ensiie
école nationale supérieure d'informatique
pour l'industrie et l'entreprise

Rapport
Projet Entrepreneurial
Running & Objets connectés

STAMINA

5 juillet 2019 - 13 septembre 2019

Kevin XU

Autres membres :
Bihe DONG
Eric WANG
Clément VEYSSIERE

Tuteur
Nicolas Brunel

Coordonnateur de projet
Sébastien Cauwet

Résumé

Dans ce rapport se trouve la synthèse du travail accompli pour le projet de start-up STAMINA. Ce projet d'une durée deux mois entre le 5 juillet et 13 septembre 2018 a été réalisé dans le cadre de la validation du cursus d'ingénieur au sein de l'ENSIIE. L'équipe ayant travaillé sur ce projet fut composé de 4 étudiants suivant le parcours Mathématiques appliquées.

Notre start-up Stamina a pour objectif d'apporter une nouvelle forme de coaching pour la course à pied. En exploitant les données récoltées par les objets connectés que les coureurs portent, nous voulons créer une application de coaching en temps réel.

Nous avons donc travaillé sur toute la partie commerciale de ce projet. Et en parallèle, nous avons travaillé sur l'élaboration de l'algorithme principale de notre application qui est la suggestion de vitesse en temps réel grâce à de l'apprentissage par renforcement.

Mots-clés

- **Sport**
- **Running/Course à pied**
- **Marathon**
- **Coureur amateur**
- **Objets connectés**
- **Coaching virtuel**
- **Temps réel**
- **Canvas**
- **Machine learning**
- **Big data**
- **Clustering**
- **Apprentissage par renforcement**
- **Processus de décision Markovien**
- **Intelligence artificielle**
- **Web scraping**

Remerciements

Tout d'abord, nous souhaitons commencer ce rapport par remercier comme il se doit les personnes qui ont contribué à ce projet.

Nous voudrions remercier en premier lieu notre tuteur Nicolas Brunel. Nous lui sommes très reconnaissant de nous avoir proposé ce projet et de nous avoir fait confiance pour le mener à bien. Il a toujours été à l'écoute pour nous aiguiller sur la partie technique du projet.

Ensuite, nous remercions bien sûr Sébastien Cauwet de nous avoir accepter pour le programme d'été de pré-incubation Sun Heat IMT Starter. Nous tenons à remercier toute l'équipe de l'IMT Starter qui nous ont accueillis avec beaucoup d'enthousiasme. De même, tous les coachs et intervenants lors des ateliers nous ont été d'une grande aide pour l'avancé du projet. Nous leur exprimons notre plus grande gratitude pour nous avoir permis de travailler dans une ambiance très chaleureuse et amicale.

Enfin, nous remercions toutes les personnes qui ont accepté de nous accorder des interviews pour partager leurs conseils et leurs expériences dans la course à pied. En particulier, nous sommes très reconnaissant envers Morhad Amdouni d'avoir accepté de nous rencontrer malgré son emploi du temps très chargé.

Table des matières

1. Introduction

a. Le besoin

La large diffusion des objets connectés dédiés au « running » a fait émerger de fortes attentes d'accompagnement haut de gamme chez les amateurs et les professionnels. Cependant, la compréhension des facteurs limitants de l'endurance ne permet pas de proposer automatiquement des stratégies de course optimale pour parcourir une distance donnée en un temps minimum sans risque de fatigue précoce.

Actuellement, les coureurs procèdent encore par essais et erreur en optant très souvent pour le maintien d'une fréquence cardiaque ou vitesse constante ce qui a été démontré comme physiologiquement non optimal en termes de santé et performance. Une autre voie est le recueil de conseils sur des sites internet ou par d'autres coureurs, qui se trouvent souvent ne pas être totalement adéquat à la personne.

b. L'idée solution

Le contexte est le suivant : La force musculaire, l'efficacité cardiaque, la résistance à l'acidose, ou encore la perception et la gestion de la douleur sont de nombreux facteurs physiologiques influant la performance du coureur. La physiologie décrit les mécanismes à l'œuvre durant un effort physique, notamment les mécanismes aérobique et anaérobique de production d'énergie qui sont déterminants dans les épreuves d'endurance, comme la course. Chaque athlète part avec un réservoir d'énergie dont la consommation ne pourra être que partiellement remplacée par la nutrition et l'hydratation.

La solution : un système d'Intelligence artificielle basé sur une modélisation stochastique par Processus de Décision Markovien et l'apprentissage par renforcement pour construire des stratégies personnalisées, adaptées au type de coureur et de course, à partir des données en temps réel disponibles grâce aux montres connectées.

2. Environnement de travail et équipe

Au début de notre projet, nous avons commencé par travailler à la BPI (Bibliothèque Publique d'Information) dans le Centre Pompidou à Paris. Ensuite, nous avons rapidement été incubé au sein de l'incubateur IMT Starter situé sur le campus de l'école Télécom SudParis.

a. IMT Starter

L'incubateur IMT Starter accompagne les meilleurs start-ups numériques. Il offre des locaux dans lesquels les jeunes start-ups comme la nôtre peuvent se développer. Pendant nos deux mois au sein de cette structure, nous avons pu profiter d'un suivi de projet et de nombreux ateliers par des intervenants extérieurs expérimentés dans le monde de la start-up.

Cette incubation a été primordiale dans l'avancée de notre projet. En effet, l'accompagnement et toutes les aides qu'on a pu bénéficier nous a permis de structurer notre projet et de cultiver le développement de nouvelles idées.

b. L'équipe

L'équipe de notre start-up est composée de jeunes étudiants ingénieurs de l'ENSIIE. C'est une petite troupe qui se connaissait déjà depuis un certain temps avec un parcours académique commun par un BAC S et une classe préparatoire MP ou PSI, arborée de différentes qualités complémentaires. De plus, trois d'entre nous allons suivre un master M2 en Data Science/Machine Learning à partir de la rentrée prochaine, ce qui est une force de notre équipe. La rigueur face à l'adversité propulsée par une curiosité intellectuelle construit le chemin dessiné grâce à un intérêt général pour la technologie. Cependant, nous manquons de compétences pour la partie commerciale du projet.

De plus, la culture et la pratique du sport sont ancrées et variées au sein de l'équipe : badminton, basket, tennis, handball, running.

Voici les rôles officiels de chaque membre de l'équipe dans la start-up :

- Kevin XU : CEO
- Clément VEYSSIERE : CTO
- Eric WANG : CMO
- Bihe DONG : Business Developper

c. Outils de travail et communication

Au sein de l'incubateur, nous avons travaillé tous les jours dans un open space avec nos propres ordinateurs portables.

