

Gdy Twój projekt związany z badaniem danych nie powiedzie się

Większość projektów data science to przedsięwzięcia wysokiego ryzyka. Próbujesz przewidzieć coś, czego nikt wcześniej nie przewidział, zoptymalizować coś, czego nikt wcześniej nie zoptymalizował lub zrozumieć dane, na które nikt wcześniej nie spojrział. Bez względu na to, co robisz, jesteś pierwszą osobą, która to robi; praca jest prawie zawsze eksploracyjna. Ponieważ naukowcy zajmujący się danymi nieustannie robią nowe rzeczy, nieuchronnie dojdiesz do punktu, w którym odkryjesz, że to, na co liczyłeś, nie jest możliwe. Wszyscy musimy zmagać się z niepowodzeniem naszych pomysłów. Niepowodzenie jest bolesne i rozdzierające serce; chcesz przestać myśleć o data science i marzyć o całkowitym opuszczeniu pola. Jako przykład rozważ firmę budującą model uczenia maszynowego w celu polecenia produktów w witrynie. Prawdopodobny przebieg wydarzeń zaczyna się od pewnego zestawu spotkań, podczas których zespół data science przekonuje kadrę zarządzającą, że projekt to dobry pomysł. Zespół wierzy, że wykorzystując informacje o klientach i ich transakcjach, może przewidzieć, co klienci chcą kupić w następnej kolejności. Kierownictwo kupuje pomysł i zapala projektowi zielone światło. Wiele innych firm ma takie modele, które wydają się proste, więc projekt powinien działać. Niestety, gdy zespół zaczyna pracować, zaczyna się rzeczywistość. Może dowiadują się, że ponieważ firma niedawno zmieniła systemy, dane transakcyjne są dostępne tylko z ostatnich kilku miesięcy. A może zespół przeprowadza eksperyment i stwierdza, że osoby, które widzą silnik rekomendacji, nie kupują niczego więcej niż osoby, które tego nie robią. Takie problemy narastają; ostatecznie zespół porzuca projekt, przerażony. Tu zdefiniujemy projekt jako porażkę, gdy nie spełnia swojego celu. W przypadku analizy projekt może zakończyć się niepowodzeniem, jeśli nie pomoże interesariuszowi odpowiedzieć na pytanie biznesowe. W przypadku problemu z uczeniem maszynowym w środowisku produkcyjnym projekt może zakończyć się niepowodzeniem, jeśli nie zostanie wdrożony lub nie będzie działał po wdrożeniu. Projekty mogą zakończyć się niepowodzeniem na wiele sposobów. Analitycy danych zwykle nie mówią o niepowodzeniu projektów, chociaż zdarza się to niezwykle często. Gdy projekt kończy się niepowodzeniem, analityk danych może czuć się bezbronny. Jeśli twój projekt się nie powiedzie, możesz pomyśleć: „Gdybym był lepszym naukowcem od danych, to by się nie wydarzyło”. Niewiele osób czuje się komfortowo, dzieląc się historiami o kwestionowaniu własnych umiejętności. U podstaw nauki o danych leży badanie i rozwój. Każdego dnia analitycy danych zbierają dane, które nigdy wcześniej nie były analizowane, i szukają trendu, który może tam być lub nie. Naukowcy zajmujący się danymi postanowili zbudować modele uczenia maszynowego na danych, w których może nie być sygnału. Nie jest możliwe, aby te zadania zawsze się powiodły, ponieważ nowe trendy i sygnały są bardzo rzadko spotykane w jakiegokolwiek dziedzinie. Jednak w dziedzinie takiej jak inżynieria oprogramowania zazwyczaj można wykonać zadanie (choć może to zająć więcej czasu i zasobów niż planowano). Ważne jest, aby zrozumieć, w jaki sposób projekty analizy danych kończą się niepowodzeniem i co robić, gdy kończą się niepowodzeniem. Im lepiej rozumiesz nieudany projekt, tym więcej niepowodzeń w przyszłości możesz uniknąć. Nieudane projekty mogą również dać wgląd w to, co się powiedzie, dzięki zbadaniu, które części projektu zadziałały. I możesz być w stanie dostosować nieudany projekt przy odrobinie pracy w coś, co może być przydatne w organizacji. W tym rozdziale omówimy trzy tematy: dlaczego projekty analizy danych kończą się niepowodzeniem, jak myśleć o ryzyku projektu i co robić, gdy projekt się nie powiedzie. Omawiamy trzy główne powody, dla których większość projektów kończy się niepowodzeniem, co z nim zrobić i jak radzić sobie z emocjami, które możesz odczuwać.

