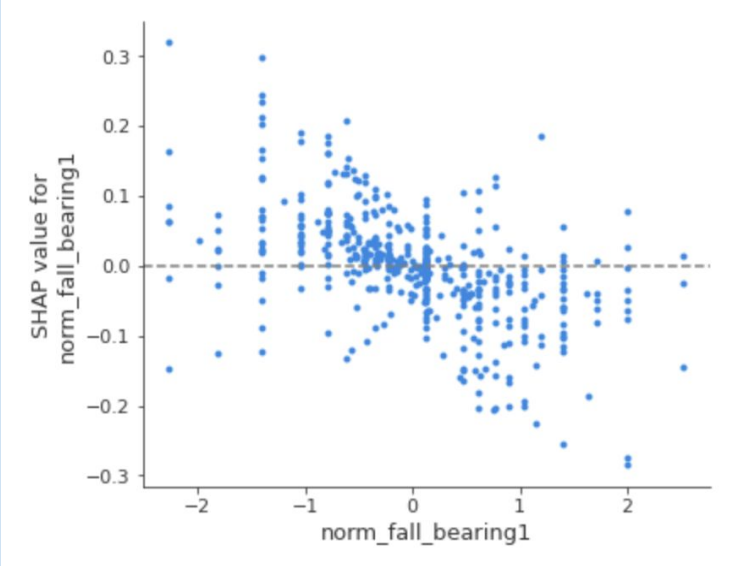


Fall bearing 1

Random Forest

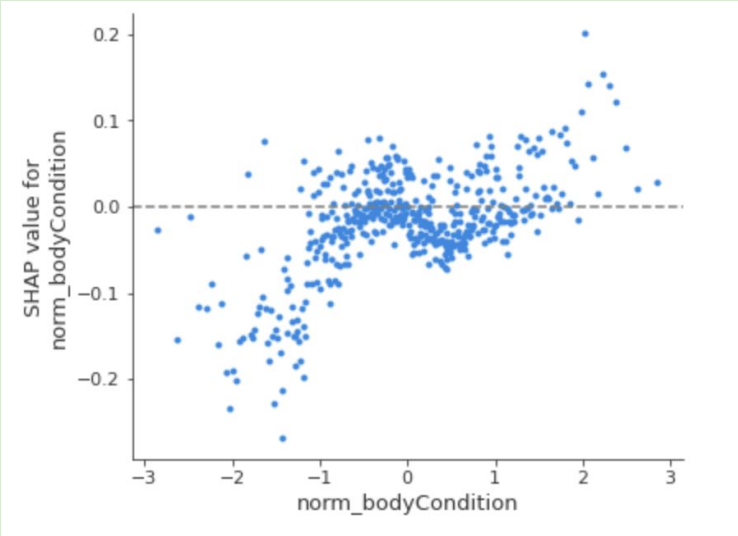


Neural Network

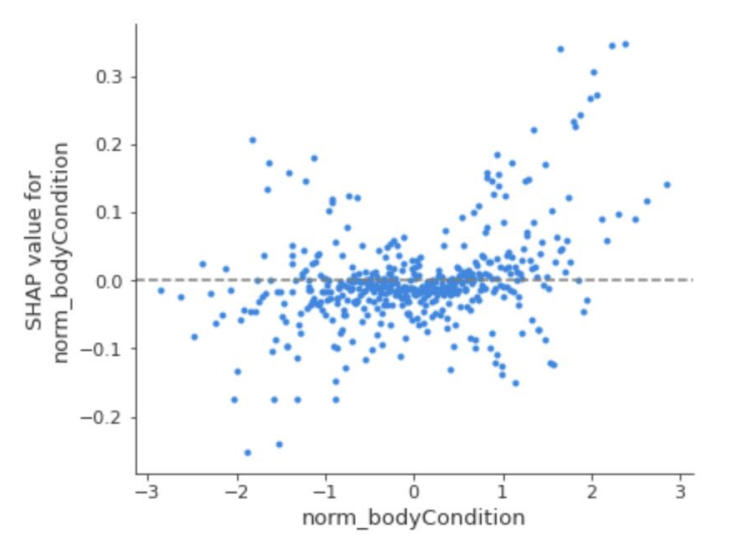


Body condition

Random Forest

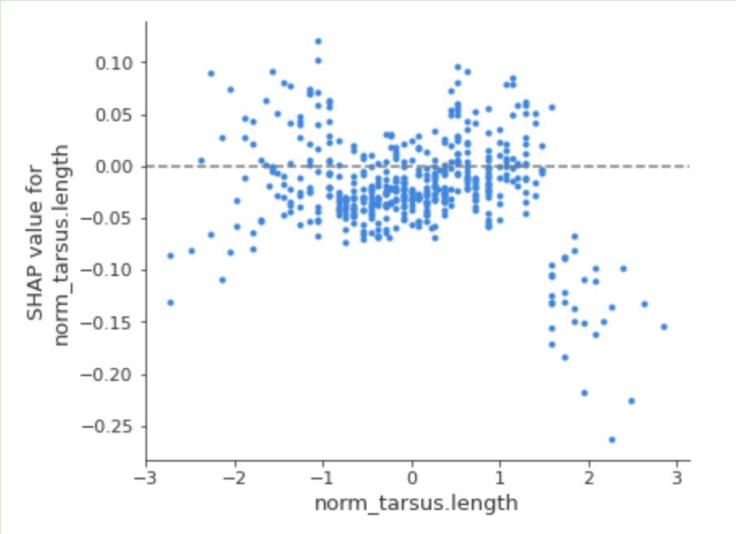


Neural Network

