

Zastosowanie regresji liniowej w przewidywaniu popytu detałicznego



Od pewnego czasu publikuję artykuły propagujące metody optymalizacji w codziennej działalności gospodarczej. Metody te nie są zwykle zbyt wyszukane i trudne. Niemniej nie wszyscy podzielają pogląd, że są one potrzebne i mogą przyczynić się do istotnej poprawy efektywności procesów produkcyjnych, ograniczyć koszty, zminimalizować błędy jakości lub zmaksymalizować zysk. Dziesięć lat temu miałem okazję zastosować zespół prostych modeli regresji liniowej, które okazały się niezwykle potrzebne w procesie planowania produkcji dużej piekarni. Co ciekawe, zlecenie na stworzenie modeli prognostycznych otrzymałem od właściciela piekarni, osoby, która niewiele wiedziała o istnieniu modeli regresji liniowej. Nie interesowała go statystyka ani ekonometria, lecz proces produkcji i sprzedaży wyrobów piekarskich.

Model regresji liniowej – co to takiego?

Wyobraźmy sobie, że chcemy zbudować model prognozujący ilość samochodów przejeżdżających przez most w jakimś dużym mieście. Wiadomo, że natężenie ruchu zależy od godziny. Inny ruch zanotujemy w godzinach szczytu, a inny w nocy lub o świcie.

Kolejnym czynnikiem jest dzień tygodnia. Nie da się porównać ruchu niedzielnego z ruchem czwartkowym. Następnym czynnikiem jest pogoda. Gdy pada deszcz, mieszkańcy miasta nie będą skłonni dojeżdżać do pracy rowerem lub chodzić pieszo.

Mamy więc trzy zmienne do naszego modelu:

x_1 – godzina pomiaru

x_2 – dzień tygodnia pomiaru

x_3 – ilość opadów atmosferycznych w momencie dokonania pomiaru

Na podstawie tych trzech zmiennych nasz model ma określić prawdopodobny poziom natężenia ruchu na moście. Tym razem nie będę wchodził w techniczne aspekty budowy modelu regresji liniowej. Powiem tylko tyle, że model buduje się sam. Naszym zadaniem jest jedynie wpisanie w pierwszej kolumnie arkusza kalkulacyjnego pełnych godzin (zmienna x_1). W drugiej kolumnie należy wpisać, jaki był wtedy dzień tygodnia, a w kolejnej, jaki był wtedy poziom opadów w mm. Najważniejsza jest ostatnia kolumna, w której należy zapisać

ilość samochodów, jakie przejechały przez most w określonej godzinie i dniu tygodnia. Ta ostatnia kolumna zawiera zmienne zależne, zwane również wartościami wynikowymi.

Resztę pracy zrobi za nas arkusz kalkulacyjny. Większość arkuszy kalkulacyjnych, również Excel, jest wyposażonych w narzędzie samodzielnie tworzące model regresji liniowej. Narzędzie mające do dyspozycji kilkaset tysięcy godzin w dniach tygodnia, w których zanotowano natężenie ruchu na moście. Na podstawie tych danych tworzony jest specjalny mechanizm matematyczny, który nauczony historią jest w stanie przewidzieć, jaki będzie poziom ruchu na moście w przyszłości, o określonej godzinie, w określonym dniu tygodnia i na podstawie prognozowanego poziomu opadów.

Tak właśnie w praktyce tworzy się najstarszy, używany powszechnie model regresji liniowej.

Problem drugiej zmiany

Jest to studium przypadku, który miał miejsce naprawdę i który stanowił poważny problem biznesowy. Pewna piekarnia miała sieć kilkunastu sklepów z pieczywem. Znajdowały się one w bezpośrednim sąsiedztwie stacji kolejowych w szeregu podmiejskich miejscowości, których mieszkańcy codziennie dojeżdżali do pracy publicznym środkiem transportu.

Sklepy z pieczywem pracowały od 6 rano do godziny 18. Tak długi okres funkcjonowania wymagał dodatkowej dostawy pieczywa w środku dnia. W pierwszej dostawie właściciel piekarni wysłał wszystkie produkty nie analizując, jaki będzie na nie popyt. Prawdziwy kłopot pojawiał się przy drugiej dostawie. Jednego dnia asortyment udawało się sprzedać prawie w całości, innego dnia większość trzeba było utylizować. Sprzedaż drugiej dostawy pieczywa była decydująca dla rentowności piekarni. Pierwsza dostawa pokrywała koszty stałe i zmienne piekarni, zaś druga wypracowywała dzienny zysk. Kiedy na drugiej zmianie pieczywa było za dużo, nie udawało się go sprzedać i musiało być ono ze stratą utylizowane. Wtedy piekarnia

ponosiła dzienna stratę na działalności. Kiedy pieczywa było za mało, klienci wychodzili ze sklepu rozczarowani i niezadowoleni, a piekarnia nie zarabiała tyle, ile mogła zarobić. Warto podkreślić, że wieczorne zakupy pieczywa były również kluczowe dla mieszkańców miejscowości. Wracając z pracy kupowali świeże pieczywo na następny dzień. Kiedy nie znajdowali w sklepach piekarni szukali go w innych sklepach.

