



## **ACCORD-CADRE DE TECHNIQUES DE L'INFORMATION ET DE LA COMMUNICATION**

**Marché multi-attributaires de faible montant**

**Développement d'un modèle prédictif de la  
fréquentation des cantines scolaires**

### **Cahier des Clauses Techniques Particulières**

## Table des matières

1. Résumé.....	3
2. Objectifs.....	3
3. Philosophie de la démarche.....	3
4. Besoin.....	4
4.1. Éclairer le processus de décision en périodes normales.....	4
4.2. Prise en compte des événements particuliers.....	5
5. Contraintes techniques.....	5
5.1. Données mises à disposition.....	5
5.2. Méthodes.....	6
5.3. Portabilité.....	6
5.4. Licence.....	7
6. Livrables attendus.....	7
6.1. Document de travail.....	7
6.2. Code source et documentation.....	7
6.3. Données.....	7
6.4. Présentation publique.....	8
7. Échéances.....	8
8. Annexes.....	8
8.1. Annexe 1 Présentation des données de fréquentation disponibles.....	8
8.2. Annexe 2 : Analyse graphique des données historiques de fréquentation par site	10
8.3. Annexe 3 : Données complètes.....	10
8.4. Annexe 4 : Licence MIT à associer au code source.....	10

## **1. Résumé**

Nantes Métropole souhaite développer un modèle prédictif de la fréquentation des cantines scolaires afin de mieux ajuster la production de la cuisine centrale et réduire le gaspillage alimentaire. La collectivité partagera pour cela les données complètes et détaillées de fréquentation quotidiennes et par établissement sur les neuf dernières années (2011-2019), ainsi que plusieurs jeux de données complémentaires (menus, vacances scolaires, grèves, effectifs des écoles...).

## **2. Objectifs**

La Ville de Nantes sert tous les jours 15000 à 16000 repas dans 87 restaurants scolaires. Comme dans toutes les collectivités, il existe une différence entre le nombre de repas anticipés et le nombre de repas produits et consommés. L'objectif du présent marché est d'utiliser les données de la cuisine centrale et d'autres données pour modéliser la fréquentation des cantines de Nantes et la prédire plus fiablement deux à trois semaines à l'avance. Ce défi poursuit plusieurs finalités, au premier rang desquelles la sobriété : il s'agit avant tout de réduire le gaspillage alimentaire que représente l'excédent de production de la cuisine centrale.

Une finalité secondaire est économique, pour optimiser les montants dépensés par la collectivité pour l'achat de matières premières. Une meilleure prévision des montants consommés peut également avoir une incidence positive sur le développement local. La cuisine centrale est en effet le principal débouché de nombreux petits producteurs de la région, qui adaptent leurs volumes de production en fonction de ses commandes et qui gagneraient à mieux anticiper cette demande.

Réaliser cette démarche dans le cadre d'une procédure d'achat innovant vise enfin à faire grandir les entreprises et laboratoires novateurs, qui, en retour, offrent aux administrations la possibilité de répondre plus efficacement aux attentes et besoins des citoyens, afin de susciter une amélioration continue des services publics.

## **3. Philosophie de la démarche**

Nantes et sa région disposent d'un écosystème luxuriant d'universités, laboratoires et acteurs privés actifs dans les sciences des données, associés notamment dans le cadre de NaonedIA, un collectif pour une intelligence artificielle éthique, responsable et populaire. Nantes Métropole propose à ces partenaires de travailler sur des données qu'elle met à disposition pour résoudre un problème concret, d'intérêt général.

## 4. Besoin

Nantes Métropole souhaite mieux anticiper la fréquentation des cantines scolaires afin de mieux ajuster la production de la cuisine centrale et réduire le gaspillage alimentaire. Le modèle prédictif devrait se concentrer en priorité sur les périodes “normales”, c’est-à-dire hors semaine de rentrée en septembre, jours de grève et sorties scolaires. La capacité du modèle à traiter ces jours exceptionnels apporterait une grande valeur ajoutée, mais demeure facultative.

