 **PROJETS DEUXIEME ANNEE : JALON 2 / S7**

Date ……13…/………11……/……2018……… Numéro de Projet P5S661 ……………..

Titre du Projet : …Mooveat ………………………………………………………………………………….…….

Encadrant du projet ………Celine Hudelot……………………………………………………………….…………….…….

Noms des élèves

..........................................................................................................................................................................

............Othmane Jebbari

............Michel Richard..................................................................................................................................................

...........Diallo Alassane .............................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

..........................................................................................................................................................................

**LE RAPPORT POUR LE JALON 2 / S7 EST A DEPOSER DANS LA SECTION TRAVAUX**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A rendre au plus tard :  27 novembre | **Les trois points clés du Jalon 2**  1/ qu'avons-nous fait chacun dans l'équipe des 5 séances de travail écoulées ?  2/ Etude documentaire   * Identifier l’information pertinente : sources, experts, documents * Recueillir l’information : lectures, entretiens, observations, mesures * Mettre en forme l’information : comparaison, synthèse, rédaction   3/ Quel impact sur notre planning et notre livrable ? | Si rendu à l’heure :  Noté sur 4 points | Si 24h de retard  Noté sur 3 points  (-1 point) | Si rendu avec une semaine de retard ou plus  Pas noté  (- 4pts) |

1. **Qu'avons-nous fait chacun dans l'équipe sur la période écoulée ?**

|  |  |  |  |
| --- | --- | --- | --- |
| Séances / Nom | Alassane Diallo | Michel Richard | Othmane Jebbari |
| 16 octobre | Recherche documentaire sur l’algorithme de prix de vente prévisionnel /Traitement de sources de donnée fournies par le client | Recherche documentaire sur l’algorithme de demande prévisionnel /Traitement de sources de donnée fournies par le client | Recherche documentaire sur l’algorithme de l’offre prévisionnel /Traitement de sources de donnée fournies par le client |
| 23 octobre | Traitement de sources de données fournies par le client | Traitement de sources de données fournies par le client | Traitement de sources de données fournies par le client |
| Travail personnel | Montée en compétences en Machine Learning (en suivant l’électif Machine Learning dispensé à CentraleSupelec et le MOOC ‘Machine Learning’ sur coursera.org | Montée en compétences en Machine Learning (en suivant l’électif Machine Learning dispensé à CentraleSupelec et le MOOC ‘Machine Learning’ sur coursera.org | Montée en compétences en Machine Learning (en suivant l’électif Machine Learning dispensé à CentraleSupelec et le MOOC ‘Machine Learning’ sur coursera.org |
| 30 octobre | simplification de l'objectif du client et recherche documentaire sur l'algorithme 3 | simplification de l'objectif du client et recherche documentaire sur l'algorithme 3 | simplification de l'objectif du client et recherche documentaire sur l'algorithme 3 |
| 06 novembre | Scope review avec le client , établissement d’un planning / Discussion avec notre mentor /  Recherche de base de donnée | Scope review avec le client , établissement d’un planning / Discussion avec notre mentor /  Recherche de base de donnée | Scope review avec le client , établissement d’un planning / Discussion avec notre mentor /  Recherche de base de donnée |
| 13 novembre | Mise au point d’un modèle de l’algorithme de l’offre et bilan avec notre mentor au sujet de la scope review et notre avancée sur le projet | Mise au point d’un modèle de l’algorithme de l’offre et bilan avec notre mentor au sujet de la scope review et notre avancée sur le projet | Mise au point d’un modèle de l’algorithme de l’offre et bilan avec notre mentor au sujet de la scope review et notre avancée sur le projet |
| décembre |  |  |  |
| décembre |  |  |  |
| Travail personnel hors créneaux | Jeudi |  |  |

1. **Etude documentaire (10 à 20 pages)**

**Introduction**

D’abord, notre projet Mooveat est un projet de Tech for good explorer. Mooveat est un site mobile permettant la géolocalisation des produits locaux issus d’une agriculture biologique et ainsi permettre aux consommateurs de détecter ces produits de très bonnes qualités en vue de les commander. Le site mobile a déjà été mis au point et il existe actuellement une version sur internet.

Toujours est-il que, le propriétaire du site mobile veut une version plus complète en ce sens qu’il désire étendre le site à d’autre utilisateur à savoir les producteurs. Ainsi afin de répondre au mieux au besoin du client, Le propriétaire du site (notre client) a ainsi fait appel à notre aide.

Notre mission consiste à :

L’objectif du projet est clairement défini par le client :

-Développer un algorithme permettant de calculer le prix de vente prévisionnel théorique des produits du consommateur

-Développer un algorithme permettant de calculer la production surfacique théorique (rendement)

-Développer un algorithme de calculer la planification des besoins théoriques à savoir entre autres : les bassins de consommation et les segments de client.

Notre livrable sera donc un prolongement de site mobile en mode professionnel, c’est à dire pour le seul bénéfice des producteurs, mais dont il sera possible pour le consommateur et les clients professionnels de passer commande sur le site de l’agriculteur.

**Méthodologie de travail**

Cette partie méthodologie permet d’expliquer notre démarche de recherche documentaire . Dans un premier temps, nous nous sommes familiarisés avec les notions abordées dans le problème à savoir les concepts de rendement surfacique théorique , prix de vente prévisionnel ainsi que les notions intervenant dans le calcul de la planification des besoins théoriques à savoir le bassin de consommation …

Une fois ces notions assimilées, nous avons décidé de nous répartir les tâches : chacun devrait s’occuper d’un algorithme en particulier :

* **Richard** de l’algorithme du calcul du bassin de consommation
* Othmane de l'algorithme du calcul du rendement surfacique
* Enfin Alassane pour le calcul prévisionnel du prix de vente

Dans ce qui suivra plus tard se trouve une étude détaillée de chaque algorithme . À savoir que cette phase de recherche documentaire est divisée en deux partie :une première phase de recherche documentaire (recherche documentaire 1) durant la première période du projet (jusqu’au 19 octobre ) ensuite une deuxième phase de recherche documentaire (recherche documentaire 2 ) qu’on a commencé à partir du 23 octobre. .

1. **RECHERCHE DOCUMENTAIRE 1**

* **Rendement surfacique théorique : recherche bibliographique**

**Principe de calcul :**

Les données acquises par télédétection et les indices de végétation dérivés peuvent être utilisés pour estimer des rendements futurs, à condition que la relation entre les indices de végétation et le rendement d'une culture donnée dans une région donnée ait été établie dans le passé. Il s'agit d'estimations "à la grosse louche" et qui ne sont pas fiables à 100%. Des prédictions plus précises des rendements futurs peuvent être obtenues en utilisant des modèles. Ceux-ci combinent les indices de végétation avec d'autres types de données (météorologie, pratiques culturales, propriétés du sol, etc). Comme ils prennent en compte des paramètres liés à la localisation des cultures, ces modèles sont mieux adaptés à la prévision de rendement que les indices de végétation seuls.

**A propos des prévisions météorologiques :**

Les prévisions du type de temps (par exemple anticyclonique hivernal, régime perturbé d'ouest…) ou de la température minimale et maximale sont de leur côté utilisables jusqu'à une quinzaine de jours, contre huit il y a vingt ans. Cet allongement des échéances est dû aux progrès de l'observation, de la modélisation et des moyens de calcul. Les efforts de recherche dans ces domaines devraient permettre de gagner encore un jour sur la prévision à l'horizon 2020.