Dlaczego projekty data science kończą się niepowodzeniem

Wygląda na to, że projekty data science kończą się niepowodzeniem z nieskończonej listy powodów. Od budżetu po technologię i zadania, których wykonanie trwa znacznie dłużej niż oczekiwano, istnieje

wiele powodów niepowodzenia. Ostatecznie te wiele rodzajów awarii można podzielić na kilka podstawowych tematów.

Dane nie są tym, czego chciałeś

Nie możesz zajrzeć do każdego możliwego źródła danych przed rozpoczęciem projektu. Konieczne jest dokonanie świadomych założeń dotyczących tego, co jest dostępne, na podstawie tego, co wiesz o firmie. Kiedy projekt się zaczyna, często dowiadujesz się, że wiele z Twoich założeń nie jest prawdziwych. Być może dane nie istnieją, nie są przechowywane w przydatnym formacie lub nie są przechowywane w miejscu, do którego masz dostęp. Jeśli na przykład przeprowadzasz analizę, aby zrozumieć, w jaki sposób wiek klienta wpływa na korzystanie przez niego z programu lojalnościowego, możesz dowiedzieć się, że klienci nigdy nie są pytani o wiek podczas dołączania do programu. Ta porażka może bardzo szybko zakończyć projekt. Ponieważ potrzebujesz danych, zanim będziesz mógł cokolwiek zrobić, tego typu problemy są pierwszymi poważnymi problemami, które się pojawiają. Często reakcją na napotkanie tego problemu są negocjacje wewnętrzne, w ramach których próbujesz objąć luki w danych. Mówisz takie rzeczy jak „Cóż, nie mamy dziesięciu lat danych, takich jak chcieliśmy, ale może rok danych wystarczy dla modelu” i masz nadzieję, że będzie jak najlepiej. Czasami takie podejście może działać, ale alternatywne rozwiązania nie zawsze są wystarczające, aby projekt był wykonalny. Kiedy wystawiasz projekt, nie zawsze masz dostęp do danych lub nawet pełne zrozumienie, czym one są (specjalny problem w konsultingu, gdzie nie masz dostępu do danych, dopóki praca nad projektem nie zostanie sprzedana). Ponadto dane mogą istnieć, ale mają krytyczną wadę, która czyni je bezużytecznymi. Dane mogą istnieć w tabeli bazy danych, ale identyfikatory klientów mogą być uszkodzone i bezużyteczne. Jest tak wiele sposobów, w jakie zbiór danych może mieć problemy, że niezwykle trudno jest sprawdzić je wszystkie przed rozpoczęciem projektu. Z tego powodu często zdarza się, że projekty data science ledwo wychodzą poza fazę uruchomienia. Im szybciej uzyskasz dostęp do danych i je eksplorujesz, tym szybciej możesz zminimalizować ryzyko nieodpowiednich danych. Najlepszym scenariuszem uniknięcia tego błędu jest pobranie próbek danych przed rozpoczęciem projektu. Jeśli nie jest to możliwe, następnym najlepszym scenariuszem jest opracowanie harmonogramu projektu wokół możliwości, że dane będą słabe. Dzięki wczesnemu etapowi „idź/nie idź” w projekcie, w którym interesariusze zgadzają się na ponowną ocenę wykonalności projektu, jest mniejsza szansa, że interesariusze będą zaskoczeni, że dane mogą być złe. Jeśli zmagasz się z brakiem dobrych danych, masz ograniczone możliwości. Możesz na przykład spróbować znaleźć alternatywne źródła danych do zastąpienia. Może nie masz danych o tym, które produkty zostały zakupione, ale wiesz, jaka ilość została wyprodukowana i możesz to wykorzystać. Problem zwykle polega na tym, że te substytuty są na tyle różne, że powodują realne problemy z analizą. Kiedy nie możesz znaleźć realnego substytutu, czasami wystarczy rozpocząć oddzielny projekt, aby zacząć zbierać lepsze dane. Dodanie oprzyrządowania i telemetrii do stron internetowych i aplikacji, tworzenie baz danych do przechowywania danych zamiast ich wyrzucania oraz wykonywanie innych zadań może pomóc zespołowi podjąć się zadania w przyszłości dzięki lepszym zebranych danym

Przykładowa awaria: Analiza statusu programu lojalnościowego

Dyrektor w dziale marketingu dużej sieci restauracji chce zrozumieć, czy klienci wydają inaczej w miarę wzrostu statusu w programie lojalnościowym firmy. Program ma poziomy srebra, złota i platyny, a dyrektor chce wiedzieć, czy ktoś, kto trafi na platynę, kupił w ten sam sposób, kiedy był tylko przy srebrze. Zespół ds. analizy danych zgadza się przyjrzeć tej prośbie, ponieważ zadanie powinno być dość proste i nie pracowali wcześniej z danymi lojalnościowymi. Są zszokowani, gdy dowiadują się, że przestarzała baza danych programów lojalnościowych nie śledzi historycznych poziomów programów - dokładnie tam, gdzie obecnie znajdują się klienci. Jeśli klient jest obecnie na poziomie platyny, nie ma możliwości sprawdzenia, kiedy był srebrem lub złotem. Dlatego analiza jest niemożliwa do wykonania.

