

AQUIMER

APRI : Développement d'un **Algorithme** utilisant l'intelligence artificielle pour faciliter la saisie de données et générer les étiquettes **PRIx** en poissonnerie

Rapport Final

Remerciements

Cette étude a été soutenue par France Filière pêche via le projet APRI « Développement d'un Algorithme utilisant l'intelligence artificielle pour faciliter la saisie de données et générer les étiquettes **PRIx** en poissonnerie » dans le cadre de l'appel à projet « Enjeux immédiats 2020 ». Le Projet APRI était porté par le pôle AQUIMER, en partenariat avec l'entreprise Hublot.

TABLE DES MATIERES

Remerciements	1
INTRODUCTION	3
1. PARTENAIRES.....	5
1.1. AQUIMER.....	5
1.2. HUBLOT	5
2. RAPPEL DES OBJECTIFS.....	5
3. RESULTATS.....	6
CONCLUSION ET PERSPECTIVES	10
REFERENCES	10

INTRODUCTION

Les points de vente français rencontrent de nombreuses difficultés dans le suivi de la traçabilité et de l'étiquetage. Et malgré l'arrivée du numérique, la situation s'aggrave encore d'année en année. Selon la DGCCRF, en 2016, 46% des contrôles ont relevé au moins 1 anomalie sur l'information consommateur (sur 1402 établissements). En 2018 (soit 2 ans plus tard), c'est 53%, soit 7 points de plus (soit une centaine d'établissements) concernés par ces erreurs.

Ce constat est aussi partagé par l'UFC-Que-Choisir, qui relève que 66% des mentions obligatoires ne sont pas conformes.

Et pourtant, les managers de rayons marée ne ménagent pas leurs efforts au quotidien pour assurer l'étiquetage le plus fiable possible. Chaque matin, les équipes des rayons prennent entre 30 minutes et 1h30 (en fonction de la taille) pour réaliser leurs étiquettes. Il s'agit de gérer entre 30 et 120 étiquettes sanitaires chaque matin, comportant une dizaine de paramètres (Engins, zone, sous-zone, noms latins ...), soit plus de 1000 vérifications unitaires chaque matin pour les rayons les plus importants. L'amende prévu lors de contrôle pouvant aller jusqu'à 500€/étiquette fausse, cette vérification est essentielle pour le bon fonctionnement du magasin.

La cible du projet est assez large, puisqu'aujourd'hui il existe plus de 3800 poissonneries indépendantes et 12000 supers et hypermarchés équipés d'une poissonnerie traditionnelle.

L'ensemble de ces acteurs rencontre au quotidien ces problématiques d'étiquetage prix. De plus l'harmonisation des informations reste complexe, on estime qu'il y a environ 7000 producteurs et 2500 bateaux en capacité de produire des étiquettes de traçabilité, toutes différentes.

Contexte Socio-Economique et réglementaire

Hublot est un logiciel de gestion des produits de la mer qui permet d'aider les poissonniers dans leurs tâches quotidiennes : rentabiliser leur rayon en réduisant la perte de matière premières, libérer un maximum de temps pour le contact clientèle. L'objectif de l'application est d'optimiser les tâches, comme les commandes fournisseurs des magasins, pour limiter le taux de casse aujourd'hui situé entre 4 et 10%. Actuellement, les utilisateurs de Hublot présentent un taux entre 3 et 4%. Également, la gestion de la traçabilité est une tâche laborieuse et extrêmement chronophage au quotidien pour les poissonneries.

De plus, la recopie manuelle des informations depuis les étiquettes sanitaires des fournisseurs peut causer des erreurs sur l'étiquette prix présentée au client.

Aujourd'hui, les consommateurs sont de plus en plus sensibles par ces problématiques de traçabilité, et la législation est devenue plus pointilleuse, or la filière halieutique a peu d'outils pour répondre à ces exigences.

En plus de ces problématiques réglementaires, il arrive que des produits soient dévalorisés, en raison d'un doute, ou d'un manque d'information de la part du poissonnier (Extra, A, B, etc.) ou alors certains labels de qualités ne peuvent pas être apposés sur le produit (Pavillon France, MSC, ASC, etc.).

L'amélioration de la traçabilité des produits de la mer, peut également apporter un grand atout aux clients finaux. Ils pourraient ainsi connaître la date de pêche, le nom du bateau, éventuellement visualiser la zone d'où vient le poisson. Ces informations peuvent permettre de promouvoir la pêche française, en mettant en avant ces informations.

