

Into the cloud... inspiré d'une histoire vraie

La protection intelligente contre le cyberharcèlement



Bodyguard



90 %

des contenus haineux sont détectés par l'application



2 %

c'est la marge d'erreur de l'algorithme (faux positifs)



+ de 2 millions

de commentaires haineux ont été supprimés en 20 mois

Le contexte

Charles Cohen a débuté la programmation informatique à l'âge de 10 ans. Onze ans plus tard, il lance sa première application mobile, baptisée Bodyguard. Derrière une idée en apparence simple se cache une mission complexe : celle de protéger les internautes en temps réel contre le cyberharcèlement.

Mais pourquoi a-t-il choisi de relever ce défi ? Simplement car aucune application de ce type n'était alors disponible. Quant à la modération des différentes plateformes, elle ne se révélait pas suffisamment efficace.

« Je n'ai jamais été victime de cyberharcèlement à proprement parler, mais j'ai grandi avec les réseaux sociaux. J'ai pu observer les dégâts engendrés par la haine en ligne. Cette dernière entraîne une restriction de la liberté d'expression et c'est ce dont j'ai souffert durant mon adolescence. Voilà pourquoi je n'ai jamais osé créer sur Internet ou m'exposer publiquement, de peur de me faire harceler. »

Charles Cohen, fondateur et CEO de Bodyguard

Le défi

Être capable d'analyser le contexte dans lequel est formulé un commentaire, ainsi que de déterminer la ou les personnes à qui il s'adresse.

Il fallait que la technologie Bodyguard puisse saisir et interpréter les états d'esprit. Une couche d'intelligence artificielle était donc indispensable pour réduire les faux positifs (des contenus détectés comme haineux alors qu'ils ne le sont pas) et augmenter la précision.

« Cette technologie devait être capable de comprendre l'ironie, le sarcasme ou encore l'humour. Le modèle prédictif, réalisé avec la plateforme de machine learning OVHcloud AutoML, m'a beaucoup aidé dans ce sens. »

Charles Cohen, fondateur et CEO de Bodyguard

Le modèle prédictif devait également permettre à la technologie d'apprécier la relation qui unit deux individus. Par exemple, l'auteur d'un commentaire est-il « abonné » à la personne à laquelle il répond ? Ceci a nécessité de la recherche, ainsi que le croisement de plus de 80 métadonnées. Parmi celles-ci, figurent le temps de réaction après publication, le pourcentage de majuscules ou encore la photo de profil.

Pour ce projet innovant, il a aussi fallu trouver l'algorithme adéquat parmi ceux proposés par scikit-learn. Cette bibliothèque open source offre des algorithmes écrits majoritairement en Python, destinés au machine learning.

Enfin, dans le cahier des charges, il était question de précision. Le modèle prédictif à réaliser ne devait pas dépasser un taux d'erreur de 10 %.



La solution

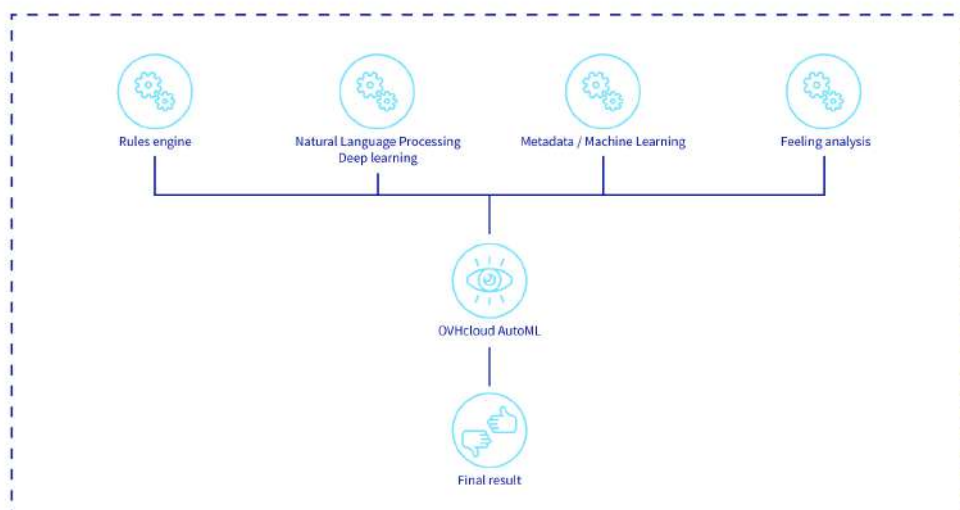
Un service managé, facile à utiliser et qui permet d'accélérer la mise en production.