La suite bureautique de Google **G Suite** a été notre principale outil de création des différents rapports et présentations. En effet, la possibilité d'édition à plusieurs des documents a été primordial.

Pour l'organisation des différentes tâches de notre projet, nous avons utilisé l'outil de gestion de projet **Trello**. Nous avions séparer les tâches en deux catégories principales : **la partie business et la partie technique**.

Pour la communication au sein de l'équipe, on avions un groupe sur l'application Messenger de Facebook. Puisque, nous travaillions tous dans la même salle, nous communiquions essentiellement à l'oral.

3. Développement commercial

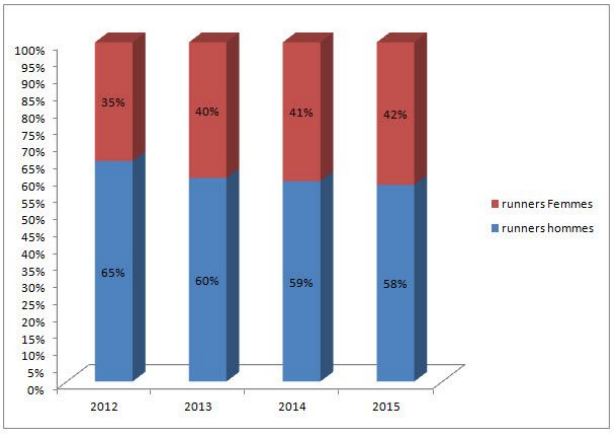
a. Etude de marché

i. Le marché du running

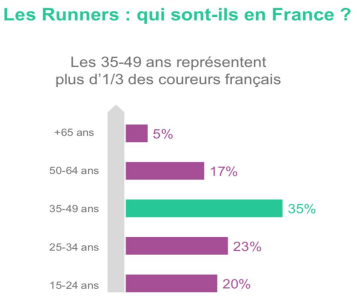
En 1979, la Fédération française d'athlétisme recensait moins de 500 coureurs, ce nombre connaît une augmentation fulgurante pour atteindre 3 millions en 2000. Aujourd'hui, on compte près de 13 millions de runners réguliers, la simple course à pied est devenue un vrai phénomène social. Par ailleurs, ces chiffres sont en constante progression avec un ratio de débutant élevé. Ainsi, le marché du running a encore de beau jour devant lui néanmoins parmi ces 13 millions de Français, qui sont-ils réellement ?

ii. Qui sont les coureurs ?

Tout d'abord, le running est une pratique qui se féminise, il y a de plus de plus de coureuses débutantes que de coureurs même si au total les hommes restent majoritaires. De plus, selon Running Expo, 62% des runneurs sont des CSP+.



Ensuite, la course est un sport **intergénérationnel** touchant l'ensemble de la population, jeune ou plus âgée. Le coureur moyen est donc âgé de 39 en 2012 et 35 ans en 2014. De plus, les trois principales motivations invoquées par les runners dans l'ordre sont : **améliorer sa condition physique, se défouler, perdre du poids.**



iii. Les objets connectées et le coureur

La simple pratique du jogging a vu son nom évolué en “running” mais ce changement s’accompagne aussi d’une révolution technologique, ainsi les “runneurs” sont devenus de vrais technophiles et s’équipent de plus en plus pour optimiser leur session d'entraînement. On recense 73% de coureurs qui portent des objets connectés dans les salons de running. De ce fait, la numérisation de la course constitue un laboratoire technologique pour les marques qui restent à l’affût pour créer de nouveaux gadgets pour améliorer les performances de ses usagers.

Cette attraction liée au gadget, au tout connecté participe au succès actuel du running, elle devient même un outil de motivation et ajoute un côté ludique à la course qui dans le fond n’a jamais changé. En fin de compte, c’est plus de “60% qui courent avec un objet connecté” d’après M.Mignon. Un runner tel un religieux, lui faut la dernière technologie.

iv. Segmentation du marché

Malgré un large marché, le niveau d’un coureur à un autre est assez hétérogène, il est donc simple de regrouper les runners, par exemple lors des départs groupés suivant leur VMA durant de marathons. Nous allons séparer le marché potentiel en trois catégories selon leurs besoins : le débutant, l’amateur et le semi-pro voire professionnel.

Le débutant

Ce groupe n’est pas le coeur de nos cibles mais constitue le plus grand nombre d’utilisateurs possibles : environ 80%. Ce segment inclut tous les novices notamment citadins ou sédentaires, du jeune au âgé qui souhaite reprendre une activité sportive. Nous visons à encadrer ces coureurs occasionnels (une course par semaine au moins). Au fur et à mesure de leur progression, on leur poussera à souscrire à un abonnement premium leur permettant de continuer à évoluer.

L’amateur

L’amateur est un coureur davantage aguerri et expérimenté, il a pour objectif de performer et s'entraîne plusieurs fois par semaine : 20% des utilisateurs possibles. Il a besoin d’outils davantage poussés et se tournera ainsi vers la formule premium.

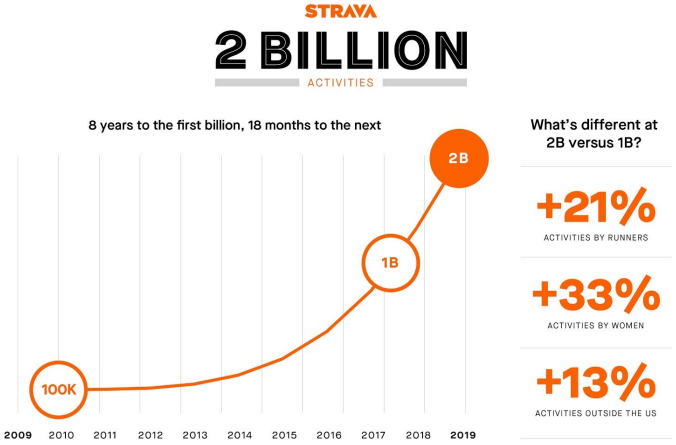
Le pro

Le coureur professionnel spécialiste des longues distances possède déjà à sa disposition coach sportif et médecin, toutefois l’indicateur de vitesse en temps réel pourrait l’intéresser. Néanmoins, les professionnels ne représenteront qu’une infime partie du marché visé.

v. Concurrence (directe et indirecte)

Cette analyse a pour but de recenser les principaux concurrents, et d'expliquer leur succès. Dans le marché très compétitif du running, parmi les nombreuses applications, seule une poignée s'offre la globalité du marché dont les plus notable sont : Strava, Runtastic, Nike Run Club. Intéressons-nous sur ces trois mastodontes du coaching virtuelle pour en dégager des directives à suivre.