Une des limites à la création d'outils dédiés à une traçabilité numérique est la multitude d'acteurs (de la TPE à la multinationale) qui utilisent des systèmes de traçabilité différents, et l'interopérabilité demeure un enjeu majeur pour la filière. AQUIMER travaille depuis de nombreuses années sur ce sujet à travers des projets d'innovation :

- « Etat des lieux : Numérisation de la traçabilité dans la chaîne logistique », financé par le dispositif TIC&PME 2010, qui a permis de faire un état des lieux de la maturité de la filière au regard de la traçabilité numérique ;
- « DEMATFLUX », financé par le FEAMP en 2008-2009, qui visait à identifier et caractériser les principaux flux du secteur des produits aquatiques et de standardiser (norme UNCEFACT) les données de traçabilité (en format ebXML)
- « Traseanet Center », financé par le FEAMP, 2009-2010, qui a défini le cadre d'une plateforme d'échange de données en termes d'architecture informatique
- « TRASEAPILOT », financé par le dispositif TIC&PME 2015 qui visait à valider la faisabilité de mettre en place une plateforme d'échange de données entre les acteurs de la filière

Toutes ces études menées par AQUIMER ont mis en lumière la nécessité de standardiser les échanges pour garantir une traçabilité numérique de l'amont à l'aval, or on déplore des difficultés à déployer un tel système à l'échelle de la filière pour plusieurs raisons :

- Ce type de système n'a de sens que si tous les acteurs y adhèrent
- Cela nécessite des investissements importants pour tous les acteurs de la filière
- Le pilotage du système doit être centralisé et la maintenance assurée par une seule instance qui n'a pas été identifiée par AQUIMER

France Filière Pêche a également récemment réalisé une étude sur les flux de données dans la filière et les conclusions semblent être les mêmes.

Un des défis à relever pour obtenir une traçabilité numérique est donc de créer des outils permettant d'extraire les données de traçabilité sans modifier les outils utilisés par les acteurs du marché. Alors que les précédents essais de numérisation nécessitaient, pour avoir un fonctionnement optimal, de partir de 0 pour atteindre 100% d'adoption de la part de tous les acteurs de la filière, il s'agit ici de développer un algorithme compatible avec 95% des étiquettes sanitaires avant même de changer les habitudes et les systèmes existants. Le levier pour y parvenir est d'utiliser l'intelligence artificielle pour extraire les données, et pouvoir procéder à la numérisation de cette manière.

De plus, il nous paraissait intéressant de partir d'un système déjà implémenté et éprouvé chez certains poissonniers (indépendants et GMS). En effet le système d'Hublot est déjà utilisé dans plusieurs hypermarchés et poissonneries, d'ailleurs ces professionnels soutiennent fortement le système (Cf les lettres de soutien). L'adoption du système par ces acteurs démontre qu'il répond déjà à une grande partie de leurs besoins et le projet APRI permettra de consolider les liens préétablis entre Hublot et ses clients, et de conquérir de nouveaux utilisateurs.

1. PARTENAIRES

Le projet rassemble deux partenaires complémentaires. L'entreprise Hublot qui a développé l'application innovante de gestion du rayon marée. AQUIMER qui rassemble parmi ses adhérents une partie de potentiels utilisateurs de la solution.

1.1. AQUIMER

AQUIMER, Association interprofessionnelle créée en 1999 et labellisée Pôle de Compétitivité national en 2005, le Pôle AQUIMER est une interface active entre les projets des entreprises et leur concrétisation. Son objectif est de contribuer à la compétitivité des entreprises du secteur des produits aquatiques en répondant à leurs besoins, en les aidant à pérenniser et développer leurs activités grâce à l'innovation, et à l'amélioration des technologies.

Pour le Pôle, il était important d'être porteur du projet APRI pour de multiples raisons. Le pôle accompagne quotidiennement des entreprises du secteur des produits aquatiques afin d'innover et de développer leur activité. De nombreux projets ont d'ailleurs été menés en matière d'amélioration de la traçabilité dans la filière, et de standardisation. La rencontre avec l'entreprise Hublot nous a convaincu de l'utilité de l'intérêt d'exploiter l'intelligence artificielle pour améliorer la traçabilité, c'est pourquoi nous avons monté le projet avec l'entreprise Hublot.

