UNIVERSITE DES ANTILLES ET DE LA GUYANE

Laboratoire LAMIA (Laboratoire de mathématiques informatique et applications)

Etude et caractérisation de l'habitat du Moqueur à Gorge Blanche sur la presqu'île de la Caravelle.

Nathalie DESSART, Philipe Hunel

1 Reconnaissance automatique de l'espèce

1.1 Objectifs

Dans le cadre du projet LIFE-CAPDOM, nous avons mené une étude de l'habitat du Moqueur à Gorge-Blanche. Cette espèce d'oiseaux est endémique de la Martinique et elle évolue sur le territoire de la presqu'île de la Caravelle. Le but de notre étude était de mettre en évidence les caractéristiques des lieux de prédilection du Moqueur à Gorge Blanche en utilisant la technologie des réseaux de capteurs sans fil, en proposant un algorithme de détection du Moqueur à Gorge Blanche. En effet, déterminer la présence du Moqueur dans tel ou tel autre milieu, permettrait alors de dresser un modèle qui caractériserait les habitats de prédilections de l'espèce. Dans un premier temps, nous avons fait un état de l'art des études menées sur le suivi d'espèces à l'aide des réseaux de capteurs sans fil et sur les méthodes de reconnaissance sonore. Cette étude bibliographique nous a permis notamment d'assoir nos choix en ce qui concerne la reconnaissance automatique de l'espèce.

1.2 Matériel utilisé

Notre objectif premier était d'utiliser les capteurs en notre possession. Cependant, étant donné que ceux-ci n'étaient pas conditionnés pour un déploiement sur le terrain, nous avons choisi d'acquérir des capteurs de même type mais déjà conditionnés pour un déploiement en milieu naturel. Par le même temps, nous avons pu acquérir des sondes capables de mesurer les facteurs abiotiques suivants : la température, la luminosité, le taux d'humidité du sol, le point de rosée et le taux d'humidité. Les capteurs utilisés (voir figure 1) sont des capteurs sans fil "eKo" [1].

Figure 1 – Capteur sans fil eKo

Figure 2 – Enregistreurs de sons Song Meter

Ces capteurs transmettent des données vers une station de base sur laquelle elles peuvent être visualisées ou sauvegardées pour une analyse ultérieure. Au niveau de la station de base il est également possible de visualiser la topologie du réseau déployé.

Le second type d'équipement (voir figure 2) utilisé pour mener cette étude sont les enregistreurs de sons "Song Meter" [2]. Ces Song Meter ont été utilisés pour récolter des échantillons de chants du Moqueur à Gorge Blanche. Les Song Meter sont des enregistreurs permettant un paramétrage fin de l'engistrement. Ils permettent notamment de choisir la fréquence d'enregistrement, le mode (stéréo ou mono), le taux de compression des enregistrements [1] et les plages horaires durant lesquelles on souhaite acquérir des données sonores.

Les enregistreurs de son Song Meter sont livrés avec un logiciel de reconnaissance sonore appelé Song Scope. Rappelons que la reconnaissance sonore se décompose en quatre étapes distinctes. Ces étapes sont montrées dans la figure 3. Les deux premières étapes consistent à acquérir le son. Celui-ci est généralement enregistré à l'aide d'équipements analogiques (micro) et est ensuite converti en numérique. La troisième étape consiste à la construction du modèle ou empreinte de référence qui permettra de caractériser le motif que l'on souhaite reconnaître. Dans notre cas il s'agit de fabriquer une empreinte de référence qui correspond au Moqueur à Gorge Blanche. Enfin, lors de la dernière étape, il s'agit de détecter le motif que l'on cherche à reconnaître en s'aidant de l'empreinte de référence.

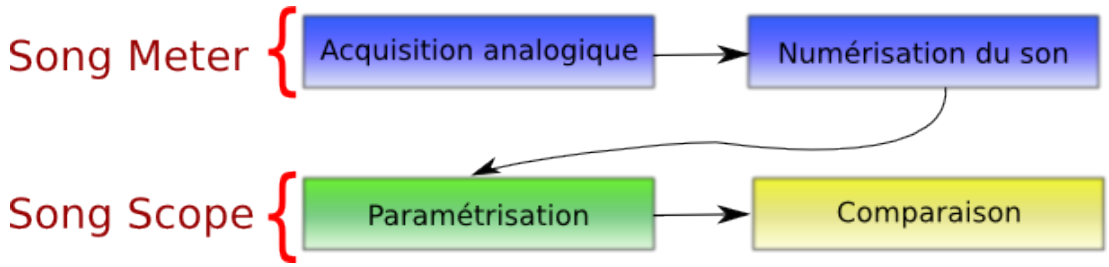

FIGURE 3 – Principe de la reconnaissance sonore

Il existe plusieurs techniques de paramétrisation :

- les méthodes spectrales (par exemple par transformée de Fourrier : FFT)
- les méthodes d'identification (par exemple par codage prédictif linéaire : LPC).

et plusieurs méthodes de comparaison :

- les méthodes statistiques[7, 10, 12, 21]
- les méthodes basées sur les réseaux de neurones [8, 16, 15]
- les méthodes de classification distribuée [9]
- les méthodes de fouille de données [21, 20].

1. Permet d'effectuer plus ou moins d'heures d'enregistrement

Ce sont les Song Meters qui permettent la réalisation des deux premières étapes alors que le logiciel Song Scope nous donne la possibilité de générer des empreintes de référence et de lancer une reconnaissance automatique. Le logiciel Song Scope est basé sur les méthodes spectrales et utilise des Tranformées de Fourrier rapides pour la phase de paramétrisation. L'algorithme de reconnaissance utilisé par ce logiciel est basé sur des méthodes statistiques et implémente les Modèles de Markov Cachés (MMC) [2].

Les modèles de Markov cachés sont des modèles stochastiques qui ont été validés par de nombreuses études[18, 4]. Ils ont été utilisés pour la reconnaissance de la parole, de l'écriture et dans divers domaines d'activité tels que ceux de l'industrie, l'économie ou encore la robotique.

Nous pensons donc que le logiciel Song Scope constitue une bonne alternative pour la reconnaissance automatique des espèces. C'est la raison pour laquelle nous avons cherché à déterminer les réglages adéquats pour la génération d'empreintes de référence à l'aide du logiciel Song Scope.

Avant de présenter les différents paramètres qui ont été retenus pour la reconnaissance automatique nous allons décrire dans la section suivante, la méthodologie adoptée lors des déploiements sur le terrain. Ensuite nous traiterons des travaux effectués sur la caractérisation de l'habitat du Moqueur à Gorge Blanche. Enfin, nous allons conclure.