Strava a été créé en 2009 et compte aujourd'hui plus de 1,2 millions d'utilisateurs réguliers et 200 000 utilisateurs premiums ainsi on estime un revenue annuel de 12 millions $ en 2016. De ce fait, il propose un abonnement à 59.99€/an.



Strava a toujours eu une base de coureurs actifs mais ce n'est que récemment que l'utilisation de l'app à littéralement explosé.

Elle est adapté aussi bien aux débutants qu'aux amateurs grâce à ses analyses détaillés de chaque session de course, les coureurs peuvent même se défier sur des même segments de course. Par ailleurs *Strava* s'apparente à un réel réseau social, où chacun possède un profil et peut amasser des followers, puis donner des "kudos" (leur "like"). Par conséquence, cela pousse le runner a alimenté sa page et mettre à jour son statut pour mettre en avant ses progressions et les challenges qu'il déverrouille au fur et à mesure. Ainsi, l'utilisateur peut partager toutes ses courses et continue d'utiliser *Strava*.

Runtastic a été créé en 2009 et compte 147 millions de membres enregistrés à ce jour, on estime à 11

millions $ de chiffres d'affaires annuels. Il possède une partie gratuite et une autre partie premium payante à partir de 59.99€/an. Ils ont attendu 2 an pour acquérir leur premier million d'utilisateurs, c'est qu'après qu'ils lancèrent leur première version premium de l'app. *Runtastic* se sert de *Facebook* comme un tremplin pour accroître leur marché, ainsi 30% des usagers se connectent sur l'app via leur profil *Facebook* et 20% partagent leur activité sur leur *Facebook*. Finalement, *Runtastic* a été racheté par firme *Adidas* en 2015 pour la modique somme de 220 millions €.

Contrairement aux applications précédentes, *Nike Run Club* est totalement gratuite, c'est un produit de Nike pour consolider leur image de marque auprès des coureurs, l'application génèrent du revenue qu'indirectement. *NRC* sert notamment de plateforme pour promouvoir les produits de running de Nike et leur événement sponsorisé. L'application a été téléchargé plus de 100 millions de fois, notamment grâce aux adeptes déjà existantes de la marque à virgule. Finalement, c'est une vraie vitrine publicitaire pour la marque *Nike*.

Dans les concurrents indirects, il est à noter que le coaching sportif se développe petit à petit mais reste un marché difficilement mesurable. De ce fait, des coureurs considèrent les applications comme pas assez suffisant et se tournent vers un coaching davantage personnalisé. On compte que les femmes ont davantage recours à des coachs. Toutefois, le prix est relativement élevé, environ 50€ une séance.

On remarque que la grande force qui constitue l'avantage compétitif des concurrents se repose sur leur base d'utilisateurs faramineuse. De même que leur programme de course à pieds intégré dans leur application possède un caractère universel susceptible de capter le plus de public possible allant du débutant, en passant par l'amateur, voire le professionnel. Toutefois, c'est aussi un handicap, du fait qu'ils proposent des programmes d'entraînement très générique, certaines personnes souhaitent davantage de personnalisation.

Ainsi, Stamina se démarquera par leur programme utilisant l'intelligence artificielle qui est une fonctionnalité inédite sur le marché. Néanmoins, il subsiste toujours une crainte que le département de recherche et développement des grandes firmes mette aussi au point une fonctionnalité similaire.

b. Les maquettes

Nous avons rapidement commencé à créer des maquettes de l'application. Après avoir élaboré un cahier des charges pour lister les fonctionnalités, ces maquettes permettent d'avoir une idée à quoi ressemblerait l'application qu'on voudrait commercialiser.

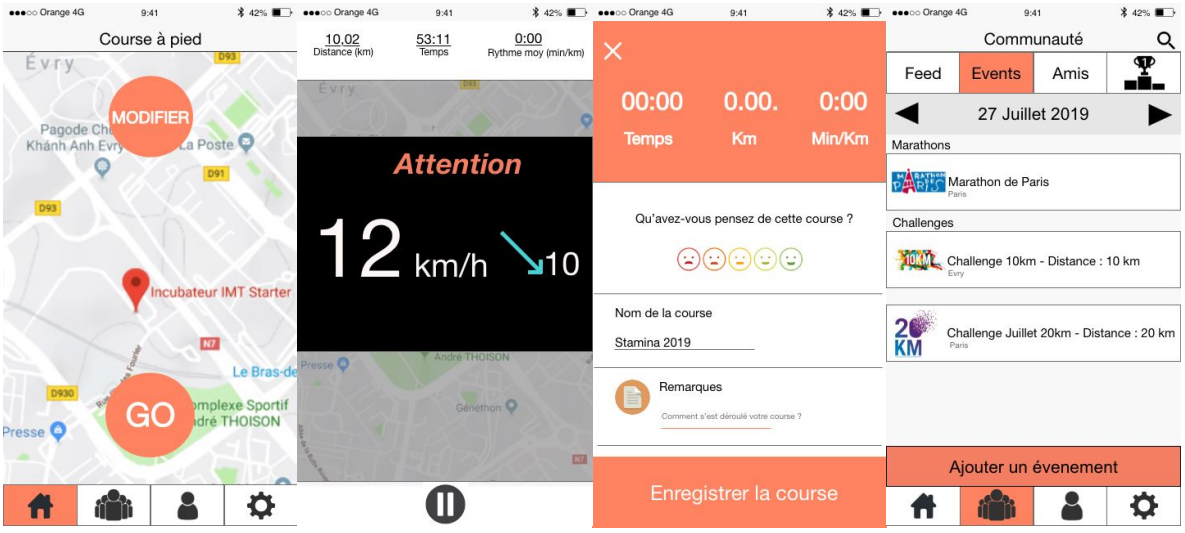
Pour réaliser ces maquettes, nous avons utilisé le logiciel de création de maquette Adobe XD qui est totalement gratuit.

Grâce à ce logiciel, nous avons les maquettes de l'interface sur la montre :

À côté de la vitesse de course actuelle, on remarque ici la petite flèche rouge dirigée vers le haut. Cette flèche signifie que le coureur doit accélérer pour atteindre la vitesse de 16 km/h.
Il y a aussi le temps écoulé depuis le début de la course et le nombre de kilomètres parcouru.

Ici, nous avons une flèche bleue qui indique au coureur de ralentir afin d'atteindre la vitesse de 7 km/h.

Ensuite, pour les maquettes de l'application sur smartphone, nous nous sommes beaucoup inspirés des applications déjà existantes dédiées au running.