**En bref : quels phénomènes peut-on prévoir ?**

* **Quelques heures à l'avance**

Des orages, des lignes de grains, des rafales, des averses, des brouillards … (taille caractéristique des phénomènes : environ 20 km) à l'échelle d'une commune.

* **1 à 5 jours à l'avance**

L'arrivée d'une tempête (taille caractéristique du phénomène : environ 2000 km) à l'échelle d'un département (1 jour à l'avance) ou d'une région (5 jours à l'avance).

* **5 à 15 jours à l'avance**

Un type de circulation atmosphérique, des indications sur le type de temps, une tendance pour la température (taille caractéristique des phénomènes : environ 7000 km).

* **3 semaines à l'avance**

Une indication sur les conditions moyennes (température, précipitations) à l'échelle de la France.

Par exemple : température moyenne probablement supérieure de 2 °C à la normale à l'échelle de la France.

* **3 mois à l'avance**

Éventuellement un signal qualitatif sur les conditions moyennes (température, précipitations), à l'échelle d'une zone comme l'Europe de l'Ouest.

Par exemple : en Europe de l'Ouest, les températures devraient être supérieures aux normales de saison pour le trimestre à venir.

Pour faire des prévisions plus justes, il est essentiel de prendre en compte certains facteurs qui influencent la croissance de la culture et le rendement à venir. Des informations telles que les conditions climatiques ou météorologiques, les propriétés du sol et les pratiques culturales peuvent être combinées avec des données récentes issues de la télédétection pour modéliser la croissance de la culture et estimer le rendement final.

En plus de données météorologiques et des pratiques culturales, des informations géographiques sont aussi inclues dans certains modèles. Par exemple, la pente des champs, couplée à des données sur les propriétés du sol, peut fournir des informations sur le comportement des eaux de surface. Ce comportement peut être très utile quand il s'agit d'estimer les quantités de fertilisants ou de pesticides qui sont acheminées par les eaux de ruissellement.

Un système d'information géographique (SIG) est également souvent utilisé pour enregistrer les informations relatives à des champs ou à des zones particulières. Un SIG permet de compiler l'estimation du rendement avec d'autres d'informations comme le(s) type(s) de culture(s), les propriétaires des champs, les propriétés du sol et des données historiques telles que les pratiques culturales d'année en année ou la succession des attaques de ravageurs.

## Mars Crop Yield Forecasting System (MYCFS):

**Description du système :**

Le système de prévision du rendement des cultures de Mars utilisé par le secteur STAT de MARS (Surveillance de l'agriculture avec la télédétection) de l'unité Agriculture et pêche du Centre commun de recherche (CCR) associe des données de télédétection à d'autres formes de données et d'informations. Le système est composé de composants indépendants qui sont combinés pour former le système. Les principales composantes comprennent le traitement des données météorologiques, un modèle simulant la croissance des cultures, le traitement des données de télédétection et le calcul des indices, ainsi que la production d’indicateurs basés sur une série chronologique de données produites par les composantes précédentes.

**Les paramètres pris en compte par le modèle :**

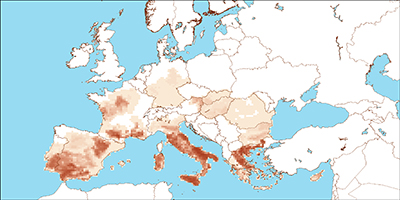
Le modèle simulant la croissance d'une culture comprend des paramètres tels que les calendriers de culture du sol, les pratiques de culture et les paramètres de culture (qui sont plus ou moins constants pour chaque culture et proviennent de la base de connaissances de la culture) et les paramètres de sol. Les données de télédétection se présentent sous forme de résolutions basse et moyenne provenant de SPOT-Vegetation / NOAA-AVHRR et de MERIS / MODIS, avec une résolution de 1 km et de 300 à 500 mètres respectivement. Ces systèmes à satellites présentent l'avantage d'une période de visite fréquente, ce qui signifie que le temps entre deux passages successifs dans la même zone est très court. SPOT-Vegetation peut couvrir la quasi-totalité de la Terre en une journée, comme le montre l’image de droite. Cela permet de sélectionner les scènes avec la couverture nuageuse minimale sur une période de quelques jours et de créer une mosaïque globale d’images avec une couverture nuageuse minimale, voire nulle.

**Les données acquises et générées par le système :**

Les produits du processus sont des bases de données et des représentations cartographiées des indicateurs de qualité agricole pour la saison. Celles-ci incluent les températures extrêmes observées à un moment de la saison de croissance, des simulations de la biomasse et du rendement produits, de l’humidité du sol, du stade de développement de la culture par rapport au stade prévu qu’elle aurait dû atteindre à un moment donné.

**Bases de données :**

**Yearly modeled crop area in EU-28 at regional level :**

****

Crop list : Soft Wheat,Grain Maize,Spring Barley,Rye,Rice,SugarBeet,Potato,FieldBeans,Soybean,Winter rapeseed,Sunflower,Fodder Maize,Winter Barley,Oats,Durum Wheat,Barley,Triticale,Wheat,Turnip rapeseed,Rapeseed,Spring rapeseed

Years : from 1975 to 2017

Variables : Modeled Crop Area, Weight Ratio

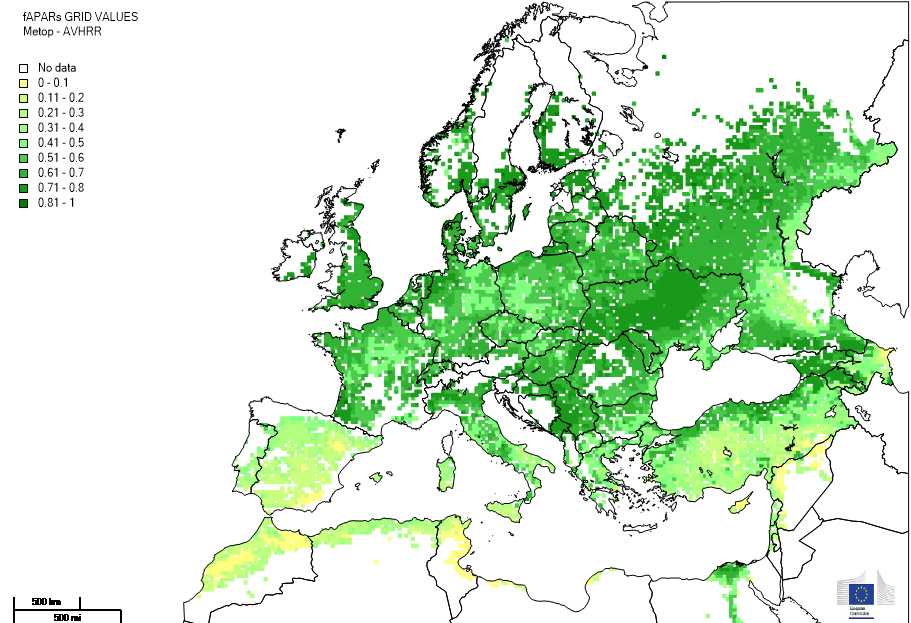
Geographic : Europa 28

Temporal resolution : Yearly

Update Frequency : Yearly

Format : csv

**Gridded remote sensing Data in Europe :**

****

Description : Remote Sensing data for vegetation monitoring in Europe. The aggregation procedure considers only pixel with a given threshold of land cover according to Genovese et al. methodology [ 2001 - Agronomie 21, 91–111].

