



جامعة محمد الخامس بالرباط  
Université Mohammed V de Rabat

**École Nationale Supérieure d'Informatique et d'Analyse des Systèmes**  
**Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur**

## **AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT**

**Monsieur Omar ELALAOUI**

Soutiendra publiquement sa thèse de Doctorat en Informatique

**Le Samedi 28 Juin à 10h00 au Grand Amphi à l'ENSIAS de Rabat**

**Intitulé de la thèse**

### **Ensemble Learning for Species Distribution Modeling: Conservation Applications for Moroccan Bird Species**

**Président :**

Pr. Bouchaib BOUNABAT, PES, ENSIAS, Université Mohammed V de Rabat

**Directeur de thèse :**

Pr. Ali IDRI, PES, ENSIAS, Université Mohammed V de Rabat

**Rapporteurs :**

Pr. Atman JBARI, PES, ENSAM, Université Mohammed V de Rabat

Pr. Hassan SILKAN, PES, Faculté de Sciences, Université Chouaib Doukkali, Eljadida

Pr. Abdelali ZAKRANI, MCA, ENSAM, Université Hassan II, Casablanca

**Examineurs :**

Pr. Mahmoud EL HAMLAOUI, MCH, ENSIAS, Université Mohammed V de Rabat

Pr. Taoufik RACHAD, MCH, ENSIAS, Université Mohammed V de Rabat



**Résumé:** Les modèles de qualité d'habitat, généralement appelés modèles de distribution d'espèces (SDM), sont largement appliqués en écologie à de nombreuses fins, notamment la conservation des espèces, la découverte d'habitats, et l'acquisition de connaissances évolutives en estimant la distribution des espèces. Ces modèles utilisent des données environnementales et des enregistrements d'occurrence d'espèces pour prédire la distribution des espèces et évaluer la qualité de leur habitat. Les algorithmes d'apprentissage automatique ont gagné en popularité dans les applications SDM grâce à leur capacité à capturer des relations complexes et non linéaires entre les espèces et leur environnement. Ces techniques surpassent constamment les approches statistiques traditionnelles en modélisant efficacement les facteurs écologiques complexes qui influencent la distribution des espèces. Cependant, malgré leurs avantages, ils semblent avoir certaines limitations dues aux différentes hypothèses et mécanismes d'apprentissage propres à chaque modèle. Par conséquent, pour remédier à cette limitation, les chercheurs ont étudié l'approche d'apprentissage d'ensemble qui consiste à combiner des apprenants individuels précis et diversifiés afin de consolider leurs avantages et de surmonter leurs faiblesses en utilisant des méthodes de combinaison telles que le vote pondéré. Cette thèse aborde ces limitations en étudiant et en évaluant diverses stratégies d'apprentissage d'ensemble pour modéliser la qualité d'habitat des espèces d'oiseaux au Maroc. La recherche explore trois techniques d'ensemble principales : les ensembles homogènes utilisant le bagging, les ensembles hétérogènes avec vote pondéré, et les ensembles basés sur la généralisation empilée. Pour les ensembles hétérogènes, deux stratégies de sélection ont été mises en œuvre : la sélection basée sur la performance, qui privilégie les apprenants de base hautement performants, et la sélection basée sur la diversité, qui privilégie une grande diversité parmi les apprenants de base. Huit algorithmes d'apprentissage automatique (SVM, KNN, MLP, GB, DT, RF, AB et QDA) ont été évalués en tant que modèles individuels et comme apprenants de base pour construire des ensembles sur sept jeux de données d'espèces d'oiseaux de deux genres : Oenanthe (traquet) et Phoenicurus (rougequeue). Pour l'évaluation de la performance des modèles, cette recherche utilise plusieurs métriques de classification, notamment la précision, le rappel (sensibilité), la précision, le score F1, l'AUC, la spécificité, le Kappa et le TSS. La méthode de classement Borda Count a été utilisée pour comparer les modèles selon plusieurs critères de performance, tandis que le test statistique Scott-Knott a été appliqué pour déterminer les différences significatives entre les modèles. Cette thèse évalue également l'impact potentiel du changement climatique sur la distribution des espèces en projetant les modèles les plus performants dans des scénarios climatiques futurs. Cette recherche contribue aux domaines de la modélisation écologique et de l'apprentissage automatique en fournissant des aperçus complets sur l'efficacité de différentes stratégies d'ensemble pour

