

CR P5: De la Donnée Capteur à l'Information

1) Description de notre sujet de projet

En chacun de nous, chaque jour, nous avons besoin de manger et de boire pour vivre. Mais maintenant, une situation grave est arrivée à la nourriture et c'est aussi la situation dans presque tous les pays du monde - le gaspillage alimentaire. Il a eu un grand impact sur notre environnement: sur l'émission de gaz à effet de serre (environ $\frac{1}{3}$ d'émission de GES provenant des activités agricoles), le gaspillage d'eau mais aussi sur la biodiversité.

En plus, dans notre pays natal Viet Nam, beaucoup de gens prennent l'habitude d'aimer beaucoup cuisiner, de ne pas tout manger, puis de les conserver au réfrigérateur ou de conserver les aliments sans oublier leur date de péremption. Le gaspillage alimentaire affecte non seulement directement notre budget, mais plus largement, il détruit également notre cadre de vie et menace la sécurité alimentaire d'aujourd'hui.

Pour ces raisons, nous avons créé une boîte intelligente qui peut stocker les aliments plus longtemps qu'un réfrigérateur. Si nous utilisons les bonnes méthodes de conservation, la durée de conservation des aliments peut augmenter de 2 à 3 fois.

Nous avons décidé une approche indirecte pour mesurer la qualité des aliments en mesurant la qualité d'air dans notre armoire. En effet, nous utilisons le capteur de température et d'humidité DHT11, le capteur de gaz O2 SEN0159 et le capteur de gaz CO2.

2) Objectifs et tâches en analyse de données:

Dans cette phase, on a pour objectif réalisé de manière supervisée:

- Nettoyage des données: Écarter les valeurs impossibles (null ou négatifs) et les valeurs aberrantes.
- Visualiser les données en fonction des fonctions d'histogramme, boxplot, courbes, etc.
- Modélisation: Visualiser l'évolution de l'état atmosphérique en fonction du temps et le lien entre les mesures.

3) Descriptions des données collectés:

No	Capteurs	Données collectées	Format des données	Quantitatives/Qualitatives	Valeurs possibles
1	Capteur température	Températures	Numériques	Quantitatives	-20 à 80 C
2	Capteur humidité	Humidité			0 à 80%
3	Capteur de gaz o2	Concentration o2			0 à 80%
4	Capteur de gaz co2	Concentration co2			0 à 8000ppm

4) Création des jeu de données:

Pour le jeu de données, nous avons importé les données dans RStudio à partir d'un fichier csv contenant de 5 variables: .

- datetime: le moment où la mesure est effectué
- temp: la température du moment en Celcius
- humid: l'humidité du moment en pourcentage
- o2: la concentration d'oxygène en pourcentage
- co2: la concentration co2 en ppm (particules par million)
- box: identifiant du frigo où les mesures sont effectués

Comme les valeurs de chaque capteur sont stockées dans sa colonne correspondante, nous n'avons pas observé la nécessité de créer des sous-ensembles.

5) Analyse statistique des données et visualisation:

- summary(): sommaire des données ce qui nous permet de visualiser les indicateurs.
- hist(): Histogramme qui nous sert à visualiser la répartition des valeurs de chaque capteurs:

```
par(mfrow=c(2,2))  
hist(p2i[,2], main = "Temperature")  
hist(p2i[,3], main = "Humidite")  
hist(p2i[,4], main = "o2")  
hist(p2i[,5], main = "co2")
```

- Boxplot: une boîte de moustache qui montrer tous les informations contenant dans summary() d'une manière visualisé

```

par(mfrow=c(2,2))
boxplot(p2i[,2], main = "Temperature")
boxplot(p2i[,3], main = "Humidite")
boxplot(p2i[,4], main = "o2")
boxplot(p2i[,5], main = "co2")

```

- Plot: générer des graphes d'évolution des données en fonction du temps

```

par(mfrow=c(2,2))
plot(p2i[,1], p2i[,2], type="l", main = "Temperature")
plot(p2i[,1], p2i[,3], type="l", main = "Humidite")
plot(p2i[,1], p2i[,4], type="l", main = "o2")
plot(p2i[,1], p2i[,5], type="l", main = "co2")

```

- Pairs: afficher le nuage des points

```

pairs(p2i)

```

6) Description de la chaîne de traitement. Justification des approches et algorithmes utilisés.