2. En anglais : Hidden Markov Models (HMM)

1.3 Déploiement terrain : méthodologie adoptée

Le principe a été le suivant, dans un premier temps nous avons déployé les équipements sur des sites a priori fréquentés par les Moqueurs à Gorge Blanche. Nous avons commencé par récolter quelques échantillons en installant un Song Meter sur un lieu désigné par les ornithologues comme étant un lieu de prédilection de l'espèce. Le Song Meter a été configuré avec les paramètres donnés dans le tableau de la figure 5. Ces premiers enregistrements ainsi obtenus servent par la suite à l'expérimentation de la détection automatique. Ils nous ont aussi permis d'obtenir les premiers échantillons des chants de moqueurs.

Par la suite nous avons installé des couples Song Meter/eKo dans un site toujours choisi du fait de la présence, a priori, du Moqueur. Les couples eKo/Song Meter ont été déployés non loin du château Dubuc (voir figure 6) [3]. Dans la figure 6, les repères jaunes indiquent l'emplacement des capteurs et des enregistreurs. Le repère rouge indique l'emplacement de la station de base. Les données récoltées périodiquement par les capteurs eKo sont ainsi acheminées vers la station de base. Les capteurs eKo ont relevé, la température, la luminosité, le point de rosée et le taux d'humidité dans le sol. La zone de déploiement correspondait à un emplacement boisé se trouvant à 150 m environ de la station de base.

La figure 4 montre la configuration adoptée lors des déploiements. Cette configuration prend en un compte le réglage des microphones et les paramètres d'enregistrement. Les équipements ont été placés à 1 m environ du sol afin de maximiser les chances de déceler la présence du Moqueur étant donné que selon les spécialistes l'oiseau passe la majeur partie de son temps au sol, où il se nourrit.

FIGURE 4 – Matériel sur le terrain

Paramètre	: **Valeur**
Compression	: wac0
Mic Bias	: L : on R : on
High Pass	: L : 1 KHz R : 1 KHz
Filter	: L : 1 KHz R : 1 KHz
gain 1	: L : 24 Db R : 24 Db
gain 2	: L : 12 Db R : 12 Db
Sample rate	: 16000
Channels	: Stereo
Gain L	: 10.5Db
Gain R	: 10.5Db
Solar	: SunriseSet
Advanced	: Default

FIGURE 5 – Paramètres du Song Meter

3. Carte Google Map

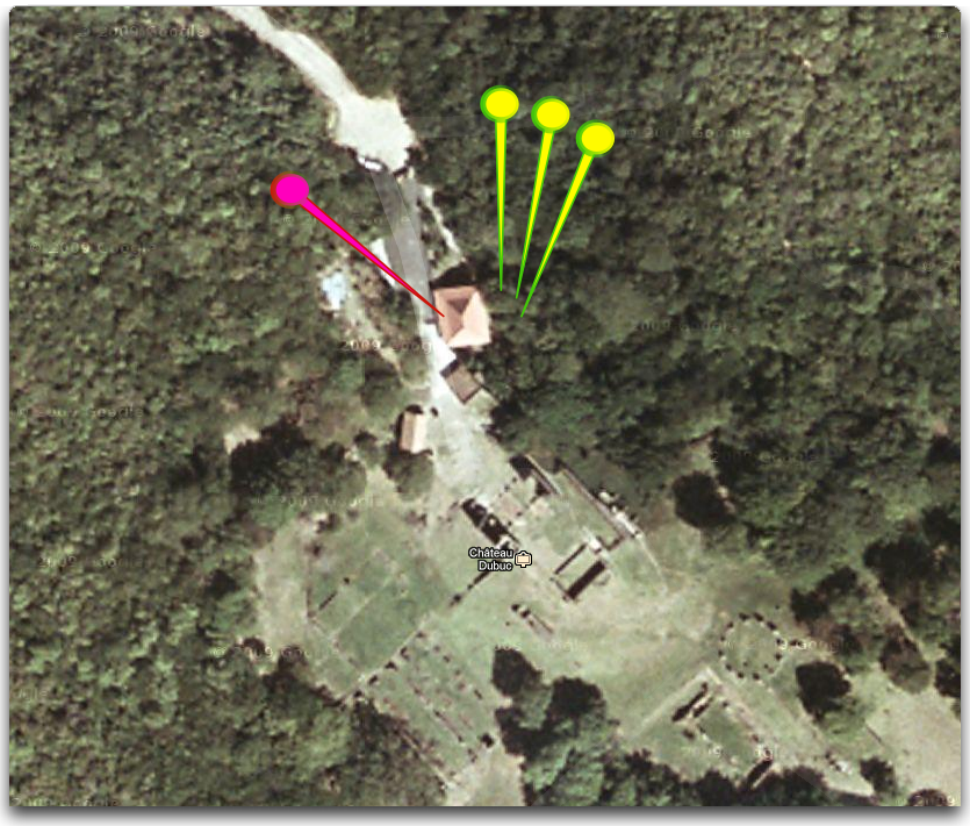

FIGURE 6 – Emplacement des couples eKo/Song Meter

Nous avons ainsi recueilli plusieurs heures d'enregistrements effectués durant les deux tranches horaires suivantes : 5h30-10h30 et 18h-19h soit un total de 1080 heures (3 Song Meter enregistrant chacun durant 6 heures et pendant 60 jours).

A titre informatif, dans l'optique d'un déploiement à plus grande échelle couvrant un territoire plus étendu, il faudrait amplifier la puissance de réception de la station de base par adjonction d'une antenne extérieure et disposer d'un plus grand nombre de capteurs eKo. En effet, le site de la presqu'île de la Caravelle est un site très boisé, et de ce fait il n'est pas possible de bénéficier de la portée radio maximale des capteurs sans fil. La topologie du réseau doit donc présenter un ensemble de nœuds suffisamment proches les uns des autres sur le terrain. Beaucoup de capteurs seraient alors utilisés seulement comme relais pour l'acheminement des données vers la station de base.

Une fois le matériel déployé sur le terrain, nous sommes capables de donner les valeurs des différents facteurs abiotiques mesurés à chaque instant, et en particulier nous obtenons la valeurs de ces facteurs quand l'espèce étudiée s'est manifestée.