Zespół data science zaleca dostosowanie systemu, ale zmiana architektury bazy danych programu lojalnościowego wymaga milionów dolarów, a w firmie jest na to niewielkie zapotrzebowanie, więc żadne zmiany nie są wprowadzane, a pomysł na analizę zostaje porzucony.

Dane nie mają sygnału

Założmy, że hazardzista zatrudnia analityka danych, który ma nadzieję wykorzystać statystyki do wygrania gry w kości. Hazardzista rzuca kostką 10 000 razy i zapisuje rzuty; następnie płaci analitykowi danych za stworzenie modelu, który przewidzi następny rzut kostką. Pomimo tego, że badacz danych dysponuje ogromną ilością danych, nie ma sposobu, aby przewidzieć, jaki rzut będzie następny, poza przypisaniem każdej stronie prawdopodobieństwa $1/6$ (jeśli kość jest sprawiedliwa). Pomimo tego, że badacz danych ma dużo danych, nie ma w tych danych sygnału, która strona zostanie przerzucona w następnej kolejności. Ten problem braku sygnału w danych jest niezwykle powszechny w nauce o danych. Założmy, że prowadzisz witrynę e-commerce i chcesz stworzyć model do przewidywania, którzy klienci będą zamawiać produkty na podstawie ich przeglądarki, urządzenia i systemu operacyjnego. Nie ma sposobu, aby przed rozpoczęciem projektu dowiedzieć się, czy te punkty danych mogą być rzeczywiście wykorzystane do przewidzenia, czy klient złoży zamówienie, czy też w danych brakuje sygnału, tak jak miało to miejsce w przypadku danych z rolki. Tworzenie modelu uczenia maszynowego do prognozowania polega na testowaniu danych, aby sprawdzić, czy zawiera w sobie sygnał, a może nie być. W rzeczywistości w wielu sytuacjach byłoby bardziej zaskakujące, gdyby był sygnał, niż gdyby nie było sygnału. Niestety brak sygnału w danych może być końcem projektu. Jeśli projekt jest zbudowany wokół próby znalezienia relacji w danych i dokonania na tej podstawie prognozy, a nie ma tam związku, prognoza nie może zostać wykonana. Analiza może nie wykazać niczego nowego ani interesującego, a model uczenia maszynowego może nie przynieść wyników lepszych niż losowy przypadek. Jeśli nie możesz znaleźć sygnału w hałasie, masz kilka możliwych sposobów:

* **Przeformułuj problem.** Możesz spróbować przeformułować problem, aby sprawdzić, czy istnieje inny sygnał. Założmy, że masz zestaw artykułów i próbujesz przewidzieć najbardziej odpowiedni artykuł dla użytkownika. Możesz sformułować problem jako problem klasyfikacji, aby spróbować sklasyfikować, który artykuł w zestawie artykułów jest najbardziej odpowiedni.

* **Zmień źródło danych.** Jeśli wydaje się, że nic nie wyciąga sygnału z danych, możesz spróbować zmienić źródło danych. Podobnie jak w przypadku poprzedniego punktu awarii polegającego na braku dobrych danych, dodanie nowego źródła danych do problemu czasami tworzy nieoczekiwany sygnał. Niestety, zwykle zaczynasz od zbioru danych, który miał największą szansę być użyteczny, więc szanse, że ta strategia Cię uratuje, są dość ograniczone.

Zazwyczaj naukowcy zajmujący się danymi, którzy utknęli w takiej sytuacji, próbują użyć mocniejszego modelu, aby znaleźć sygnał. Jeśli regresja logistyczna nie może dać sensownej prognozy, próbują losowego modelu lasu. Jeśli losowy model lasu nie działa, próbują sieci neuronowej. Każda metoda staje się bardziej czasochłonna i bardziej złożona. Chociaż te metody mogą być przydatne do uzyskania dokładniejszych prognoz, nie mogą zrobić czegoś z niczego. Najczęściej, jeśli najprostsza metoda nie jest w stanie wykryć żadnego sygnału, te bardziej złożone nie będą w stanie. Dlatego najlepiej zacząć od prostych metod modelowania, aby zweryfikować wykonalność projektu, a następnie przejść do bardziej złożonych i czasochłonnych metod, niż zaczynać od złożonych i iść prostszymi. Nie zgub się, spędzając miesiące na budowaniu coraz bardziej skomplikowanych modeli, mając nadzieję, że być może następny będzie tym, który uratuje projekt.