Tout d'abord, nous avons éliminé les mesures certainement impossibles dans le jeu de données, donc il y a 14 entrées où la valeur de co2 est inférieure à 1. Puis nous avons converti les valeurs dans la colonne datetime de type "string" au type "datetime" de format "DD-MM-YY hh:mm:ss"; colonne "temp" et "humid" de type "string" au type "numérique" afin de traiter des données. Vu que toutes les mesures venaient d'un même box, nous avons éliminé la colonne "box".

Comme la fonction principale du programme est de visualiser l'évolution des indices de l'air, nous avons concentré sur la création des graphiques afin d'indiquer l'influence des éléments externes. Nous avons aussi cherché les influences mutuelles entre la température, l'humidité et la concentration des gaz.

7) Analyse des résultats obtenus

Premièrement, les résultats obtenus par les fonctions summary, hist et boxplot nous indiquent que toutes les valeurs sont dans la domaine acceptable. En particulier, quand on analyse les jeux de données de 28/05 et 02/06, on observe une variation légère de toutes les valeurs dans une même période (chaque période durant au maximum 45 minutes se sont séparées de 30 minutes à plusieurs heures), alors nous ne pouvons rien conclure ici. Pourtant, quand nous analysons les jeux de données de 05/06 dont les mesures sont collectées pendant 2 heures, nous observons bien une variation de concentration de CO2. Il existe des pics dans une courte période suivant une concentration stable de CO2. Nous déduisons que ces valeurs sont des valeurs aberrantes.

Quand nous avons fait un graphique des valeurs de o2 et co2, on obtient une graphe de forme inconnue, ce qui nous aide à conclure qu'il n'y a aucun lien distingué entre ces deux valeurs. De même pour les autres valeurs.

```
summary(p2i[,2:5])
```

temp	humid	o2	co2
Min. : 0.00	Min. : 0.00	Min. : 0.00	Min. : 0
1st Qu.:28.00	1st Qu.:30.00	1st Qu.:12.46	1st Qu.:279
Median :28.00	Median :30.00	Median :12.46	Median :286
Mean :28.01	Mean :29.92	Mean :12.41	Mean :285
3rd Qu.:28.00	3rd Qu.:30.00	3rd Qu.:12.46	3rd Qu.:288
Max. :29.00	Max. :31.00	Max. :23.17	Max. :515
NA's :1	NA's :1		

