

TEST - TDTI

Durée : 3 heures – Tous les documents sont autorisés

« Google's mission is to organise the world's information and make it universally accessible and useful. »

Google

« La connaissance s'acquiert par l'expérience. Tout le reste n'est que de l'information. »

Albert Einstein

« La **route nationale 346**, ou **RN 346**, est une [route nationale française](#) qui correspond actuellement à la voie rapide de la banlieue-est de [Lyon](#), nommée **Rocade est**, reliant les deux interruptions de l'autoroute [A 46](#) entre [Vaulx-en-Velin](#) et [Saint-Priest](#). »

Wikipedia

« L'information n'est pas la connaissance ».

Frank Zappa

PARTIE I. – ETUDE DE CAS

Affecté(e) à la DIR Centre Est au 1^{er} Aout 2015, vous êtes partie prenante d'une étude visant à expérimenter différentes solutions fournissant de l'information temps de parcours (TP) sur un réseau. Une expérimentation sur la rocade Est de Lyon a été envisagée (RN346). Plus précisément, vous allez considérer la portion entre l'échangeur de Manissieux (avec l'A43) au PR41 et le diffuseur du Grand Large (au PR33), voire jusqu'à l'échangeur du Nœud des Iles (avec l'A42) au PR27. C'est le lieu d'une congestion récurrente les soirs de semaine, entre les PR39 et le PR35 (point haut) du fait d'une rampe ralentissant les PL., dans le sens sud-nord. Cette section est également équipée en boucles électromagnétiques et infrastructures suffisantes (PMV, branchements électriques) pour l'accueil de nouveaux systèmes de recueil.

Après un appel d'offres international et une mise en concurrence en bonne et due forme, votre service sélectionne les solutions suivantes :

1. SENSYS¹ : Les capteurs magnétomètres de Sensys Networks ;
2. BLIP² : Les capteurs Bluetooth de Blip Systems ;
3. TOM-TOM³ : La solution trafic temps réel de Tom Tom ;
4. BLUEVIA⁴ : Les capteurs Bluetooth de Bluevia
5. LAPI⁵ : Une solution de Lecture Automatique de Plaques d'Immatriculation (LAPI).

¹ <http://www.sensysnetworks.com/products/sensor/>

² <http://www.blipsystems.com/bliptrack-bluetooth-traffic-sensor/>

³ https://www.tomtom.com/en_gb/licensing/products/traffic/

⁴ <http://www.neavia.com/bluevia/>

⁵ <http://www.smartplanet.fr/smart-technology/le-radar-polonais-iskip-detecte-les-voitures-volees-et-pese-les-vehicules-17931/>

TRAVAIL A REALISER :

1) Analyse comparative des solutions sélectionnées

- a) Décrire de façon synthétique les spécificités, complémentarités, similarités entre les différents systèmes de recueil de données choisis
- b) Expliciter les risques et dangers lors de l'installation et la configuration des capteurs sur la zone expérimentale
- c) Donner pour chacun des systèmes, la ou les configuration(s) optimale(s) permettant un recueil de qualité (nombre de dispositifs à installer, rayons de détection, interférences possibles, etc.)

2) Evaluation du système de mesure des TP

L'évaluation d'un tel système recouvre essentiellement trois questions

- i) Celle de la référence : par rapport à quoi la précision de la mesure est vérifiée
 - ii) Celle de la formulation des exigences de précision une fois la référence définie
 - iii) Celle de la robustesse aux aléas, c'est-à-dire la proportion des situations où les exigences de performance pourront être satisfaites
- a) Quelles sont d'après vous les principales exigences attachées à la mesure des temps de parcours ?
 - b) A chacune des exigences évoquées plus haut, relatives aux systèmes de mesure des TPS, associer le critère ou les critères permettant d'évaluer la conformité du système.
 - c) Parmi les solutions proposées quelle est, d'après vous, celle qui pourra être vue comme la « vérité » terrain, i.e. la référence ? Justifier votre réponse
 - d) Proposer un plan d'expérience (définition des campagnes de vérification des TP, conditions de comparaison, situations de trafic choisies)
-

PARTIE II. – TP SOUS R

En octobre 2015, les premiers systèmes de recueil ont pu être installés et testés et les premières analyses de données sont lancées. La rocade Est a été découpée en sous-segments qui sont délimités par des capteurs, Bluetooth ou autre. Le jeu de données à analyser comporte des enregistrements d'usagers enregistrés par des capteurs Bluetooth sur une journée entière. Nous nous focalisons en particulier sur un sous-segment de route (délimité par deux capteurs Bluetooth) près de Pusignan. Le tableau individus-variables (à chaque ligne correspond l'enregistrement d'un usager sur le sous segment) comporte les variables suivantes :

V1 : vitesse moyenne de l'usager sur l'ensemble de son itinéraire

V2 : écart type de la vitesse de l'usager sur l'ensemble de son itinéraire

V3 : vitesse de l'usager sur le sous segment

Le jour de l'enregistrement, un événement quelque peu inhabituel a eu lieu. Une fois par an, l'amicale des vélosolèx réunis de Pusignan organise sa traditionnelle randonnée de cyclotourisme. En empruntant exceptionnellement la rocade Est, les cyclomotoristes ont irrésistiblement et, malgré eux, créé une classe d'automobilistes à part sur une section essentiellement fréquentée par des automobilistes. L'objectif de ce TP est de détecter ces deux classes dans votre jeu de données à l'aide des outils de statistique descriptive vues en TDTI. En particulier, vous répondrez aux questions suivantes :

- 1) Peut-on distinguer facilement les cyclomotoristes des automobilistes à partir de la seule observation des distributions de vitesses sur le sous-segment ?
- 2) À partir des 3 variables fournies, esquisser une méthode de filtrage permettant de ne conserver que les individus identifiés comme des automobilistes. (Pensez à justifier, succinctement, le choix des 3 variables introduites quant à l'objectif de séparation des divers types d'usagers [Rôle de chacune des variables, identifier (aide : *hist*) et déterminer les variables discriminantes ...].)
- 3) En travaillant sur les vitesses moyennes (V1) et les écarts types moyens (V2), montrer de manière brute comment les deux classes apparaissent clairement.
- 4) Rappeler, en quelques mots, le principe de la classification par la méthode des *kmeans*
- 5) Effectuer un **clustering** à l'aide des *kmeans* pour retrouver ces deux classes de façon automatique (effectuer cette classification sur le jeu de données normalisé à l'aide de la fonction *scale*). Représenter sur un même graphe les 2 classes obtenues.
- 6) Tracer la distribution des vitesses sur le sous-segment en fonction des 2 classes créées auparavant. Superposez les 2 répartitions (avec *hist*, utiliser *add*). Vérifier la pertinence de l'existence des 2 classes grâce à un test statistique bien choisi.
- 7) Modéliser ces deux répartitions à l'aide de lois normales. Évaluez l'adéquation des 2 distributions à une loi statistique à l'aide d'un test (utiliser les fonctions *curve* et *dnorm*)