Przykładowy błąd: Wykrywanie błędów na stronie z danymi sprzedażowymi

Hipotetyczna firma e-commerce ma problem: na stronie ciągle pojawiają się błędy i błędy. Co gorsza, błędy nie zawsze są wykrywane przez DevOps lub zespół inżynierów oprogramowania. Pewnego razu błąd został wykryty przez zespół marketingowy, który zauważył, że dzienne przychody są zbyt niskie. Kiedy marketing zauważył błąd zamiast DevOps lub inżynierii, to zła sytuacja. Zespół zajmujący się analizą danych postanowił wykorzystać statystyczne techniki kontroli jakości danych sprzedaży, aby otrzymywać alerty, gdy przychody są tak niskie, że w witrynie musi być jakiś błąd. Mają listę dni, w których wykryto błędy i historyczne dane o przychodach. Wykorzystanie sprzedaży do przewidywania błędów wydaje się proste. Niestety, wiele powodów, dla których przychody mogą się codziennie zmieniać, sprawia, że wykrywanie błędów jest prawie niemożliwe. Przychody mogą być niskie z powodu dnia tygodnia, pory roku, promocji marketingowych, wydarzeń globalnych lub wielu innych rzeczy. Chociaż marketing był kiedyś w stanie zobaczyć błąd, tego faktu nie można było uogólnić, ponieważ nie było na to sygnału w danych.

Klient nie chciał tego

Bez względu na to, jak dokładny jest model lub analiza, liczy się to, że zapewnia wartość interesariuszom. Analiza może zawierać wyniki, które są niezwykle interesujące dla analityka danych, ale nie dla przedsiębiorcy, który o to poprosił. Model uczenia maszynowego może tworzyć bardzo dokładne prognozy, ale jeśli ten model nie zostanie wdrożony i używany, nie zapewni dużej wartości. Wiele projektów dotyczących nauki danych kończy się niepowodzeniem, nawet po wykonaniu prac związanych z nauką danych. Ostatecznie analiza, model lub pulpit nawigacyjny w zakresie nauki o danych jest produktem. Projektowanie i tworzenie produktu to praktyka, w którą wiele osób włożyło setki lat zbiorowej myśli. Mimo to każdego roku wydaje się miliardy dolarów na tworzenie produktów, których ludzie nie chcą. Od New Coke po Google Glass, niektóre głośne produkty nie trafiają do klientów, a niektóre niskoprofilowe też nie. Tak jak Microsoft i Nokia mogły włożyć wiele wysiłku w stworzenie systemu Windows Phone, którego klienci ostatecznie nie kupili, tak analityk danych może tworzyć produkty, które nie są używane. Uniwersalne wytyczne dotyczące tworzenia produktów, które spodobać się klientom, to spędzanie dużej ilości czasu na rozmowach i pracy z klientami. Im lepiej rozumiesz ich potrzeby, pragnienia i ich problemy, tym większe prawdopodobieństwo, że stworzysz produkt, którego chcą. Dziedziny badań rynku i badań doświadczeń użytkowników to różne sposoby rozumienia klienta, poprzez ankiety i grupy fokusowe w badaniach rynku lub historie użytkowników, persony i testy w badaniach doświadczeń użytkowników. Wiele innych dziedzin wymyśliło własne metody i stosuje je od lat. Pomimo całego dobrego myślenia, jakie zrobili ludzie, nauka o danych jako dziedzina jest szczególnie podatna na niepowodzenie z powodu niezrozumienia potrzeb klientów. Z jakiegoś powodu naukowcy zajmujący się danymi czują się znacznie bardziej komfortowo patrząc na tabele i wykresy, niż wychodzą i rozmawiają z ludźmi. Wiele projektów związanych z nauką danych zakończyło się niepowodzeniem, ponieważ naukowcy zajmujący się danymi nie włożyli wystarczającego wysiłku w rozmowy z klientami i interesariuszami, aby zrozumieć, jakie były ich prawdziwe problemy. Zamiast tego naukowcy zajmujący się danymi zaczęli budować ciekawe modele i badać dane. W rzeczywistości ta sytuacja jest jednym z głównych powodów, dla których zdecydowaliśmy się poświęcić część zarządzania interesariuszami. Mamy nadzieję, że już lepiej rozumiesz, jak myśleć o relacjach z interesariuszami, czytając ten rozdział, ale jeśli go pominąłeś, może powinieneś to sprawdzić. Jeśli znajdziesz się w sytuacji, w której masz produkt, który nie wydaje się łądować, najlepszą rzeczą, jaką możesz zrobić, to porozmawiać z klientami. Nigdy nie jest za późno na rozmowę z klientami. Niezależnie od tego, czy Twój klient jest interesariuszem biznesowym, czy klientami Twojej firmy, pomocna może być komunikacja i zrozumienie. Jeśli Twój produkt nie jest dla nich przydatny, czy mogą powiedzieć, dlaczego tak nie jest? Czy możesz potencjalnie rozwiązać problemy, dodając nowe funkcje do produktu? Może mógłbyś zmienić analizę, dołączając do niej inny zbiór danych. Może mógłbyś ulepszyć model uczenia maszynowego, dostosowując format danych