Figure 1. Sommaire des données

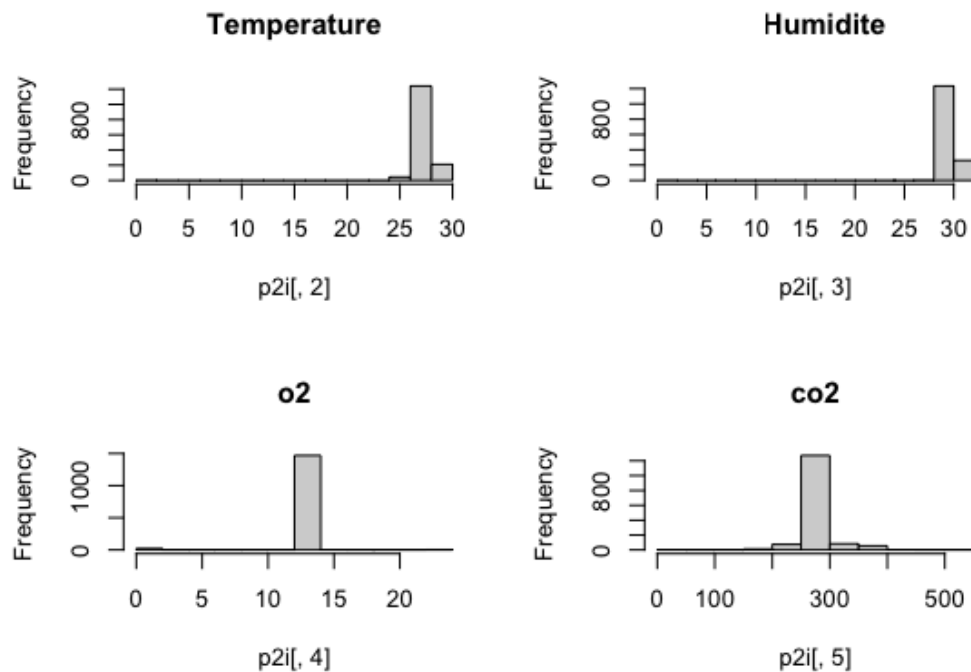


Figure 2. Histogramme des données

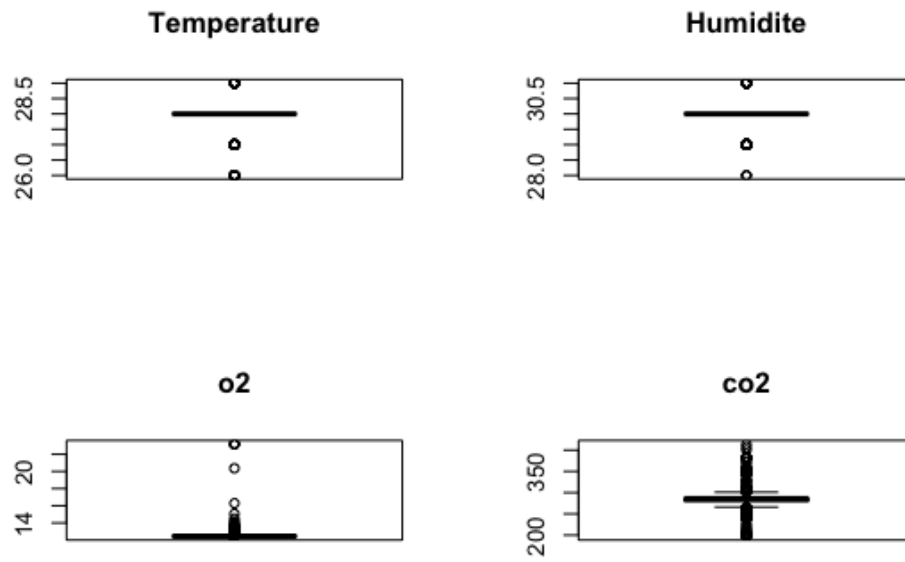


Figure 3. Boxplot des donnés

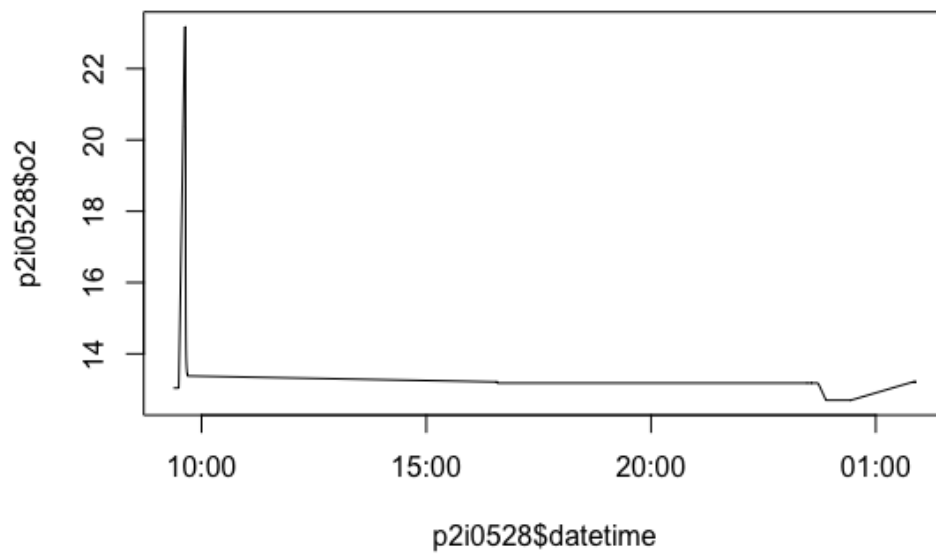


Figure 4. Concentration o2 le 28/05 en fonction du temps

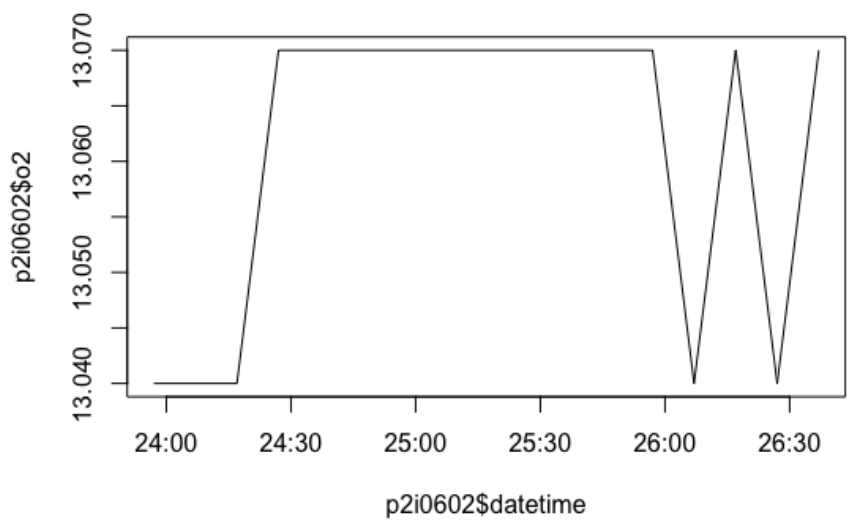


Figure 4. Concentration o2 le 02/06 en fonction du temps