### 4.1. Éclairer le processus de décision en périodes normales

Trois paramètres entrent en ligne de compte pour quantifier la production de repas pour les cantines scolaires : les menus, le grammage par type de convive et le nombre de convives attendus. Les menus sont définis pour une période intervacances, lors de la période précédente. Les grammages sont déterminés par type de convives classés en trois catégories : 1) enfants en première ou deuxième section de maternelle, 2) enfants en grande section ou élémentaire et 3) adultes. Le nombre de convives est estimé à trois étapes :

- **Etape 1** : Un mois avant le jour en question, le responsable logistique effectue une première estimation approximative. Celle-ci servira à adresser un bon de commande aux fournisseurs, afin que ceux-ci disposent d’un ordre de grandeur pour préparer les commandes.
- **Etape 2** : Deux à trois semaines avant, chaque responsable de restaurant scolaire demande un nombre précis de repas pour le jour en question. La cuisine centrale agrège ces demandes et ajuste les commandes auprès de ses fournisseurs. C’est ce nombre de repas consolidé qui déterminera le volume de nourriture livré et facturé par les fournisseurs, puis préparé par la cuisine centrale deux à six jours avant pour le jour en question.
- **Etape 3** : Le jour en question, neuf chauffeurs-livreurs de la cuisine centrale déposent les repas dans les écoles, en deux tournées successives, de 6h à 9h environ. En arrivant à l’école entre 8h30 et 9h, les enfants s’inscrivent, ou non, pour le déjeuner. L’équipe d’encadrement de chaque école fait la somme des convives effectivement inscrits pour le repas du midi et contacte la cuisine centrale si cette somme dépasse le nombre de repas demandés trois semaines auparavant. Chaque chauffeur-livreur dispose de 20 repas supplémentaires dans son camion, qu’il apporte aux écoles où des repas manquent. Les sites qui ont reçu trop de repas appellent la cuisine centrale pour qu’ils soient récupérés. Les chauffeurs viennent alors récupérer les repas excédentaires pour les apporter là où il en manque.

Le besoin porte donc avant tout sur un modèle prédictif pouvant servir d’aide à la décision à l’étape 2, afin d’ajuster les commandes et la production planifiées deux à trois semaines

à l'avance, avec les informations disponibles à ce stade. Un besoin secondaire porte sur une modélisation plus précise avec des paramètres connus la veille du jour en question (ex. météo ou grèves) et qui permettrait de prendre d'adopter des mesures correctives.

## **4.2. Prise en compte des événements particuliers**

Trois types d'événements induisent des difficultés particulières pour l'analyse : les grèves, les rentrées de septembre et les sorties scolaires non anticipées. Les grèves sont de nature et d'ampleur hétérogènes et elles affectent autant la distribution que la collecte de données sur les repas effectivement consommés. Le nombre de repas à préparer pour les premières semaines de septembre est estimé début juillet par les responsables de site, en fonction du nombre d'élèves inscrits pour l'année suivante qui sont enregistrés avant les vacances d'été. Mais de nouvelles inscriptions surviennent souvent en juillet ou en août, qui ne sont généralement validées que quelques jours avant la rentrée. Généralement après les vacances de Pâques, les enseignants organisent des sorties scolaires à la journée. Lorsque l'enseignant oublie de prévenir la responsable du restaurant scolaire des repas sont livrés en trop.

La capacité du modèle à traiter ces jours exceptionnels apporterait une grande valeur ajoutée, mais demeure secondaire.