wyjściowych lub szybkość jego działania. Nigdy się nie dowiesz, dopóki nie porozmawiasz z ludźmi. To również wpisuje się w koncepcję minimalnie opłacalnego produktu (MVP), który jest intensywnie używany w tworzeniu oprogramowania. Chodzi o to, że im szybciej produkt zacznie działać i wprowadzić go na rynek, tym szybciej można uzyskać informację zwrotną na temat tego, co działa, a co nie, a następnie powtórzyć te opinie. W nauce o danych im szybciej działa dowolny model lub jakakolwiek analiza, tym szybciej możesz pokazać go klientom lub interesariuszom i uzyskać ich opinię. Spędzanie miesięcy na iterowaniu modelu uniemożliwia uzyskanie tej opinii. Im lepiej rozumiesz klientów przez cały proces projektowania i budowania swojej pracy, tym mniej prawdopodobne jest, że odniesiesz porażkę od klienta, który nie będzie chciał produktu. A jeśli w ten sposób poniesiesz porażkę, najlepszym rozwiązaniem jest rozpoczęcie komunikacji, aby spróbować znaleźć rozwiązanie.

Przykładowe nieporozumienie: przewidywanie wartości kampanii sprzedażowej i marketingowej

Rozpoczęto projekt w firmie zajmującej się sprzedażą detaliczną, aby stworzyć model uczenia maszynowego do przewidywania, jaki zwrot z inwestycji (ROI) przyniosą przyszłe kampanie reklamowe. Zespół ds. analizy danych zdecydował się zbudować model po tym, jak zobaczył, jak bardzo zespoły ds. marketingu i sprzedaży zmagają się z tworzeniem arkuszy kalkulacyjnych Excel, które przewidywały ogólną wartość. Załóżmy, że wykorzystując uczenie maszynowe i modelowanie na poziomie klienta, zespół analityków danych stworzył model oparty na Pythonie, który dokładniej przewidywał ROI kampanii. Później zespół zajmujący się analizą danych odkrył, że jedynym powodem, dla którego zespoły ds. marketingu i sprzedaży stworzyły arkusze kalkulacyjne Excel z prognozami zwrotu z inwestycji, było nakłonienie działu finansowego do podpisania ich. Zespół finansowy odmówił pracy z czymkolwiek innym niż Excel; Python był dla nich zbyt czarną skrzynką. W związku z tym narzędzie nie zostało użyte, ponieważ zespół ds. analizy danych nie wziął pod uwagę potrzeb klienta. Nie było potrzeby jak najdokładniejszego przewidywania; była to przepowiednia, która przekonałaby zespół finansowy, że kampanie są finansowo wykonalne.

Zarządzanie ryzykiem

Niektóre projekty są bardziej ryzykowne niż inne. Zbieranie danych, z którymi zespół pracował wcześniej, i tworzenie standardowego pulpitu nawigacyjnego w standardowy sposób, prawdopodobnie zakończy się sukcesem. Znalezienie nowego zestawu danych w firmie, zbudowanie wokół niego modelu uczenia maszynowego, który będzie działał w czasie rzeczywistym, oraz wyświetlenie go klientowi w przyjemnym dla użytkownika interfejsie to bardziej ryzykowny projekt. Jako naukowiec zajmujący się danymi masz w każdej chwili pewną kontrolę nad ryzykiem. Jedną z ważnych kwestii związanych z ryzykiem jest liczba projektów, nad którymi pracujesz jednocześnie. Jeśli pracujesz nad pojedynczym ryzykownym projektem i ten projekt się nie powiedzie, poradzenie sobie z tym niepowodzeniem może być dość trudne. Jeśli jednak będziesz w stanie pracować nad kilkoma projektami jednocześnie, będziesz w stanie zminimalizować ryzyko. Jeśli jeden z tych projektów się nie powiedzie, masz inne projekty, na których możesz się oprzeć. Jeśli jeden projekt jest niezwykle złożonym modelem uczenia maszynowego, który ma ograniczone szanse powodzenia, możesz jednocześnie pracować nad prostszym tworzeniem kokpitów i raportowaniem; następnie, jeśli projekt uczenia maszynowego nie powiedzie się, interesariusze mogą nadal być zadowoleni z raportów. Posiadanie wielu projektów może być również korzystne z punktu widzenia wykorzystania. Projekty z zakresu nauki o danych mają wiele początku i końca, od oczekiwania na dane po oczekiwanie na odpowiedź interesariuszy, a nawet oczekiwanie na dopasowanie modeli. Jeśli z jakiegoś powodu utkniesz w jednym projekcie, będziesz mieć możliwość zrobienia postępu w innym. Może to nawet pomóc w przypadku blokad umysłowych; rozpraszanie się, gdy utkniesz, może być świetnym sposobem na odświeżenie myślenia. Innym sposobem na ograniczenie ryzyka jest wprowadzenie do projektu punktów wczesnego zatrzymania. Idealnie, projekt, który wydaje się, że może się nie powieść,