Les trois premières maquettes sont les interfaces qu'un coureur verrait s'il veut lancer une session de course, durant une session et enfin l'enregistrement de la course.

Ces maquettes ne sont vraiment que des ébauches préliminaires de l'application. Grâce aux interviews, nous avons pu recueillir beaucoup d'avis pour améliorer ces maquettes.

c. Interviews

Afin d'obtenir plus de vision sur l'état du marché, les intérêts et les soucis des coureurs, nous nous adonnons à l'exercice de l'interview dans le but de fournir un outil adapté aux besoins des clients. Nos interviews se sont opérés sur plusieurs plans pour obtenir une plus grande diversité d'avis :

- ➔ interviews sur les pistes de course
- ➔ dans les magasins de sport (Décathlon, Go Sport)
- ➔ en ligne sur les groupes Facebook de running
- ➔ démarchage d'interviews auprès d'athlètes.

i. Les interviews sur les pistes de course

Dans un premier temps, nous sommes directement partis sur les lieux de pratique de course à pied. Les pistes de course à Paris sont beaucoup fréquentées par des coureurs amateurs ou professionnels. Nous avons alors réussi à obtenir plusieurs interviews de coureurs avant ou après leur séance.

Il en ressort que les coureurs sont globalement divisés en deux. Il y a ceux qui courent pour la forme, la santé, maigrir et puis d'autres qui sont orientés vers la performance.
Les premiers ont souvent des courses simples et monotones, avec ou sans smartphone pour les accompagner dans leur course. Ils seraient au plus intéressés par une application de course gratuite sur leur smartphone, potentiellement quelques conseils sur leur course puisqu'ils organisent eux-même leur plan de course.
Les derniers ont des accessoires plus pratiques (smartwatch) et sont plus demandants de conseils d'amélioration et de prévention. Avoir une application est un must pour le suivi de leurs courses et de leurs activités. Certains font partis de groupes de running ou sont/étaient inscrits dans un club d'athlétisme.

ii. Les interviews en grande surface

Les distributions spécialisées dans les équipements de sport comme Décathlon et Go Sport sont des lieux où la plupart des sportifs de tous niveaux se retrouvent. Nous étions donc parti à la recherche de clients potentiels dans les rayons dédiés à la course à pied.

Les gens que de nous avons rencontré dans ces lieux courent pour la majorité pour la forme et sont pour la plupart des coureurs occasionnels et courent à peine une fois par semaine. On retiendra surtout qu'une infime partie utilise des applications de running et encore moins de personnes portaient des montres connectées.

iii. Les interviews sur les groupes Facebook

Les coureurs font très souvent partis de communauté comme les groupes Facebook. On retrouve alors des groupes comme *Communauté Garmin Connect* qui regroupent des utilisateurs de montres Garmin. Le groupe avec le plus de membres à son actif *Course a pied "Esprit running"* compte plus de 12 000 membres. Ces personnes sont dans ces groupes pour partager leur passion pour le running et donner/recevoir des conseils.

Nous avions donc commencé dès le début du projet à contacter des coureurs sur Facebook. Contrairement aux personnes rencontrées dans les stades ou en magasins, ces personnes ont surtout pour objectif la performance. En outre, elles possèdent majoritairement des montres connectées.
Certains coureurs sont inscrits dans des clubs et ont donc des coachs pour les aider directement. Les autres se basent sur leurs ressentis et en analysant leurs performances à partir des données enregistrées par des appareils pour établir des programmes d'entraînement détaillés.
Les personnes interviewées ont souvent l'intention de participer à des compétitions ou ont même déjà plusieurs courses à leur actif. Par ailleurs, ces coureurs pratiquent souvent d'autres sports en dehors de la course à pied.

En conclusion, les interviews sur Facebook nous ont permis d'avoir une vision assez large de l'univers de la course à pied.

iv. Morhad Amdouni

Nous avons aussi cherché à recueillir l'avis de coureurs professionnels, ainsi nous avons contacté plusieurs pros avec plus ou moins de succès, finalement Morhad Amdouni a accepté de nous recevoir et nous a proposé une entrevue dans un café aux Champs-Elysées le 14 août.

Tout d'abord, nous lui avons questionné sur la nature et fréquence de ses sessions d'entraînement. Il en ressort qu'il a toute une équipe derrière lui pour suivre son état physique et ses progressions. Par ailleurs, la VO2 max joue un rôle essentiel dans les performances du coureur mais que la motivation et la confiance en ses propres capacités sont primordiales pour finir de longue course. En effet, le mental est un facteur important et difficilement mesurable qui influe grandement sur un exercice aussi exigeant que le marathon, de ce fait il faut croire en ses ressources pour continuer d'avancer.

Ensuite, nous nous sommes concentrés sur l'aspect software et hardware, c'est-à-dire s'il avait recours à des applications de running ou autre objets connectés. Ainsi, il avait une montre Garmin au poignet qu'il s'en servait seulement pour consulter son allure, quant aux applications, il n'en utilisait aucune. De son point de vue d'expert, il ne ressentait pas le besoin de ce genre de gadget, et notre fonctionnalité d'IA laissa un avis assez mitigé aussi.

Enfin, nous avons présenté notre maquette de l'application, il nous a donné plusieurs points à améliorer. Il fallait mettre en avant notamment le critère de posture du corps car le positionnement du corps sur la durée agit fortement sur l'endurance du coureur. La prévention de blessure est aussi un enjeu majeur pour le bien être du coureur.

Finalement, il était très intéressé par notre projet et souhaite intégrer notre équipe à apporter son expertise si le projet devient concret.

d. Stratégie marketing

Dans l'optique d'avoir des finances durables à partir du business model précédent, nous élaborons les actions marketing suivantes :

- Page web vitrine de présentation avec une vidéo et recueil d'emails des intéressés pour un test d'une version bêta
- Présence aux courses et tout autres événements autour du running (gratuit et renforce le contact clientèle)

- **Facebook :** Feed Facebook nourri par nous-mêmes au départ puis par les utilisateurs pour forger une communauté (courses, conseils, …)