La couche logicielle

Charles a choisi d'utiliser OVHcloud AutoML, une plateforme d'apprentissage automatique, distribuée et évolutive. Cette solution Software as a Service (SaaS) lui a permis d'automatiser les processus de création, de déploiement ainsi que de requête des modèles de machine learning. Elle lui a également donné la possibilité d'intégrer des algorithmes open source, tels que ceux proposés par scikit-learn.

De plus, OVHcloud AutoML a accéléré la phase de développement. Dix jours ont été nécessaires pour la création du modèle prédictif de Bodyguard et vingt pour l'élaboration du modèle de méta-learning, qui analyse la relation entre l'auteur du contenu et celui du commentaire.

Grâce à ces modèles, la détection de la technologie Bodyguard a augmenté de 12,5 %, passant de 80 à 90 %. Le nombre de faux positifs a lui diminué de 50 %, passant de 6 à 3 %.



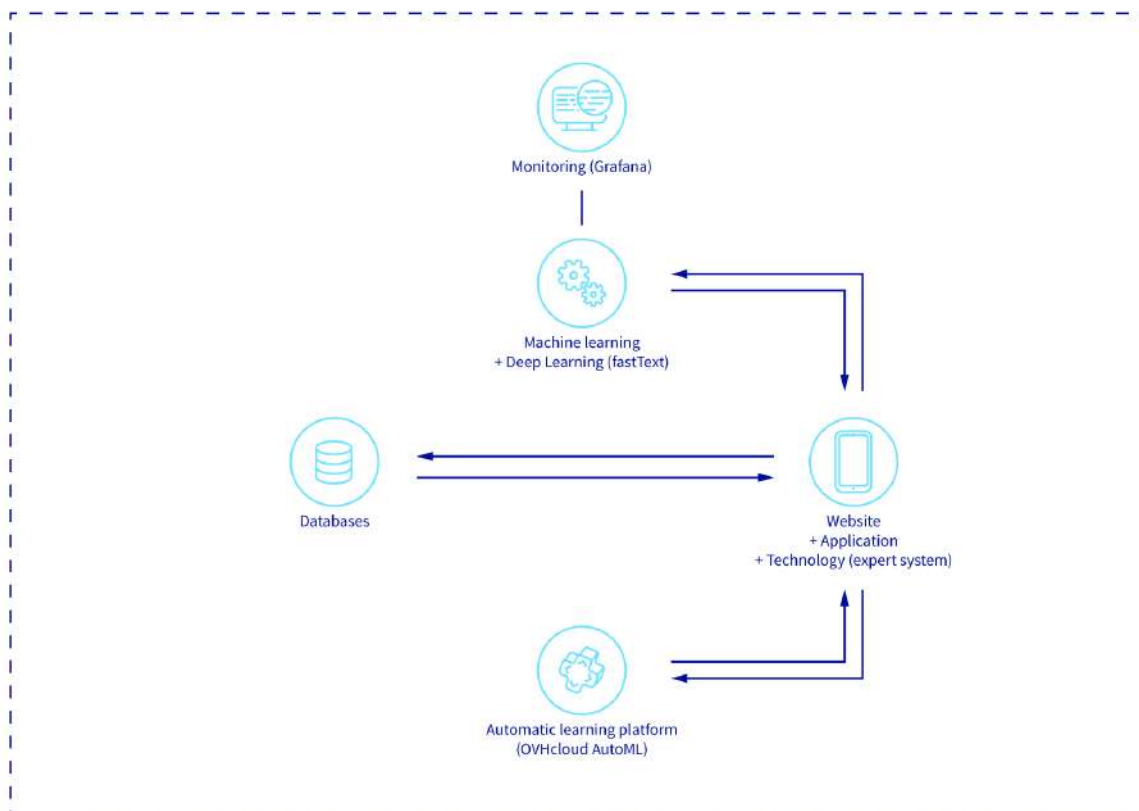
Côté monitoring, Charles a opté pour Logs Data Platform associé au logiciel Grafana. Il peut ainsi surveiller les performances de son infrastructure et de ses bases de données. Il s'en sert également pour mesurer les indicateurs clés de performance (KPI) : nombre d'utilisateurs, volume de contenus haineux supprimé en temps réel, quantité de requêtes API, etc.

