Into the cloud... based on a true story

Inteligentna ochrona przed cyberprzemocą

OVHcloud

Bodyguard

90%
treści zawierających mowę nienawiści wykrywanych przez aplikację

2%
to margines błędu algorytmu (wyniki fałszywie pozytywne)

ponad 2 miliony
komentarzy zawierających mowę nienawiści zostało skasowanych w ciągu 20 miesięcy

W skrócie

Charles Cohen zajął się programowaniem w wieku 10 lat. Jedenaście lat później uruchomił swoją pierwszą aplikację mobilną o nazwie Bodyguard. Za pozornie prostą ideą kryje się ambitna misja: ochrona użytkowników Internetu w czasie rzeczywistym przed cyberprzemocą.

Skąd przyszedł mu do głowy taki pomysł? Odpowiedź jest bardzo prosta - na rynku nie istniała aplikacja tego typu. Wiele pozostawiała również do życzenia skuteczność moderowania różnych platform.

> *"Nigdy nie doświadczyłem osobiście cyberprzemocy, ale byłem świadkiem wielu sytuacji na portalach społecznościowych, w których ludzie wyrządzali sobie krzywdę, stosując mowę nienawiści. Mowa nienawiści ogranicza również swobodę wypowiedzi i z tego powodu cierpiałem jako nastolatek. Nigdy nie ośmieliłem się publikować niczego w Internecie, ponieważ obawiałem się, że stanę się celem hejtu."*
>
> **Charles Cohen, założyciel i CEO Bodyguard**

Wyzwanie

Analiza kontekstu komentarza oraz identyfikacja osoby/osób, do których jest on kierowany.

Opracowaliśmy technologię Bodyguard w taki sposób, aby nasze rozwiązanie wychwytywało i interpretowało wydźwięk emocjonalny wypowiedzi. Warstwa sztucznej inteligencji była zatem niezbędna, aby ograniczyć wyniki fałszywie pozytywne (treści wykryte jako mowa nienawiści, podczas gdy są one neutralne) i zwiększyć dokładność wyszukiwania oraz interpretacji.

> *"System powinien rozumieć ironię, sarkazm lub żart. Bardzo mi w tym pomógł model predykcyjny opracowany przy użyciu platformy AutoML od OVHcloud służącej do obsługi projektów Machine Learning."*
>
> **Charles Cohen, założyciel i CEO Bodyguard**

Zgodnie z naszym założeniem model predykcyjny powinien również umożliwiać systemowi ocenę rodzaju relacji między dwiema osobami. Na przykład: czy autor komentarza „obserwuje" osobę, której odpowiada? Wymagało to przeprowadzenia badań oraz porównania ponad 80 metadanych. Badane parametry to między innymi czas reakcji od momentu publikacji wpisu, procent tekstu zapisanego drukowanymi literami czy zdjęcie profilowe.

Konieczne było również wybranie odpowiedniego algorytmu spośród algorytmów udostępnianych przez scikit-learn. Scikit-learn to biblioteka open source oferująca algorytmy napisane w języku Python przeznaczone do zastosowań związanych z uczeniem maszynowym.

Wreszcie, bardzo istotnym wymaganiem była precyzja. Przyjęliśmy założenie, że poziom błędu w modelu predykcyjnym nie może przekroczyć 10%.

Rozwiązanie

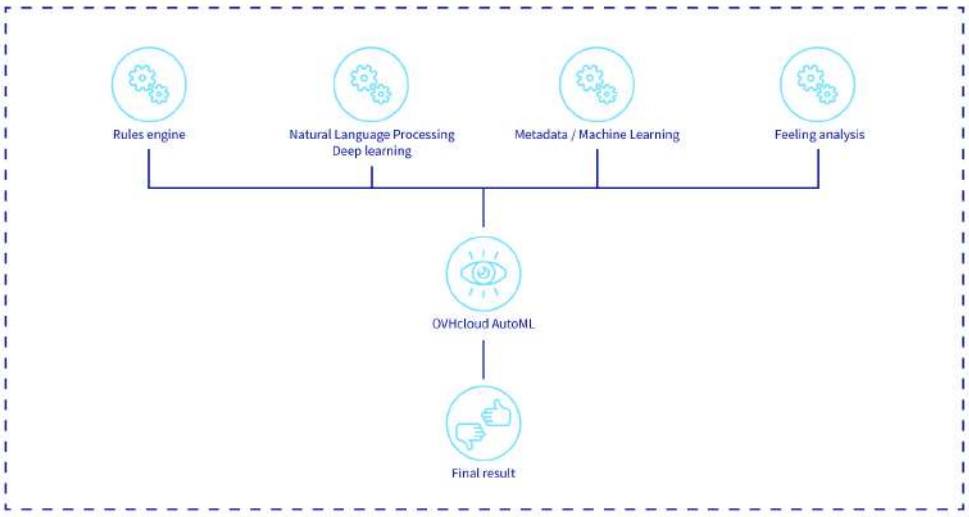
Zarządzana usługa - łatwa w użyciu i przyspieszająca wdrożenia.