## **5. Contraintes techniques**

### **5.1. Données mises à disposition**

Les jeux de données suivants seront fournis par Nantes Métropole aux chercheurs acceptant de collaborer à ce défi :

- pour chaque cantine et chaque jour ouvré depuis 2011, on dispose du nombre de repas commandés et livrés par la cuisine centrale (étape 2 ci-dessus), ainsi que du nombre de repas effectivement consommés dans les écoles (étape 3). Cela représente 106287 points de données entre le 03/01/2011 et le 19/12/2019. Le fichier ne contient pas de désagrégation par type de convive (catégories énumérées au 4.1). Une description plus poussée de ces données est fournie dans la section 8.1).
- pour chaque journée depuis 2011, les menus des cantines scolaires. Ce fichier contient pour chaque date l'intitulé des plats proposés dans le "menu standard" et le "menu allergique". Un aperçu de ce jeu de données pour les deux derniers mois est constamment tenu à jour sur le site Open data de Nantes Métropole.
- le calendrier des vacances scolaires pour l'académie de Nantes saisi manuellement à partir des archives du calendrier scolaire de l'éducation nationale.
- le nombre d'enfants inscrits dans chaque école de la ville de Nantes depuis 2011.

- les jours de grèves : des mouvements de grève peuvent affecter la livraison ou l'organisation des repas dans les cantines. Ces événements sont pour l'instant documentés sous forme d'événements (c'est-à-dire date des jours de grève). Une étude est en cours pour déterminer s'il est possible de collecter des informations complémentaires pour mieux caractériser ces journées de grève.

Par ailleurs, les chercheurs sont encouragés à mobiliser notamment les jeux de données suivants :

- Le calendrier des jours fériés, disponibles sur le portail de données ouvertes de l'Etat français ;
- Les dates des fêtes religieuses (cf. la liste des principales fêtes recensées par l'administration publique) ;
- Les données météorologiques, comme les relevés de la station météo de Nantes Bouguenais disponibles toutes les trois heures depuis le 1er janvier 2010 sur la base de données publiques SYNOP ;

## **5.2. Méthodes**

Le pouvoir adjudicateur manifeste un intérêt particulier pour des algorithmes d'intelligence artificielle (forêt aléatoire, réseaux de neurones...), mais des approches bayésiennes ou des modèles déterministes (ARIMA, lissage exponentiel, décomposition...) sont aussi envisageables. Si le titulaire emploie un algorithme d'intelligence artificielle, une attention particulière devra être apportée à prévoir et décrire dans la proposition technique une méthode d'apprentissage qui n'obère pas les propriétés de généralisation du modèle (risques de sur-ajustement/overfitting). Si le titulaire emploie une méthode déterministe/bayésienne, il devra prévoir et décrire dans la proposition technique les tests d'erreur qui seront utilisés pour qualifier la fiabilité du modèle.

## **5.3. Portabilité**

Le contenu du modèle devra être encapsulé dans un conteneur (docker ou autre) avec l'environnement nécessaire à son exécution ou s'appuiera sur un moteur statistique facile à installer localement ou sur un serveur (p. ex. R ou Python) selon un procédure précisément documentée. Le modèle devra prendre en entrée et produire en sortie des données dans un format ouvert, structurées selon un schéma explicitement décrit dans la documentation associée. La documentation associée devra être didactique et accessible à des personnes disposant de compétences techniques limitées.

## **5.4. Licence**

Conformément à la Loi pour une République numérique, le pouvoir adjudicateur mettra le code source développé par le titulaire du marché sous un régime de licence de logiciel libre, à savoir la Licence MIT (cf. annexe 2).

Le titulaire du marché s'engage donc à réaliser le logiciel, dans des conditions qui permettent au pouvoir adjudicateur de mettre ce logiciel à disposition de tiers, dans les conditions posées par la licence annexée au présent marché.

## **6. Livrables attendus**

### **6.1. Document de travail**

Le titulaire remettra un document de travail décrivant les données mobilisées, le modèle utilisé, les résultats et des tests mesurant la précision et la capacité prédictive de ces derniers. Ce document sera évalué à l'aune de cinq critères : caractère novateur de la proposition, clarté et caractère didactique du document de travail, validité scientifique de l'approche, fiabilité des résultats et faisabilité d'une réutilisation par la collectivité pour améliorer le service public. Le pouvoir adjudicateur publiera ou encouragera la publication de ce document sur un support libre d'accès (revue scientifique, arXiv.org, hal.archives-ouvertes.fr...).

