

Radzenie sobie z trudnymi wyzwaniami

Omówimy szereg skomplikowanych wyzwań, których trudno uniknąć, a które będą wymagały odpowiedniego zestawu taktyk, aby skutecznie sobie z nimi radzić. Mówiąc dokładniej, pokażę Ci, co musisz zrobić, aby podejmować właściwe decyzje, jeśli chodzi o skuteczne i spójne pozyskiwanie danych i zarządzanie nimi, konfigurowanie środowiska nauki o danych, zarządzanie ograniczeniami prawnymi związanymi z danymi i algorytmami, których potrzebujesz dla Twojej firmy, a także przygotowanie się na gwałtowne zmiany w obszarze data science jako całości, które z pewnością nadejdą.

Pobieranie danych stamtąd do tego miejsca

Kiedy firma decyduje się rozpocząć podróż, aby stać się napędzaną danymi, koncentruje się naturalnie na samych danych, co nieuchronnie prowadzi do większej świadomości rzeczywistej różnorodności danych potrzebnych do uzyskania pełnej proaktywnej i opartej na danych kontroli nad bieżącą działalnością firmy. Co więcej, firmy szybko zdają sobie sprawę, że aby wyjść poza to, co jest obecnie możliwe, zbiory danych muszą stać się jeszcze bardziej zróżnicowane. W tym momencie wiele firm zaczyna zdawać sobie sprawę, że dane, które mają zasadnicze znaczenie dla rzeczywistego wykorzystania danych, mogą w rzeczywistości należeć do kogoś innego lub znajdować się w innym kraju, z innymi przepisami dotyczącymi danych. W tej sekcji wyjaśniono, jak strategicznie podejść do takich praktycznych wyzwań w ramach pozyskiwania danych.

Obsługa zależności na danych należących do innych

Radzenie sobie z zastrzeżonymi danymi jest nieuniknionym, ale możliwym do opanowania wyzwaniem, przed którym stoi każda firma dążąca do pełnego wykorzystania danych. Zwykle dzieje się tak, że zidentyfikowałeś i dokładnie określiłeś wszystkie potrzebne dane w swojej strategii danych, a kiedy zaczniesz zastanawiać się, jak strategicznie podejść do przechwytywania danych, zdajesz sobie sprawę, że masz problem z własnością danych. Jeśli korzystasz tylko z danych generowanych z Twojego wewnętrznego środowiska IT, masz oczywiście mniejszy problem. Jeśli jednak tak jest, to prawdopodobnie Twoja firma nie jest tak naprawdę oparta na danych we właściwym tego słowa znaczeniu. Firma oparta na danych wyjaśnia, w jaki sposób wykorzystywane są jej produkty i/lub usługi oraz jak działa w prawdziwym życiu ustawienia, nie tylko w środowisku laboratoryjnym. I za każdym razem, gdy zaczynasz korzystać z danych generowanych przez życie w prawdziwym świecie, napotykasz problem z własnością danych.

O jakich danych mówię? Przede wszystkim dotyczy to danych należących do Twoich klientów, ale może również obejmować dane należące do klientów Twoich klientów, w zależności od tego, w jakiej firmie się znajdujesz. Musisz poświęcić trochę czasu, aby naprawdę zrozumieć szczegółowy kontekst danych, które potrzeba. Może odnosić się do kwestii prywatności danych, ale nie musi. Może po prostu być tak, że dane, których potrzebujesz, aby lepiej zrozumieć wyniki lub potencjał Twojej firmy, należą do kogoś innego.

Nie zniechęcaj się, jeśli chodzi o kwestie własności. Większość sytuacji można rozwiązać z prawnego punktu widzenia, jeśli chcesz otwarcie zająć się nimi z właścicielami danych, wyjaśniając, dlaczego potrzebujesz danych i jak będziesz je traktować po ich posiadaniu. Wszystko sprowadza się do zdobycia zaufania co do tego, w jaki sposób i w jakim celu dane będą wykorzystywane. (Nie zaszkodzi również sprecyzować, w jaki sposób Twoja praca może, jeśli to możliwe, wnieść wkład z powrotem do właścicieli danych.) Pod koniec dnia musisz mieć absolutną pewność, że rozumiesz (i przestrzegasz) ograniczenia prawne mające zastosowanie do każdego rodzaju danych, z których zamierzasz korzystać. Korzystanie z danych musi być również regulowane w drodze umowy ze stroną będącą właścicielem

danych, w tym praw, jakie Twoja firma ma w związku z dostępem do danych, przechowywaniem i użytkowaniem w czasie.