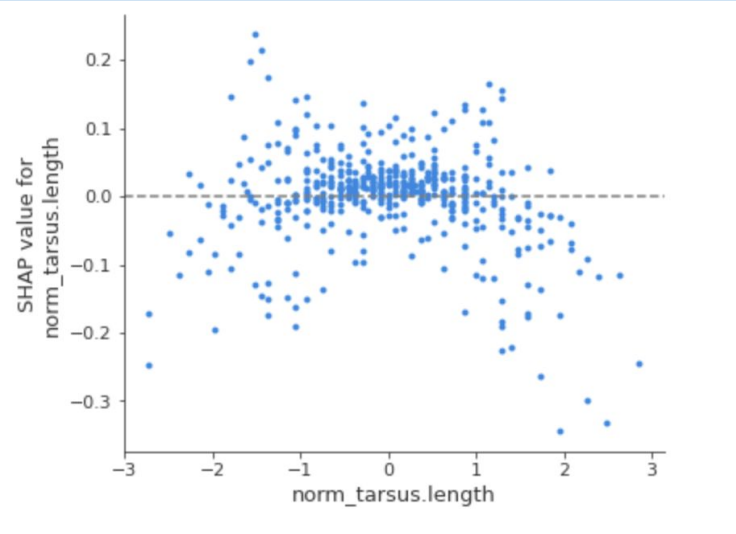


Tarsus length

Random Forest

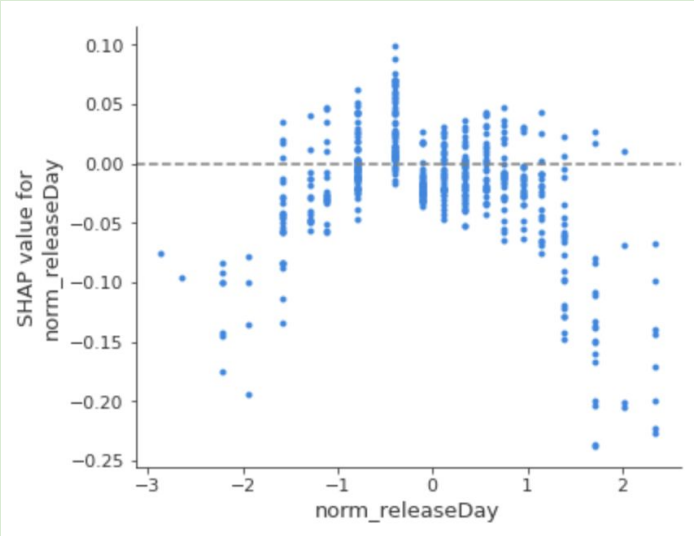


Neural Network

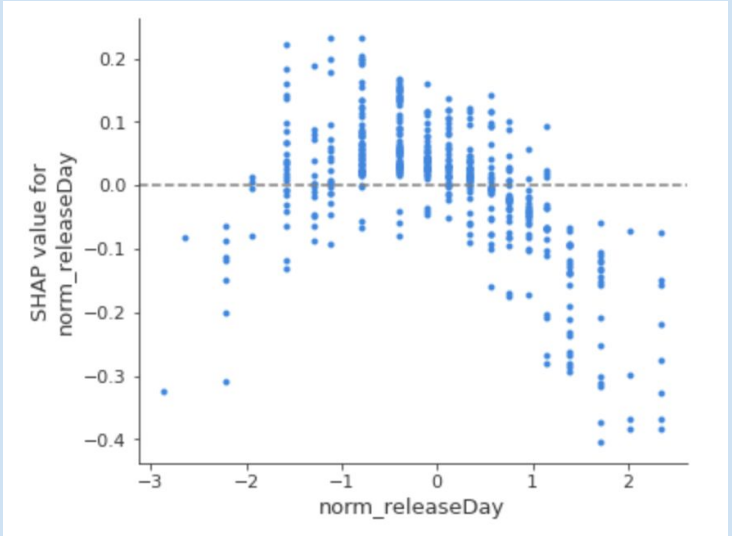


Release day

Random Forest

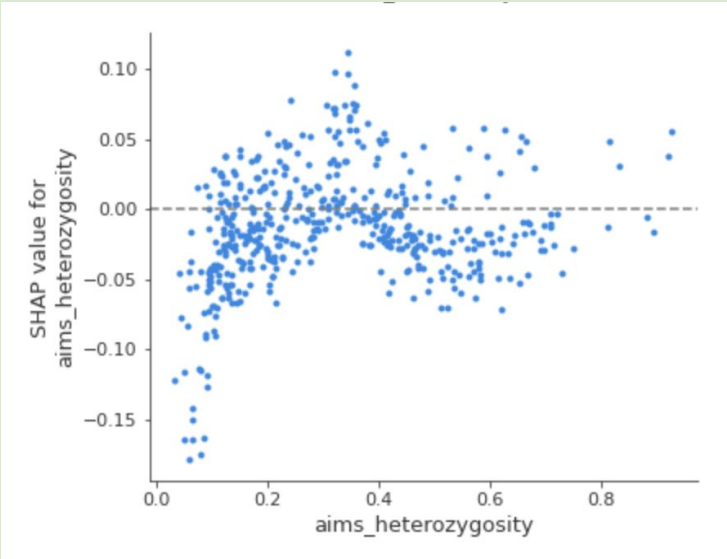


Neural Network