Voici un schéma présentant le parcours marketing

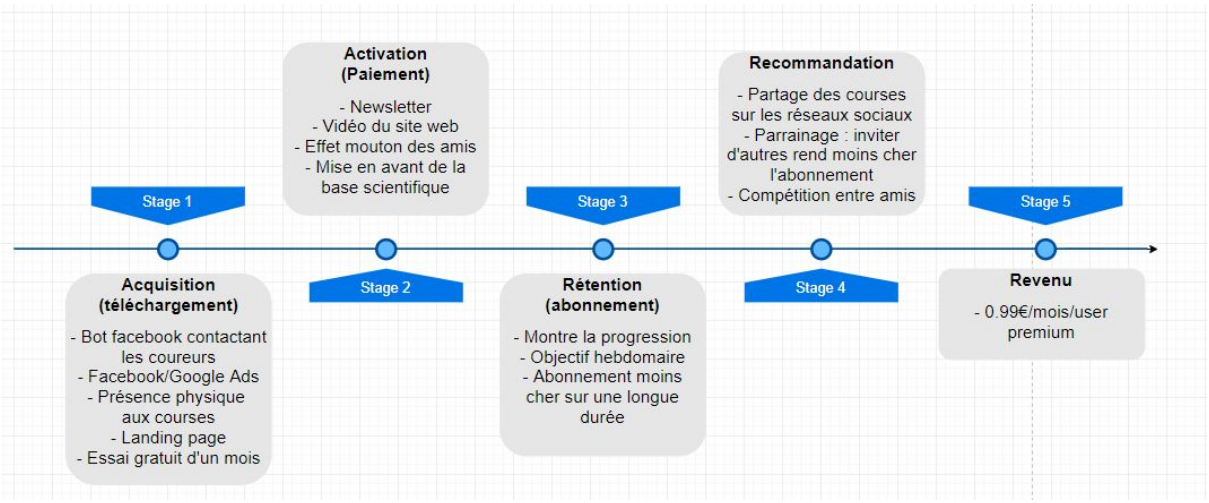


Fig x. AARRR de la stratégie marketing

4. Développement technique

a. Modélisation générale du problème

La fonctionnalité principale de notre application est le coaching virtuel en temps réel. Nous voulons aider un coureur à courir de manière optimale pour qu'il utilise correctement ses ressources sur une distance donnée.

Lors d'une session de course, un coureur passe par différents états physiologiques qui se peuvent être caractérisés par les facteurs physiologiques comme la fréquence cardiaque, la vitesse, le nombre de pas par minute, etc … Ces données recueillies en temps réel grâce aux montres connectées permettent une analyse relativement poussée de l'état physique d'un coureur. C'est pour cela que dans un premier temps, nous devons trouver le bon nombre d'états.

Ensuite, la solution que nous voudrions proposer : un système d'Intelligence artificielle basé sur une modélisation stochastique par Processus de Décision Markovien et l'apprentissage par renforcement pour construire des stratégies personnalisées, adaptées au type de coureur et de course. Cette méthode a été révélée efficace par les travaux de polytechniciens sur le football. Cet ensemble de concepts sera expliqué plus tard.

b. Outils

Nous avons utilisé divers outils lors du développement technique, aussi bien des outils hardwares que software.

Du côté hardware, nous avons eu la chance de pouvoir obtenir deux montres et deux ceintures garmin financées par le CNRS pour la durée de notre projet, avec qui M. Brunel est en contact.
Ces objets nous ont permis de voir les possibilités proposées par les applications existantes et de comprendre mieux les besoins d'un coureur au moment de sa course.
Pour le développement de l'algorithme, nous avons utilisé les langages Python et R.

c. Scraping des données

Pour entraîner notre modèle, nous avions besoin des données captées par les objets connectés dédiés au sport.
Nous avons pu récupérer ces données de deux manières différentes :

- À partir de nos propres courses en utilisant le matériel fournis par M. Brunel
- En minant les données de courses de personnes les publiant sur des application de running

Pour pouvoir travailler sur notre modèle algorithmique avec un nombre conséquent de données, nous avons préféré scraper des courses publiées en ligne.
Pour ce faire, nous avons utilisé le module Selenium de Python et nous nous sommes concentrés sur le site Strava.

Nous avons donc conçu un algorithme de webscraping permettant de récupérer toute les courses d'au moins 5 km d'un utilisateur de Strava sur une année.

Lors du lancement du script python, il est demandé de rentrer l'id du coureur dont on veut récupérer les courses et l'année ciblée. Ensuite, grâce aux fonctions de Selenium, le script lance un navigateur web, se connecte à *Strava* via un compte créé spécialement pour l'occasion, navigue jusqu'à la page de l'utilisateur cible et liste les liens menant aux pages des courses de l'année choisie. Finalement, pour chaque course, le script vérifie que les données telles que le rythme cardiaque ou la cadence sont bien présentes puis enregistre ces données dans un fichier csv qui sera rangé dans un dossier nommé selon l'id du coureur.

d. Préparation des données

i. Nettoyage

Les données récupérées sont sous format de caractères, inexploitables par des programmes standards de visualisation ou de modélisation. Le nettoyage d'une course consiste dans notre cas à enlever les unités et en transformant les chiffres sous format caractères en format numérique.

Fig : Screenshot du scraping

Cela peut se faire tout de suite après le scraping sur le site de *Strava* ou après coup, indépendamment du scraping. Compte tenu que l'opération de scraping dure déjà assez longtemps (~2min par course sur un historique de course annuel) et qu'il peut y avoir des erreurs, le choix retenu est de nettoyer les données en dehors du scraping. Les NA sont simplement enlevés. Les données étant continues, il y a très peu d'information perdue.
A mesure que le scraping se scale, le nettoyage doit suivre la tendance pour maintenir une efficacité optimale. Ce qui se faisait au départ sur une course doit alors se faire sur un ensemble de fichiers de course et enfin sur un ensemble de dossiers d'individus comprenant plusieurs courses.

ii. Segmentation des courses

Pour simplifier la modélisation, nous avons segmenté les sessions de courses. Avec les données nettoyées de *Strava*, nous avons regroupé tous les relevés de la course sur chaque

kilomètre. Ce découpage par tronçon de 1 km permet de réduire drastiquement la quantité de données d'une course et donc de pouvoir les manipuler plus aisément.

Pour chaque tronçon, nous avons pris les moyennes de chaque variable. Par exemple, une course d'une distance de 10 km contient au alentour de 300 relevés de données. Cette course devient alors après découpage une course de 10 échantillons.