Prawa i przepisy mają w zwyczaju zmieniać się w czasie. Ostatnio obserwuje się tendencję do dalszego zwiększania ograniczeń w celu ochrony prawa jednostki do własnych danych. Jednym z ostatnich przykładów jest dość restrykcyjne Ogólne Rozporządzenie o Ochronie Danych i Rozporządzenia (RODO) przyjęte przez Unię Europejską (UE), które weszło w życie w maju 2018 r. Biorąc pod uwagę niedawne informacje o niewłaściwym wykorzystywaniu danych przez podmioty takie jak Cambridge Analytica i Facebook, USA a Kanada zdecydowanie przygląda się przepisom podobnym do unijnego RODO. Wszystko, co pomaga chronić prawo jednostki do prywatności, jest najlepsze, ale pamiętaj tylko, że sposób, w jaki dziś traktujesz przepisy dotyczące prywatności, najprawdopodobniej będzie zupełnie inny w najbliższej przyszłości. Dlatego należy strategicznie i proaktywnie przemyśleć konfigurację infrastruktury i potrzeby dotyczące danych, aby zapewnić uwzględnienie tego typu ograniczeń w bieżącym i zmieniającym się środowisku nauki o danych.

Zarządzanie transferem danych i obliczeniami ponad granicami kraju

Jeśli Twoja firma ma oddziały w wielu różnych krajach lub prowadzi działalność (a tym samym ma wielu klientów) w wielu krajach, jednym z głównych wyzwań, z jakimi możesz się zmierzyć, jest zarządzanie danymi, które muszą przekraczać granice międzynarodowe. Musisz dokładnie rozważyć szereg różnych aspektów układanki danych, jeśli Twoja firma ma komponent międzynarodowy. Oto lista głównych obaw:

* **Legalność:** Ograniczenia prawne dotyczące przenoszenia danych przez granice to kwestia, której firma musi przestrzegać. Prawa i przepisy różnią się w zależności od kraju, więc możliwe są różne rozwiązania w zależności od kraju, w którym prowadzisz działalność. Ograniczenia są również różne w zależności od rodzaju danych, które wyprowadzasz z kraju. Dane zawierające dane osobowe są zwykle znacznie trudniejsze do przeniesienia niż dane niewrażliwe. Łamanie przepisów związanych z przesyłaniem danych może być dość kosztowne i może poważnie wpłynąć na markę firmy, jeśli zostanie stwierdzone, że naruszyłeś zaufanie klientów.

* **Podjęcie do przesyłania danych:** odnosi się do tego, w jaki sposób faktycznie wykonujesz transfer danych. Jest to zazwyczaj dość kosztowne, a także różni się w zależności od kraju. W zależności od ilości przesyłanych danych i częstotliwości przesyłania danych, możesz wynajmować przestrzeń w istniejącej infrastrukturze łączności i łączach danych lub - jeśli nie możesz spełnić swoich wymagań dotyczących takich aspektów jak pojemność, bezpieczeństwo lub wyłączność - zainwestować w własne linki.

* **Możliwości lokalnych obliczeń i przechowywania:** Jeśli możesz przechowywać dane i przeprowadzać analizę w kraju, w którym dane zostały przechwycone, możesz obniżyć koszty i zwiększyć szybkość dostawy. Jednak, aby ta konfiguracja działała wydajnie, musisz odpowiednio przemyśleć, jak będzie wyglądać Twoja rozproszona architektura obliczeniowa. Co zostanie zrobione gdzie? a gdzie będą przechowywane np. dane źródłowe? Czy będzie centralny punkt przechowywania danych i globalnej analizy, czy tylko rozproszone konfiguracje? To, jak odpowiesz na te pytania, zależy w dużej mierze od rodzaju prowadzonej działalności i tego, jak wygląda konfiguracja w różnych krajach.