Parallèlement aux données récoltées sur le terrain, un spécialiste, en la personne de Jonhatan Priam, a effectué des relevés terrains qui nous ont permis par la suite de mettre en évidence des occurrences de présence du Moqueur sur le terrain. L'expert nous a rapporté huit occurrences de présence du Moqueur à Gorge Blanche. Nous avons alors écouté les bandes sonores (enregistrement de 6 heures par jour durant 15 jours) et nous avons repéré sur celles-ci les chants du Moqueur. Le tableau 1 fait le parallèle entre les relevés effectués sur le terrain lors de la première phase de déploiement et les échantillons trouvés sur les bandes. Chaque ligne du tableau correspond à une date et heure données pour des données récoltées par un Song Meter précis. La position sur la bande correspond à l'occurrence du chant entendu sur le terrain. Pour chaque occurrence nous avons vérifié, sur la bande sonore, si elle était en adéquation avec un chant de Moqueur à Gorge Blanche. Si oui, l'occurence est alors à l'état « Confirmé », sinon elle est à l'état « Non exploitable » si le bruit est trop important dans le signal et à l'état « Non confirmé » si aucun Moqueur n'a été entendu.

Table 1 – Moqueur Gorge Blanche identifé sur le terrain

Date	**Horaires**	**Position bande**	**SM**	**État**
19 juillet	6h32, 6h34	1h04, 1h06	SM-1	Non exploitable
20 juillet	7h17, 7h30, 7h34	1h49, 2h02	SM-1	Non exploitable
18 juillet	7h08, 7h15	1h41	SM-2	Confirmé, retenu
19 juillet	7h29	1h31	SM-2	Confirmé, retenu
21 juillet	8h00	2h32	SM-2	Confirmé, non retenu
27 juillet	9h04	3h36	SM-2	Non confirmé
01 aout	9h08	3h40	SM-3	Confirmé, retenu
04 aout	9h54	4h26	SM-3	Confirmé, non retenu

Parmi les occurrences confirmées, certaines n'ont pas été retenues du fait de leur faible qualité sonore. En effet, nous avons évalué si la qualité auditive de chacune des occurrences était suffisante. Cette évaluation qualitative nous a amené à classer les différents chants observés en trois classes d'audibilité :

- Classe 1 : Échantillon de qualité médiocre
- Classe 2 : Échantillon de moyenne qualité
- Classe 3 : Échantillon de bonne qualité

La mise en place de ces classes nous a permis d'annoter les échantillons sonores de façon plus fine et de construire plusieurs empreintes de référence. La section suivante décrit le travail réalisé pour déterminer des paramètres optimaux nécessaire à la reconnaissance automatique à l'aide du logiciel Song Scope.

1.4 Paramètres pour la reconnaissance automatique

Le but ici, est d'automatiser la reconnaissance du Moqueur afin de minimiser au maximum le travail d'écoute des spécialistes. Pour ce faire nous avons généré et testé plusieurs empreintes de références sur plusieurs enregistrements afin de déterminer les meilleurs paramètres pour la reconnaissance. Ces empreintes ont été construites à partir des échantillons retenus à la phase de déploiement (cf tableau 2). Certaines empreintes ont également été générées à partir d'échantillons de meilleure qualité dans le but de déterminer dans quelle mesure les échantillons utilisés pour la construction d'empreintes doivent être proche des enregistrements pour lesquels on cherchera à reconnaître l'espèce.

En d'autres termes, nous voulons savoir si le chant d'un oiseau capturé et enregistré dans des conditions optimales permet de construire de bonnes empreintes de référence. La question se pose étant donné que les enregistrements sur le terrain comporte du bruit introduit par les éléments naturels (pluie, vent, autres espèces animales) et par l'activité humaine (marcheurs, touristes).

Table 2 – Classification des chants de Gorge Blanche

Fichier	Nombre d'occurences	Classe	Position sur la bande
SM2-20110719_053000.wav	10 ⇒ 6 échantillons	2 et 3	1h03,38
SM2-20110718_053000.wav	7 ⇒ 7 échantillons	1 et 2	1h41,36
SM3-20110801_053000.wav	1 ⇒ 1 échantillons	1	3h40,49

La mise en place d'une empreinte de référence repose sur le réglage de 19 paramètres concernant à la fois la phase de paramétrisation et la phase de comparaison. Il a donc fallu chercher à déterminer les bonnes valeurs de ces paramètres pour la détection du Moqueur. Les paramètres requis pour la construction d'empreintes de référence et leurs significations sont donnés dans le tableau 3.

Les paramètres 2 et 3 permettent d'ajuster les réglages pour la Transformée de Fourrier. Le taux d'échantillonnage (paramètre 1) doit correspondre à celui utilisé pour l'acquisition des échantillons utilisés.

Les paramètres 5 et 6 permettre de sélectionner une plage de fréquence donnée afin de pouvoir cibler la plage de fréquence de l'espèce étudiée ou de s'en rapprocher. Ils dépendent de la fréquence du chant du Moqueur, qui se situe dans les fréquences basses. Cette observation, nous permet d'éliminer les fréquences hautes et de réduire les fréquences basses dans le but d'éliminer du bruit dans le signal.

TABLE 3 – Paramètres requis pour la construction d'empreintes de référence

1. Sample rate	⇒ taux d'échantillonage
2. FFT overlap	⇒ facteur de chevauchement des fenêtres FFT
3. FFT size	⇒ taille des fenêtres FFT
4. Background filter	⇒ permet d'appliquer un filtre pour l'atténuation du bruit
5. Frequency minimum	⇒ fréquence minimum de la plage de fréquence choisie
6. Frequency range	⇒ fréquence maximum de la plage de fréquence choisie
7. Playback speed	⇒ vitesse de lecture de l'enregistrement
8. Amplitude gain	⇒ taux d'amplification du signal
9. Algorithm	⇒ choix de l'algorithme de classification
10. Max song	⇒ durée maximale du chant de l'espèce
11. Max syllabe	⇒ durée maximale d'un mot dans le chant
12. Max syllabe gap	⇒ durée maximale d'un blanc entre deux mots
13. Dynamic range	⇒ rapport signal sur bruit
14. Maximum complexity	⇒ nombre d'états maximum
15. Maximum resolution	⇒ taille des vecteurs acoustiques
16. Max sample delay	⇒ temps de propagation d'une onde sonore
17. Minimum quality	⇒ qualité minimum requise pour la classification
18. Minimum score	⇒ score minimum requise pour la classification
19. Show top	⇒ nombre de résultat à afficher

Les paramètres 10, 11 et 12 ont été déterminés par rapport à la structure du chant du Moqueur dont la forme spectrale du chant est donné en figure 7. Nous voyons plusieurs occurences du chant du Moqueur, et en étudiant la forme d'un chant (Song) (voir figure 8) nous voyons qu'il est constitué de plusieurs mots (Syllabe) séparés par des blancs (Gap). Nous avons déterminé la durée moyenne de chacune de ces trois grandeurs afin d'ajuster les trois paramètres correspondants. Le chant du Moqueur à Gorge Blanche est constitué d'un ensemble de mots courts séparés par des périodes de silence moyennement courtes.