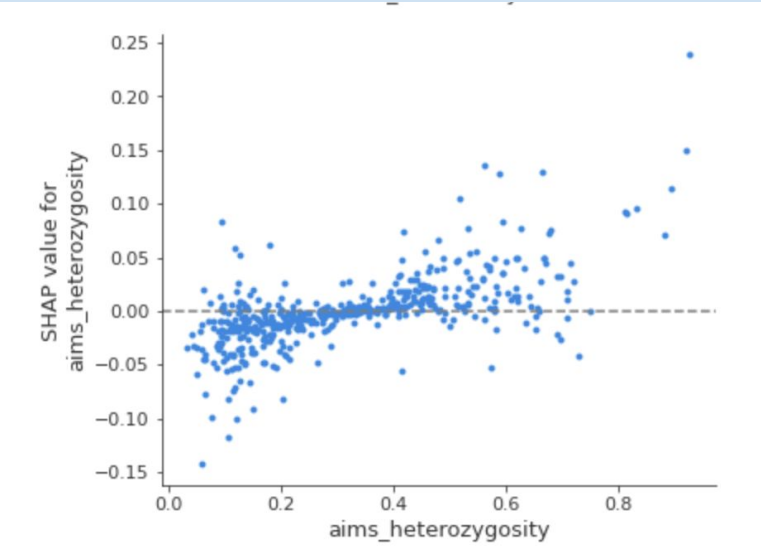


Heterozygosity

Random Forest



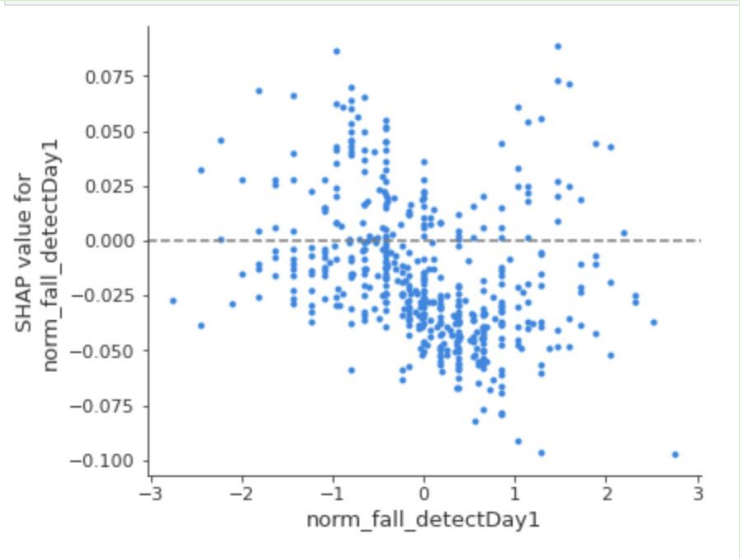
Neural Network

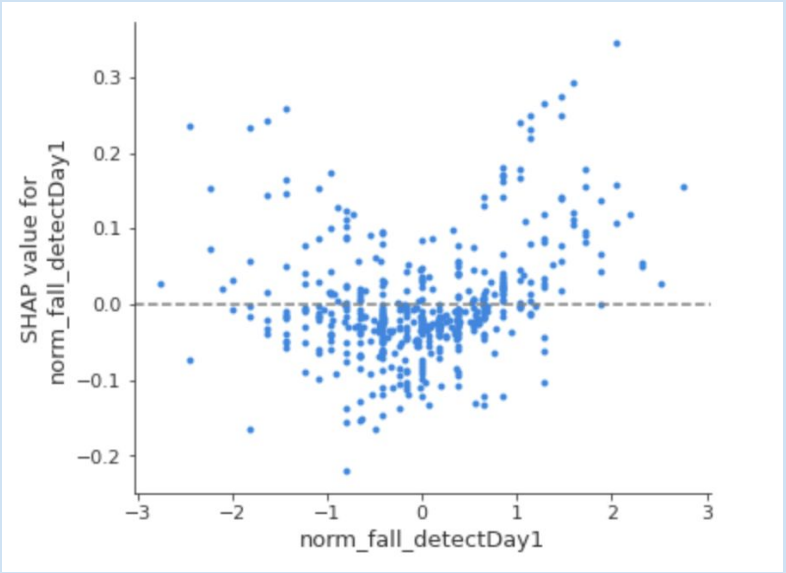


Random Forest

Detect day

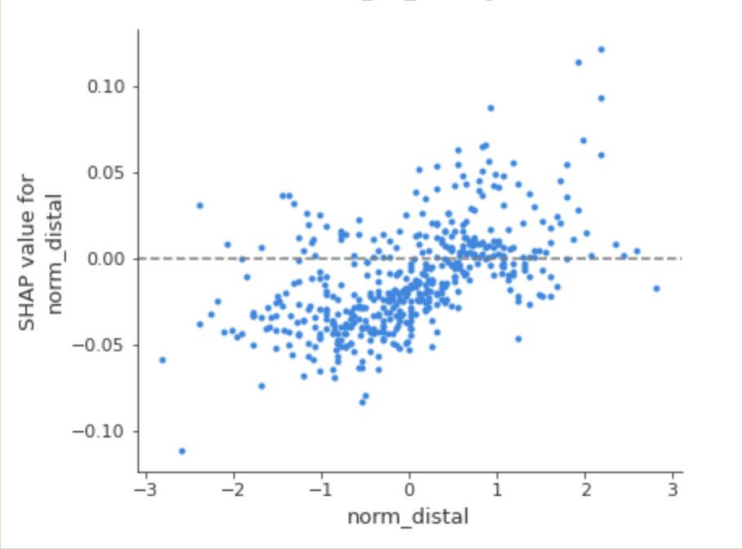
Neural Network