Zarządzanie spójnością danych w całym środowisku nauki o danych

Zapewnienie spójności danych w różnych częściach środowiska nauki o danych może wydawać się prostym zadaniem, ale jest znacznie trudniejsze, niż się wydaje. Po pierwsze, ten obszar wydaje się być bardziej złożony niż powinien, pochłaniając więcej czasu i zasobów niż pierwotnie szacowano. Potrzeba spójności obejmuje takie aspekty, jak zarządzanie danymi i formaty danych, ale także spójne

etykietowanie danych - na przykład przy użyciu identyfikatorów klientów z wielu różnych źródeł, aby umożliwić korelację różnych typów danych powiązanych z tym samym klientem. Wyzwanie polega na tym, że istnieje wbudowana sprzeczność w zakresie infrastruktury między umożliwieniem korzystania ze specjalnych narzędzi, aby umożliwić naukowcom i inżynierom danych bycie innowacyjnym i wydajnym, a jednocześnie zapewniającym spójność danych. Dzieje się tak, ponieważ wyspecjalizowane narzędzia są zoptymalizowane pod kątem rozwiązywania określonych problemów, ale albo nie utrzymują spójnego formatu, albo nie współpracują dobrze z innymi narzędziami potrzebnymi w przepływie end-to-end. Zoptymalizowane, wyspecjalizowane narzędzia do uczenia maszynowego po prostu nie są dobre w graniu razem z innymi, podobnymi wyspecjalizowanymi narzędziami, które rozwiązują porównywalne lub inne sąsiadujące problemy. Ale czy to naprawdę takie złe? No cóż, może to prowadzić do prawdziwych problemów, w zależności od tego, jak wiele swobody w realizacji architektonicznej i między zespołami zostanie dopuszczona. Oto kilka przykładów problemów, które mogą wynikać z braku spójności w środowisku AI:

* Rozwiązania ad hoc: każda sprawa jest traktowana jako odosobniony problem, który musi zostać rozwiązany w tej chwili, aby zespół mógł iść do przodu. Wynik? Bez długoterminowego rozwiązania i bez uczenia się między zespołami.

* Zwiększony koszt: gdy musisz powielić możliwości narzędzia, aby poradzić sobie z brakiem spójności, lub gdy musisz wbudować możliwości w zakupione narzędzia, aby zapewnić tylko podstawową spójność, koszty te sumują się.

* Nie działa kompleksowo: niespójności mogą wystąpić, gdy infrastruktura jest wdrażana u kilku dostawców chmury, co utrudnia lub uniemożliwia przesyłanie danych i zachowanie spójności danych w różnych środowiskach zwirtualizowanych.

Ponieważ kierownictwo firmy nie może wymusić i może nie chcieć wymusić spójności danych w całej organizacji jako zasad firmy, musi użyć innych środków, aby zachować spójność danych od początku do końca. Jednym ze sposobów jest upewnienie się, że wszystkie zespoły przestrzegają właściwych i odpowiednich wytycznych dotyczących oceny i zakupu nowych narzędzi, które zawierają konkretne dyrektywy związane z spójnością danych. Wyraźnie motywujące, dlaczego jest to kluczem do pomyślnej realizacji strategii analizy danych. Ważne jest również, aby zastanowić się, jakie limity są potrzebne dla każdej firmy, w zależności od rodzaju działalności, jej celów i tak dalej. Trzymaj się linii, jeśli chodzi o spójność danych: w przeciwnym razie możesz skończyć z kłopotliwą i kosztowną implementacją nauki o danych, daleko odbiegającą od produktywnego środowiska nauki o danych, na które liczyłeś.

Zapewnienie wyjaśnialności w AI

Wyjaśnialna sztuczna inteligencja (XAI), zwana również przezroczystą sztuczną inteligencją, obejmuje zdolność do wyjaśnienia, w jaki sposób algorytm osiągnął konkretny wgląd lub wniosek, który skutkuje podjęciem określonej decyzji o podjęciu działania. Chociaż jest to ważny aspekt, który należy wziąć pod uwagę w ramach ewolucji sztucznej inteligencji, nie jest to łatwe do rozwiązania technicznie, zwłaszcza jeśli sztuczna inteligencja działa w czasie rzeczywistym, a tym samym wykorzystuje dane strumieniowe, które nie zostały zapisane. Aby przybliżyć ten punkt do domu, wyobraź sobie, jeśli chcesz, że nie możesz tego wyjaśnić twojemu klientowi, dlaczego maszyna podjęła określoną decyzję - decyzję, której nie podjąłbyś na podstawie własnego doświadczenia. Co wtedy mówisz klientowi? Zajmowanie się sztuczną inteligencją, którą można wyjaśnić, staje się coraz ważniejsze z punktu widzenia naszej ludzkiej zdolności do lepszego zrozumienia, dlaczego i jak sztuczna inteligencja działa w określony sposób. Innymi słowy, co można zrozumieć, badając, w jaki sposób maszyna uczy się, przetwarzając te ogromne ilości danych z wielu wymiarów, szukając określonych wzorców lub odchyleń? Co takiego