Convaincu des nombreuses applications des systèmes déployés dans le projet APRI dans la filière, il était primordial pour AQUIMER de monter en compétence sur la compréhension de l'intelligence artificielle afin de pouvoir sensibiliser les entreprises à l'utilité d'intégrer une telle technologie pour gagner du temps, du confort de travail et de la compétitivité.

.

1.2. HUBLOT

Hublot est une start-up française qui propose une solution de gestion pour le rayon marée dans la grande distribution et chez les indépendants. Son objectif est de simplifier la vie du chef de rayon, d'améliorer le suivi de traçabilité sanitaire, de rentabiliser le rayon en apportant tous les outils nécessaires.

Hublot a conçu un logiciel de gestion du rayon poissonnerie, qui permet de soulager le chef de rayon sur les aspects suivants : Gestion, Traçabilité, Hygiène, Commerce, Étiquette Prix.

Afin d'améliorer ce logiciel sur la partie « étiquette prix » Hublot a souhaité intégrer de l'intelligence artificielle dans son système. L'objet du projet APRI est donc d'utiliser l'intelligence artificielle pour reconnaître automatiquement les informations clés de traçabilité, à partir de photos d'étiquettes sanitaires, afin de générer les étiquettes prix sans qu'il y ait de saisies de la part du chef de rayon.

2. RAPPEL DES OBJECTIFS

Les objectifs de ce projet sont nombreux, mais tous rejoignent par un objectif principal : La numérisation de la traçabilité dans la filière pêche.

- traçabilité sans faille
- gain de temps

- respect de la législation
- transparence pour le client
- intégration facile avec les SI (systèmes d'information)
- mise en valeur des produits/acteurs français/de qualité

Les bénéfices du projet à court terme sont principalement dirigés vers les clients directs de l'entreprise HUBLOT, notamment leur fournir une version fonctionnelle de l'outil qui permettra de générer automatiquement leurs étiquettes prix et donc d'améliorer à la fois l'efficacité de leur travail.

L'information au consommateur sera également fiabilisée par l'intermédiaire de ce nouvel affichage. Les erreurs seront ainsi évitées, et les produits de qualité mis en avant à travers l'impression des labels reconnus sur l'étiquette sanitaire directement sur l'étiquetage. Le manager de rayon pourra aussi faire le choix de mettre en avant les produits français à travers un liseré Bleu-Blanc-Rouge autour de l'étiquette.

À moyen/long terme, les retombées du projet bénéficieront à toute la filière des produits de la mer, en assurant une meilleure traçabilité sans impliquer d'investissement temporel supplémentaire pour les différents acteurs. Les consommateurs seront également bénéficiaires de ce projet, car ils auront accès à des données de traçabilité plus précises et transparentes.

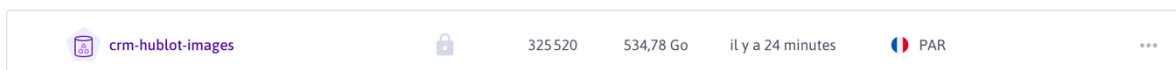
3. RESULTATS

T1 : Collecte de données, reconfiguration du prototype, achats et configuration des serveurs d'entraînement IA

L'objectif étant de limiter les erreurs d'étiquetage en rayon poissonnerie, il fallait apporter des améliorations au prototype afin qu'il soit plus opérationnel. Pour cela une étape indispensable était la collecte de données (via des photos d'étiquettes sanitaires). En effet afin de pouvoir entraîner le système d'intelligence artificielle (abordé en tâche T2), il faut détenir un nombre considérable de photos. Via le prototype, la récolte et le stockage des données nécessaires ont été entamés (en effet plus de 50k photos d'étiquettes de traçabilité, et 10k étiquettes prix au format numérique Hublot).

Avec la crise COVID et les pénuries d'approvisionnement en composants, il a fallu pour Hublot se fournir auprès de fournisseurs proposant du matériel reconditionné.

Grâce à l'achat d'un nouveau PC et la location d'un espace de stockage complémentaire, le nombre de données récoltées a considérablement augmenté puisqu'à ce jour 325k photos ont été collectées. Sachant que pour réaliser un apprentissage efficace, le plafond se situe à environ 200k.



T2 : Phase de développement et d'entraînement de l'intelligence artificielle

La seconde étape indispensable était d'entraîner des systèmes d'intelligence artificielle pour qu'ils soient efficaces.

Des améliorations ont été réalisées à quatre niveaux :

Sur le modèle pré-entraîné CRAFT :

Le modèle de données qui était utilisé dans le système de détection de texte (CRAFT), est un modèle pré-entraîné très générique, et il nécessitait des ajustements.

