Ruprecht-Karls-University Heidelberg

Institute of Geography

Report of machine learning project:

Predicting flight behavior of lesser kestrel using supervised machine learning

Presented by Leon Faller

Lecturers: Anna Buch and Niko Kolaxidis

Registration number: 4186748

Deadline: 15.09.2025

Table of contents

1. Introduction

The field of machine learning has developed rapidly in recent years, and with it, the potential applications have expanded. One of these is the use of machine learning to analyze animal movement data. Many animal species are increasingly threatened by global change, including migrant birds, whose behavior can be affected by environmental changes (ZURELL et al. 2018). This project, “Predicting flight behavior of lesser kestrel using supervised machine learning” attempts to develop a reproducible approach for a movement pattern analysis as behavior prediction. In the following report, the focus is from a methodological and application perspective.

In machine learning, a distinction is made between unsupervised and supervised learning. Here, supervised learning was used to predict the classified label “flight behavior”. This label is divided into three classes: resting, flying, and migrating, and is based on the variable "ground speed". The speed threshold of the classification is determined based on the daily distances traveled during migration by various Lesser Kestrel individuals (MAURIZIO 2019). Spatial and temporal features are used primarily to try to predict the correct behavior class. This provides insights into the model's predictive power and the potential impact of various factors. By implementing a machine learning approach, the aim is to build a reproducible workflow for behavior prediction that could be adapted to other bird species in Movebank or similar datasets.

2. Data

For the following analysis, only data from the Movebank repository were used. This is a dataset with movement data from 17 lesser kestrel individuals with 380,796 location values between November 13, 2022, and July 24, 2025 (BUSTAMANTE 2025). For each bird, there are several location values per day. The migration area extends from Spain to Senegal. In addition to the standard variables, an additional variable containing the ESA Land Cover Classification was downloaded from Movebank. (Movebank Dataset).

The features hour, day, and month were extracted from the variable Timestamp to incorporate them into the model and analyze possible temporal impacts. As an additional spatial feature, the migration area was divided into three zones using the kmeans clustering method. This provides a rough distinction between breeding grounds, transit areas, and overwintering areas. This results in the following data structure for subsequent supervised learning.

Target classes = resting/flying(searching)/migrating (derived from ground-speed (km/h))

Features: categorical = ['month', 'day', 'hour', 'Landcover_Classification', 'geo_zone'] numerical = ['heading (degrees)', 'location-long (WGS84)', 'location-lat (WGS84)']

For the variables ground-speed, location-long(lat), individual rows were identified as outliers using the interquartile method and subsequently extracted from the dataset. Since this is a very large dataset and the model's runtime would suffer significantly, the dataset was reduced to 10% of the total dataset. Care was taken to ensure that the three labels were equally represented to avoid imbalance. A lot of data is lost this way, but there are still approximately 35,000 data points, and the focus can be on being able to predict all three classes well. A time-based split is then performed with 80% train data and 20% test data. This prevents data leakage in the model, since we want to predict future behavior and the model shouldn't learn from future inputs.

3. Model

For the analysis, a comparison between 3 models RandomForest, XGBoost and LogisticRegression is chosen. Logistic regression is a linear model and uses L2 regularization. It has the advantage of being fast and easy to train. Since the dataset contains fewer linear relationships, this model performs the worst in such analyses, but it is a good baseline for model comparability. RandomForest and XGBoosting are both ensemble methods with decision trees. Both are better suited for nonlinear relationships and have higher complexity than linear models. RandomForest uses bagging and is less computationally intensive than XGBoost. In comparison, XGBoost uses gradient boosting and often performs best on average in performance comparisons, even in animal behavior contexts. However, it also requires the most fine-tuning for the hyperparameters (BERGEN et al. 2023). With these three models, different strengths and weaknesses are covered for comparison, thus enabling a more comprehensive analysis.

A pipeline was used for model training. This includes a preprocessor that scales the numerical variables and encodes categorical variables to make them usable for the models. A parameter sampler with 10-fold iteration is then used to search for the hyperparameters, and the best model with the best parameters is selected based on the F1 score. Other approaches, such as randomized search with cross-validation or grid search, would produce more stable results but require longer runtimes. This was also attempted, but did not lead to significant changes in the results. Therefore, the runtime-efficient approach with a parameter sampler is preferred. The following hyperparameters were used for the different models.

```
"RandomForest": {
    'model__n_estimators': [50, 100, 150],
    'model__max_depth': [5, 10, None],
    'model__min_samples_split': [2, 5, 10]
```

"**XGBoos**t": {

'model__n_estimators': [50, 100, 150],

'model__max_depth': [3, 5, 10],

'model__learning_rate': [0.05, 0.1, 0.2],

'model__colsample_bytree': [0.33, 0.66, 1.0]

"**LogisticRegression**": {

'model__C': [0.1, 1.0, 10, 20],

'model__max_iter': [500, 750, 1000],

n_estimators and max_depth are important parameters for ensemble learners to gain an overview of potential overfitting. With RandomForest, higher n_estimators result in less overfitting, and XGBoost results in higher overfitting, although this also depends heavily on the learning rate. This is because XGBoost is trained sequentially. In logistic regression, the degree of regularization is crucial. Smaller values mean stronger regularization and thus less overfitting.