Period : From: 11/03/2007 - To: 31/12/2016

Variables : Mean Pixel value

Geographic : Europa 28

Temporal resolution : Ten-daily (each ten days)

Format : csv

**Gridded Agro-meteorological Data in Europe :**

****

Description : CGMS database contains meteorological parameters from weather stations interpolated on a 25x25 km grid. Meteorological data are available on a daily basis from 1975 to the last calendar year completed, covering the EU Member States, neighbouring European countries, and the Mediterranean countries.

Period : From: 01/01/1975 - To: 31/12/2017

Variables : maximum air temperature (°C),

minimum air temperature (°C),

mean air temperature (°C),

mean daily wind speed at 10m (m/s),

vapour pressure (hPa),

sum of precipitation (mm/day),

potential evaporation from a free water surface (mm/day),

potential evapotranspiration from a crop canopy (mm/day),

potential evaporation from a moist bare soil surface (mm/day),

total global radiation (KJ/m2/day),

Snow Depth(cm)

Geographic : Europa 28

Temporal resolution : daily

Format : csv

**Description des bases de données fournies par le client :**

**Consommation des ménages (projections) :** étude de l’insee qui donnent la proportion des

populations par catégorie d’âge (0-19,20-59,60 et +,80 et +) pour les années de 2005 à 2030 pour

chaque région française (base de données en xls)

**Valorisation des services écosystémiques :** rapport de l’inra qui comporte une évaluation des services rendus par les écosystèmes agricoles pour mieux les gérer (pas de base de données)

**Prix du foncier agricole :** Trois cartes et trois enseignements clés de l’évolution du foncier agricole (Des prix des terres libres 2017, l’évolution des prix des terres libres de 2016 à 2017 et l’évolution des prix des terres de 2007 à 2017, tout ça par département) (base de données sous forme de carte géographique, peut-être convertie en csv)

**Données agraires diverses :** statistiques et données diverses (variétés en vigne (régionales), Base de Données Géographique des Sols de France à 1/1 000 000(base de données en xls),etc…)( Que des données régionales )

**Agreste :** données de rendement surfacique pour les produits agricoles (hors viandes) par région + abattages contrôlés (production de viandes par type) par mois de 1999 à 2018

**Prix produit** (méthode de calcul de marge, volumes et valeurs par types de circuits) ( données par région )

**Poids carbone d’un aliment :** ( Données pas très utiles pour notre algorithme de calcul de rendement surfacique )

**Impact socio-économique des circuits courts :** ( beaucoup de données qualitatives

Impact nutritionnel d’un produit agricole : composition nutritionnelle par aliment (apport

énergétique, protéines, glucides …)

**Foncier, exploitations agricoles (agreste) :** Les principaux résultats chiffrés au format tableur des recensements agricoles 2010, 2000 et 1988 par département, canton et commune. Nombre

d’exploitations, travail (en unité de travail annuel), superficie agricole utilisée (SAU), superficie en

terres labourables, en cultures permanentes, superficie en herbe.

**Synthèse :**

Tous les paramètres qu’on juge utile dans la prédiction des rendements surfaciques sont disponibles dans les bases de données qu’on possède, sauf les données chiffrées sur les rendements surfaciques passés mais qu’on peut éventuellement lier en relation avec les indices de végétation qu’on possède déjà. Notre algorithme de machine learning se fera en deux étapes : l’entraînement du modèle sur une partie de la base de données des indices de végétation passés, tester le modèle sur une autre partie de cette base, et enfin appliquer le modèle sur les indices de végétation qu’on souhaite prédire. Tout ça en intégrant tous les autres paramètres (météo, irrigation, …)

* **Prix de vente prévisionnel**

..........................................................................................................................................................................

La prévision des ventes est toujours une opération difficile en raison des nombreuses variables et inconnues qui entrent en ligne de compte. La tâche devient d’autant plus complexe lorsque vous devez établir des prévisions de ventes pour un nouveau produit car, dans ce cas, vous ne pouvez pas fonder vos estimations sur des données antérieures.

En dépit de ces difficultés, vous devez faire des prévisions de ventes afin de pouvoir planifier les ressources nécessaires pour répondre à la demande réelle, notamment en matière de stocks, de main-d’œuvre et de liquidités. Les prévisions de ventes représentent également un outil important pour évaluer le rendement de vos activités, de vos ventes et de votre marketing.

Prévisions basées sur les ventes de produits existants

La méthode la plus commune pour prévoir la demande pour un nouveau produit consiste à se baser sur les ventes de produits existants. Cette méthode est particulièrement utile lorsque le nouveau produit est en fait une nouvelle version d’un produit existant, mais d’une couleur, d’une taille ou d’une saveur différente.

Dans ce cas, il y a de bonnes chances que les ventes de votre nouveau produit ressemblent à celles du produit existant, qu’il s’agisse du vôtre ou de celui d’une autre entreprise. C’est d’autant plus vrai si vous ne comptez apporter aucun changement majeur au marketing ou à la distribution du produit.

Par contre, la prévision des ventes devient plus complexe lorsque vous lancez un produit entièrement nouveau. Dans ce cas, vous devrez consacrer plus d’efforts à l’étude de marché pour en savoir plus sur le marché potentiel de votre nouveau produit.

Utilisez des techniques d’étude de marché abordables

Les grandes entreprises utilisent des techniques d’étude de marché sophistiquées, mais il existe des moyens plus économiques qui pourront vous aider à établir les prévisions de ventes de votre nouveau produit.

Mme Rikely, qui aide des entreprises de Vancouver à améliorer leurs ventes vous offre les conseils suivants pour établir vos prévisions de ventes.

Informez-vous auprès de votre équipe des ventes

Vos représentants connaissent bien votre marché, ainsi que les produits de vos concurrents. Votre équipe du service à la clientèle pourra également vous éclairer sur le potentiel d’un nouveau produit. Discutez avec eux de votre projet et demandez-leur de vous aider à estimer le nombre d’unités que vous pourrez vendre au cours des premiers mois ainsi que le taux de progression des ventes. De plus, s’ils se sentent impliqués, ils pourraient ressentir une plus grande mobilisation à l’égard du lancement et de la réussite du produit.

Sondez d’autres sources d’information

Discutez avec des clients, des fournisseurs et des partenaires de vente (par exemple des détaillants ou des distributeurs) de confiance, pour leur demander comment ils entrevoient les chances de succès de votre nouveau produit durant sa première année. «Vous pouvez leur décrire votre projet et votre nouveau produit et leur demander ce qu’ils en pensent, suggère Mme Rikely. C’est incroyable toute l’information que vous pourrez obtenir des gens si vous leur posez la question.»