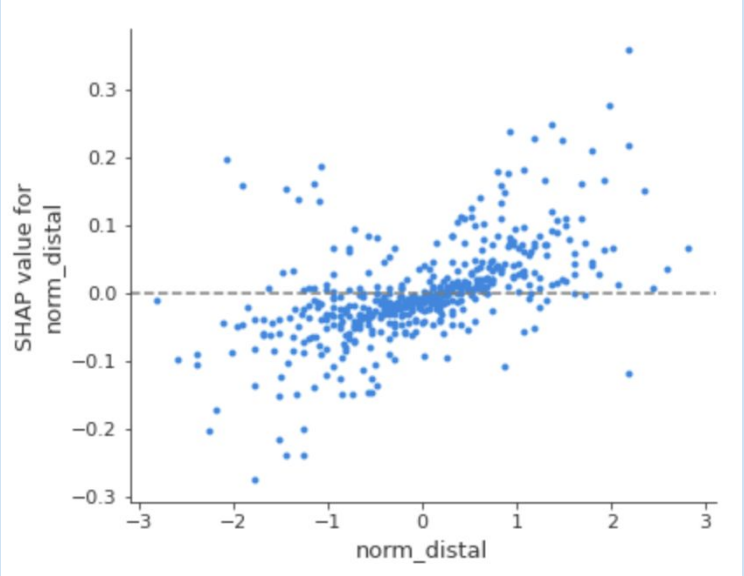


Distal

Random Forest

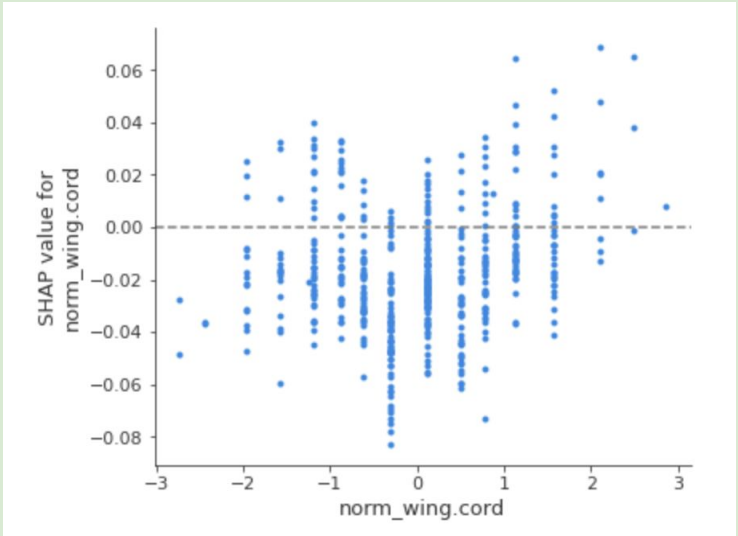


Neural Network

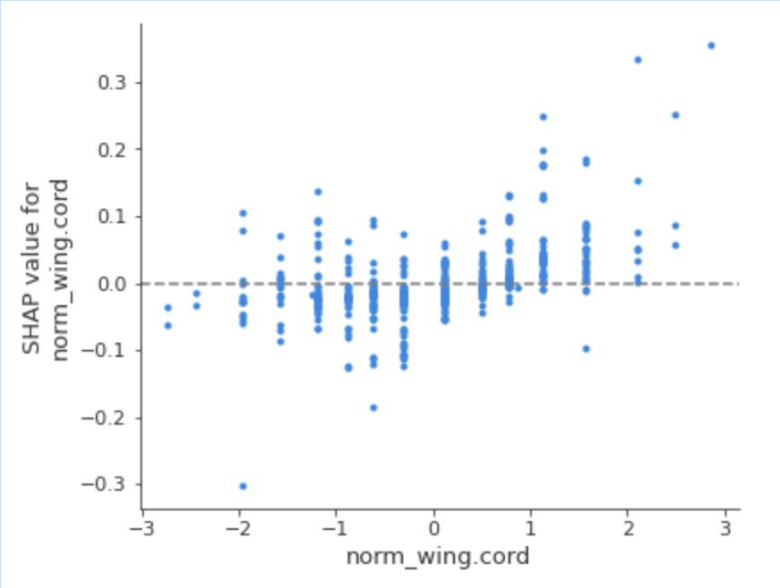


Wing cord

Random Forest

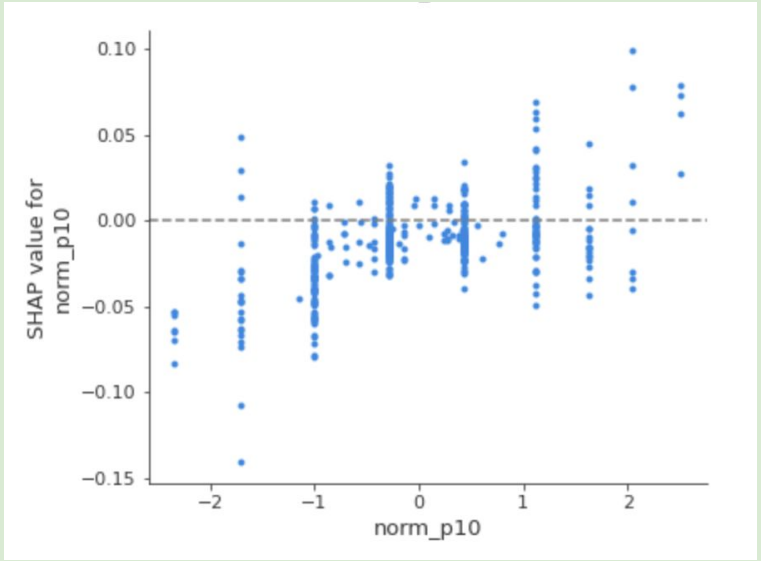


Neural Network