4. Results

The training for all 3 models only took about 39 seconds, as the focus was also on keeping the run time low. A confusion matrix with a report was printed for all models, which can be viewed in more detail in the notebook. Here, we focus on the best model XGBoost, where the estimators are low and learning rate is high, leading to fast learning. The maximum depth and colsample bytree are also low, leading to better generalization and less overfitting. The overall F1-Scores were as follows:

RandomForest: 0.570

XGBoost: 0.578*{n_estimators: 50, max_depth: 3, learning_rate: 0.2, colsample_bytree: 0.33}*

LogisticRegression: 0.477

The ensemble learners are significantly better than logistic regression, but overall, no model can perform a high-performance classification based on the selected features. Nevertheless, XGBoost at least achieved the best F1-Score of 0.578. Class 0 resting performed best, achieving a value of 0.85. However, the poorer recall value of 0.49 makes up for a good F1-Score. class 0 resting, was predicted best with an F1 score of 0.62, slightly better than class 2 migrating, with 0.61. On the other hand, the class 1 flying performed worst with 0.50 (see figure 1).

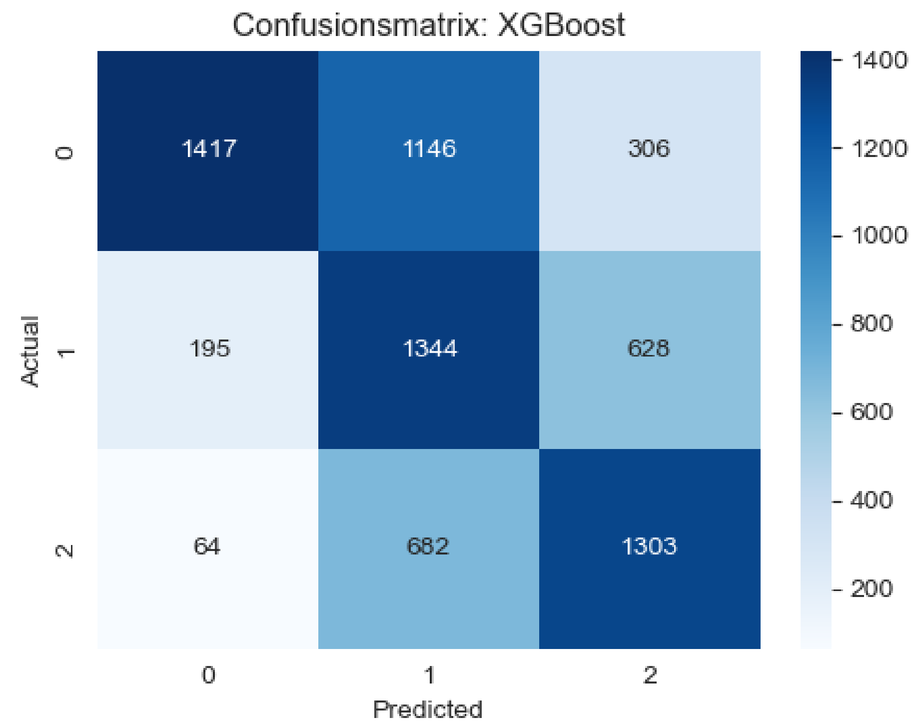

Figure 1: Confusionmatrix of best model XGBoost

The permutation importance provides a broader overview of the feature importances and shows the relative importance of the features on the model. In this case, it is based on the change in the F1-Score when the feature is randomly permuted and thus destroys the value of the feature. Heading has the greatest influence on the model, and without it, the XGBoost would perform about 0.07 units worse, reaching only 0.571. Location-Lat is in second place with about 0.04, and hour is in third place with about 0.02. Location-Long, month, geozone, and day follow with smaller values. Land cover classification shows no impact at all (see figure 2).