L'implémentation a été réécrite dans un outil différent (Pytorch -> Pytorch Lightning), ce qui permet d'avoir plus de paramètres. Ainsi, nous avons pu mieux paramétrer l'algorithme d'entraînement afin qu'il soit plus précis et plus performant.

De plus, le code a été amélioré, ce qui permet d'être plus recentré sur le métier.

Concernant l'(ex)Tesseract (OCR) :

Le système de reconnaissance de caractère (Tesseract) a été remplacé par un nouveau système (IA en Pytorch-Lightning, inspiré des travaux de ClovaAI.), plus précis et plus rapide.

Cette intelligence artificielle a dû être entraînée, pour cela trois phases ont été nécessaires :

Phase 1 : Des tests de performance et de taux de reconnaissance ont d'abord été effectués avec le modèle fourni par Clova.

Phase 2 : Ensuite des tests ont été réalisés via l'IDCAR, qui fournit des données de référence (données génériques) pour l'entraînement des IA à la reconnaissance de caractère.

Phase 3 : Finalement des tests ont été réalisés avec des données spécifiques aux problématiques la filière. (Fond de différente couleur, polystyrène, étiquette mouillée etc...)



Deux phases complémentaires sont à prévoir :

Phase 4 : Pour continuer à suivre l'état de l'art autour des technologies de l'OCR.

Phase 5 : Allègement du modèle (qui ne peut tourner actuellement que sur des serveurs spécifiques) et revue de son architecture afin de pouvoir l'utiliser via mobile

Toutes ces améliorations et les nouveaux outils ont été mis en place de manière opérationnelle pour améliorer la performance, le temps de traitement d'une image a été **divisée par quatre**. On constate une grande amélioration de la reconnaissance avec ce nouveau modèle, des améliorations pourront être apportées pour la reconnaissance de certains caractères spéciaux et certaines majuscules.

Le Parser

Il permet d'analyser tout le texte contenu dans une image en cherchant le mot qui nous intéresse.

L'objectif est de trouver des mots malgré les différentes fautes d'orthographe, abréviations, ou imprécisions (nombreuses coquilles, des lettres inversés (L et I), des mots incomplets et sans structure de phrase) de la part de l'OCR. Ce domaine de recherche est assez peu développé (NER Noisy) :

NER (Named entity recognition) : retrouver la/les nature(s) d'un mot

Nom / Verbe / Complément

Poisson / Découpe / Nom Latin / Taille

Lieu / Pays / Zone

Nous avons donc recherché un algorithme capable de faire ce travail, nous avons trouvé un modèle qui semble pertinent sur les données de tests. Cette recherche a pris beaucoup de temps car il existe très peu de travaux sur le sujet.

L'intelligence artificielle a été développée en Pytorch-Lightning, comme le reste du projet, afin d'assurer une cohérence dans le développement.

Les articles consultés et exploités par l'équipe lors du développement sont dans la partie référence de ce rapport.

L'application

L'intelligence artificielle a pu être intégrée à l'application Hublot. Elle figure maintenant comme une fonctionnalité à activer par les utilisateurs s'ils le souhaitent. Cette fonctionnalité a été adoptée par la majorité d'entre eux.

Vidéo explication de démonstration autour de l'utilisation de l'IA dans Hublot : <https://www.youtube.com/watch?v=BdNhJtZ5Op0>

T3 : Mise en test dans les points de vente, retour d'expérience, Évaluation et retours consommateurs / organisme de contrôle

T3.1 Evaluation et retour des utilisateurs

Afin d'évaluer les améliorations apportées au système des tests dans les points de vente étaient nécessaires. Hublot a pu se rendre chez ses clients afin de présenter la nouvelle fonctionnalité (« activer l'IA »).

Axes d'améliorations identifiés sur l'étiquetage en lui-même :

Les retours sont très positifs, quelques axes d'amélioration sont suggérés notamment sur certains produits comme les crevettes roses. En effet, pour les crevettes, certaines informations, notamment relatives aux calibres sont difficiles à retrouver (30/50 vs 40/60 etc). Aussi compte tenu des provenances diverses de ces produits, certaines étiquettes sont difficiles à déchiffrer. Des entraînements complémentaires doivent être réalisés pour améliorer le système sur ces produits « difficiles ».