Songez à utiliser les études primaires

Les études de marché primaires font appel à certaines techniques, comme les [sondages](https://www.bdc.ca/fr/articles-outils/technologie/applications-gratuites-peu-couteuses/pages/outils-sondage-ligne-gratuits-peu-chers.aspx), l’organisation de groupes de discussion et l’observation des clients. Elles peuvent vous permettre d’obtenir de précieux renseignements sur la demande potentielle des clients pour votre nouveau produit, mais elles exigent un investissement en temps et en argent. Pour que cet investissement en vaille la peine, il est recommandé de recourir à un professionnel spécialisé dans les études de marché pour vous guider.

Commencez par un projet pilote

Il est souvent judicieux de tester votre produit à petite échelle avant de le proposer à tous vos clients potentiels. Vous pourriez demander à quelques-uns de vos représentants des ventes de vérifier la réaction du marché à votre produit, explique Mme Rikely. «Utilisez les commentaires recueillis par les membres de votre équipe pour adapter votre produit et votre marketing. Ensuite, lancez le produit à grande échelle.»

Surveillez vos résultats et rectifiez vos prévisions s’il y a lieu

Vos prévisions des ventes initiales pour un nouveau produit reposent sur plusieurs hypothèses, et vous devrez les rectifier dès que vous obtiendrez des résultats réels de vente.

Vous devez donc adopter une approche rigoureuse pour effectuer le suivi de vos ventes sur une base mensuelle. Pendant les premiers mois, vous obtiendrez ainsi de précieux renseignements sur la tarification du produit, sa production et la réaction générale des clients à son égard, explique Mme Rikely.

«Vous devez déterminer chaque mois si vous êtes sur la bonne voie par rapport à vos prévisions. Si ce n’est pas le cas, vous devez comprendre le plus tôt possible les raisons pour lesquelles vos résultats diffèrent de vos prévisions. Ainsi, vous pourrez prendre les mesures nécessaires pour corriger rapidement le tir plutôt que d’attendre la fin de l’année, lorsqu’il sera trop tard.»

Les recherches rigoureuses rapportent

Mme Rikely ajoute qu’une recherche méticuleuse avant le lancement d’un nouveau produit et durant sa première année sur le marché peut être extrêmement utile pour établir des prévisions de ventes qui contribuent au succès du nouveau produit.

«Parfois, notre instinct nous permet de viser juste, mais ce n’est pas toujours le cas, poursuit-elle. C’est toujours une bonne chose de pouvoir appuyer nos prises de décisions sur des données et des résultats réels et d’injecter une certaine dose de discipline dans le processus.»

**Synthèse**:

Vu l’ensemble des information que je viens d'énumérer et la non disposition de données nécessaire pour l'implémentation de l’algorithme concerne( nous avons passé la majorité de notre temps à chercher toutes sorte d’information pouvant nous aider à réaliser l'algorithme ) ,mon équipe et moi avons finalement décidé de ne pas nous focaliser sur ce algorithme . Il s’apparente beaucoup avec l’algorithme de calcul des besoins prévisionnels qui qu’en a lui requiert aussi de nombreuses données afin d'être implémenté . En sommes, nous nous focaliserons finalement sur l'algorithme premier( rendement ) et notre décision convient le client qui a décidé de s’occuper de l'algorithme de calcul des besoins prévisionnels et du prix ( du côté demande ) .

* **Algorithme de calcul des besoins prévisionnels**

Les recherches pour cet algorithme on dans les faits mené à à des résultats qui sont aussi utiles pour les autres algorithmes, ce que nous expliciterons à la fin. Pour chacune des études sont résumés les points essentiels intéressants pour nous. En fin de section une synthèse est effectuée notamment vis-à-vis de l’impact de cette recherche sur notre vision de l’algorithme.

*Analyse étude de PWC sur l’alimentaire*

Intéressant sur la compréhension des facteurs de segmentation des clients et de l’attrait vers les circuits courts mais très peu de données concrètes ce qui est ce qui nous intéresse vraiment pour entraîner notre modèle.

Lien vers beaucoup d’acteurs du domaine.

*Rapport alimentation Harry’s Interactive:*

Chiffres sur les pourcentages d’achats par catégorie de produit et de commerce dans lequel ils sont vendus et par catégorie de personnes.

Chiffres sur les tendances alimentaires par catégorie de produits (pains, volailles, céréales, etc.)

En croisant les deux points précédents, il sera possible d’avoir les tendances alimentaires par type de consommateurs ce qui serait intéressant elles sont ensuite croisées avec les données de l’INSEE sur la population. Cependant, il nous faudrait des données d’une granularité plus grande pour que ce soit réellement intéressant par rapport au besoin client.

*Reco 2018 par Kantar food 360 :*

Plus ou moins un mélange des deux études précédentes, toujours un manque de granularité au niveau des données.

*Convertisseur Alimentaire par Terres de liens (Normandie)*

But du convertisseur alimentaire :Estimation de la surface agricole nécessaire pour approvisionner en bio une population (ici Normandie).

Chiffres sur la consommation des habitants pris depuis l’étude de l’INCA 3 (voir plus loin).

2 types de convertisseurs, le deuxième prend en compte la saisonnalité ce qui est intéressant pour nous car la demande va évidemment varier en fonction de la saison.

Les rendements sont considérés fixes pour chaque type de céréale ce qui peut être à améliorer notamment si l’on se base sur le fait que Mooveat cherche au final à développer les 3 algorithmes : cette estimation pourra donc être modifiée grâce à l’algorithme de prévision des rendements surfaciques.

La consommation calculée est la consommation globale, il semble relativement aisé de retraduire cette manière de faire pour une échelle plus basse en se servant des données apportées par les études de *Harry’s Interactive* et *Inca 3*

Chiffres sur l’élevage (quantité de lait utilisée pour l’allaitement, quantité utilisée pour des produits laitierss), les céréales, fruits, porc et les légumes notamment en terme de quantité de produits bruts utilisés pour un produit fini (ex : carcasse -> viande, lait -> beurre).

Première estimation des pertes en matières premières pour la production de produits à valeur ajoutée comme le beurre ou le pain.

Prise en compte de la rotation (mise en friche avec de la luzerne). Il reste à le comprendre plus en profondeur.

*Autonomie Alimentaire : Utopies*

En terme d’autonomie alimentaire, la plupart des villes sont en dessous de 5% (on varie entre 0,2% et 8,1% pour Avignon).

Avantages de l’autosuffisance alimentaire : réduction des émissions de CO2, moins de dépendance et de vulnérabilité, sécurisation des approvisionnements, meilleure qualité et traçabilité des produits, création d’emplois, mieux comprendre la consommation, développement de lien social.

Autonomie alimentaire définie comme le rapport de la production agricole locale consommée dans une aire urbaine sur la quantité importée.

La production agricole local est redistribuée de plusieurs manières : éventuellement industries agro-alimentaires puis restauration locale et circuits de distribution locaux pour enfin finir dans les ménages.

La plupart de la production agricole (97%) dans les aires urbaines est utilisée à l’exportation.

En moyenne seulement 15,5€ par an consommés dans les aires urbaines en pdt locaux !!

En théorie on devrait pouvoir en moyenne atteindre une auto-suffisance de 50% !