### **6.2. Code source et documentation**

Le code d'analyse devra également être transmis et sera publié avec les documents de présentation pour assurer la reproductibilité des analyses. Les fichiers de scripts composant le code source seront explicitement commentés, afin de faciliter la consultation du code source par d'autres développeurs.

Le code sera accompagné d'une documentation décrivant pas à pas la manière dont l'outil doit être utilisé. En particulier, la documentation donnera des indications didactiques sur la procédure à suivre pour installer le programme, formater et charger les données nécessaires à son fonctionnement, et obtenir et exploiter les résultats.

### **6.3. Données**

À l'issue du travail, les données remises au titulaire par le pouvoir adjudicateur seront publiées sur le portail open data de Nantes Métropole. Les données modifiées, enrichies ou ajoutées par le titulaire pour permettre le fonctionnement du modèle prédictif seront remises au pouvoir adjudicateur et publiées également sur le portail open data de Nantes Métropole.

## **6.4. Présentation publique**

Le titulaire réalisera une présentation publique de son travail. À cette fin, il préparera un diaporama et un poster (imprimable format A0) synthétisant les points saillants inclus dans le document de travail (décrit au 6.1) et il présentera ces éléments lors de sessions dont les modalités précises seront précisées ultérieurement.

## **7. Échéances**

Dès que possible : transmission par le pouvoir adjudicateur au titulaire les données préparées (déjà disponible)

06/11/2020, livraison provisoire : transmission électronique par le prestataire au pouvoir adjudicateur des livrables 6.1 , 6.2 et 6.3.

23/11/2020 : transmission électronique par le pouvoir adjudicateur au titulaire de ses questions et commentaires.

01/12/2020, livraison révisée : transmission électronique par le titulaire au pouvoir adjudicateur des livrables 6.1, 6.2 et 6.3 intégrant les questions et commentaires du pouvoir adjudicateur, ainsi que des livrables 6.4.

avant le 14/12/2020 : présentation publique, à une date et selon des modalités qui seront précisées ultérieurement.

14/12/2020 (au plus tard), livraison finale : remise de tous les éléments consolidés.