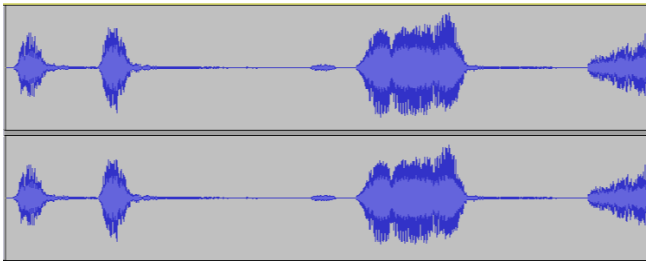

FIGURE 7 – Chants Moqueur

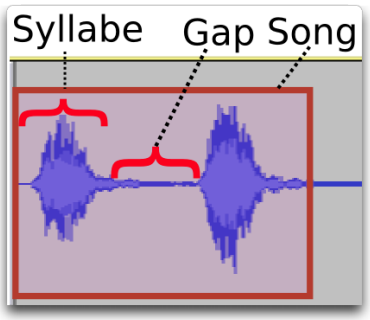

FIGURE 8 – Chant Moqueur (détail)

Nous avons testé plusieurs empreintes de référence sur plusieurs enregistrement afin de déterminer les paramètres adéquats. Lors de nos tests nous nous sommes rendus compte que générer une empreinte de référence à partir d'un son épuré obtenu dans des conditions optimales n'était pas la bonne démarche. En effet, le modèle de reconnaissance ainsi généré ne correspond pas à des enregistrements pour lesquels la qualité des chants enregistrés sera plus ou moins variable et plus ou moins sujet au bruit. Il faut prendre en compte les bruits environnant et la distance entre les oiseaux et les enregistreurs Song Meter. Par contre, l'introduction de quelques échantillons épurés dans le modèle en addition d'échantillons typiques permet une amélioration de celui-ci.

Le tableau 4 donne les réglages que nous avons retenu à la suite de cette étude.

Table 4 – Paramètres des empreintes de référence

1. Sample rate	⇒ 16000
2. FFT overlap	⇒ 1/2
3. FFT size	⇒ 256
4. Background filter	⇒ 1s
5. Frequency minimum	⇒ 10
6. Frequency range	⇒ 20
7. Playback speed	⇒ Normal
8. Amplitude gain	⇒ 0
9. Algorithm	⇒ 2.0
10. Max song	⇒ 1500
11. Max syllabe	⇒ 150
12. Max syllabe gap	⇒ 50
13. Dynamic range	⇒ 15
14. Maximum complexity	⇒ 16
15. Maximum resolution	⇒ 4
16. Max sample delay	⇒ 64
17. Minimum quality	⇒ 80
18. Minimum score	⇒ 80
19. Show top	⇒ 1 match

La section suivante présente l'analyse des données grâce à un algorithme de fouille de données. Analyse qui permet d'extraire des connaissances à partir des données récoltées sur le terrain.

2 Caractérisation de l'habitat : Fouille de données

2.1 Principe général

Lors des phases de déploiement nous avons été à même de déterminer des moments durant lesquels le Moqueur a chanté et des moments durant lesquels le moqueur n'a pas chanté. Nous avons alors appliqué des techniques de fouilles de données afin de caractériser l'habitat du Moqueur. Le but a été de savoir quelles étaient les conditions (valeurs des différents facteurs abiotiques) favorables à la manifestation du Moqueur. Cette analyse nous a permis de mettre en exergue comment certains des facteurs abiotiques mesurés étaient déterminant dans la présence ou l'absence du Moqueur à Gorge Blanche. Avant de présenter les différents résultats obtenus, nous allons voir dans quelle mesure une telle analyse peut être utile pour la caractérisation de l'habitat du Moqueur à Gorge Blanche.

L'acquisition d'un volume important de données permet par la suite d'appliquer des méthodes de fouille de données [4]. Le but est alors d'extraire une connaissances basée sur l'ensemble des données acquises. En d'autres termes, il s'agit de trouver des motifs [5] pertinents qui nous apportent une nouvelle connaissance. Dans le cas du suivi de l'espèce les motifs pertinents nous permettent d'identifier les conditions favorables à la manifestation de celle-ci. Il existe différentes techniques de fouille de données [3, 14, 17]. Nous pouvons citer entre autres, les réseaux de neurones[25, 11, 24] , les algorithmes génétiques[6, 19, 24] ou encore les arbres de décision [5, 17, 22]. Nous avons choisi d'utiliser une technique basée sur l'élaboration d'un arbre de décision, arbre qui permet d'avoir un rendu visuel relativement compréhensible, même pour les néophytes.

Nous avons utilisé une implémentation de l'algorithme C4.5 [6][13, 23], basé sur la technique des arbres de décisions. Pour cette technique, il s'agit de posséder un jeu de données que nous étiquetons par des classes que nous avons identifiées. L'algorithme permet ensuite de dégager des règles de classification pour chacune des classes. Une fois ces règles mises en évidence, de nouvelles données peuvent être classées de façon automatique.

Prenons l'exemple de la figure 9, dans lequel les données ont été classées suivant deux classes : « Présence du Moqueur » et « Absence du Moqueur ». L'arbre nous dit que le Moqueur est absent si la température est inférieure ou égale à 28°C. Par contre, si la température est strictement supérieure à 28°C, il faut alors regarder la valeur de la luminosité. Si cette dernière est inférieure à 20 alors le Moqueur ne se manifeste pas et dans le cas contraire le Moqueur se manifeste. En d'autres termes si nous possédons de nouvelles données elles pourront être automatiquement classées afin de déterminer si le Moqueur se manifeste ou non dans de telles conditions. Typiquement, cela nous permettrait de choisir des habitats adéquats pour la réintroduction du Moqueur à Gorge Blanche.