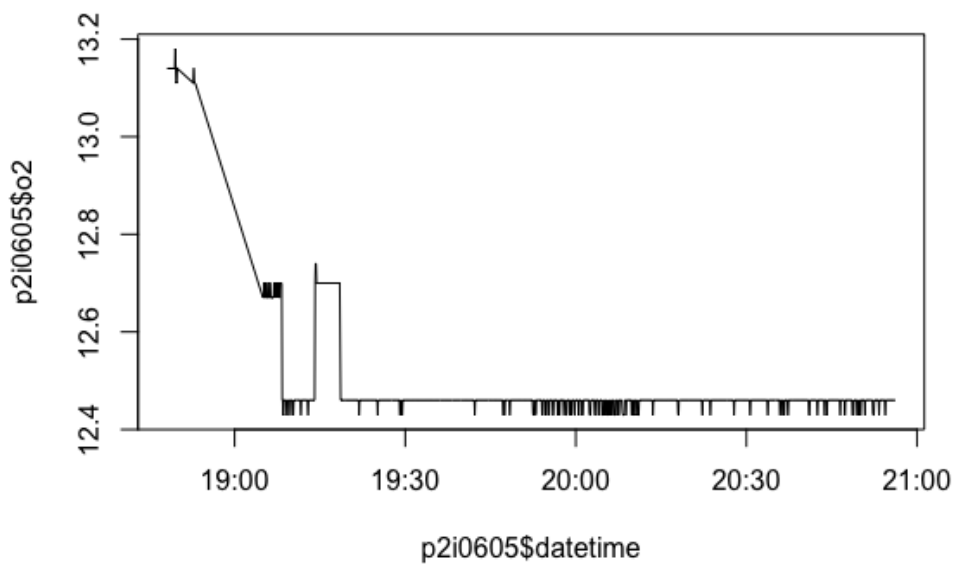


Figure 4. Concentration o2 le 05/06 en fonction du temps

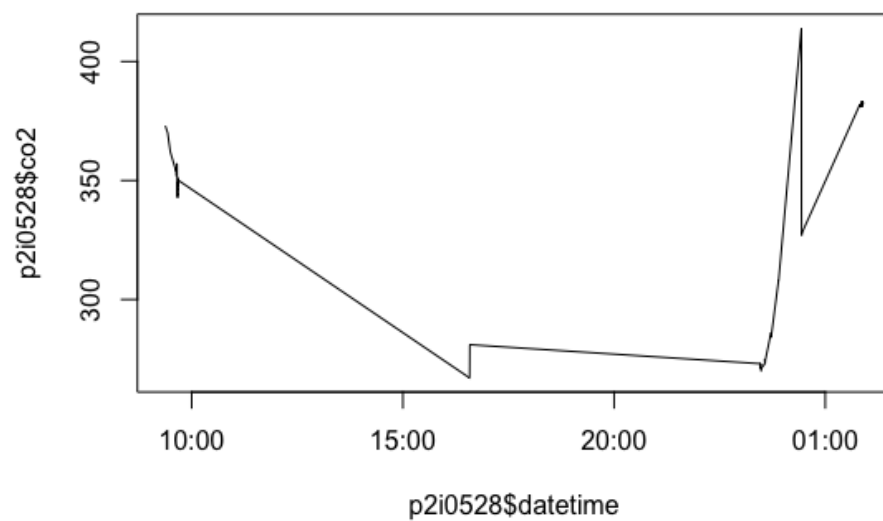


Figure 4. Concentration co2 le 28/05 en fonction du temps

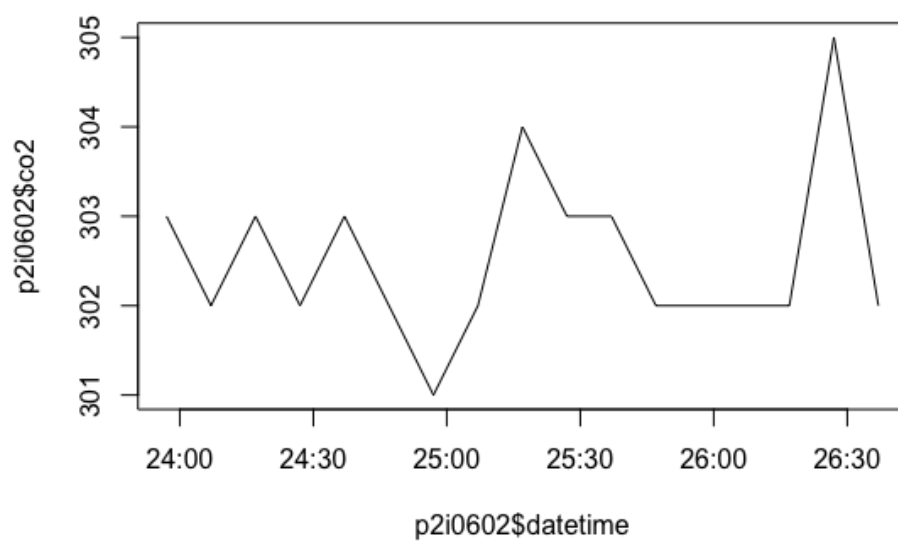


Figure 4. Concentration co2 le 02/06 en fonction du temps

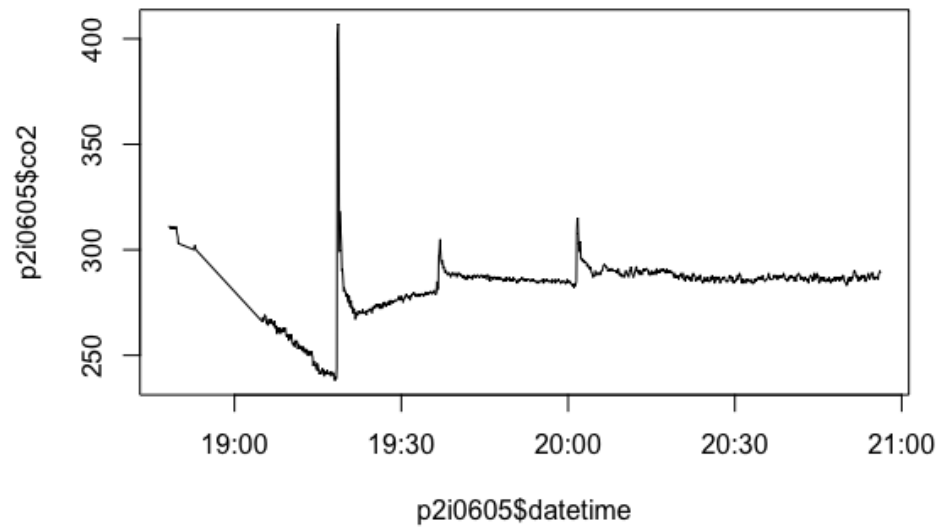


Figure 4. Concentration co2 le 05/06 en fonction du temps