powinien być zaprojektowany z oczekiwaniem, że jeśli do pewnego momentu nie powiedzie się, zostanie odcięty. Na przykład w projekcie, w którym nie jest jasne, czy dane istnieją, projekt można określić zakresem, tak aby jeśli po miesiącu poszukiwań nie można było znaleźć dobrych danych, uważa się je za niewykonalne i zagubione. Jeśli oczekiwanie, że może się nie udać, zostanie przedstawione wcześniej, zakończenie projektu jest mniej zaskakujące i mniej kosztowne. W pewnym sensie wcześniejsze zakończenie projektu kodyfikuje fakt, że nauka o danych to badania i rozwój. Ponieważ nauka o danych jest wypełniona tak wieloma niewiadomymi, sensowne jest planowanie z uwzględnieniem możliwości, że w miarę uczenia się więcej poprzez prace eksploracyjne pomysł może się nie udać. Chociaż warto minimalizować ryzyko w portfolio projektów, nie chcesz go całkowicie usuwać. Nauka o danych polega na podejmowaniu ryzyka: prawie każdy wystarczająco interesujący projekt będzie miał mnóstwo niepewności i niewiadomych. Te ryzykowne niewiadome mogą wystąpić, ponieważ nikt wcześniej nie korzystał z nowego zestawu danych, nikt w firmie nie próbował wcześniej określonej metodologii lub interesariusz pochodzi z części firmy, która nigdy wcześniej nie korzystała z nauki o danych. Wiele cennych wkładów w badania danych w firmach pochodzi od ludzi, którzy próbują czegoś nowego, a jeśli jako naukowiec ds. danych starasz się unikać projektów, które mogą zakończyć się niepowodzeniem, unikasz również potencjalnie dużych sukcesów. Chociaż w tym rozdziale omówiono wiele przyczyn niepowodzenia projektów analizy danych, zespoły zajmujące się analizą danych mogą ostatecznie ponieść porażkę, nie podejmując wystarczającego ryzyka. Rozważ zespół ds. analizy danych, który wymyśla nowe pomysły i raporty dotyczące projektów, uważa je za udane, a następnie zatrzymuje się, aktualizując i odświeżając tylko poprzednie prace. Chociaż te projekty mogą nie kończyć się niepowodzeniem, ponieważ dostarczają pracę firmie, zespół ten przegapiłby potencjalne nowe obszary nauki o danych.

Co możesz zrobić, gdy Twoje projekty zawodzą?

Jeśli Twój projekt związany z analizą danych nie powiódł się, nie oznacza to, że cały czas spędzony nad nim został zmarnowany. We wcześniejszej części opisaliśmy kilka potencjalnych działań, które możesz podjąć, aby odwrócić projekt. Ale nawet jeśli nie ma możliwości, aby projekt odniósł sukces, nadal możesz podjąć kroki, aby jak najlepiej wykorzystać to, co z niego zostało. W kolejnych sekcjach przedstawiamy kilka strategii radzenia sobie z emocjami, gdy projekt się nie powiedzie.

Co zrobić z projektem

Chociaż projekt mógł się nie powieść, prawdopodobnie nadal można wiele z niego zyskać, zarówno w zakresie wiedzy, jak i technologii. Poniższe kroki mogą pomóc Ci zachować wiele z tych korzyści.

NAUCZONE LEKCJE DOKUMENTACYJNE

Pierwszą rzeczą, którą należy zrobić z projektem, który się nie powiódł, jest ocena, czego można się z niego nauczyć. Oto kilka ważnych pytań, które należy zadać sobie i zespołowi

* Dlaczego się nie udało? To pytanie wydaje się prawie oczywiste, ale często nie można zrozumieć, dlaczego projekt się nie powiódł, dopóki nie cofniesz się i nie spojrzysz na szerszy obraz. Prowadząc dyskusję ze wszystkimi osobami zaangażowanymi w projekt, możesz lepiej zdiagnozować, co poszło nie tak. Firma Etsy spopularyzowała koncepcję nienagannego post mortem – dyskusji prowadzonej po tym, jak coś się nie powiodło, w której zespół może zdiagnozować problem bez obwiniania osoby. Myśląc o problemie jako spowodowanym błędem w sposobie pracy zespołu (zamiast błędów danej osoby), masz większe szanse na znalezienie rozwiązania. Bez strachu przed karą ludzie będą chętniej mówić otwarcie o tym, co się wydarzyło.