Explicitation des moyens pour y arriver (diapo 10 et 11 et 12)

Explication du modèle de calcul LOCAL SHIFT diapo 13

*Autonomie alimentaire en agro-écologie :*

<http://www.civam.org/index.php/agriculture-durable/agriculture-projets-locaux/548-agroecologie-transformer-les-proteagineux-pour-gagner-en-autonomie-alimentaire>

Etude sur la cuisine de produits que l’on ne consomme pas habituellement : cuire les protéagineux pour en enlever les facteurs antinutritionnels. Cela nous indique principalement qu’il y a de nombreux types de produits auxquels on ne pense pas que l’on pourrait intégrer dans notre modèle. Mais ce ne sera pas quelque chose à prendre en compte dès le début,éventuellement lorsque l’on aura déjà bien développé notre modèle.

*Outil Agribalyse d’ADEME :*

Programme concentré sur la construction de données **ICV/ACV (Analyse du Cycle de Vie) agricoles**  dans le but de déterminer l’impact environnemental des productions agricoles. L’intérêt principal est au final que cet outil fait le lien avec de nombreux outils qui servent aux acteurs du secteur de l’agriculture;

Projet Ecoalim

Logiciels Open LCA et Sima Pro

Ecoivent v3.2

ACV Analyse du cycle de vie

ACTA, ART, Arvalis, Cirad, Ctifl, insitut de l’élevage, ITB, ifip, IFV, INRA, ITAVI, CETIOM, UNIP, Terre d’innovation, MEDDE, ACTA, Agrosolutions, In vivo, CIRAD, AGROSCOPE, UNIP

Permet d’alimenter la Base de Carbone, FoodGest, Base impacts (tous de l’ADEME)

Description des principaux contenus

Documents sur la construction de la BDD qui permettent de comprendre un minimum quelles informations on trouve dans la base de données (cependant le schéma relationnel est vraiment très complexe et mériterait une étude plus approfondie pour la comprendre).

Schéma de l’update Agribalyse 2

Principaux cheptels – nombres d’élevages et de cheptels (AGRESTE)

Nombre de cheptels sur la France entière 2016 et 2017

Part des ovins et caprins sur les 5 (respt. 3) principales régions

Graphes sur les quantités de vaches laitières, allaitantes, exportations, vaches bovines au niveau de la France et de l’UE

Idem pour la production porcine : encore une fois ce sont des données qui ne sont pas d’une granularité assez grande pour nous.

*TEF 2017 (INSEE) : Budget des ménages en 2017 :*

Tendances générales : le vrai intérêt en comparaison avec toutes les autres études serait pour la prédiction et non la partie training de notre algorithme de Machine Learning : on pourra prévoir le budget des ménages en fonction des tendances déjà connues.

Chiffres sur l’alimentation séparés en : pain et céréales, viandes, poissons et fruits de mer, lait fromages et œufs, boissons non-alcoolisés, boissons alcoolisées, tabac, hôtels cafés et restaurants. ***A croiser avec la BDD INSA***

Comparaison de l’évolution du pouvoir d’achat et de la consommation des ménages : ***à croiser avec une BDD sur le pouvoir d’achat de 2011 à 2018.***

Tableaux sur l’évolution des prix des différents aliments par années uniquement (dommage que ce ne soit pas par saison) avec quelques graphiques mais qui ne font que retraduire les données des tableaux.

*c\_2015\_marche\_achats\_menages :* [*http://www.agencebio.org/clone-de-le-marche-de-la-bio-en-france*](http://www.agencebio.org/clone-de-le-marche-de-la-bio-en-france)

Evolution chiffre d’affaires des ventes bio

Evolution de la part de la vente dans les types de circuits de 2011 à 2015 (grandes distribution, Artisans-Commerçants, Vente directe et comparaison avec la consommation alimentaire des ménages.

Se servent de données provenant des sites Agence Bio et ANDI (cf ci-après)

Répartition en valeurs des achats de produits bios : parts importantes des boissons alcoolisées (et aussi sans alcool) 2016

Répartition des achats des ménages pour leur consommation à domicile de produits bio par catégorie par circuits en 2015

Part de marché des différents circuits de distribution par aliment en 2015

Part de l’approvisionnement de produits bios (76,6% de pdt bios consommés viennent de France en 2015), ils sont donc vraiment intégrés dans les circuits-courts

Les vins représentent 2/3 des exportations bios françaises

<http://www.agencebio.org/la-bio-en-france>

Evolution des opérateurs et des surfaces certifiées bio 1995 2017

Surfaces en bio en 2017 et évolutions par rapport à 2017 en fonction des différentes productions végétales

De même pour ce qui est de la production animale (on remarque qu’elle est bien séparée du reste, ce sera donc probablement des éléments à considérer séparément du reste voir même modifier le modèle de l’algorithme).

Part de la surface agricole utile en bio territorialement (encore une fois probablement pas assez granulaire pour nous : plus ou moins par département) :Il serait intéressant de croiser avec les surfaces utiles totales pour avoir une estimation de la production en bio puis regarder le rendement surfacique pour les surfaces bios et non-bios.

<http://www.agencebio.org/comprendre-le-consommateur-bio>

Chiffres sur la consommation du bio

Tendances et envies des acheteurs bios, peut-être intéressant pour prédire les évolutions futures

Chiffres sur la Restauration collective en bio et explications sur le fonctionnement

**En résumé : pour tout ce qui est des chiffres sur les marchés bios rdv sur Agence Bio**

*Etude INCA 3*

Résumé de l’étude :

Données recueillies : les consommations d’aliments, de boissons et de compléments alimentaires, l’activité physique et la sédentarité, les caractéristiques anthropométriques (poids, taille) ; les caractéristiques socio démographiques et le niveau de vie, dont l’insécurité alimentaire; les habitudes alimentaires: lieux et occasions de consommation, autoconsommation d’aliments produits par le ménage ou par un proche (potager, etc.), consommation d’aliments prélevés dans la nature (chasse, pêche, cueillette), mode de production des aliments (produits transformés, agriculture biologique, etc.), etc.; les pratiques potentiellement à risque au niveau sanitaire: préparation, conservation des aliments, température du réfrigérateur, consommation de denrées animales crues ;le traitement à domicile de l’eau destinée à l’alimentation humaine ; connaissance des comportement en matière d’alimentation

Les données ont été développées en 44 groupes à partir de la nomenclature de l’EFSA (FoodEx 2) + données de la table CIQUAL

Résultats : **consommation alimentaire et apports nutritionnels de la population** (Part totale en kg puis évolution de la consommation en % de chaque aliment par catégorie par rapport à INCA 2)/ Occasions et lieux de consommation (% de la population çà avoir telle ou telle habitude) / Apport nutritionnel / Origine du choix des aliments / Comportements alimentaires subdivisés en : consommation de denrées crues (hausse, % de la pop concernée), consommation d’eau de puits privés en % des ménages, durées de conservation des denrées, matériau au contact des denrées/ Activités physiques et sédentarité/ Connaissance de l’alimentation

Etude Inca 3 à proprement parler à partir de la diapo 35.

Étude à proprement parler :

Consommation alimentaire selon âge et sexe p97 0-17/18-79 +comparaison inca 2 ce qui est intéressant car une fois de plus cela pourrait nous permettre de grossièrement estimer les variations dans l’avenir.