4. En anglais : "datamining"
5. En anglais : patterns
6. L'implémentation J48

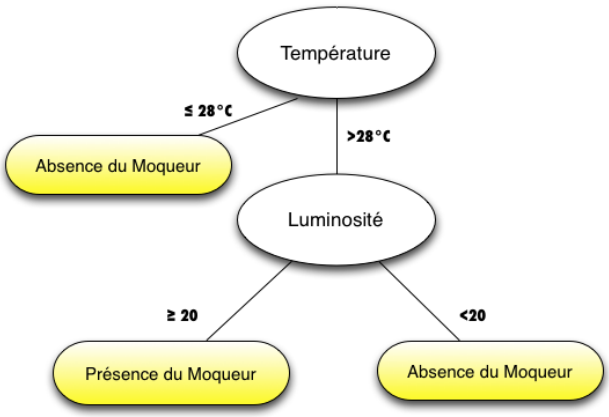

FIGURE 9 – Exemple d'arbre de décision

Imaginons que nous voulions introduire l'espèce dans un nouvel habitat, nous avons le choix entre l'habitat A et l'habitat B. La procédure à suivre est la suivante :

1. Relever les facteurs abiotiques de l'habitat A
2. Relever les facteurs abiotiques de l'habitat B
3. Appliquer l'arbre de décision aux données collectées pour l'habitat A
4. Appliquer l'arbre de décision aux données collectées pour l'habitat B
5. Choisir parmi les deux habitats potentiels celui qui correspond le mieux

A titre d'exemple, si les relevés répertoriés dans le tableau 5 sont ceux faits pour les habitats A et B. L'idée est d'analyser les résultats fournis une fois que l'arbre de décision de la figure 9 a été appliqué à ces différents relevés.

TABLE 5 – Exemples de rélevés effectués pour les habitats A et B

Habitat	Température	Luminosité	Classe
A	24	28	Absence du Moqueur
A	18	40	Absence du Moqueur
A	30	30	Présence du Moqueur
B	30	20	Présence du Moqueur
B	28	20	Absence du Moqueur
B	29	50	Présence du Moqueur

Le choix du nouvel habitat peut alors se faire sur la base de plusieurs critères. Un critère peut être le taux α de vecteurs qui correspondent à la classe « Présence du Moqueur ». Un deuxième critère pourrait permettre d'apprécier « comment l'habitat correspondrait à celui du Moqueur ». En effet, si pour un habitat potentiel donné tous les vecteurs d'enregistrement sont étiquetés « Présence du Moqueur », il est alors judicieux de voir la distance de ce vecteur au vecteur seuil. Dans notre exemple le vecteur seuil V_s est (29,20), c'est à dire que se sont les valeurs minimales pour lesquelles l'habitat correspond à un habitat propice au Moqueur. Lorsque nous considérons le vecteur (30,30) relevé pour l'habitat A et le vecteur (30,20) relevé pour l'habitat B, ces deux vecteurs correspondent à la classe « Présence du Moqueur » mais le vecteur (30,30) correspond mieux étant donné que l'on s'éloigne du vecteur seuil V_s. Nous pouvons alors basé le choix de l'habitat sur un ensemble de critères. Ici, nous pouvons combiner les deux critères énoncés et obtenir la règle suivante :

$$\alpha \geq S_\alpha \wedge \frac{\sum_i^N d(V_i, V_s)}{N} \leq D_s \qquad (1)$$

Dans la formule (1), α représentent le taux de vecteurs qui correspondent à un habitat favorable et S_α représente le seuil minimal acceptable pour que l'habitat considéré soit un bon candidat. La distance D_s correspond à la distance minimale acceptable pour que l'habitat considéré soit un bon candidat. Ici, la distance moyenne des vecteurs qui correspondent à un habitat favorable au vecteur seuil V_s doit être supérieure ou égale à D_s.

Des règles plus complexes peuvent être élaborées selon l'ensemble de facteurs abiotiques mesurés et selon le critère qu'il est préférable de privilégier. La section suivante donne les résultats obtenus grâce à une analyse de type fouille de données.

2.2 Résultats

Dans l'étude menée sur le Moqueur à Gorge Blanche, nous avons identifié deux classes : *Présence du Moqueur* et *Absence du Moqueur*, nommées respectivement « Présence » et « Absence » dans l'arbre de la figure 11 par soucis de lisibilité. Nous sommes partis d'un ensemble de vecteurs d'entrée qui nous donnait des moments de présence et des moments d'absence du Moqueur à Gorge Blanche, ainsi que les valeurs correspondantes pour les facteurs abiotiques. La figure 10 présente un aperçu des données utilisées pour la fouille de données. Le jeu complet était constitué de 130 enregistrements. Pour chacun des enregistrements, nous avons la date et l'heure ainsi que la valeur des différents facteurs abiotiques. De plus, chaque enregistrement correspond à une des deux classes définies au préalable. Le résultat obtenu par ce type de fouille de données est un arbre de décision qui permet alors de déterminer les caractéristiques précises de chacune des classes mises en jeu. Les nœuds représentent alors un attribut particulier (par exemple dans notre cas la température, ou le taux d'humidité du sol) et les feuilles représentent les classes. Un tel arbre, nous permet d'associer les valeurs des attributs à la classe à laquelle ils appartiennent.

date	heure	Humidité du sol	Humidité	Témpérature	Point de rosée	Luminosité	Classe
01-août	9h10	23,67	97,3	26,63	25,54	4,32	présence
25-août	7h15	35,96	89,33	27,51	25,58	2,52	présence
29-août	7h34	36,55	92,68	26,19	24,88	1,12	présence
02-sept	7h40	21,76	100,57	26,11	25,27	1,23	présence
05-sept	7h40	18,58	100,37	25,85	25,85	1,73	présence
12-sept	08h00	29,27	101,44	26,78	26,78	3,51	présence
15-sept	08h45	29,27	99,95	27,83	27,81	12,06	présence
19-sept	09h30	28,89	101,62	26,46	26,46	13,77	présence
30-juil	7h20	30,53	86,2	25,99	24,79	1,52	présence
21-juil	15h00	19,68	80,28	28,79	24,69	34,51	absence
26-juil	09h00	26,61	99,35	25,15	25,01	0,24	absence
03-août	09h00	30,86	92,63	27,15	25,85	3,36	absence
04-août	21h00	30,99	88	28,66	26,22	1,61	absence
16-août	03h00	31,63	97,71	25,51	25,36	0	absence
23-août	09h00	34,54	96,55	25,29	24,98	0,09	absence
23-août	09h00	29,03	98	24,89	24,89	0	absence
02-sept	09h00	21,76	100,57	26,11	26,11	1,23	absence
30-août	09h00	35,81	92	24,58	24,43	0,74	absence

FIGURE 10 – Aperçu des données.

Ainsi, l'utilisation de cette technique de fouille de données, nous a permis de savoir pour quel type de milieux (valeurs des facteurs abiotiques) le Moqueur à Gorge Blanche se manifeste. Nous avons obtenu une classification des lieux où le Moqueur est présent et des lieux où il ne l'est pas. Ce qui nous permet donc de caractériser le type d'habitat qui lui est favorable.

