Cześć! Witam Cię na kolejnej lekcji. Dzisiaj przedstawię Ci przykłady problemów, jakie możemy rozwiązać z użyciem machine learningu. Zaczynajmy!

Pierwszym zagadnieniem jest analiza churn, czyli przewidzenie ryzyka odejścia klienta od firmy w celu wcześniejszej reakcji i zapobiegnięciu tej sytuacji. Zacznijmy od problemu biznesowego. Jest to odpowiedzenie na pytanie czy klient zrezygnuje z usług. Jest to formuła biznesowa, którą w Data Science należy przekształcić na konkretną miarę techniczną. W tym wypadku możemy z taką miarę uznać fakt czy klient zrezygnuje w najbliższych 30 dniach. Aby takie rozwiązanie stworzyć potrzebujemy zmiennych czy danych historycznych. Mogą to być na przykład Liczba reklamacji w ostatnich 6 miesiącach, Kwota abonamentu, Liczba lat współpracy, Czy Klient składał wypowiedzenie w przeszłości. Rozwiązanie oparte o machine learning może zmniejszyć liczbę odejść i tym samym zwiększyć dochody firmy.

Innym zagadnieniem jest popyt na produkt oraz uzysk z przeprowadzonej kampanii reklamowej czy promocyjnej. Problemem biznesowym jest tutaj odpowiedzenie na pytanie jaka będzie sprzedaż produktu. Tutaj również należy technicznie doprecyzować te miarę, może to być na przykład liczba sztuk sprzedaży produktu w sklepie w danym dniu. Pomocne w rozwiązaniu tego zagadnienia mogą być takie zmienne jak cena produktu, aktywna akcja reklamowa wraz z jej parametrami informacje o występowaniu święta w najbliższych dniach czy pora roku. Model może pomóc zoptymalizować produkcję, łańcuch dostaw a w przypadku reakcji na promocje, wpłynąć pozytywnie na zyski.

Wróćmy do zagadnień, których rozwiązaniem jest odpowiedzenie na pytanie „czy coś się wydarzy”. Tym razem zajmiemy się innym, popularnym problemem czyli defaultem kredytowym, który jest brakiem zdolności do uregulowania zobowiązania w określonym czasie. Problem biznesowy to informacja czy klient będzie spłacać zobowiązanie kredytowe. W praktyce taka zmienna może być zdefiniowana jako terminowo spłacenie pierwszych 12 rat kredytu. Zmienne, które mogą pomóc wyjaśnić to zjawisko to na przykład Liczba spłaconych kredytów w historii, Liczba otwartych rachunków kredytowych, Kwota wypłaty/ Rodzaj zatrudnienia, Liczba opóźnień w spłacie powyżej 1 dnia w ostatnich 6 miesiącach. Warto tu dodać, że rozwiązanie machine learning będzie tutaj stosowane na moment składania wniosku kredytowego, a w wyniku zgoda lub jej brak na udzielenie. Dzięki takiemu modelowi bank może liczyć na wyższe zyski i rentowność portfela klientów.

Teraz zajmiemy się innym typem zagadnienia, związanego z obrazami. Problem do rozwiązania to stworzenie samochodu,który będzie sam się prowadził. Tutaj zmienny do wyjaśnienia może być wiele np. System dobrze rozpoznaje znaki drogowe, system w odpowiednim momencie podejmuje decyzje o zatrzymaniu pojazdu. Uogólniając, system podejmuje prawidłową reakcję w x% sytuacji. Dane, które mogą pomóc w rozwiązaniu tego zagadnienia to Rejestracja poprawnych zachowań kierowców, Obrazy ze znakami drogowymi, Rejestracja niepoprawnych zachowań kierowców, Obrazy ruchu drogowego. Zysk biznesowy to przede wszystkim nowatorskie rozwiązania i konkurencyjność, które przyciągają klientów.

Ostatnim przykładem będzie segmentacja klientów. Chcemy odpowiedzieć na pytanie Czy mogę podzielić moich klientów na różne segmenty w celu większej personalizacji działań? Tu nie będziemy mówić o technicznym odpowiedniku problemu biznesowego, ale o metryce sukcesu zastosowanego rozwiązania, którą będzie sukces w stworzeniu istotnie różnych segmentów klientów. Zmiennymi, które możemy użyć w segmentacji są na przykład Wiek, Liczba aktywnych produktów, Kwota zakupów w ostatnim roku, Liczba kontaktów z firmą w ostatnim roku. Ten problem jest inny niż omawiane wcześniej, gdyż nie chcemy prognozować przyszłości i nie mamy konkretnej miary technicznej odpowiadającej na problem. Tu chcemy podzielić zbiorowość na podgrupy o podobnych cechach. Jednak i tu możemy mówić o zysku biznesowym. Dzięki segmentacji i większej personalizacji, możemy dotrzeć z ofertą do większej liczby klientów i tym samym zwiększyć dochody.

Na zakończenie chciałbym pokazać Ci pułapki, które czekają Cię w świecie data science i machine learning. Polecam Ci wrócić do nich po zakończeniu kursu. Pierwsza pułapka Uczenie na danych z przyszłości, tzw. data leak. Oznacza to, że korzystamy z informacji, której na moment użycia modelu nie posiadamy, bo dopiero się wydarzy. Często taka zmienna może dać nam bardzo dobry model, jednak nieużywalny. Przykładem może być na przykład terminowa spłata pierwszej raty w modelu na default kredytowy.

Druga pułapka to Skupienie na jednym elemencie procesu, zapominając o całościowym obrazie i celu zadanego problemu. Tworzenie modeli ML to proces bazujący na wielu elementach. Musimy pamiętać o założeniach i procesach biznesowych, sposobach zbierania danych i wielu innych.

Kolejne to Występowanie nagłych, niespodziewanych zdarzeń (anomalii) niemożliwych do przewidzenia z wyprzedzeniem.

Innym aspektem, który może utrudnić modelowanie jest Mnogość założeń i ograniczeń biznesowych/procesowych.

Warto też wspomnieć o tym, że dla odbiorcy najczęściej ważny jest prognozowany uzysk biznesowy, którego Wysokość jest trudna do zmierzenia.

Inna ciekawa pułapka to model idealny, Czy on istnieje? Jeżeli taki znajdziecie, bądźcie ostrożni.

Kolejne to Modelowanie oczywistych zjawisk. Po co nam właściwie model, skoro jesteśmy pewni co się stanie? To tak jakby prognozować churn, kiedy klient już złożył wypowiedzenie. Jest już za późno.

Na koniec chciałbym zwrócić Twoją uwagę na fakt, że modele uczą się na podstawie zachowania grupy podobnych lub takich samych przypadków. Dlatego nawet bardzo dobre modele mogą się mylić rozpatrując pojedynczy przypadek.

Dziś omówiliśmy przykłady użycia machine learningu oraz pułapki jakie na Ciebie czekają. W kolejnym wideo pokażę jak wygląda proces tworzenia rozwiązania opartego o Data Science.