



DataScientest • com

SHIELD

Safety Hazard Identification and
Emergency Law Deployment

CAHIER DES CHARGES

1) Contexte et objectifs

Problématique

La **sécurité routière** figure parmi les **missions** assurées par les forces de l'ordre (FO : gendarmerie nationale, sécurité publique, Préfecture de police de Paris, Compagnie Républicaine de Sécurité), aux côtés d'autres acteurs publics. [1] En effet, tout usager impliqué dans un accident corporel de la circulation routière survenu sur le réseau routier ouvert à la circulation publique et impliquant au moins un véhicule doit en avertir les FO, qui dépêchent une unité sur place pour effectuer un constat. [2, 3]

Afin d'optimiser l'allocation des ressources, il est utile de **prioriser les interventions** en fonction de la **gravité** des accidents, établie à partir des conséquences médicales. Actuellement, à notre connaissance, lorsqu'un accident est signalé aux FO, ils en déterminent la gravité de manière « subjective » à partir des informations reçues ; l'appréciation pourrait alors varier d'un agent à un autre selon différents facteurs (p. ex., expérience, connaissance du terrain).

Solution

Nous proposons de remplacer la démarche ci-dessus par **une évaluation objective, précise et rapide de la gravité des accidents**, permettant ainsi de fluidifier leur gestion, rationaliser la prise de décision et maximiser l'impact des interventions.

Cette évaluation sera effectuée par une application baptisée **SHIELD** (pour « Safety Hazard Identification and Emergency Law Deployment »), dont le fonctionnement se décline en **5 étapes principales** (**Table 1**).

Table 1. Le fonctionnement de SHIELD en 5 étapes principales.

Étape	Tâche	Acteur
E1	Saisie des caractéristiques de l'accident dans l'interface	Agent des FO
E2	Transmission des caractéristiques à un modèle de machine learning	API
E3	Réalisation d'une prédiction de la gravité de l'accident	Modèle
E4	Transmission de la prédiction à l'interface	API
E5	Affichage de la prédiction	Interface

En plus d'améliorer la prise en charge des accidents **en amont**, SHIELD allégera aussi la charge administrative des agents **en aval**. En effet, pour chaque accident corporel, les FO doivent remplir un Bulletin d'Analyse des Accidents Corporels (BAAC). [3] Aujourd'hui, ce processus long et complexe commence avec la saisie d'une première version dite « pré-BAAC », comportant les 10 premiers champs obligatoires du BAAC, **après** l'envoi d'une unité sur place. SHIELD profite de la saisie des caractéristiques de l'accident effectuée à l'étape E1—correspondant aux 28 premiers champs du BAAC—pour pré-remplir le pré-BAAC **avant** l'envoi d'une unité sur place : un gain de temps pour les FO, qui pourront ainsi finaliser leurs BAAC plus rapidement.

Table 2 présente un comparatif des deux flux.

Table 2. Comparatif entre le statu quo et SHIELD.

Étape	Objet	Statu quo	SHIELD
F1	Accident	Signalement aux FO	Signalement aux FO
F2	Caractéristiques	-	Saisie dans l'application (E1)
F3	Évaluation de la gravité	Subjective	Objective (E3)
F4	Équipe dépêchée	Généraliste	Adaptée
F5	Pré-BAAC	Saisie (10 champs)	Vérification (28 champs)
F6	BAAC	Finalisation	Finalisation

Hypothèse de départ

Afin de faciliter la réalisation de l'application compte tenu de diverses contraintes, nous partons de l'hypothèse que l'agent des FO qui reçoit le signalement à l'étape

E1 soit en mesure de recueillir les 28 champs qui serviront de variables d'entrée au modèle.

Commanditaire

Unité de coordination de la lutte contre l'insécurité routière (UCLIR)

Utilisateurs

FO responsables de la gestion du trafic routier

Administrateur

Une équipe composée de data scientists et de machine learning engineers est chargée du développement, du déploiement et de la maintenance de l'application.

Contexte d'intégration

L'application sera déployée sur un serveur cloud.

Support d'utilisation

Interface graphique accessible de manière sécurisée *via* un navigateur web et implémentée *via* Streamlit

2) Modèle

Type de modèle

Modèle de classification utilisant des algorithmes d'apprentissage Random Forest

Fonctionnement

Le modèle analyse les caractéristiques d'un accident (28 variables d'entrée ; cf. annexe A) pour en prédire la priorité (variable cible) selon **Table 3**.

Table 3. Modalités de la gravité des accidents.