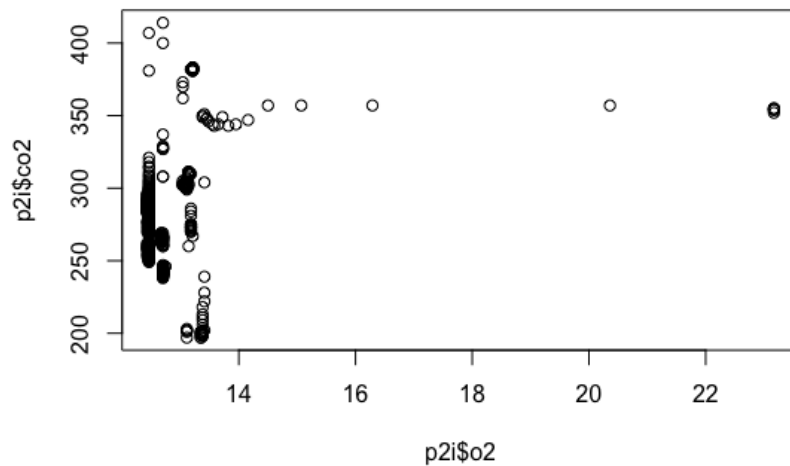


Figure 4. Concentration co2 en fonction de concentration O2

8) Difficulté rencontrés

- Nous avons besoins beaucoup de temps réfléchir aux méthodes et outils statistiques utilisés pour nos donnés afin de pouvoir l'interpréter
- Nos données couvrent seulement une courte période de temps chaque fois que nous allumons les capteurs.
- Il y avait des difficultés de formatage des dates