Apports en énergie aussi selon le sexe et l’âge : ce sont des données qui ne nous intéressent pas vraiment pour notre partie du projet mais qui pourraient servir par la suite pour Mooveat pour développer de nouvelles fonctionnalités dans le but de permettre aux utilisateurs d’avoir une meilleure compréhension de leur alimentation.

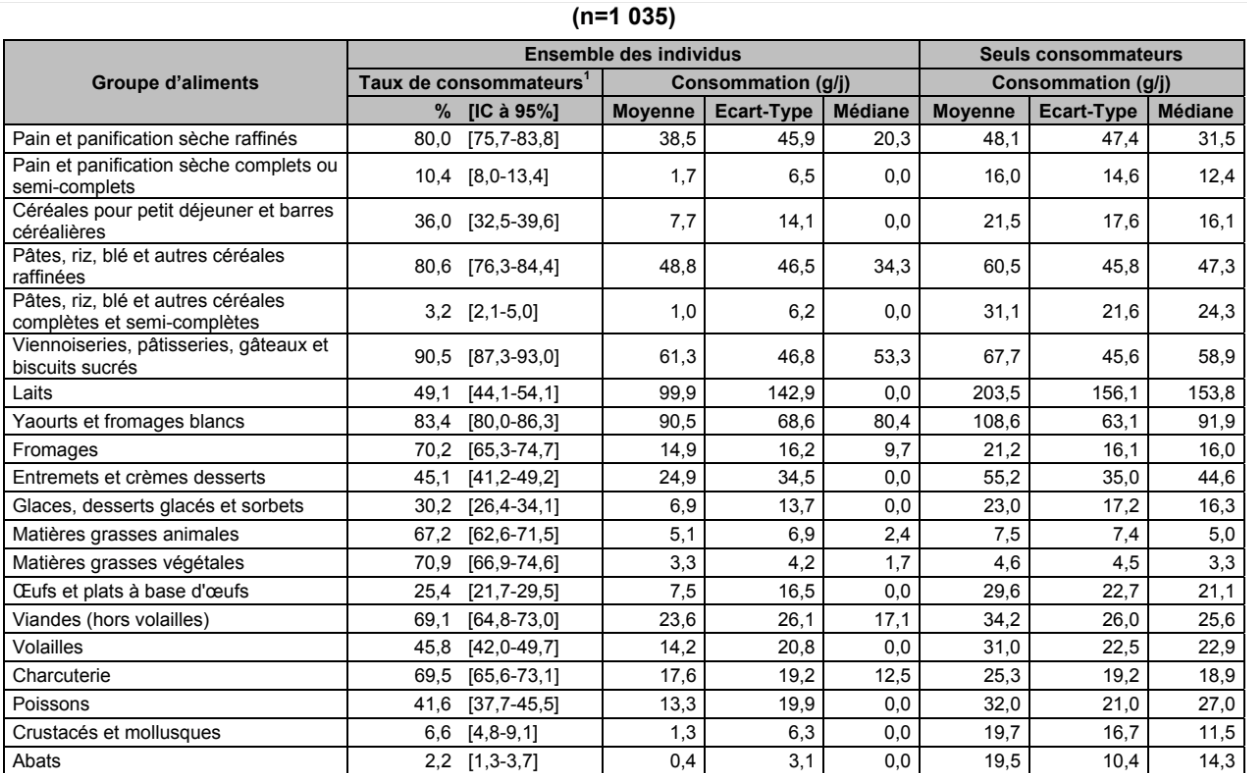
Comportements et pratiques alimentaires par âge et par sexe : les pratiques et comportements alimentaires sont intéressantes notamment pour estimer quelle part de la consommation se retrouve dans quel point de vente (restauration collective, points de ventes collectifs, etc.)

Chiffres sur les pratiques de préparation

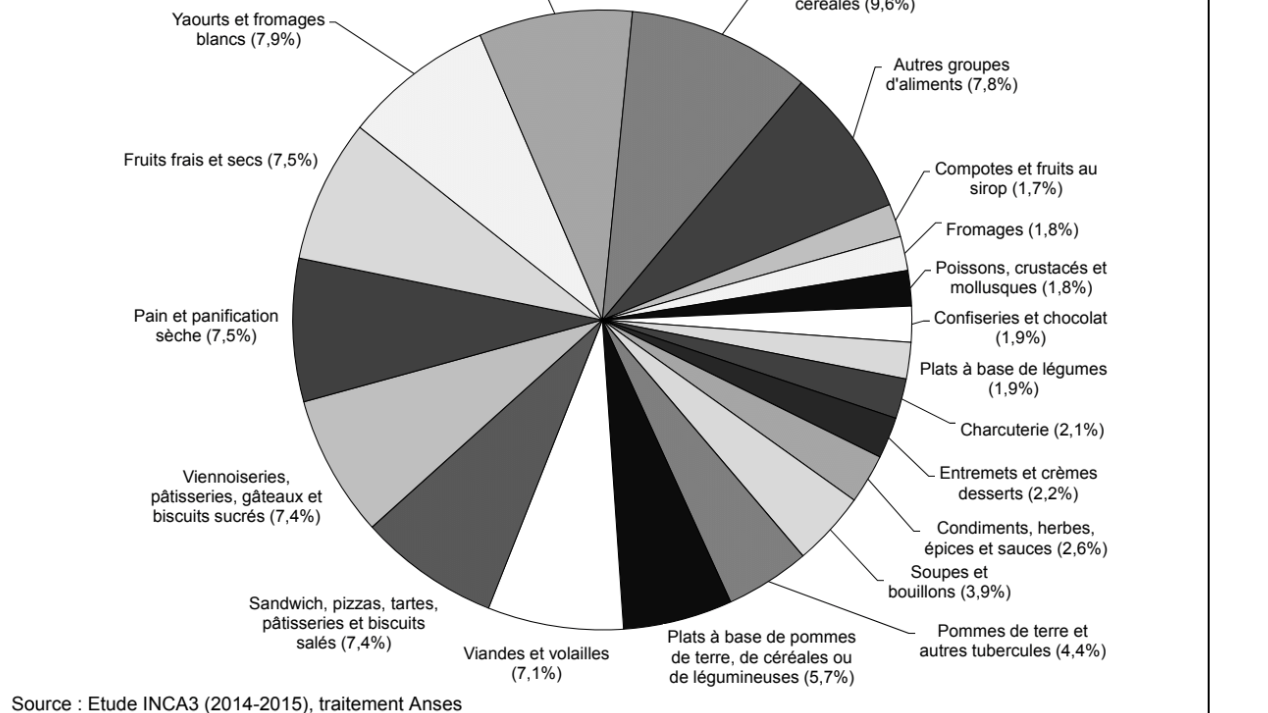
Chiffres sur les connaissances et attitudes en alimentation : peu utile pour notre aspect du projet mais utile pour le projet global de Mooveat

Chiffres sur l’activités physiques et la sédentarité (idem).

Pour avoir une idée de ce à quoi ressemblent les données et leur granularité voici quelques un des tableaux et diagrammes qui résument cela :



Consommation globale par catégorie d’ingrédients

****

Consommation pour les adolescents de 11 à 17 ans réparties en catégorie d’aliments.

En résumé, la base de données est extrêmement complète est avec une forte granularité, en la croisant avec la base de données de l’INSEE sur le recensement de la population il est possible de cartographier l’offre sur le territoire. Il ne manque que la détermination des zones de chalandises.