La couche matérielle

L'infrastructure de Bodyguard est composée de trois instances Public Cloud :

- une pour les bases de données ;
- une autre pour la technologie et les modèles de machine learning ;
- une dernière pour les robots chargés de faire fonctionner l'application mobile, en récupérant les commentaires et en les analysant grâce à la technologie.

Et pour réaliser ses sauvegardes, Charles utilise un autre service proposé par le Public Cloud d'OVH : Cloud Archive. Ce dernier lui permet de conserver ses informations sur le long terme à un coût réduit, tout en garantissant la sécurité ainsi que la récupération de ses données.



Le résultat

Il aura fallu deux ans à Charles pour développer l'algorithme d'apprentissage final et l'intégrer dans une application mobile gratuite, disponible sur Android et iOS depuis octobre 2017. Aujourd'hui, Bodyguard supprime en temps réel les commentaires haineux sur YouTube, Instagram, Twitter, Twitch et Mixer.

En juillet 2019, ce garde du corps virtuel a séduit plus de 40 000 utilisateurs et atteint un taux de satisfaction de 97 %. Quelques raisons de ce succès :

- 90 % des contenus haineux détectés par l'application ;
- 2 % seulement de marge d'erreur (faux positifs) ;
- + de 2 millions de commentaires haineux supprimés en 20 mois.

Prochainement, l'application sera traduite en anglais et en espagnol. Une nouvelle solution nommée « Bodyguard pour les familles » verra également le jour pour alerter immédiatement les parents en cas de cyberharcèlement de leurs enfants.

À terme, Charles aspire à se positionner en tant que cloud provider de solutions de modération automatique par l'IA. Pour cela, il mettra à disposition sa technologie sous le nom de « Bodyguard pour les entreprises », via une API. Elle s'adresse à tous ceux qui souhaitent se protéger, protéger leurs utilisateurs, leur image, leur réputation et leurs employés.

« À ce propos, notre plateforme à destination des développeurs (developers.bodyguard.ai) est déjà disponible pour permettre à chacun d'utiliser notre technologie. »

Charles Cohen, fondateur et CEO de Bodyguard

OVHcloud est un fournisseur mondial de cloud hyperévolutif (hyperscale) qui offre aux entreprises une valeur et des performances de référence dans le secteur. Fondé en 1999, le groupe gère et entretient 30 datacenters sur quatre continents, déploie son propre réseau mondial de fibre optique et contrôle l'ensemble de la chaîne d'hébergement. S'appuyant sur ses propres infrastructures, OVHcloud propose des solutions et des outils simples et puissants qui mettent la technologie au service des entreprises tout en révolutionnant la façon dont travaillent nos plus d'un million de clients à travers le monde. Le respect des personnes, la liberté et l'égalité des chances pour l'accès aux nouvelles technologies ont toujours été des principes solidement ancrés dans l'entreprise. « *Innovation for freedom* ».