Voici ici par exemple le tableau contenant les dix tronçons de la le session 106 :

session	distance	heart_rate	cadence_running	speed	elevation
106	461.5385	100.6154	172.7692	388.0385	4.846154
106	1427.5862	111.5172	169.9310	342.0690	5.862069
106	2425.0000	111.1786	170.8571	326.7500	5.964286
106	3427.5862	109.8621	170.5517	319.2414	6.103448
106	4425.0000	106.1786	170.0000	323.6071	6.000000
106	5424.1379	107.8621	170.5517	322.0345	6.034483
106	6437.9310	116.6552	170.3448	325.7586	6.206897
106	7435.7143	112.1071	169.3571	335.0357	6.142857
106	8421.4286	110.3929	170.0714	333.4643	6.214286
106	9280.9524	115.2381	168.9524	343.3333	5.666667

e. Clustering

Après avoir découpé les sessions de courses en tronçons de 1 km, il faut classifier ces tronçons en terme d'état physiologique. On veut trouver les tronçons pour lesquels le coureur a un état physiologique commun.

Tout d'abord, nous avons décidé de prendre une centaine de courses pour un seul coureur donné. En effet, les états physiologiques des humains sont très variables d'individu en individu. De fait, cela nous a été semblé plus raisonnable de commencer par trouver les états d'une personne.

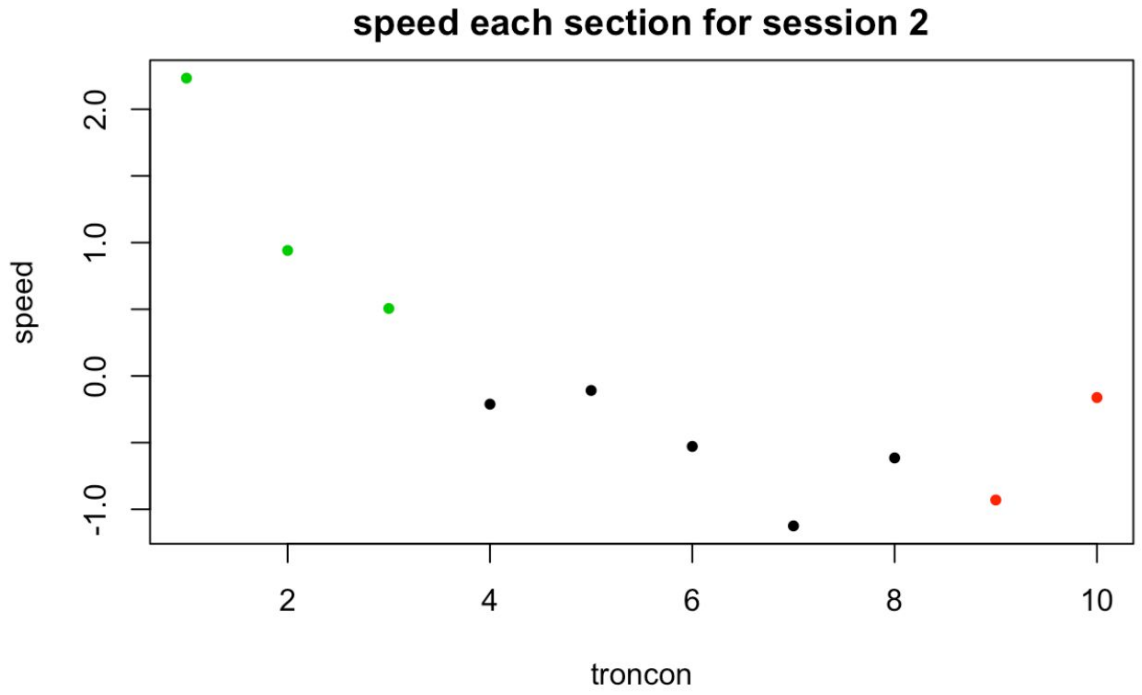
Pour cela, nous avons regroupé tous les tronçons obtenu après la segmentation de chaque course. Puis nous avons cherché à caractériser les états correspondant à chaque section.

i. La méthode des k-means

Le partitionnement en k-means permet de diviser un ensemble de points en k groupes en utilisant la distance Euclidienne. Cette méthode d'apprentissage non supervisé est l'une des plus basiques mais efficace. Les données sont évidemment normalisées avant leur traitement par l'algorithme de partitionnement.

Dans un premier temps, nous avons classifié les différents tronçons en n'incluant pas la variable *distance*. Cependant, cette variable a une importance assez primordiale dans la caractérisation de l'état de fatigue d'une personne. En effet, lorsque cette variable est assez faible au début, le coureur aura normalement plus d'énergie à dépenser.

Voici le résultat du clustering[25] qu'on obtient sur une des session de course de 10 km d'un coureur :



On peut distinguer les différents états physiologiques par les trois couleurs. Ici, on voit les vitesses moyennes sur chaque segment de la course de 10 km.

ii. La recherche du meilleur nombre d'états

Pour déterminer le meilleur nombre de clusters, nous avons eu recours à la méthode du coude ou en anglais *Elbow method.* Ainsi, sous R, nous avons implémenté la fonction ***bestK,*** qui calcule pour chaque cluster leur "The total within-cluster sum of square (WSS)". Par ailleurs, on considère la droite passant par le premier WSS et le dernier WSS, et dès la distance entre la droite et WSS est inférieure au précédent WSS, on retourne le nombre de cluster correspond ce WSS. On trouve ainsi le meilleur nombre d'états.

Grâce à cette méthode, on trouve un total de 3 classes pour la méthode des k-means.

f. Modèle (RL et MDP)

Étant majoritairement étudiants en mathématiques appliquées, le principe du reinforcement learning (RL ou apprentissage par renforcement en français) nous est connu comme une des branches du machine learning (intelligence artificielle). Quant au Markovian Decision Process (MDP ou Processus de décision markovien), c'est un concept qui est proche des chaînes de Markov étudié dans le cursus ingénieur de l'ENSIIE. Le défi de taille est la compréhension de ces concepts et surtout leur implémentation dans le cadre de notre projet. Ainsi une majeure partie du temps à nos débuts consiste en une phase d'apprentissage. Rapidement, voici une présentation des deux fondations de notre service.

- Le RL est de la programmation dynamique qui entraîne un algorithme reposant sur un système de récompenses (potentiellement négatives). Il apprend en interagissant avec son environnement ici posé par le MDP.
- Un Processus de Décision Markovien est composé d'un ensemble d'états et d'actions sur lesquelles se forment des probabilités de transition (d'un état à un autre) et de récompenses.