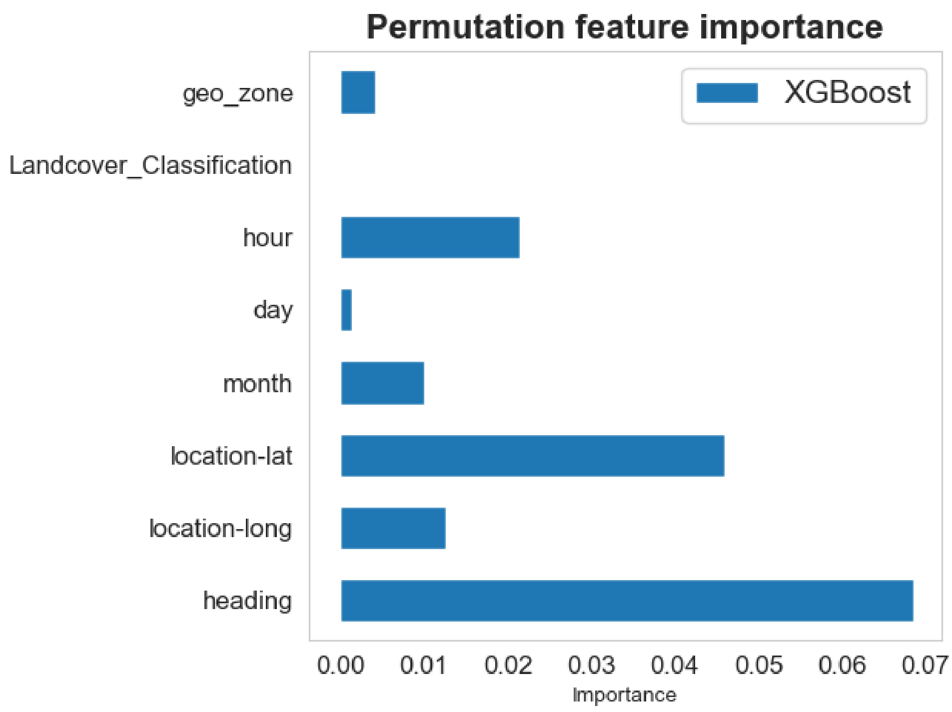

Figure 2: Permutation feature importance based on F1-Score

5. Discussion/conclusion

The model comparison for the prediction of flight behavior provides an overview of the influence of temporal/spatial and environmental features on the target class. For example, it can be seen that the resting class can be predicted best. This may be because the nighttime values are dominated by the resting class, allowing the model to predict this better. However, the total values are insufficient to extract scientifically meaningful results, making this model unsuitable for other scientific applications. The basic structure of the model is efficient, but the selected features and the target class would need to be optimized to achieve better results. Other studies show that the predictions can be significantly better with large and optimized datasets, but then the problems of the curse of dimensionality, overfitting, and runtime become more pronounced. In one study, the Random Forest runtime alone took 52 minutes, but achieved an accuracy of 86% (BERGEN et al. 2023). However, datasets are naturally not perfectly comparable, and it is unclear how good a prediction of the flight behavior of the Lesser Kestrels based on feature data can actually be.

In this model, the focus was on generalization and runtime efficiency, which is why the hyperparameters for XGBoost were retained. Assuming that other datasets are similarly structured, this model could also be tested for them. A lot of preprocessing, such as label balancing, was performed before the actual model training, so the model comparison is more suitable for balanced dataset. In this case, the model could also be suitable for other datasets to obtain a quick, automated overview of the input on flight behavior. Movebank also offers moveapps for providing your own code. However, this would require compatibility with R and some code would need to be repurposed in a more general way.

6. Short review of the project

The project was very useful for deepening the previous content of the course. Working on your own project helps you understand the difficulties and importance of the preparatory work involved in model training. For me, finding a good dataset with meaningful attributes and features was challenging in the area of animal movement analysis. Since the entire topic of machine learning can be very complex, I tried to orient myself on the workflow of previous exercises. Overall, I definitely learned a lot, but i would only be confident implementing simple machine learning applications in industry.

Bibliography

Bergen, Silas; Huso, Manuela M.; Duerr, Adam E.; Braham, Melissa A.; Schmuecker, Sara; Miller, Tricia A.; Katzner, Todd E. (2023): A review of supervised learning methods for classifying animal behavioural states from environmental features. In: *Methods Ecol Evol* 14 (1), S. 189–202. DOI: 10.1111/2041-210X.14019.

Bustamante, Javier (2025): (EBD) Lesser Kestrel (Falco naumanni) Senegal, MERCURIO-SUMHAL. Movebank Data Repository. Online available at https://www.movebank.org/cms/webapp?gwt_fragment=page=studies,path=study2398637362.

Maurizio, Sara (2019): Broad-front migration leads to strong migratory connectivity in the lesser kestrel (Falco naumanni). In: *Journal of Biogeography* 46 (12).

Zurell, Damaris; Graham, Catherine H.; Gallien, Laure; Thuiller, Wilfried; Zimmermann, Niklaus E. (2018): Long-distance migratory birds threatened by multiple independent risks from global change. In: *Nature climate change* 8 (11), S. 992–996. DOI: 10.1038/s41558-018-0312-9.