

## **Prédiction des embouteillages à Antananarivo (Tana) en fonction des variables explicatives de chaque route.**

**Objectif :** Créer un modèle de prédiction des embouteillages pour différents noms de routes à Antananarivo, où l'utilisateur entre le nom d'une route et l'IA prédit le niveau de congestion sur cette route (trop embouteillé, embouteillage moyen, pas d'embouteillage), en tenant compte des facteurs influençant le trafic.

### **Variables explicatives :**

1. **Heure de la journée** (ex : 7h, 17h)
2. **Jour de la semaine** (ex : lundi, vendredi)
3. **Température** (de l'air et ressentie)
4. **Humidité** (niveau d'humidité de l'air)
5. **Niveau de pollution de l'air**
6. **Événements spéciaux** (manifestations, fêtes, conférences)
7. **Accidents de la route** (nombre d'accidents dans la zone)
8. **Conditions météorologiques** (pluie, neige, etc.)
9. **Volume de trafic précédent** (volume de trafic des heures précédentes)
10. **Proximité des zones commerciales ou marchés** (congestion autour des zones à forte activité commerciale)
11. **Mauvais état de la route** (ex : zones de routes endommagées, travaux, nids-de-poule)

### **Démarche :**

1. **Dataset :** Il faudra créer un **dataset** avec les 50 routes de Tana, où chaque route a des informations sur les variables explicatives. Pour chaque route, on aura des données historiques de trafic (ex : niveau d'embouteillage à différentes heures), ainsi que les valeurs associées aux autres variables (température, pollution, état de la route, etc.). Cela pourrait être un fichier `.csv` ou une base de données.
2. **Prédiction :** L'utilisateur entre un **nom de route**, et le modèle analyse les variables explicatives associées à cette route pour déterminer le niveau de congestion. Le modèle peut prédire :
  - 🚦 **"Trop embouteillé"** si le volume de trafic est très élevé.
  - 🚦 **"Embouteillage moyen"** pour un trafic modéré.

🚦 "Pas d'embouteillage" si le trafic est faible.

## Format des données (dataset) :

Le dataset pourrait avoir cette structure :

	A	B	C	D	E	F	G	H	I	J	K
1	Route	Heure	Jour de la semaine	Température	Humidité	Pollution	Accidents	Conditions météorologiques	Volume de trafic	État de la route	Niveau d'embouteillage
2	Route A	7h	1 (lundi)	28°C	60%	35 µg/m³	2	Pluie	500	Mauvais état	Trop embouteillé
3	Route B	17h	5 (vendredi)	30°C	70%	50 µg/m³	0	Soleil	1000	Bon état	Embouteillage moyen
4	...	...	...	...	...	...	...	...	...	...	...

## Étapes de l'exercice :

1. **Préparation du dataset** : Collecter les données historiques pour les 50 routes, en incluant les variables explicatives pertinentes et le niveau d'embouteillage pour chaque heure/jour.
2. **Prétraitement des données** : Nettoyer les données, traiter les valeurs manquantes, normaliser les variables, et préparer les données sous forme de séquences temporelles adaptées à l'apprentissage par LSTM.
3. **Création du modèle LSTM** : Entraîner un modèle LSTM avec les données pour prédire les niveaux de congestion.
4. **Prédiction** : Lorsque l'utilisateur entre le nom d'une route, le modèle effectue une prédiction basée sur les variables explicatives associées à cette route et retourne le niveau d'embouteillage (par exemple : "Trop embouteillé", "Embouteillage moyen", "Pas d'embouteillage").

Il s'agit essentiellement d'un problème de prédiction basé sur des séries temporelles (pour chaque route) et l'apprentissage supervisé avec un modèle LSTM. Le défi majeur sera la **préparation du dataset**, car il faudra collecter suffisamment de données pour chaque route et pour chaque variable explicative. Une fois ces données collectées et traitées, le modèle pourra faire des prédictions sur les embouteillages.

**Ce projet pourrait avoir un grand impact sur l'optimisation du trafic à Tana et pourrait être utile pour les autorités locales, les entreprises de transport et les habitants.**