Warstwa oprogramowania

Charles wybrał OVHcloud AutoML, platformę służącą do uczenia maszynowego - rozproszoną, skalowalną i działającą w modelu SaaS (Software as a Service). Dzięki temu rozwiązaniu Bodyguard zautomatyzował proces tworzenia i wdrażania modeli Machine Learning. Dało mu to również możliwość zintegrowania algorytmów, takich jak te udostępniane przez scikit-learn.

Ponadto, platforma OVHcloud AutoML przyczyniła się do znaczengo przyspieszenia fazy rozwoju. Bodyguard stworzył model predykcyjny w ciągu dziesięciu dni, natomiast opracowanie modelu metalearningowego, analizującego relacje między autorem treści a autorem komentarza, zajęło dwadzieścia dni.

Dzięki tym modelom poziom wykrywalności gwarantowany przez technologię Bodyguard wzrósł o 10% - z 80% do 90%, natomiast liczba wyników fałszywie pozytywnych spadła o 50% - z 6% do 3 %.



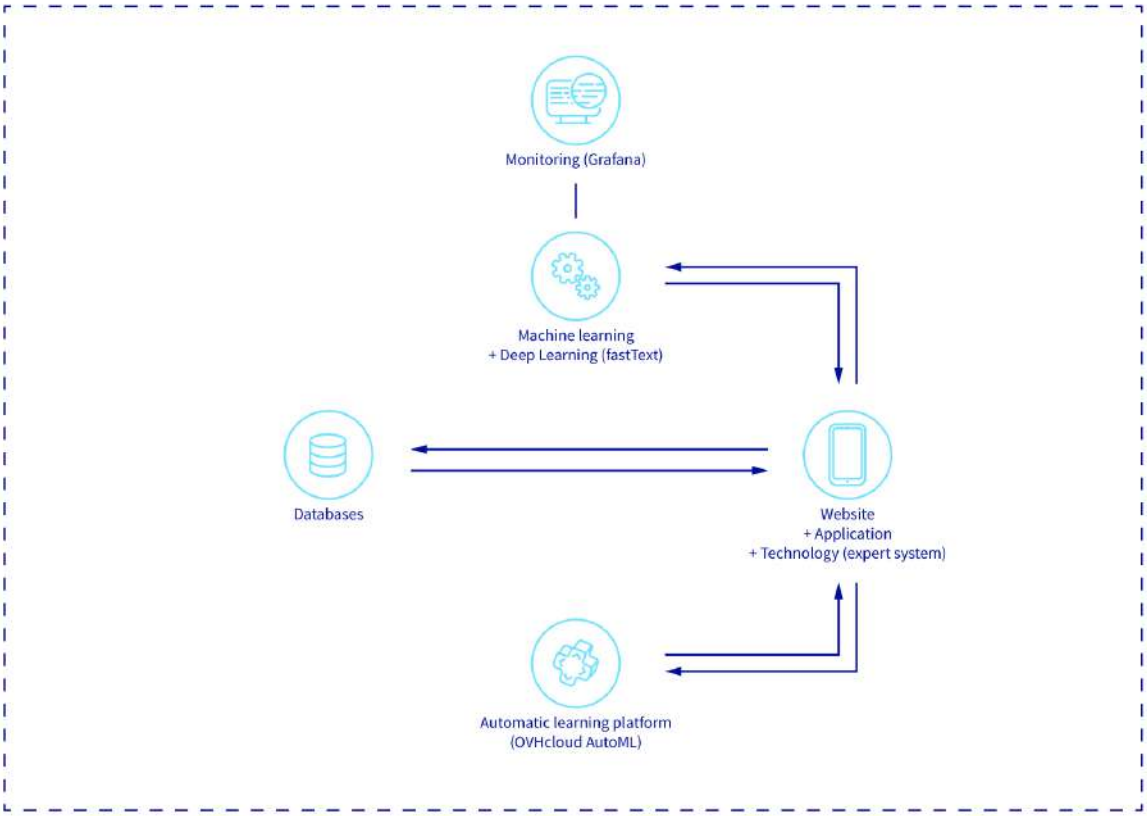
Jeśli chodzi o monitoring, Charles Cohen wybrał rozwiązanie Logs Data Platform powiązane z oprogramowaniem Grafana, dzięki czemu może śledzić wydajność infrastruktury oraz baz danych. Platforma służy również do pomiaru kluczowych wskaźników efektywności (KPI), takich jak: liczba użytkowników, liczba treści zawierających mowę nienawiści usuwanych w czasie rzeczywistym, liczba zapytań wysyłanych do API, itd.