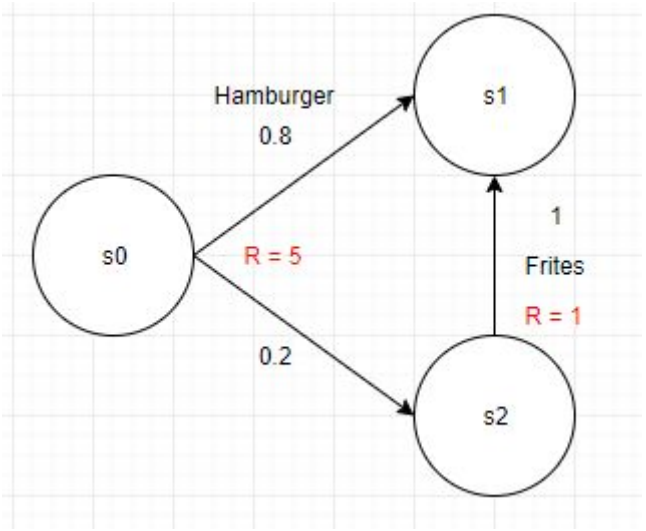


Fig x. Exemple de Processus de Décision Markovien
s0,s1,s2 = états , Nom = Action , Nombre = probabilité, R = Récompense

Ici la matrice de transition serait la suivante : $\begin{pmatrix} 0 & 0.8 & 0.2 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

Les lignes correspondent aux états de départs et les colonnes aux états d'arrivées.

Trois modèles sont réalisés. Rappelons que le but est de **résoudre le processus de décision markovien.** Nous allons voir différentes approches.

Premièrement, nous modélisons le processus (états, actions, récompenses, probabilités).

- **Les états** (caractérisés par : la fréquence cardiaque, la vitesse, #pas par minutes) sont déterminés par le clustering précédent sur les données de courses grâce au k-means. **Un état virtuel est ajouté** pour caractériser l'état "au bout du rouleau". Virtuel car les coureurs s'arrêtent avant d'y arriver. Ce serait l'état dans lequel il ne faudrait surtout pas tomber.
- **Les actions** sont représentées par la vitesse moyenne du tronçon. Toutes les vitesses des coureurs ne sont pas listées car il y en aurait énormément. Par le tracé d'histogramme des vitesses de plusieurs personnes, on remarque un schéma répétitif qui montre des pics sur 4-5 vitesses (⟶ le nombre d'actions et leur valeur)

- La récompense est calculée par l'opposé du temps nécessaire pour parcourir le tronçon. En ce sens, plus le tronçon est parcouru rapidement, plus la récompense est élevée. Un temps de 500 secondes procurerait une récompense de -500 et un temps de 1000 sec donnerait -1000. Le but est de maximiser la récompense totale. Une alternative serait de prendre l'inverse du temps.
- Les probabilités de transition ont fait l'objet de discussion sur la méthode pour les construire en collant le plus à la réalité.

i. Premier modèle : Package ReinforcementLearning sur R

Un premier modèle a été réalisé grâce au package R *ReinforcementLearning* (par Nicolas Pröllochs).

Nous avons tout d'abord effectué un premier modèle à partir de données factices et simplifiées pour tester la faisabilité d'un tel algorithme. Nous avons donc commencé avec trois vitesses disponibles et cinq états de fatigue du coureur. Nous avons créé les trois matrices de transitions d'états pour chaque vitesse et fabriqué un environnement à partir de ces dernières pour pouvoir construire le modèle en assignant des récompenses à chaque fin de tronçons en fonction de la vitesse et en pénalisant le fait de terminer un tronçon dans le cinquième état de fatigue, l'état interdit correspondant au cas où le coureur n'est plus capable de continuer sa course.

Après avoir réussi à obtenir des résultats encourageants grâce aux données simplifiées, nous sommes passés aux données de vrais coureurs obtenues depuis Strava.

ii. Deuxième modèle : Programmation dynamique à horizon fini

Sachant que le coureur sait à l'avance la distance à parcourir et qu'il progresse sans faire demi-tour, un peu de recherche nous amène vers la programmation dynamique à horizon fini. A l'instar des algorithmes de résolution de MDP couramment utilisés, il y a ici un nombre fixe d'étapes (le nombre de kilomètres) puisque le coureur avance toujours et ne prends pas de détours.

Dans le document http://researchers.lille.inria.fr/~munos/papers/files/bouquinPDMIA.pdf écrit par le "Groupe PDMIA", Frédérick GARCIA donne le pseudo code suivant pour résoudre le problème. Le principe est de partir de la fin et d'optimiser l'action précédente qui donnerait la meilleur récompense. L'opération est répétée jusqu'à l'état initial.

Algorithme 1.1 : Programmation dynamique à horizon fini

$V_0 \leftarrow 0$
pour $n \leftarrow 0$ **jusqu'à** $N-1$ **faire**
  pour $s \in S$ **faire**
    $V^*_{n+1}(s) = \max_{a \in A}\{r_{N-1-n}(s,a) + \sum_{s'} p_{N-1-n}(s'|s,a)V^*_n(s')\}$
    $\pi_{N-1-n}(s) \in$
    $\operatorname{argmax}_{a \in A}\{r_{N-1-n}(s,a) + \sum_{s'} p_{N-1-n}(s'|s,a)V^*_n(s')\}$
retourner V^*, π^*

Fig x. Pseudo code pour la programmation dynamique à horizon fini (Frédérick GARCIA)

Le modèle du MDP doit être très simplifié pour appliquer cet algorithme. Les récompenses doivent impérativement être sues sur tout le long du parcours à tout moment de la course. Autrement dit, la capacité du coureur est déterminée avant même le début de la course. Cela implique de fabriquer ou estimer les temps utilisés (récompenses) pour le kilomètre (tronçon) et va à l'encontre de l'idée première de la suggestion en temps réel du projet.

5. Partie personnelle

Cette partie personnelle vient en complément à la partie commune réalisé avec les membres de mon équipe sur le projet entrepreneurial STAMINA. Je décris ici mes contributions et mes ressentis personnels vis à vis de ces deux mois de travail.

Mon poste officiel dans la start-up était celui de CEO. Lorsqu'il a fallu se répartir les rôles, mes camarades ont décidé rapidement que ce rôle m'était le plus approprié. Je n'ai donc pas vraiment eu le choix que d'endosser ce poste. Je n'ai pas forcément eu envie d'avoir ce poste début mais ce fut tout de même une bonne expérience. C'est alors pourquoi je me suis occupé de toute la gestion de la répartition des tâches qui sont apparues au fur et à mesure de l'avancement du projet.

a. Contribution sur la partie commerciale

En tant que CEO, j'ai beaucoup travaillé sur la partie commerciale pour bien comprendre notre positionnement sur le marché du running. Bien que la solution fut la raison pour laquelle nous avons commencé ce projet, le fait d'avoir pu travailler sur la vision commerciale de notre start-up m'a été très enrichissant. En effet, je me suis intéressé aux différents parties du développement commercial dans son ensemble comme à l'étude de marché ou encore à la stratégie marketing. En outre, j'ai participé à l'élaboration des maquettes avec d'autres membres de l'équipe.