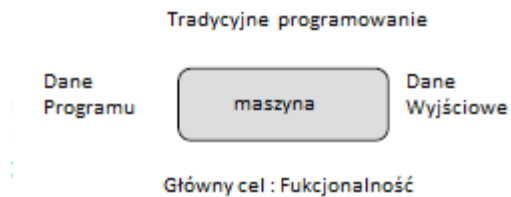
maszyna wykrywa i rozumie, że przeoczyłeś lub zinterpretowałeś inaczej lub po prostu nie byłeś w stanie wykryć? Jakie wnioski można z tego wyciągnąć? Z etycznego punktu widzenia wyjaśnialność AI będzie jeszcze ważniejsza, gdy analitycy danych zaczną budować bardziej zaawansowaną sztuczną inteligencję, w której działa wiele różnych algorytmów. Będzie to klucz do dokładnego zrozumienia, co maszyny interpretują, a także jak przebiega proces podejmowania decyzji przez maszynę. Znajomość tych informacji ma kluczowe znaczenie dla utrzymania się na szczycie ram polityki niezbędnych do ustalenia granic tego, co maszyna powinna, a czego nie powinna robić, a także tego, w jaki sposób te polityki należy rozszerzyć lub być może ograniczyć w przyszłości. Z jednej strony z czysto egzystencjalnej perspektywy, a z drugiej z potrzeby utrzymania przez ludzi kontroli nad inteligentnymi maszynami, które są budowane, nie można po prostu postrzegać sztucznej inteligencji jako czarnej skrzynki. (Wyzwanie związane z czarną skrzynką w sztucznej inteligencji odnosi się do potrzeby zapewnienia, że gdy algorytm podejmuje decyzję w oparciu o techniki zastosowane do trenowania algorytmu, proces podejmowania decyzji musi być przejrzysty dla ludzi. Przejrzystość algorytmu jest możliwa, gdy wiele z bardziej podstawowych technik uczenia maszynowego – na przykład uczenie nadzorowane – jest używanych, ale jak dotąd nikt nie znalazł jeszcze sposobu na uzyskanie przejrzystości, jeśli chodzi o algorytmy oparte na technikach głębokiego uczenia. wyjaśnij, dlaczego podjęto określoną decyzję, gdy coś poszło nie tak. Właściwym przykładem jest autonomiczny samochód, w którym działa kilka algorytmów, które współpracują ze sobą i (miejmy nadzieję) przestrzegają wstępnie zdefiniowanych zasad postępowania w określonych okolicznościach. Wszystko działa zgodnie z planem, ale wtedy następuje zupełnie nieznanne i nieoczekiwane zdarzenie i samochód podejmuje nieoczekiwaną akcję, która powoduje wypadek. W takich sytuacjach ludzie na ogół naturalnie oczekiwaliby, że nie będzie jakiś sposób na wydobycie informacji z autonomicznego samochodu o tym, dlaczego podjęto tę konkretną decyzję - stąd oczekują wyjaśnień w sztucznej inteligencji.

Poza technicznymi, etycznymi i egzystencjalnymi przyczynami zapewnienia wyjaśnialności sztucznej inteligencji, istnieje teraz również powód prawny. Ogólne rozporządzenie o ochronie danych (RODO) UE zawiera klauzulę, która wymaga algorytmicznej interpretacji. W tej chwili te wymagania nie są zbyt surowe, ale z czasem prawdopodobnie zmieni się to dramatycznie. Żądanie RODO wymaga teraz możliwości wyjaśnienia działania algorytmu na podstawie następujących pytań:

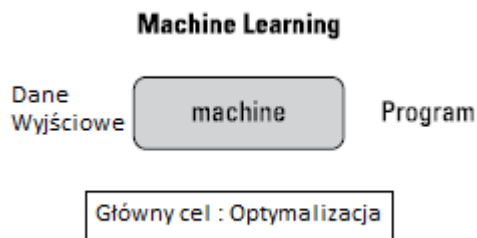
- * Jakie dane są używane?
- * Jaka logika jest używana w algorytmie?
- * Jaki proces jest używany?
- * Jaki wpływ ma decyzja podjęta przez algorytm?