<https://www.avnir.org/FR/Logiciels-ACV-et-Base-de-donnees-264.html>

-> Ecoinvent : immense BDD sur l’impact environnemental des ressources et produits qui ne nous est utile qu’à le calcul de cet impact car il nous permet de savoir comment est produit chaque ressource et produit fini et quels sont les paramètres à prendre en compte.

-> GABI LCI data : données sur plus de 12 000 cycles de vie, pas en open source.

-> World food database

<https://www.universalis.fr/encyclopedie/agrometeorologie/1-influence-du-climat-sur-la-production-agricole/>

Utile pour la compréhension de l’impact des différents paramètres météorologiques sur la production agricole (précipitations, vent, pression hydrostatique, neige, etc.)

*Classification de l’intérêt des différentes données pour les algorithmes :*

Algo offre :

Convertisseur alimentaire Terres de liens (Normandie) : notion de rotation + chiffres sur ce qui sert à faire quoi comme produit pourra permettre de transformer nos rendements surfaciques en quelque chose de plus concret c’est à dire en produits vendus.

Utopies : chiffres sur l’exportation/importation en zone urbaine.

CIVAM : de nouvelles solutions pour utiliser des produits que l’on n’utilisait pas avant : modification probable de l’offre, prendre en compte de nouveaux types de produits.

Principaux chetptels sur Agreste : gros pb : seulement par zones.

Agence bio : chiffres sur la part des surfaces bios et leur évolution, part de la surface agricole bio. Il serait intéressant de croiser avec la surface agricole utile totale pour avoir une estimation puis avoir une idée des différences de rendement avec les normaux.

INCA 3 : mode de production des différents produits qui peut-être utile.

Ecoivent 2.1 : utile aussi pour ce qui est du mode de production des différents

Algo demande :

Etude pwc : chiffres sur les habitudes alimentaires et surtout compréhension de l’attrait vers les circuits courts qui pourrait permettre de prévoir les tendances

Rapport alimentation HI site : Croisement des données “chiffres sur les pourcentage d’achat par catégorie de points de ventes et de personnes” et “Chiffres sur les tendances alimentaires au niveau des produits” pourrait permettre d’avoir accès aux tendances alimentaires par types de consommateurs.

Convertisseur alimentaire Terres de Liens : calcul de la surface agricole nécessaire par habitant et spécification de la consommation moyenne (rq : INCA 3 probablement plus intéressant).

Utopies : Quelques chiffres sur la consommation en zone urbaine.

Budget des ménages en 2011

Agence bio : chiffres sur la consommation des ménages en bio

INCA 3 : tout est à croiser avec le budget des ménages de l’INSEE qui offre une granularité assez proche de celle de l’INCA 3 mais ne permet pas de

Synthèse

Si l’ensemble des données pour calculer la demande géographique et proposer un modèle de l’algorithme , il nous manque des données pour pouvoir prédire la distribution des ménages dans le futur et ainsi obtenir la demande future (exode rural notamment et croissance démographique précise). Par ailleurs, l’aspect machine Learning ne semble pas essentiel dans cette partie de l’algorithme. Là où le Machine Learning serait utile serait dans la détermination des zones de chalandises et l’influence de la concurrence dans celle-ci au travers de paramètres qui dépendent de la localisation (proximité d’une route, d’une zone touristique, etc.) ainsi que de la concurrence. Cette influence pourrait-être déterminée à partir données sur les ventes des différents points de vente mais cependant elle semble compliquée à obtenir sur une base de données assez grande à part éventuellement sur du crowdsourcing mais cela prendrait trop de temps par rapport à la durée de notre projet.

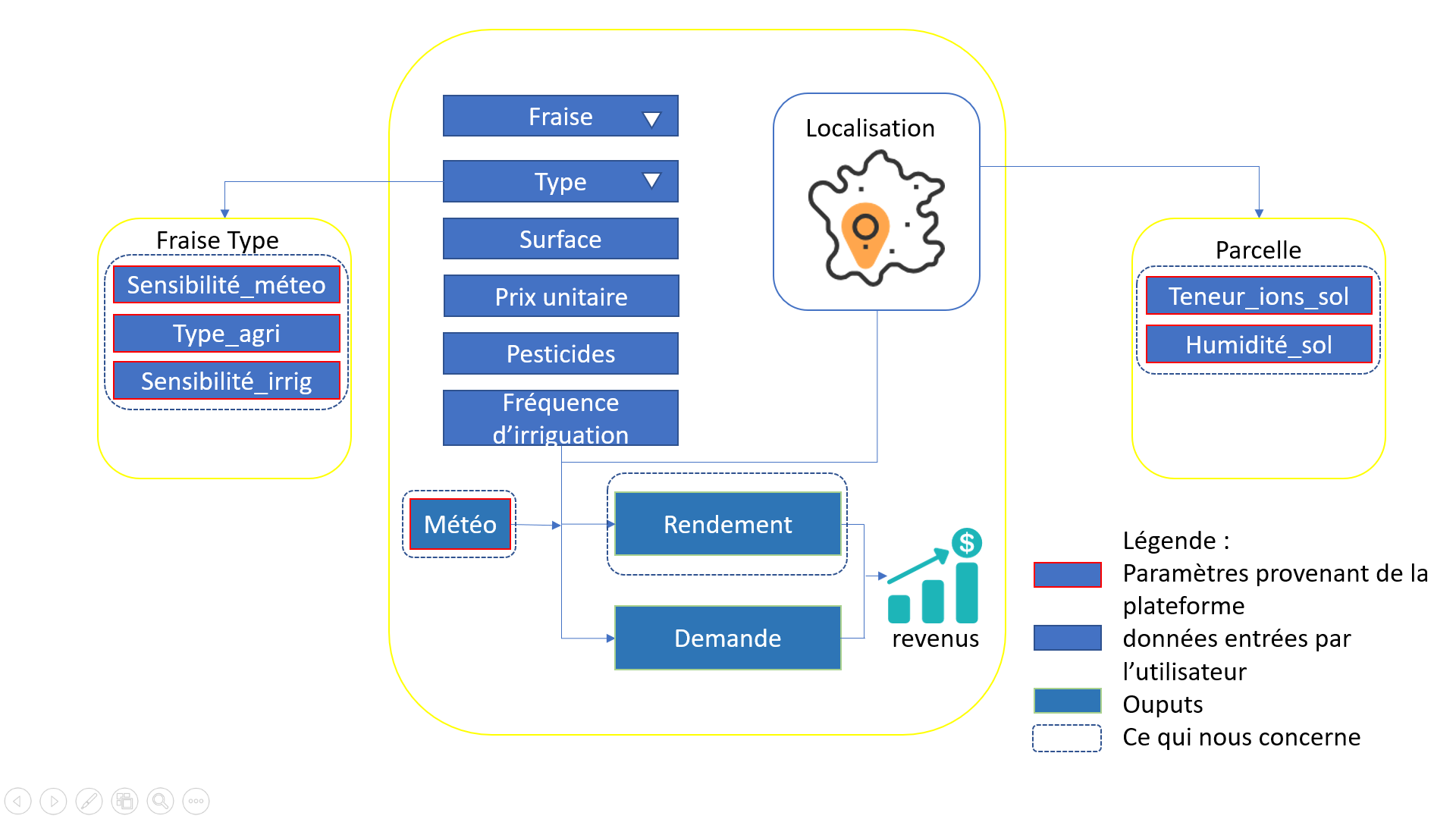
Remarque finale : nous n’avons pas parlé des recherches sur les zones de chalandises car elle se compose de nombreuses petites recherches séparées. Ce que nous en avons retenu est que les modèles les plus simplifiés (en ne prenant en compte que le temps de trajet, sans prendre en compte la visibilité, le fait que ce soit proche d’un lieu de passage, de la concurrence) sont gratuits sur internet. Cependant, la détermination réelle nécessite des connaissances très poussées en économie que nous n’avons pas. C’est pour cela qu’il aurait pu être intéressant de travailler avec du Machine Learning sur cette partie.