Nous avons par ailleurs réalisé plusieurs interviews en binôme avec Bihe DONG comme l'interview de Morhad Amdouni ou d'autres interviews sur le terrain. J'ai proposé au tout début de commencer à chercher des coureurs à interviewer sur les différentes communautés Facebook. Cette idée fut apparu peut-être pas assez sérieuse pour certains membres de l'équipe mais c'est comme ça que nous avons réussi à recueillir le plus d'informations et d'avis de coureurs. Et aussi, c'est par ce biais que nous avons pu organiser une interview avec Morhad Amdouni.

b. Contribution sur la partie technique

En parallèle au travail sur la partie commerciale, j'ai aussi étudié la partie technique du projet. En effet, pour l'élaboration de l'algorithme de coaching en temps réel, il a fallu entreprendre beaucoup de recherche sur les différents concepts scientifiques comme l'apprentissage par renforcement.

Tout d'abord, j'ai participé au développement du partitionnement des données récupérées grâce au Web Scraping. Afin de pouvoir obtenir des données d'apprentissage pour notre algorithme, il a fallu traiter les sessions de courses et déterminer les différents états physiologiques d'un individu. Cette partie fut assez ardue car les données récupérées n'étaient pas parfaitement propres. Afin de trouver la meilleure façon de classifier les données, j'ai également mis en oeuvre différentes configurations possibles.

Ensuite, j'ai étudié quelques modèles envisageables pour notre algorithme de coaching. À partir des données obtenu après le clustering, j'ai essayé de construire des modèles en utilisant des packages sous R. Mais ces modèles ne furent pas très concluants.

c. Ressentis personnels

L'idée de cette start-up, proposée par Nicolas Brunel, m'est apparu au début très intéressante. Je pense que j'étais personnellement assez motivé car le monde des start-ups m'était déjà assez familier. En effet, j'ai réalisé mon précédent stage au sein d'une start-up parisienne. Cependant, tout le processus de concrétisation d'une idée en start-up m'était encore assez flou. Par ailleurs, la solution que notre start-up veut proposer, le coaching pour le running en temps réel, m'intriguait aussi beaucoup. De fait, les concepts scientifiques intervenant dans la création de la solution m'intéressaient fortement.

Sur le plan du travail en équipe avec mes collègues, je trouve que cette expérience entrepreneurial fut très fructifiante car elle se différencie des projets réalisés dans le cadre scolaire. Je pense avoir considérablement appris en ce qui concerne les différents mécanismes qui régissent un travail d'équipe efficace et cohésif. À certains moments, nous avons pu manquer de motivation face au travail sur la partie commerciale. En effet, nous sommes tous les 4 plus confortables sur la partie scientifique du projet. Et de fait, le manque de communication au sein de l'équipe nous a possiblement dessoudé par moment. Mais en tant que CEO, j'ai essayé d'entretenir au maximum une ambiance agréable et motivante. Pour tout ce j'ai entrepris, comme les travails de recherche scientifique ou le travail sur la partie commerciale, j'ai réalisé des compte-rendus afin que les informations soient partagées entre les différents membres de l'équipe. Et j'ai invité mes collègues à faire de même afin d'assurer la meilleure communication possible.

Conclusion sur le projet

Au niveau du développement commercial, nous avons pu définir correctement notre problématique et la solution que vous voudrions proposer. Grâce à une analyse du marché et de la concurrence détaillée, le positionnement de notre start-up sur le marché du running nous est assez clair.

Ensuite, sur la partie technique, le Web Scraping a permis de récolter une quantité non négligeable de données pour développer notre solution. Ces données sont par la suite classifiées à destination de l'algorithme de coaching en temps réel. C'est alors vers la fin du projet que nous avons obtenu un modèle algorithmique qui semble fonctionnel pour la suggestion de vitesse en fonction de l'état du coureur. Néanmoins, ce modèle est à tester en conditions réelles sur le terrain. Nous aurions voulu au moins réussir à proposer un MVP au bout de ces deux mois de travail.

Pour la suite, nous ne pensons pas poursuivre ce projet. En effet, avec la reprise des cours en master pour chacun, nous n'avons pas prévu pour le moment de nous lancer dans cette voie.

Conclusion personnelle

Dans l'ensemble, je pense avoir pu travailler sur les différentes parties de ce projet. La quantité de travail que j'ai fournie pour la partie commercial fut à mon avis plus importante que pour la partie technique. Cette expérience entrepreneuriale m'a permis découvrir les aspects commerciales des start-ups qui m'étaient jusque là encore mystérieux. Grâce aux ateliers et aux séances de coaching programmés par l'incubateur IMT Starter, j'ai considérablement appris et développé ma vision entrepreneuriale des choses. De plus, ce projet a éveillé en moi l'intérêt pour les possibilités offertes par la voie entrepreneuriale. Ainsi, ce projet ne m'a été que bénéfique et restera une très bonne expérience.

Bibliographie

Processus de Décision Markovien
https://fr.wikipedia.org/wiki/Processus_de_d%C3%A9cision_markovien

Cours de Reinforcement Learning par David Silver
https://www.youtube.com/watch?v=2pWv7GOvuf0&list=PLqYmG7hTraZDM-OYHWgPebj2MfCFzFObQ

Package R : ReinforcementLearning
https://cran.r-project.org/web/packages/ReinforcementLearning/vignettes/ReinforcementLearning.html

Programmation dynamique à horizon fini
http://researchers.lille.inria.fr/~munos/papers/files/bouquinPDMIA.pdf

Q-learning
https://towardsdatascience.com/simple-reinforcement-learning-q-learning-fcddc4b6fe56

Nombre d'abonnés premiums sur Strava
https://www.quora.com/How-many-strava-premium-users-are-there

Facts and figures about Runtastic
https://www.runtastic.com/career/facts-about-runtastic/

Nike run club gamification
https://www.reallygoodux.io/blog/nike-run-club-gamification

Owler : site de collecte d'informations concurrentielles
https://www.owler.com/portfolio

Marché du running en France
https://blog.fitmyrun.fr/levolution-du-running-depuis-2009-pratiques-marche-et-tendances/
https://www.filieresport.com/wp-content/uploads/2017/04/Filièresport-n°47-Dossier-Running-chiffres.pdf

Running et technologie
https://www.capital.fr/entreprises-marches/running-le-boom-de-la-course-a-pied-et-des-innovations-technologiques-1220734

Annexe

Lien du GitHub : https://github.com/Kevin-WKX88/project_stamina.git