* Co można było zrobić, aby zapobiec awarii? Kiedy zrozumiesz czynniki, które przyczyniły się do niepowodzenia, zrozumiesz, jak uniknąć podobnych sytuacji w przyszłości. Jeśli na przykład dane nie były wystarczająco dobre, aby projekt zadziałał, niepowodzeniu można było zapobiec dzięki dłuższej fazie eksploracji. Tego rodzaju lekcje pomagają Twojemu zespołowi rozwijać się i dojrzewać.

* Czego dowiedziałeś się o danych i problemie? Nawet jeśli projekt zakończy się porażką, często uczysz się rzeczy, które będą wartościowe w przyszłości. Być może dane nie miały w sobie sygnału, ale aby dojść do tego punktu, nadal musiałeś dołączyć do kilku nowych zestawów danych; teraz możesz łatwiej wykonać te same połączenia w innych projektach. Te pytania mogą pomóc w przeprowadzeniu burzy mózgów na temat możliwych rzeczy, które można uratować z projektu i wymyśleniu alternatywnych pomysłów na projekt. Organizując spotkanie, na którym zespół pracuje nad tymi pytaniami, a następnie zapisując wyniki we wspólnej lokalizacji, zyskasz znacznie więcej korzyści z nieudanego projektu.

ROZWAŻYĆ PRZEKSZTAŁCENIE PROJEKTU

Chociaż sam projekt mógł być porażką, mogą istnieć sposoby na przekształcenie go w coś użytecznego. Jeśli na przykład próbujesz zbudować narzędzie do wykrywania anomalii w przychodach firmy i to się nie powiedzie, nadal możesz używać tego samego modelu jako całkiem przyzwoitego narzędzia do prognozowania. Całe firmy zostały zbudowane na wzięciu pomysłu, który okazał się porażką, i przekształceniu go w coś udanego. Zmiana produktu wymaga dużo komunikacji z interesariuszami i klientami. Zasadniczo jesteś z powrotem na początku procesu projektowania produktu, próbując znaleźć dobre zastosowanie dla swojej pracy. Rozmawiając z interesariuszami i klientami, możesz zrozumieć ich problemy i sprawdzić, czy Twoja praca jest przydatna do czegoś nowego.

ZAKOŃCZ PROJEKT (WYTNIJ I URUCHOM)

Jeśli nie możesz obrócić projektu, najlepszą rzeczą, jaką możesz zrobić, to go zakończyć. Poprzez definitywne anulowanie projektu pozwalasz sobie i zespołowi przejść do nowej, bardziej obiecującej pracy. Bardzo łatwo jest naukowcowi danych chcieć kontynuować pracę nad projektem w nieskończoność w nadziei, że pewnego dnia to zadziała. (Istnieją tysiące algorytmów; w końcu jeden zadziała, prawda?) Ale jeśli utkniesz, próbując coś zadziałać, skończysz niepotrzebnie wysiłek. Poza tym nie jest fajnie pracować nad tym samym do końca czasu! Choć cięcie projektu jest trudne, bo wymaga przyznania, że nie jest już warte wysiłku, na dłuższą metę się opłaca.

KOMUNIKUJ SIĘ Z INTERESARIUSZAMI

Naukowiec zajmujący się danymi powinien komunikować się ze swoimi interesariuszami w trakcie projektu z zakresu nauki o danych, ale powinien zwiększyć ilość komunikacji, jeśli projekt się nie powiedzie. Chociaż ukrywanie ryzyka i problemów przed interesariuszami może być wygodne, aby ich nie rozczarować, wpadnięcie w sytuację, w której interesariusz jest zaskoczony niepowodzeniem projektu, może mieć katastrofalne skutki dla kariery. Informując interesariuszy, że pojawiają się problemy lub że projekt nie może już iść do przodu, zachowujesz przejrzystość wobec interesariuszy i wzbudzasz zaufanie. Po tym, jak pomożesz im zrozumieć stan projektu, możesz wspólnie podjąć decyzję o kolejnych krokach. Jeśli nie masz pewności, jak komunikować problemy z interesariuszem, Twój menedżer powinien być dobrym źródłem informacji. Mogą albo przeprowadzić burzę mózgów nad podejściem do przekazywania wiadomości, albo potencjalnie przejąć inicjatywę w jej dostarczaniu. Różne osoby i organizacje lubią, gdy wiadomości są dostarczane na różne sposoby, od arkuszy kalkulacyjnych przedstawiających problemy z kodowaniem kolorami zielonym/żółtym/czerwonym po rozmowy przy kawie. Twój menedżer lub inne osoby z Twojego zespołu powinny mieć wgląd w to, co działa najlepiej. Jako naukowiec zajmujący się danymi często odczuwasz niepokój, komunikując, że

projekt kończy się niepowodzeniem; czujesz się bardzo emocjonalnie wrażliwy i myślisz, że jesteś w sytuacji słabości. Chociaż zdarzają się sytuacje, w których wiadomości są odbierane słabo, inne osoby często są chętne do pomocy w rozwiązywaniu problemów i podejmowaniu decyzji o dalszych krokach. Po ogłoszeniu niepowodzenia projektu możesz poczuć ulgę, a nie cierpienie.