L'arbre de décision de la figure 11 correspond au résultat obtenu par la fouille de donnée. Nous pouvons observé que la présence du Moqueur est d'abord déterminée par le taux de luminosité. Les deux premiers niveaux nous permettent de voir que le Moqueur ne se manifeste pas lorsque la luminosité est très faible, et a fortiori la nuit. Si par contre les conditions de luminosité sont réunies, il faut alors examiner la valeur du facteur « Point de rosée ». Nous voyons alors qu'une valeur élevée de ce facteur suffit à conclure à la présence du Moqueur. Ainsi de suite, l'ensemble de l'arbre peut être lu en utilisant la même procédure.

Après l'analyse complète de l'arbre obtenu (figure 11), il apparaît que la luminosité est le facteur le plus déterminant dans le sens où le Moqueur ne se manifeste que le jour (luminosité>1.07). Il ne se manifesterait pas le nuit, comme nous l'avons mentionné ci-dessus. Par contre, nous pouvons également constater que le Moqueur n'a pas besoin d'un taux de luminosité très important pour se manifester. Cette observation corrobore les observations des ornithologues qui ont décrit les lieux de prédilection des Moqueurs comme étant des sous-bois ombragés.

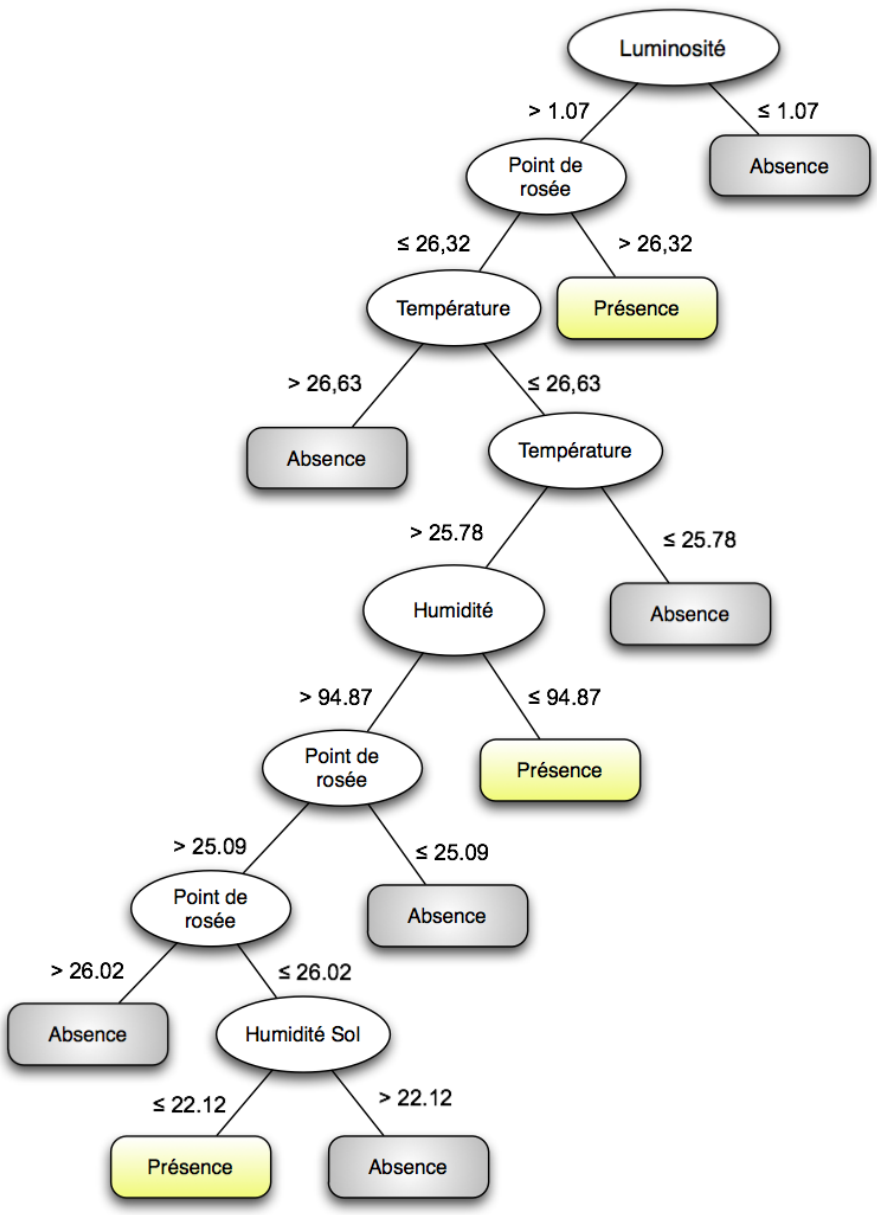

FIGURE 11 – Modèle 1 - Résultats d'analyse de fouille de données.

Nous voyons également que Le Moqueur privilégierait les zones assez chaudes car on ne le trouve que dans des endroits où la température est comprise entre 25°C et 26°C. Il serait intéressant ici de mener une analyse plus fine en relevant des données dans des zones dont la température ambiante serait comprise entre 25°C et 40°C, par exemple afin d'affiner cette connaissance et de voir si le Moqueur peut évoluer dans un habitat dont la température moyenne est, par exemple, de 32°C.

Les valeurs des taux d'humidité ambiant, "Point de rosée", et taux d'humidité dans le sol nous indiquent que le Moqueur apprécie les milieux très humides. En effet, la valeur du point de rosée est fortement corrélée à celle de l'humidité. La valeur du point de rosée est obtenue par la formule (2). Dans cette formule le point de rosée P_r est fonction de l'humidité relative RH et de la température T.

$$P_r = \frac{C_2 * f(T, RH)}{C_1 - f(T, RH)} \qquad \text{avec } f(T, H) = \frac{C_1 * T}{C_2 + T} + ln(RH/100) \tag{2}$$

Si nous observons la deuxième apparition du point de rosée dans l'arbre de décision de la figure 11, à ce niveau, nous constatons que si la valeur de ce facteur abiotique est inférieur à 25,09°C alors la Moqueur ne se manifeste pas. Par contre, si la valeur est supérieure à 25,09°C il y a des chances que l'espèce se manifeste. Dans le premier cas, la valeur du point de rosée s'éloigne de la température mesurée (en effet nous atteignons ce nœud de l'arbre quand la température est supérieure à 25,78°C et dans le second cas elle s'en rapproche. Cela signifie que plus l'humidité est importante et plus nous avons de chance que l'espèce se manifeste. En effet, selon la formule (2), plus l'écart entre le point de rosée et la température est faible et plus l'humidité relative est importante.