Axes d'améliorations identifiés sur l'utilisation :

En magasin, l'utilisation est parfois compliquée à cause d'une connexion internet peu stable, il est donc à terme important d'intégrer l'intelligence artificielle directement sur la tablette.

Cette axe d'amélioration a été identifié dans la partie T2 .

Voici un retour utilisateur du E Leclerc Atlantis : <https://youtu.be/bRfsebQsliU>

T3.2 Evaluation et retour des parties prenantes

Le contact avec les instances de contrôle a été assez difficile à réaliser en pratique. En effet, AQUIMER a sollicité plusieurs fois les inspecteurs de la DDPP pour avoir l'occasion de leur présenter les travaux du projet APRI, sans obtenir de retour favorable.

Hublot a pu quand même échanger avec un contrôleur de la DDPP. L'échange a permis de valider l'étiquetage de l'application et la traçabilité électronique. Des échanges ont également eu lieu sur les tailles réglementaires des espèces et les futures évolutions de la réglementation.

T4 : Mise à disposition de l'algorithme, publications OpenSource, configuration et dimensionnement des serveurs de production

La mise à disposition d'une partie du code open-sourcé a été réalisée via la plateforme Github de Hublot. Il reste à développer la documentation pour améliorer la lisibilité du code.

La prochaine étape à réaliser sera de réaliser des communications plus larges et de mettre en place une API (Application Programming Interface) afin que les utilisateurs puissent connecter leur système existant à Hublot.

T5 : Échanges avec les autres acteurs de la filière, retour d'expérience (REX) et présentation

Enfin d'avoir le retour d'expérience des acteurs de la filière, l'application Hublot a été présentée à l'ABAPP (Association Bretonne des Acheteurs des Produits de la Pêche), qui regroupe plusieurs criées Bretonnes. L'ABAPP a trouvé la solution Hublot très intéressante, même s'ils possèdent déjà leur système de traçabilité interne pour l'ensemble des Criées Bretonnes. Des perspectives ont été envisagées par la suite pour élargir à d'autres produits.

Afin d'avoir le retour des mareyeurs et dans le but d'étendre le champ d'application du système Hublot, le projet APRI doit être présenté à l'UMF, Peter Samson, avait déjà entendu parler du logiciel et était très intéressé pour avoir une présentation plus approfondie, une visio doit être organisée ces prochains mois.

Le système a été présenté à l'entreprise Chaluts des 2 ports (mareyeur). Leur problématique pourrait être résolue par l'algorithme développé par Hublot mais elle nécessite un investissement conséquent qui n'est pas réalisable pour l'instant d'un point de vue humain chez Hublot et d'un point de vue financier chez le mareyeur.

Le projet APRI et ses développements ont été présentés à la centrale Ouest E-Leclerc et la solution a été déployée dans la moitié des E-Leclerc de la Scaouest en 2021.

Une diffusion plus large des résultats du projet APRI est prévue par AQUIMER courant d'année 2022.

CONCLUSION ET PERSPECTIVES

Le projet APRI a permis des avancées significatives en termes d'intégration de l'intelligence artificielle pour la reconnaissance d'étiquettes.

En effet, d'un point de vue technique, l'entraînement et l'amélioration de l'intelligence artificielle a permis d'améliorer significativement la reconnaissance d'étiquettes et permet donc un gain de temps important pour les chefs de rayon.

De plus la nouvelle fonctionnalité 'activer l'intelligence artificielle' fait l'unanimité auprès des utilisateurs du logiciel Hublot car elle permet de réduire considérablement le temps affecté à l'élaboration des étiquettes prix. Les présentations aux divers acteurs, et notamment à d'autres maillons de la filière (acheteurs, mareyeurs) montrent une réelle reproductibilité de la solution dans d'autres domaines, ce qui confirme les perspectives identifiées au dépôt du projet.

Dans l'immédiat, les travaux concernant l'amélioration du système vont être poursuivis (notamment l'allègement du système pour voir utiliser l'intelligence artificielle directement sur la tablette, continuer l'entraînement de l'intelligence sur des produits spécifiques ou « difficiles »).

La diffusion des résultats du projet va se poursuivre par le biais de communications ciblées.

A plus long terme, le projet APRI a permis de confirmer le potentiel de la solution pour le mareyage et d'autres maillons de la filière.

REFERENCES

Liste des papiers de recherches exploités par l'équipe lors du développement (NER NOISY):

<https://ash-sunday-4f4.notion.site/86773a9fb18c411a9c8a55cb9cf283a5?v=f6a0894158634f4098e5116aba50d173>