Warstwa sprzętowa

Infrastruktura Bodyguard zbudowana jest z trzech instancji Public Cloud:

- jedna służy do zarządzania bazami danych
- druga do obsługi technologii oraz modeli Machine Learning
- trzecia do obsługi robotów odpowiedzialnych za uruchomienie aplikacji mobilnej poprzez pobieranie komentarzy i ich analizowanie.

Do wykonywania kopii zapasowych Charles Cohen wybrał usługę Public Cloud: **Cloud Archive**. Umożliwia ona długoterminowe przechowywanie danych w niższej cenie i daje gwarancję bezpieczeństwa oraz odzyskania danych w przypadku ich utraty.



Korzyści

Charles Cohen opracował w ciągu dwóch lat ostateczny algorytm uczenia maszynowego i zintegrował go z darmową aplikacją mobilną dostępną od października 2017 r. na urządzeniach z systemem Android i iOS. Dzisiaj Bodyguard usuwa w czasie rzeczywistym komentarze o negatywnym zabarwieniu w serwisach, takich jak YouTube, Instagram, Twitter, Twitch i Mixer.

W lipcu 2019 rozwiązanie to przyciągnęło ponad 40 000 użytkowników, a poziom zadowolenia wyniósł 97 %. Przyczyny tak spektakularnego sukcesu były następujące:

- 90% treści zawierających mowę nienawiści wykrytych przez aplikację
- zaledwie 2% marginesu błędu (wyniki fałszywie pozytywne)
- ponad 2 mln komentarzy usuniętych w ciągu 20 miesięcy.

W przyszłości aplikacja zostanie przetłumaczona na język angielski i hiszpański. Wprowadzone zostanie również nowe rozwiązanie o nazwie „Bodyguard for Families", które ma na celu natychmiastowe zaalarmowanie rodziców o cyberprzemocy stosowanej wobec ich dzieci.

Docelowo, Charles Cohen zamierza pozycjonować swoją firmę jako dostawcę rozwiązań w chmurze wykorzystujących sztuczną inteligencję do automatycznej moderacji. W tym celu udostępni swoją technologię pod nazwą „Bodyguard dla przedsiębiorstw" za pośrednictwem API. Rozwiązanie to przeznaczone jest dla wszystkich, którzy chcą chronić siebie, swoich pracowników oraz użytkowników swoich rozwiązań, a także swój wizerunek oraz reputację.

> *"W tej chwili dostępna już jest nasza platforma przeznaczona dla deweloperów (developers.bodyguard.ai), którzy mogą swobodnie korzystać z naszej technologii."*
>
> **Charles Cohen, założyciel i CEO Bodyguard**

OVHcloud jest globalnym i wiodącym w Europie dostawcą chmury, zarządzającym 400 000 serwerów w 30 własnych centrach danych na czterech kontynentach. Od dwudziestu lat Grupa wykorzystuje zintegrowany model, który zapewnia jej pełną kontrolę nad łańcuchem wartości: począwszy od projektowania własnych serwerów, poprzez zarządzanie należącymi do niej centrami danych, po budowanie i utrzymywanie własnej globalnej sieci światłowodowej. To unikatowe podejście umożliwia OVHcloud wspieranie, w sposób niezależny, wszystkich potrzeb 1,5 miliona klientów z ponad 130 krajów. OVHcloud oferuje klientom rozwiązania najnowszej generacji, łączące wysoką wydajność, przewidywalną cenę i pełną kontrolę nad danymi, wspierając w ten sposób ich nieograniczony rozwój. "Innovation for freedom".