Les résultats obtenus sont satisfaisants car le modèle généré permettrait de classer correctement 87% des enregistrements (voir tableau 6) présents dans un jeu de données. Le nombre et le pourcentage d'enregistrements bien classés et mal classés sont donnés dans le tableau 6. Ce tableau nous montre également le taux de « vrais positifs » et de « faux positifs » obtenus pour chacune des deux classes. Les « vrais positifs » correspondent à des enregistrements définis comme appartenant à une classe C considérée et qui appartiennent réellement à cette classe. Les « faux positifs » correspondent à des enregistrements définis comme appartenant à la classe C et qui n'appartiennent pas réellement à cette classe.

Les valeurs que nous avons ici nous indiquent que le modèle détermine correctement la présence du Moqueur. Cependant, elles nous indiquent également que le modèle doit être affiné en introduisant de nouvelles mesures pour lesquelles l'espèce a été repérée.

Table 6 – Pertinence du modèle obtenu (Modèle 1)

Classement	**Nombre**	**Pourcentage**
Correct	115	87%
Incorrect	16	12%
Classe	**Taux de vrais positifs**	**Taux de faux positifs**
Présence	0.467	0.069
Absence	0.931	0.533

Les résultats précédent sont issus d'une analyse qui n'incluait pas le moment (heure des manifestations). Nous avons également effectuer une fouille de donnée en incluant le paramètre « heure ». L'arbre de décision obtenu est donné en figure 12. Ici, nous remarquons aussi que la luminosité est une donnée clef et qu'en-dessous d'un certain seuil le Moqueur ne se manifeste pas. Ce même arbre nous indique que le Moqueur à Gorge Blanche se manifesterait plutôt le matin jusqu'à 9 heures. Le taux de classification correcte pour ce modèle est de 98% (voir tableau 7), et aucun faux positif n'est à dénombrer pour la classe « Présence ». Le modèle est de ce fait pertinent.

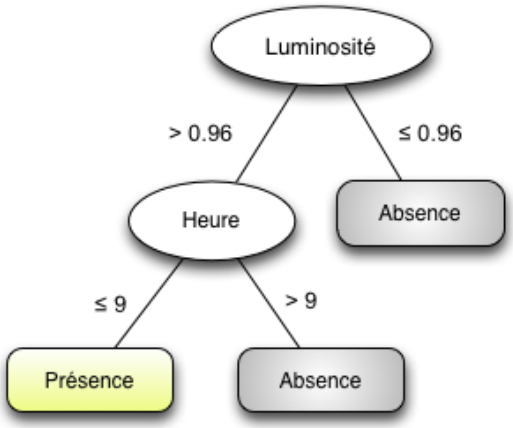

FIGURE 12 – Modèle 2 - Arbre de décision : avec l'heure

TABLE 7 – Pertinence du modèle obtenu (Modèle 2)

Classement	**Modèle 2**	**Modèle 3**
Correct	98%	90%
Incorrect	1.4%	9.8%
Classe	**Taux de VP • Taux FP**	**Taux de VP • Taux FP**
Présence	0.933 • 0	0.8 • 0.071
Absence	0.986 • 0.053	0.929 • 0.901

Étant donné que le point de rosée est calculable à partir de l'humidité relative et de la température, nous avons reconduit l'analyse faite pour le modèle 1, en ne prenant pas en compte le point de rosée. La dépendance entre des variables correspondantes aux grandeurs mesurées peut effectivement introduire un biais dans le modèle généré.

Comme nous pouvons le constatée en tableau 7, le modèle 3 est une amélioration du modèle 1 puisqu'il permet 90% de bonne classification. Nous avons également un taux de vrais positifs qui est doublé pour la classe « Présence » et un taux de vrais positifs qui reste correct pour la classe « Absence ». De plus, le taux de faux négatifs est très faible pour les deux classes.

Le modèle 3 est donné en figure 13. Ce modèle nous apporte les même informations que le modèle 1. Par contre, il nous confirme que le taux d'humidité joue un rôle primordial.

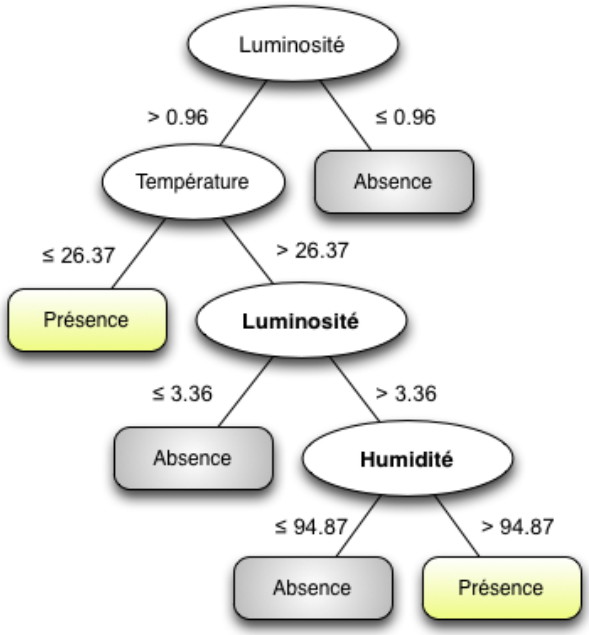

FIGURE 13 – Modèle 3 : Arbre de décision : sans point de rosée

Cette première analyse des jeux de données récoltées nous a donc permis de dresser le constat suivant : le Moqueur à Gorge Blanche se manifeste en journée et préfère les zones relativement humides et chaudes.

3 Conclusion

L'étude de l'habitat du Moqueur à Gorge Blanche nous a permis d'une part de déterminer les conditions selon lesquelles le Moqueur à Gorge Blanche se manifeste, et d'autre part de mettre en évidence des paramètres correctes pour la génération d'empreintes de référence.

Nous avons pu déterminer que le Moqueur se manifeste le jour et privilégie les zones fortement humide. Ces observations sont issues de l'analyse des jeux de données en notre possession, et correspondant aux différents enregistrements effectués sur le terrain. Il faudrait enrichir le jeu de donnée afin d'affiner le modèle. Par exemple, effectuer des enregistrement sur d'autre périodes de l'année, nous permettrait de constater d'autres comportement de l'oiseau. Rappelons que les enregistrements couvrent une période de trois mois et qu'il serait intéressant d'avoir des enregistrements en pleine période de Carême, période durant laquelle le temps est plus chaud et moins humide qu'en Août par exemple.