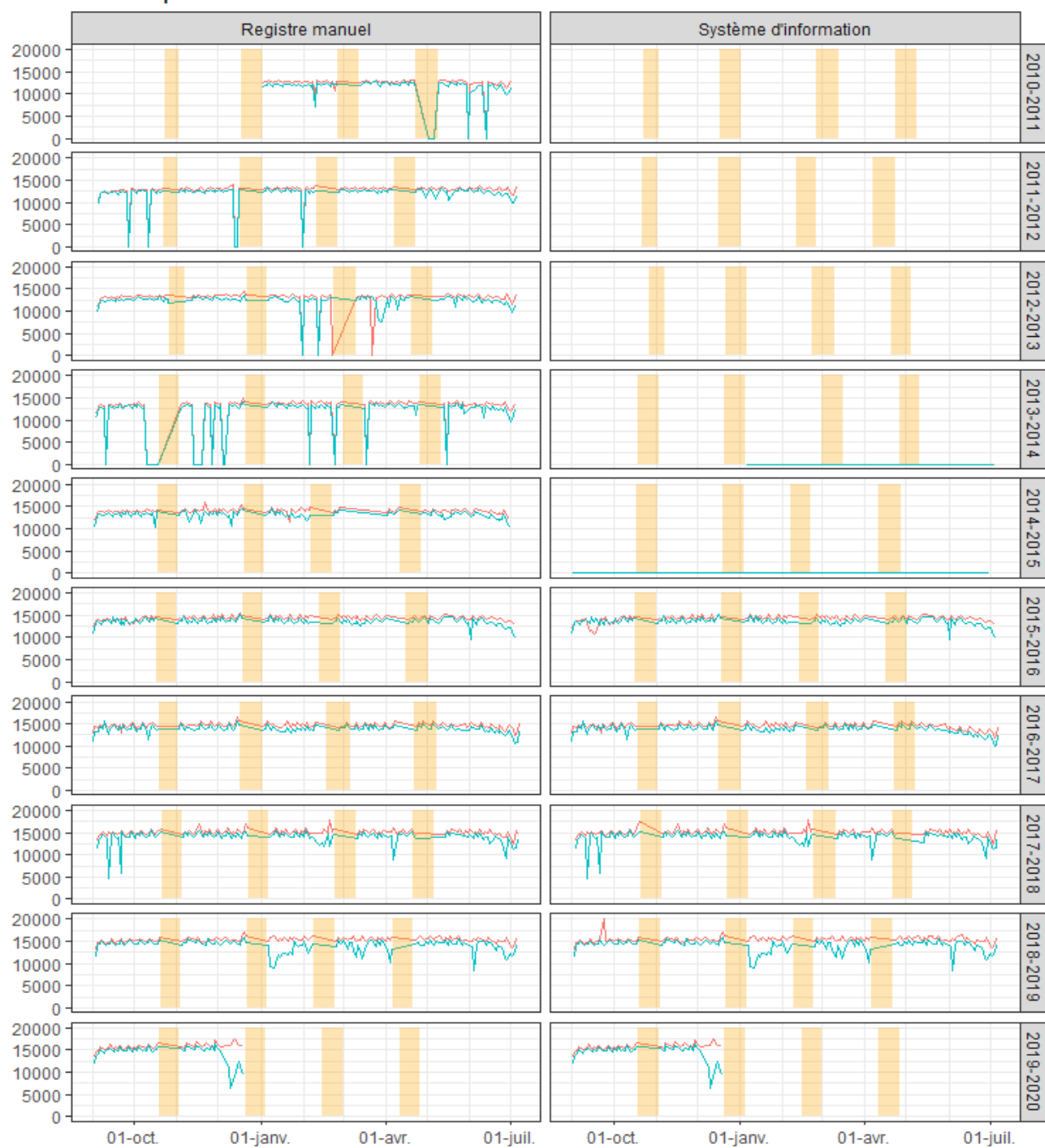
## **8. Annexes**

### **8.1. Annexe 1 Présentation des données de fréquentation disponibles**

La complétude, la fiabilité et la volatilité des données d'historique disponibles pour la variable à prédire déterminent généralement la performance d'un modèle prédictif. Pour permettre aux participants d'apprécier ces conditions, nous représentons ci-dessous l'ensemble des cantines scolaires la somme quotidienne du nombre de repas prévus (c'est-à-dire préparés) d'une part, et du nombre de repas effectifs (c'est-à-dire consommés) d'autre part. On présente les données issues de deux sources connexes : le système d'information de la cuisine centrale (logiciel Fusion, de l'éditeur Salamandre) et le registre manuel dans lequel les données de Fusion sont copiées, pour ensuite être corrigées par les responsables de la production et de la distribution des repas.



## Somme pour l'ensemble des cantines



Nombre de repas : — Prévision (produit) — Réel (consommé) Événements : ■ Vacances scolaires

### Excédent de production (en %)

Année scolaire	Excédent (Registre manuel)	Excédent (Système d'information)
2010-2011	5.3	NaN
2011-2012	4.7	NaN
2012-2013	5.1	NaN
2013-2014	5.3	NaN
2014-2015	6.1	NaN
2015-2016	5.2	4.5
2016-2017	5.3	5.3
2017-2018	7.6	7.7
2018-2019	8.8	9.1
2019-2020	7.9	7.9

## 8.2. Annexe 2 : Analyse graphique des données historiques de fréquentation par site

Nous incluons en annexe une désagrégation de ces mêmes informations établissement par établissement.

## 8.3. Annexe 3 : Données complètes

Le fichier zippé contient deux sous-dossiers

- Données : Ce dossier contient l'ensemble des données mentionnées au 5.1.
- Code\_source\_preparation : Ce dossier contient l'ensemble des scripts (R) utilisés pour la préparation et la vérification de la qualité des données ;

## 8.4. Annexe 4 : Licence MIT à associer au code source

Se référer à la dernière version de la version MIT référencée sur le site officiel : [www.data.gouv.fr/fr/licences](http://www.data.gouv.fr/fr/licences)