Radzenie sobie z różnicą między uczeniem maszynowym a tradycyjnym programowaniem oprogramowania.

Jest dość dobrze ugruntowane i powszechnie przyjęte w branży oprogramowania, jaka jest rzeczywista różnica między tradycyjnym programowaniem a uczeniem maszynowym. Jednak jeśli chodzi o to, jak należy poradzić sobie z tą różnicą, nie ma zgody. Biorąc pod uwagę ten podział, chcę poświęcić czas na wyjaśnienie, co należy wziąć pod uwagę, jeśli chodzi o podejścia do wdrażania, a także jak radzić sobie z tymi różnymi punktami widzenia pod względem aspektów rozwoju, a także środowiska produkcyjnego. Ale najpierw pozwól, że zacznę od przyjrzenia się, o co chodzi w kłótni. Tradycyjne podejście programistyczne, pokazane na rysunku, wymaga wcześniejszego podjęcia decyzji, jak rozwiązać określony problem za pomocą opracowywanego programu. Głównym celem programisty jest zbudowanie wymaganej funkcjonalności.



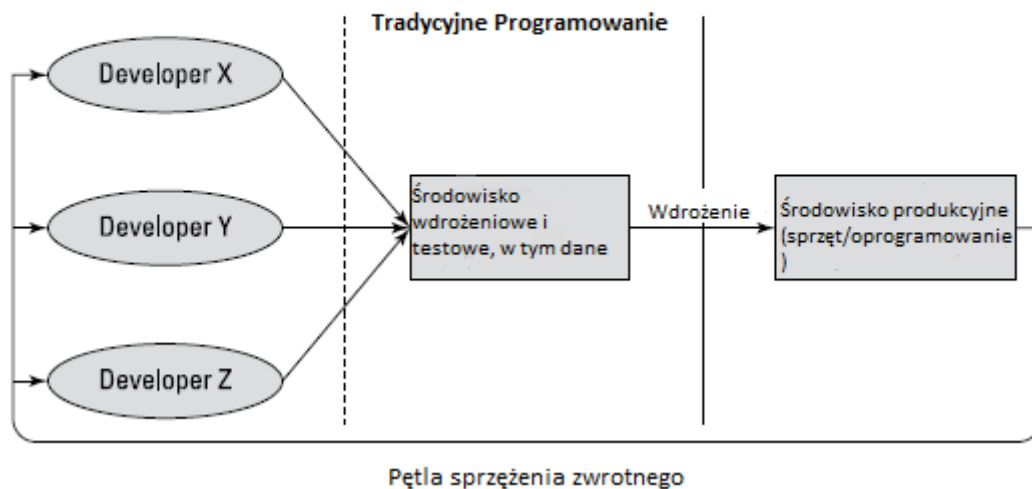
Na podstawie danych i programu maszyna wykonuje analizę dokładnie tak, jak chcesz, niezależnie od tego, czy jest to najbardziej zoptymalizowany sposób rozwiązania problemu. Założenie jest takie, że programista (a nie maszyna) najlepiej wie, jak rozwiązać problem. Z drugiej strony, jeśli chodzi o rozwój uczenia maszynowego, punktem wyjścia jest umożliwienie maszynie znalezienia najlepszego rozwiązania, gdy ustalisz granice, których danych użyć i jaki wynik osiągnąć - i nic więcej. Zakłada się, że w tych warunkach maszyna znajdzie najbardziej zoptymalizowany program do rozwiązania problemu.