la modélisation de la distribution des espèces. Le cadre méthodologique développé dans cette thèse offre des conseils pratiques aux chercheurs et aux conservationnistes cherchant à mettre en œuvre des techniques de modélisation avancées pour l'évaluation de la qualité des habitats et la planification de la conservation des espèces face au changement environnemental.

**Mots clés:** Modèles de Distribution d'Espèces, Qualité d'Habitat, Apprentissage d'Ensemble, Apprentissage Automatique, Ensembles Homogènes, Ensembles Hétérogènes, Généralisation Empilée, Conservation des Oiseaux, Changement Climatique, Modélisation Écologique, Sélection Basée sur la Performance, Sélection Basée sur la Diversité, Méthode Borda Count, Test Scott-Knott.

**Abstract:** Habitat suitability models, usually referred to as species distribution models (SDMs), are widely applied in ecology for many purposes, including species conservation, habitat discovery, and gain evolutionary insights by estimating the distribution of species. These models use environmental data and species occurrence records to predict the distribution of species and assess their habitat suitability. Machine learning algorithms have gained significant popularity in SDM applications due to their ability to capture complex, non-linear relationships between species and their environments. These techniques consistently outperform traditional statistical approaches by effectively modeling the intricate ecological factors that influence species distributions. However, despite their advantages, they seemed to have some limitations due to the different assumptions and learning mechanisms for each model. Therefore, to address this limitation, researchers investigated the ensemble learning approach, which consists of combining single learners that are accurate and diverse to consolidate their advantages and overcome their weaknesses using a combination method such as weighted voting. This thesis addresses these limitations by investigating and evaluating diverse ensemble learning strategies for modeling the habitat suitability of bird species in Morocco. The research explores three primary ensemble techniques: homogeneous ensembles using bagging, heterogeneous ensembles with weighted voting, and stacked generalization-based ensembles. For heterogeneous ensembles, two selection strategies were implemented: performance-based selection, which prioritizes high-performing base learners, and diversity-based selection, which prioritizes high diversity among base learners. Eight machine learning algorithms (SVM, KNN, MLP, GB, DT, RF, AB, and QDA) were evaluated as single models and as base learners for constructing ensembles across seven bird species datasets from two genera: *Oenanthe* (wheatear) and *Phoenicurus* (redstart). For model performance assessment, this research uses multiple classification metrics, including accuracy, recall (sensitivity), precision, F1-score, AUC, specificity, Kappa, and TSS.



جامعة محمد الخامس بالرباط  
Université Mohammed V de Rabat

The Borda Count ranking method was used to compare models across multiple performance criteria, while the Scott-Knott statistical test was applied to determine significant differences between models. This thesis also assesses the potential impact of climate change on species distributions by projecting the best-performing models under future climate scenarios. This research contributes to both ecological modeling and machine learning fields by offering extensive information about the effectiveness of different ensemble strategies for species distribution modeling. The methodological framework developed in this thesis offers practical guidance for researchers and conservationists seeking to implement advanced modeling techniques for habitat suitability assessment and species conservation planning in the face of environmental change.

**Keywords:** Species Distribution Models, Habitat Suitability, Ensemble Learning, Machine Learning, Homogeneous Ensembles, Heterogeneous Ensembles, Stacked Generalization, Bird Conservation, Climate Change, Ecological Modeling, Performance-based Selection, Diversity-based Selection, Borda Count, Scott-Knott Test.