BAAC Rubrique « gravité »		Modèle Conversion	Modèle Prédiction	
<i>Code</i>	<i>Signification (état usager)</i>	<i>Code converti</i>	<i>Code binarisé</i>	<i>Signification (priorité intervention)</i>
1	Indemne	1	NL	Prioritaire
2	Tué	3	1	Prioritaire
3	Blessé hospitalisé	4	1	Prioritaire
4	Blessé léger	2	0	Non prioritaire

La signification de l'état des usagers est définie par l'arrêté du 27 mars 2007 relatif aux conditions d'élaboration des statistiques relatives aux accidents corporels de la circulation [4] :

- **Indemne** : impliqués non décédés et dont l'état ne nécessite aucun soin médical (du fait de l'accident)
- **Tué** : personnes qui décèdent du fait de l'accident, sur le coup ou dans les 30 jours qui suivent l'accident
- **Blessé hospitalisé** : victimes hospitalisées plus de 24 heures
- **Blessé léger** : victimes ayant fait l'objet de soins médicaux mais non admises à l'hôpital ou ayant été admises à l'hôpital 24 heures au plus

Métriques d'évaluation

Notre variable cible est la priorité de l'intervention, dans le cadre d'une classification binaire. Elle est déséquilibrée: 66% des accidents sont non prioritaires (catégorie blessés légers), 34% sont prioritaires (blessés hospitalisés, tués). Les prédictions sont globalement meilleures sur la classe négative (intervention non prioritaire, f1-score = 0.84) que sur la classe positive (f1-score = 0.65). Ainsi, plutôt que de choisir une métrique d'accuracy pour suivre l'évolution de la qualité des prédictions du modèle, nous choisissons de mesurer sa qualité globale via le **f1-score macro avg** (ici à 0.74).

Nous suivrons également le **temps d'entraînement** et le **temps de prédiction**.

3) Bases de données

Bases de données utilisées

- **Accidents** : [bases de données annuelles des accidents corporels de la circulation routière](#) fournies par [l'Observatoire national interministériel de la Sécurité routière \(ONISR\)](#)
- **Utilisateurs** : user_db.json (*provisoire*)
- **Logs** : un fichier *.jsonl pour chacun des endpoints 4 à 9 de l'API

Gestion des données

- **Par l'ONISR** : mise à jour annuelle de la base de données des accidents
- **Par SHIELD** : extraction de 4 fichiers CSV correspondant aux 4 rubriques du BAAC (caractéristiques, lieux, véhicules, usagers) pour chaque année depuis 2005 pour l'entraînement / le ré-entraînement du modèle
- **Par SHIELD** : utilisation de pipelines de prétraitement pour nettoyer et transformer les données

4) API

Fonctionnalités

- Interaction avec les bases de données pour récupérer les données d'entrée pour l'entraînement / le ré-entraînement du modèle
- Utilisation du modèle pour prédire la gravité des accidents (E2, E4)
- Enregistrement des prédictions dans des logs
- Authentification des utilisateurs

Endpoints

1. **/status** : vérification du fonctionnement de l'API
2. **/register** : inscription d'un utilisateur
3. **/remove_user** : suppression d'un utilisateur
4. **/predict_from_test** : prédictions de priorité à partir des données test
5. **/predict_from_call** : prédictions de priorité à partir de données saisies
6. **/train** : entraînement du modèle avec de nouvelles données
7. **/update_data** : mise à jour de la base de données des accidents
8. **/label_prediction** : labellisation d'une prédiction enregistrée
9. **/update_f1_score** : mise à jour du F1 score

5) Testing et monitoring

Tests unitaires

- Bon fonctionnement du modèle lors de l'entraînement
- Bon fonctionnement du modèle lors de la prédiction
- Bon fonctionnement des endpoints de l'API
- Bon fonctionnement du processus d'ingestion de nouvelles données

Monitoring

- Évaluation de la performance du modèle sur l'ensemble du jeu de test
- Ré-entraînement périodique du modèle si les performances sont insuffisantes
- Ré-entraînement sur un échantillon des données les plus récentes pour optimiser les ressources

6) Schéma d'implémentation

Composants

- **GESTION DES UTILISATEURS**
 - **Base de données des utilisateurs** : stocke les identifiants des utilisateurs
 - **Authentification** : authentifie les utilisateurs
 - **Autorisation** : autorise les utilisateurs
 - **FO** : accèdent à l'application pour obtenir des prédictions sur la gravité des accidents
 - **Admins** : gèrent et surveillent l'application
- **MATIÈRE PREMIÈRE**
 - **Base de données des accidents** : stocke les données sur les accidents
- **MLOPS**
 - **Modèle** : effectue des prédictions sur la gravité des accidents
 - **Transformation** des données saisies
 - **Preprocessing** des nouvelles données d'entraînement
 - **Entraînement** du modèle
 - **Validation** des performances du modèle
 - **Déploiement** du nouveau modèle
 - **Testing** : effectue des tests
 - **Monitoring** : surveille l'API et le modèle
- **COMMUNICATION**
 - **API**
- **FRONTEND**
 - **Interface** : pour interagir avec l'application

Figure 1 présente le schéma d'implémentation de SHIELD.