Co więc oznaczają te odrębne podejścia w kontekście środowiska programistycznego i produkcyjnego? Jednym z głównych aspektów do rozważenia jest to, że tradycyjne programowanie obejmuje znacznie bardziej rygorystyczny proces. Jest oparty na regułach i zgodny z predefiniowanymi zasadami projektowania. Z drugiej strony, punkt wyjścia do rozwoju uczenia maszynowego jest znacznie bardziej odkrywczy i otwarty. Jak można się domyślić, będzie to miało dość znaczący wpływ na to, jak należy skonfigurować środowisko programistyczne. Niektóre firmy mają tendencję do bagatelizowania wpływu konfiguracji środowiska programistycznego i tego, jaki będzie to miało wpływ na produktywność analizy danych. Jeśli zaczniesz od tego punktu widzenia, możesz dojść do wniosku, że możesz użyć tej samej (lub podobnej) konfiguracji infrastruktury zarówno dla tradycyjnego środowiska programistycznego, jak i środowiska nauki o danych. Nic nie może być dalsze od prawdy – przyjęcie takiego podejścia oznacza, że stawiasz poważne bariery na drodze do osiągnięcia celu, jakim jest pożyteczna sztuczna inteligencja/uczenie maszynowe. Tradycyjne programowanie jest znacznie bardziej restrykcyjne, jeśli chodzi o to, jakich języków programowania użyć i jakie zasady zastosować do danego zadania. Ma to oczywiście wpływ na sposób konfiguracji zarówno środowiska programistycznego, jak i produkcyjnego. Rysunek przedstawia graficzną reprezentację tego, jak przebiega tradycyjne programowanie.

Programowanie może odbywać się w oderwaniu od środowiska testowego danych i programowania

Wdrożenie oprogramowania lub środowiska produkcyjnego odseparowane od środowiska programistycznego i testowego

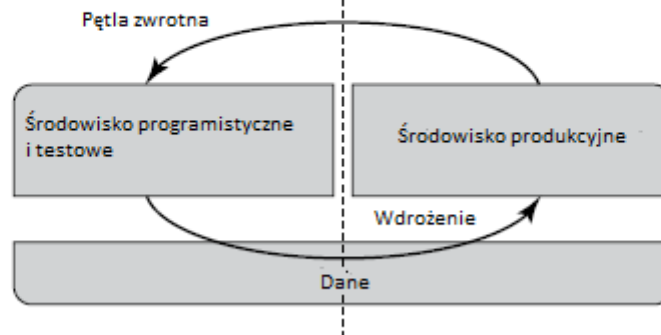


Jak widać po lewej stronie rysunku, tradycyjne programowanie może odbywać się niezależnie od danych oraz środowiska programistycznego i testowego. Nie musi się to odbywać osobno, ale faktem jest, że można to zrobić w izolacji – nawet na laptopie w kawiarni - a następnie zintegrować z innym kodem w środowisku deweloperskim i testowym. W tym momencie do modelu można dodać dane w celu uzyskania pożądanego wyniku. Rysunek pokazuje również, że wdrażanie oprogramowania odbywa się w oddzielnym środowisku (na przykład w produkcji oprogramowania/sprzętu lub podobnym środowisku produkcyjnym) poza środowiskiem programistycznym i testowym. Przechodząc ponownie do podejścia do uczenia maszynowego, musisz pamiętać, że eksploracyjny i uczący się charakter rozwoju uczenia maszynowego wymaga, aby konfiguracja dostępna dla analityków danych była niezwykle elastyczna. Wydajne zarządzanie danymi, łatwy dostęp do danych i różnorodne specjalistyczne narzędzia do uczenia maszynowego muszą być łatwo dostępne. Nikt nie wchodzi w proces z predefiniowanymi pojęciami, której dokładnie techniki uczenia maszynowego użyć, ponieważ wszystko to musi zostać zbadane, a najbardziej zoptymalizowane rozwiązanie może stać się jasne dopiero po rozpoczęciu procesu. Jak pokazuje poniższy rysunek, rozwój uczenia maszynowego nie może odbywać się w izolacji i bez danych. Wszystko zaczyna się i kończy na danych w przepływie rozwoju uczenia maszynowego, ponieważ same dane szkolą model pod kątem zoptymalizowanego projektu. Aby to zadziałało, oczywiście musisz mieć stały przepływ danych, co oznacza, że potrzebujesz stabilnego potoku danych - najlepiej zwirtualizowanego, który zapewnia większą elastyczność infrastruktury w czasie.