Radzenie sobie z negatywnymi emocjami

Zapomnij na chwilę o projekcie i firmie: musisz też pomyśleć o własnym samopoczuciu. Porażka projektu jest emocjonalnie trudna! To najgorsze! Jeśli nie będziesz ostrożny, nieudany projekt może być prawdziwym drenażem i prześladować Cię długo po zakończeniu projektu. Rozważając sposób, w jaki reagujesz na porażkę i historię, którą o niej tworzysz, możesz przygotować się na bardziej długofalowy sukces. Naturalny monolog wewnętrzny na zakończenie nieudanego projektu brzmi: „Gdybym był tylko lepszym naukowcem zajmującym się danymi, projekt by nie zakończył się porażką”. Ta myśl jest błędna: większość projektów związanych z nauką o danych kończy się niepowodzeniem, ponieważ nauka o danych jest z natury oparta na próbowaniu rzeczy, które nigdy nie zadziałają. Większość wielkich naukowców zajmujących się danymi była zaangażowana w projekty, które się nie powiodły, a nawet prowadziła takie projekty. Obarczając siebie winą za nieudany projekt i możliwe niedociągnięcia w zakresie analizy danych, obciążasz siebie całym projektem. Ale jak omówiono wcześniej w tym rozdziale, istnieje wiele powodów, dla których projekty z zakresu analizy danych kończą się niepowodzeniem i bardzo rzadko problemem są kompetencje naukowca danych. Bardzo często martwisz się, że projekt się nie powiedzie z twojego powodu, ale ten niepokój jest w twojej głowie i nie jest odzwierciedleniem rzeczywistości. Jeśli pozwolisz sobie na porażkę i zaakceptujesz, że porażka nie jest oznaką słabości, będziesz bardziej w stanie uczyć się na tym doświadczeniu. Być pewnym siebie i swoich umiejętności ułatwiają myślenie o porażce i o tym, co się do niej przyczyniło, ponieważ nie będzie to tak bardzo bolało. Biorąc to pod uwagę, umiejętność bycia pewnym siebie i posiadania porażki to taka, która wymaga czasu, cierpliwości i praktyki, więc nie zdziw się, jeśli walczysz o pewność siebie. W porządku! Kluczową kwestią jest tutaj to, że najlepszą rzeczą, jaką możesz zrobić dla siebie, gdy projekt się nie powiedzie, jest zrozumienie, że porażka nie jest odzwierciedleniem twoich umiejętności. Projekty kończą się niepowodzeniem z przyczyn niezależnych od Ciebie, a Ty będziesz w stanie przejść od porażki. Im bardziej jesteś w stanie trzymać te rzeczy blisko siebie, tym łatwiej będzie zaakceptować porażkę. Zakończymy ten rozdział metaforą nauki o danych. Początkujący i młodszy badacze danych często myślą o profesjonalnym naukowcu danych jak o architekcie budynków. Początkujący architekt może projektować proste domy, a doświadczony architekt może budować drapacze chmur, ale jeśli któryś z nich zawali się, jest to porażka w karierze. Podobnie, jednym ze sposobów spojrzenia na analityka danych jest to, że budują coraz bardziej złożone modele, ale jeśli jeden się nie powiedzie, jego kariera jest zagrożona. Po przeczytaniu tej części mamy nadzieję, że uznasz, że nie jest to dokładny model profesjonalnego naukowca danych. Lepszą metaforą jest to, że analityk danych jest jak poszukiwacz skarbów. Poszukiwacz skarbów wyrusza na poszukiwanie zagubionych kosztowności, a jeśli będzie miał szczęście, znajdzie je! Początkujący poszukiwacz skarbów może szukać standardowych towarów, ale doświadczony poszukiwacz znajdzie najbardziej legendarny skarb. Badacz danych jest bardziej jak poszukiwacz skarbów; wyszukują udane modele, a od czasu do czasu ich modele i analizy działają! Chociaż starszy analityk danych może pracować nad bardziej skomplikowanymi lub skomplikowanymi projektami, wszyscy ciągle zawodzą, a to tylko część pracy.