7) Références

- [1] « Les acteurs publics de la Sécurité routière ». Consulté le : 1 mars 2024. [En ligne]. Disponible [ici](#).
- [2] C. route, art. R. 231-1.
- [3] « Guide de production du fichier BAAC ». 16 décembre 2021. Consulté le : 1 mars 2024. [En ligne]. Disponible [ici](#).
- [4] A., 27 mars 2007, relatif aux conditions d'élaboration des statistiques relatives aux accidents corporels de la circulation, NOR : EQU0700567A.

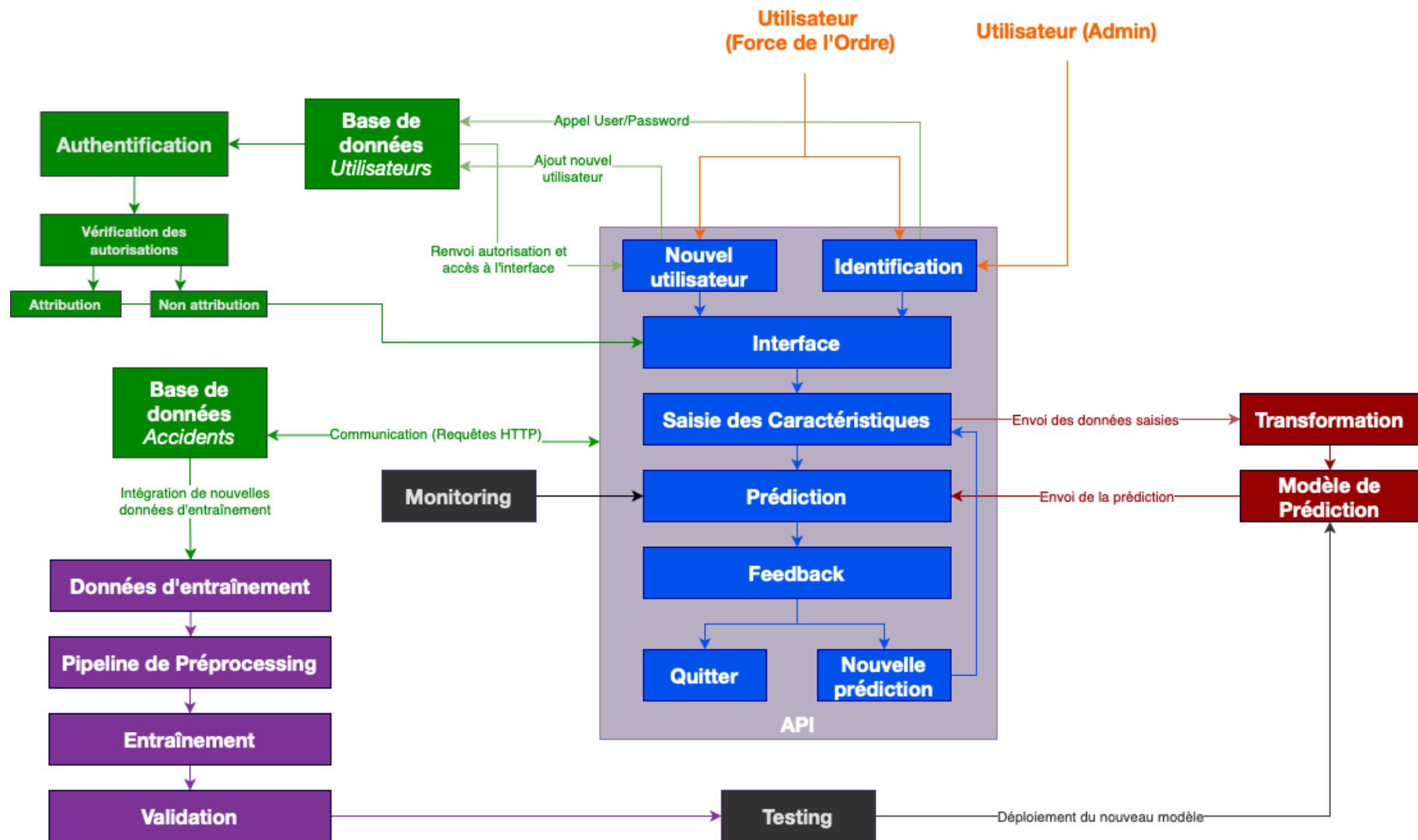


Figure 1. Schéma d'implémentation de SHIELD.