**3/ Quel impact sur notre planning et notre livrable ?**

**Confirmation / modification livrable**

A partir de fin octobre, en accord avec nos recherches documentaires et nos discussions avec notre client nous avons décidé de nous aligner sur l’algorithme de modélisation de l’offre (détermination des rendements surfaciques. En effet, c’était le,seul algorithme où nous avions des bases de données assez précises (bases de données d’Agri4Cast) car nous avions besoin de données au niveau d’une parcelle pour obtenir la granularité que le client souhaite (par variété de fruit/légume/céréale). Nous nous sommes donc logiquement tourné vers cet algorithme. Par ailleurs, il n’est pas influencé par des données économiques qui peuvent être extrêmement compliqués à prévoir et paramétrer mais dépend seulement de paramètres physiques, environnementaux et biologiques que nous avons plus de chance de comprendre

Début novembre, la scope review nous a permis de vérifier que nous nous étions bien compris sur ce que nous voulions. Le livrable que nous avions prévu est donc resté de produire un back-end opérationnel qui permettent de donner en fonction de paramètres entrés par l’agriculteur le rendement surfacique prévisionnel d’une parcelle. Voici un tableau résumant la vision de ce livrable :



Nous avons aussi compris que ce livrable pourrait se déterminer de deux manières différentes :

La première serait en implémentant un algorithme de Machine Learning si nous pouvons trouver une base de données nous donnant les rendements passés pour pouvoir entraîner notre algorithme.

La deuxième serait qu’à partir de recherches documentaires, de rencontres avec des experts nous puissions proposer un modèle qui nous permettrait de déterminer le rendement en fonction des divers paramètres que nous prendrons en compte (agrométéorologie, qualité du sol, pesticides etc.).

Le premier est risqué car si nous ne trouvons pas la base de données sur les rendements ou un moyen alternatif de les trouver (à travers du traitement d’images satellites par exemple) nous n’aboutirons à rien. Or la base de données dont nous aurions besoin doit être extrêmement précise : il nous faut le rendement mais au niveau d’une parcelle et non pas une donnée agrégée sur plusieurs parcelles. Cela rend une telle base de données difficile à trouver. Cependant c’est l’approche qui d’un point de vue calendaire et de formation nous intéresse le plus car elle nous permet de monter en compétence en Machine Learning et de faire d’une pierre deux coups (cf planning : projet final de ML).

Le second est tout aussi ambitieux car il demande une expertise que nous n’avons pas : nous devons accumuler une quantité d’informations différentes dont nous ne maîtrisons pas forcément les concepts. Ce travail aura donc aussi pour but de trouver les bonnes personnes qui puissent nous donner les informations sur le besoin. Cependant, il semble plus facile de rendre un livrable qui soit un minimum utile pour notre client même s’il ne donne pas forcément ce que l’on voulait initialement (il est plus simple de procéder par étapes).

Par ailleurs, cette scope review nous a aussi permis de proposer des alternatives possibles pour ce que nous ferons une fois le projet du cours d’introduction de Machine Learning rendu :

* s’il se révèle que nous pouvons opter pour la première possibilité nous aurons terminé l’algorithme au plus tard le 6 janvier. Deux possibilités s’offrent à nous : soit développer aussi la partie Front-End soit chercher comment réutiliser l’algorithme que nous venons de produire pour les 2 autres algorithme qui intéressent notre client.
* sinon une nouvelle option s’ajoute : il se pourrait que nous ayons encore besoin de peaufiner notre modèle et l’influence des paramètres, nous pourrions donc utiliser le temps restant pour le faire.

Par ailleurs, face à la forte granularité demandée, nous nous sommes mis d’accord que nous allions procéder sur environ 3 produits agricoles particuliers mais assez éloignés (par exemple lait, fraises et blé) puis de voir ensuite comment on pouvait relier nos prédictions pour chacun des produits les unes aux autres.

**Confirmation / modification planning**

Comme décrit dans la partie méthodologie du travail et recherche documentaire, notre première phase de recherche sur les 3 algos nous a mené à choisir l’algorithme de détermination de l’offre potentielle d’une parcelle ce qui a fondamentalement changé notre manière de travailler car cela nécessitait bien plus de communication entre nous pour pouvoir se répartir les tâches de manière à atteindre efficacement les objectifs que nous nous étions fixés.

Notre deuxième phase de recherche documentaire nous a permis de nous rendre compte qu’une grande partie de notre projet était de trouver les base de données qui nous permettront d’entraîner notre algorithme et nous n’avions pas prévu que cela prenne autant de temps. Si nous comptions déjà faire la montée en compétences en Machine Learning en dehors du temps de travail pour le projet inno nous nous sommes rendus comptes que la phase de compréhension du sujet et de recherche de données prendrait plus de temps que celui déjà alloué au projet inno et que nous avions besoin d’un autre moment dans la semaine pour mettre en commun nos idées/avancées. Nous avons donc décidé de nous rajouter une séance de 3h le jeudi car nous sommes tous libres.

Suite à notre entretien avec Michael (notre mentor Latitudes) le 13 novembre, nous avons décidé que la manière la plus efficace de répondre à notre livrable serait de commencer à coder avec l’indice de végétation plutôt que le rendement surfacique puis d’adapter si l’on trouve ou que l’on arrive à construire la base de données sur les rendements. Nous avons donc décidé de changer notre approche face au problème : on répartit le travail différemment avec Alassane et Othmane qui commencent à coder l’algorithme renvoyant l’indice de végétation et Michel qui continue la recherche documentaire et essaie de contacter des experts pour avoir plus d’informations sur les différents paramètres influant sur le rendement et les base de données à disposition. Nous avons effectué cette répartition car Michel est plus intéressé par la compréhension de l’univers des possibles apportés par le Machine Learning que par la montée en compétences tandis qu’Alassane et Othmane favorisent cette montée en compétences.Cela a donc modifié notre planning en scindant en deux les objectifs mais qui doivent tout de même se concorder. Par ailleurs, elle donne un avantage notable dans la construction du modèle de l’algorithme : nous comptons commencer par utiliser quelques paramètres essentiels puis l’améliorer le modèle au fur et à mesure en rajoutant des paramètres. En travaillant ainsi, nous pouvons prévoir dans quel ordre ajouter les paramètres pour parvenir à notre objectif final et ainsi ne pas prendre en compte des paramètres inutiles.