Dans le même ordre d'idée, il est nécessaire que les experts puissent identifier de nouvelles manifestations de l'espèce afin de pouvoir enrichir la base d'échantillons utilisés pour la constuction des empreintes de référence.

L'étude précédente menée à partir d'un algorithme de fouilles de données va également être conduite à l'aide d'une analyse statistique basée sur l'étude des corrélations entre variables. Cette analyse statistique, actuellement en cours, nous permettra de confirmer ou de préciser nos résultats.

Ces travaux devraient nous permettre de déployer un système de réseau de capteurs réactifs pour la surveillance de l'environnement du Moqueur à Gorge Blanche. En effet, ayant identifié les facteurs abiotiques favorables à la présence du Moqueur, nous allons chercher à mettre en place un système commandant l'enregistrement des chants lorsque le contexte environnemental est propice. Bien sûr, ce système pourra être évolutif, dans le sens où il pourra s'adapter aux nouveaux comportements du Moqueur si celui-ci était amené à modifier son comportement à cause de facteurs extérieurs comme la présence d'un prédateur ou la perturbation liée à une présence humaine trop importante.

Nous remercions les différents acteurs qui ont permis la réalisation de ce projet, à savoir le Parc Naturel Régional de la Martinique, les collègues géographes, informaticiens, mathématiciens et ornithologues et la Ligue pour la Protection des Oiseaux. Un remerciement particulier à M. Stattner pour sa collaboration.

Références

[1] http ://www.memsic.com/.

[2] http ://www.wildlifeacoustics.com/.

[3] Srikant R. Agrawal, R. Fast algorithms for mining association rules. *VLDB Conference*, 1994. cited By (since 1996) 5.

[4] M. A. Anusuya and S. K. Katti. Speech Recognition by Machine, A Review. January 2010.

[5] Weiss S. Ap, C. Data mining with decision trees and decision rules. *Future Generation Computer Systems*, 13(2-3) :197–210, 1997. cited By (since 1996) 61.

[6] Lopes H.S.-Freitas-A.A. Araujo, D.L.A. A parallel genetic algorithm for rule discovery in large databases. *Proc 1000 IEEE Systems, Man and Cybernetics Conf*, 3 :940–945, 1999. cited By (since 1996) 5.

[7] T. S. Brandes. Feature Vector Selection and Use With Hidden Markov Models to Identify Frequency-Modulated Bioacoustic Signals Amidst Noise. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(6) :1173–1180, 2008.

[8] Jinhai Cai, D. Ee, Binh Pham, P. Roe, and Jinglan Zhang. Sensor Network for the Monitoring of Ecosystem : Bird Species Recognition. In *Proc. 3rd Int. Conf. Intelligent Sensors, Sensor Networks and Information ISSNIP 2007*, pages 293–298, 2007.

[9] Xu Cheng, Ji Xu, Jian Pei, and Jiangchuan Liu. Hierarchical distributed data classification in wireless sensor networks. In *Proc. IEEE 6th Int. Conf. Mobile Adhoc and Sensor Systems MASS '09*, pages 10–19, 2009.

[10] Chih-Hsun Chou, Chang-Hsing Lee, and Hui-Wen Ni. Bird Species Recognition by Comparing the HMMs of the Syllables. In *Proc. Second Int. Conf. Innovative Computing, Information and Control ICICIC '07*, 2007.

[11] Shavlik J.W. Craven, M.W. Using neural networks for data mining. *Future Generation Computer Systems*, 13(2-3) :211–229, 1997. cited By (since 1996) 63.

[12] Mark Gales and Steve Young. The application of hidden Markov models in speech recognition. *Found. Trends Signal Process.*, 1 :195–304, January 2007.

[13] Xu L.T. Jiang, X. C4.5 decision tree display algorithm design. *Computer Engineering and Applications*, 4 :93–94, 2003. cited By (since 1996) 1.

[14] Symphor-J.-E. Nock R. Poncelet P. Laur, P.-A. Statistical supports for mining sequential patterns and improving the incremental update process on data streams. *Intelligent Data Analysis*, 11(1) :29–47, 2007. cited By (since 1996) 5.

[15] A. L. McIlraith and H. C. Card. Birdsong recognition with DSP and neural networks. In *Proc. IEEE Communications, Power and Computing WESCANEX 95*, volume 2, pages 409–414, 1995.

[16] A. L. McIlraith and H. C. Card. Bird song identification using artificial neural networks and statistical analysis. In *Proc. IEEE 1997 Canadian Conf. Electrical and Computer Engineering*, volume 1, pages 63–66, 1997.

[17] J.R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1) :81–106, 1986. cited By (since 1996) 3434.

[18] Gailius Raškinis and Danutė Raškinienė. Building medium-vocabulary isolated-word lithuanian hmm speech recognition system. *Informatica*, 14 :75–84, January 2003.

[19] Lei-H. Shi, X.-J. A genetic algorithm-based approach for classification rule discovery. volume 1, pages 175–178, 2008. cited By (since 1996) 3.

[20] E. Vilches, I. A. Escobar, E. E. Vallejo, and C. E. Taylor. Data Mining Applied to Acoustic Bird Species Recognition. In *Proc. 18th Int. Conf. Pattern Recognition ICPR 2006*, volume 3, pages 400–403, 2006.

[21] Erika Vilches, Ivan A. Escobar, Edgar E. Vallejo, and Charles E. Taylor. Targeting Input Data for Acoustic Bird Species Recognition Using Data Mining and HMMs. In *Proceedings of the Seventh IEEE International Conference on Data Mining Workshops*, ICDMW '07, pages 513–518, Washington, DC, USA, 2007. IEEE Computer Society.

[22] Jing X.-ZiMian-H. HongLli L. Xiong, W. Data analysis and prediction of securities based on decision tree algorithm. pages 83–87, 2009. cited By (since 1996) 0.

[23] Ai T.-H. Yu, M. Study of rs data classification based on rough sets and c4.5 algorithm. volume 7492, 2009. cited By (since 1996) 1.

[24] Yuchang Lu-Chunyi-Shi Yuanhui, Zhou. Combining neural network, genetic algorithm and symbolic learning approach to discover knowledge from databases. volume 5, pages 4388–4393, 1997. cited By (since 1996) 5.

[25] G. Peter Zhang. Neural networks for data mining. In Oded Maimon and Lior Rokach, editors, *Data Mining and Knowledge Discovery Handbook*, pages 419–444. Springer, 2010.