Rozwój ML nie może odbywać się w izolacji i bez danych. Stabilny potok danych ma kluczowe znaczenie

Machine Learning

W przypadku rozwiązań innych niż brzegowe preferowane jest posiadanie środowisk programistycznych i produkcyjnych w ramach tej samej infrastruktury



W przypadku zwirtualizowanych środowisk produkcyjnych uczenia maszynowego, które nie są zaimplementowane na urządzeniach brzegowych (wewnątrz urządzeń IoT, takich jak telefon komórkowy, samochód, zegarek, lodówka lub inne rodzaje urządzeń, które są połączone i na których może działać algorytm ML), staraj się, aby środowisko programistyczne i produkcyjne było blisko lub w ramach tej samej konfiguracji infrastruktury. Ułatwia to produktywność uczenia maszynowego podczas przechodzenia między rozwojem a produkcją, z szybszymi i wydajniejszymi pętlami sprzężenia zwrotnego w ramach korzyści. Zyskujesz również korzyści w zakresie efektywności kosztowej, gdy nie musisz duplikować infrastruktury, ponieważ obie są oparte na tym samym potoku danych.

Zarządzanie szybką ewolucją technologii AI i brakiem standaryzacji

Technologie AI/ML nieustannie ewoluują i stają się coraz bardziej zaawansowane. Wraz ze wzrostem wydajności obliczeniowej, takie technologie mogą teraz dostosowywać się do działania na mniejszej powierzchni sprzętowej. Te postępy przesuwają analitykę, ML i realizację sztucznej inteligencji również na brzeg, co oznacza, że algorytm ma wsparcie obliczeniowe do działania wewnątrz urządzenia, a nie że urządzenie po prostu dostarcza dane do algorytmu działającego zdalnie w chmurze. To dobry trend, ponieważ pozwoli społeczeństwu na szersze wykorzystanie inteligencji maszynowej w środowiskach systemowych oraz miliardach urządzeń mobilnych i innych połączonych podmiotach. Jednak jeden obszar nie nadąży za wszystkimi szybkimi zmianami: standaryzacja ML/AI. Brak standaryzacji nie jest oczywiście czymś, co Ty lub pojedyncza firma możesz rozwiązać, ale ważne jest, aby być świadomym tej sytuacji w ramach swojej strategii analizy danych. I oczywiście pod koniec dnia wszyscy naukowcy zajmujący się danymi mają obowiązek dążyć do większej standaryzacji w uczeniu maszynowym i sztucznej inteligencji. Ale tylko dlatego, że nie istnieje jeszcze żadna oficjalna, międzynarodowa standaryzacja, nie oznacza to, że nie istnieją żadne inicjatywy. Standardy, które są dostępne, opierają się w większości na czymś, co często określa się mianem standaryzacji de facto, wywodzącej się z wpływowych inicjatyw open source koordynowanych przez uniwersytety, takie jak UC-Berkeley (laboratorium AMP i laboratorium RISE) oraz firmy takie jak Google (Google Beam) i AT&T (Acumos). Inny trend, który można wykryć, dotyczy rosnących obaw, jeśli chodzi o dostęp do danych osobowych (i korzystanie z nich) z nieprzejrzystych lub nawet ukrytych powodów. Doprowadziło to do zaostrzenia przepisów w różnych krajach, ale doprowadziło również do bardziej toczących się dyskusji na temat potrzeby zwiększenia regulacji i narzucenia standaryzacji związanych z etyką AI. Miejmy nadzieję, że ten pozytywny trend będzie nadal skłaniał ludzkie społeczeństwo do lepszego wyobrażenia sobie – jako grupy – jak powinna wyglądać przyszłość wykorzystania sztucznej inteligencji. Oczywiście ten trend ma swoje minusy. Ponieważ teraz jest tak mało standaryzacji dostępnej, aby oprzeć się na inwestycjach w

naukę danych, musisz wziąć pod uwagę możliwość, że będziesz musiał wprowadzić poważne zmiany w swojej infrastrukturze, gdy nowe standardy w końcu pojawią się w najbliższej przyszłości. Najgorszy scenariusz? Być może będziesz musiał powtórzyć ten proces kilka razy, a nawet całkowicie przemodelować całą infrastrukturę. Moja rada dla Ciebie? Nieustannie śledź trendy i wypatruj wszelkich oznak, że nowe przepisy lub regulacje lub inicjatywy standaryzacyjne spływają na dno - zwłaszcza te o otwartym kodzie źródłowym. Przygotuj się na dostosowanie swojego podejścia do analizy danych do nadchodzących zmian.