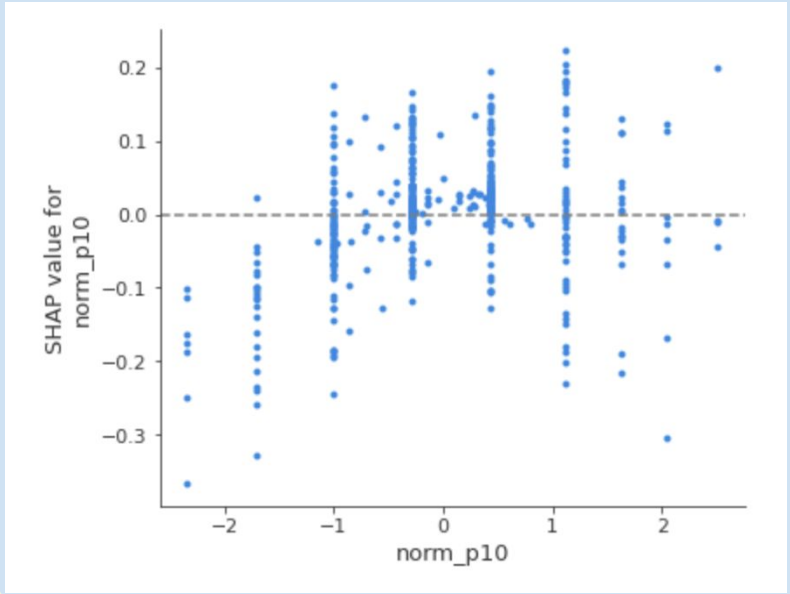


p10

Random Forest

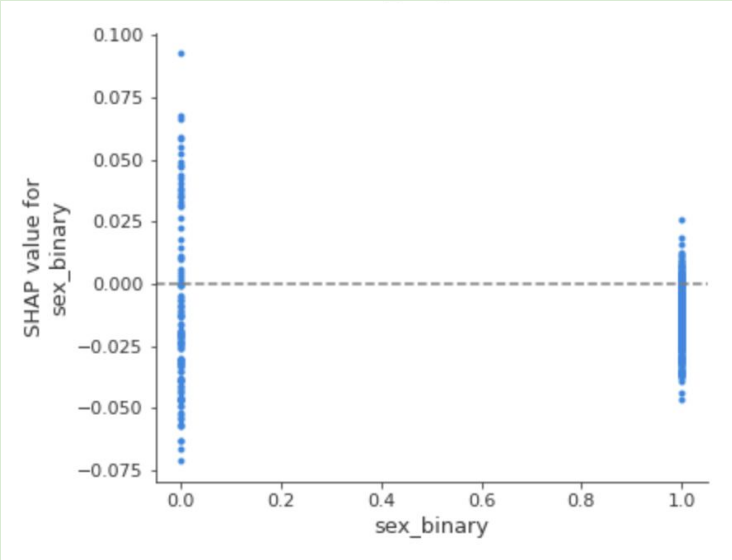


Neural Network

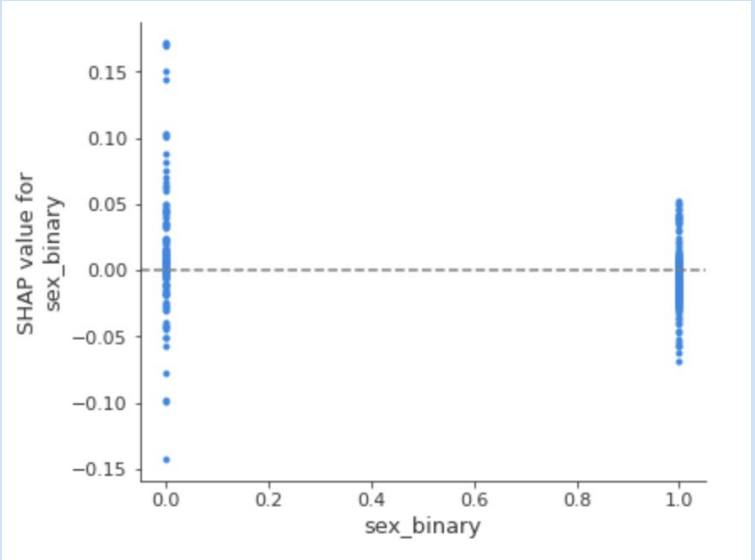


Sex binary

Random Forest



Neural Network



References

1. Ruegg, K. C., & Smith, T. B. (2002). Not as the crow flies: a historical explanation for circuitous migration in Swainson's thrush (Catharus ustulatus). Proceedings of the Royal Society of London. Series B: Biological Sciences, 269(1498), 1375-1381.
2. Delmore, K. E., & Irwin, D. E. (2014). Hybrid songbirds employ intermediate routes in a migratory divide. Ecology Letters, 17(10), 1211-1218.
3. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
4. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357.
5. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.