Co decydowało o popycie na pieczywo?

Pracownicy piekarni zauważyli, że popyt na pieczywo w godzinach popołudniowych zależał od dnia tygodnia. Wzmożony występował w poniedziałki, środy i soboty. Zauważyli również, że na poziom popytu pośredni wpływ miała pogoda, pora roku oraz miesiąc, w którym prowadzona była sprzedaż. Pracownicy odpowiedzialni za planowanie produkcji brali te zmienne pod uwagę przy planowaniu produkcji. Nie zawsze jednak potrafili właściwie zinterpretować jednocześnie wiele czynników, które tworzyły konfiguracje wzajemnych zależności. Proces planowania przez „wrócenie” był bardzo pracochłonny i frustrujący. Piekarnia miała kilkanaście sklepów, prognozę trzeba było prowadzić dla kilkudziesięciu produktów w każdym z tych sklepów codziennie i trzeba było robić to stosunkowo szybko. W akcie rozpaczy właściciel wpadł na pomysł, aby produkować na drugą dostawę tyle pieczywa, ile udawało się sprzedawać dokładnie w tym samym dniu tygodnia rok temu. Ten pomysł doprowadził kierownika produkcji na skraj załamania nerwowego. Musiał on codziennie rano wchodzić do historii systemu sprzedaży i odszukiwać właściwy dzień sprzed roku. Następnie wypisywać poziomy sprzedaży z każdego z trzydziestu produktu w każdym z kilkunastu sklepów. W porę pomocy udzieli informatycy, którzy zebrali wszystkie dane z kilku lat w jeden arkusz kalkulacyjny. Od tej pory kierownik produkcji mógł ułożyć plan produkcji nawet na kilka tygodni do przodu. Niestety nie była to efektywna metoda. Sprzedaże sprzed wielu lat niepokojąco różniły się w sprzedaży konkretnych dni. Postanowiono więc bazować na średnich wieloletnich dla pojedynczych dni.

Czy warto okrywać koło na nowo?

Aby precyzyjnie stworzyć model regresji liniowej wystarczy arkusz kalkulacyjny, w którym będą znajdowały się zmienne opisujące oraz zmienna wynikowa.

Pracownicy piekarni w trakcie swoich prób przepowiadania przyszłości uporządkowali historię wielkości sprzedaży. Każdy wiersz oznaczał pojedynczą transakcję dokonaną w określonym czasie, dniu tygodnia i zakończoną określoną ilością sprzedanego pieczywa. Jak pamiętamy, pracownicy piekarni wskazywali, że pośredni wpływ na sprzedaż miała pogoda. Teraz wystarczyło dołączyć do danych historyczne informacje o pogodzie, aby uzyskać kompletny zestaw danych do stworzenia modelu regresji liniowej. Uzupełniono więc bazę o dane pogodowe, takie jak temperatura, ciśnienie oraz wielkość i rodzaj opadów. Dane ze sprzedaży przedpołudniowej zostały usunięte z bazy, która następnie została rozdzielona na 30 oddzielnych zestawów odpowiadającym 30 wspomnianym produktom. W rezultacie powstało 30 oddzielnych modeli. Modele regresji liniowej mają postać prostego równania, który z łatwością można zapisać jako funkcję w arkuszu kalkulacyjnym.

Każdy z 30 asortymentów sprzedawanych na drugiej zmianie miał własny model prognostyczny oparty na tysiącach transakcji z poprzednich lat. Modele te były zapisane jako funkcje arkusza kalkulacyjnego. Wystarczyło wpisać dzień tygodnia, miesiąc oraz kilka informacji z prognozy pogody, aby dowiedzieć się, jaka będzie przyszła sprzedaż w wybranej lokalizacji.

Ten prosty arkusz modeli regresji liniowej prognozował sprzedaż z dokładnością 65-82%, co wystarczało, aby uzyskać zysk.

Opisana sytuacja zaistniała naprawdę, co pokazuje, że metody naukowe nie są oderwane od rzeczywistości. Mając do dyspozycji podobne bazy sprzedaży oraz współczesną technologię składającą się z nowoczesnych modeli prognostycznych oraz sieci neuronowych można wykrywać niezwykle współzależności, koincydencje i trendy. Na podstawie wielkości i różnorodności pojedynczych zakupów można grupować klientów, typować ich preferencje lub wrażliwość na promocje czy podwyżki cen. Dane sprzedażowe stanowią skarbnicę wiedzy, która kiedyś w przyszłości może zostać skutecznie wykorzystana.

Wojciech Moszczyński
współpraca Ewa Moszczyńska

Wojciech Moszczyński – absolwent Katedry Ekonometrii i Statystyki Uniwersytetu Mikołaja Kopernika w Toruniu, specjalista z zakresu ekonometrii, *data science* oraz rachunkowości zarządczej. Specjalizuje się w optymalizacji procesów produkcyjnych i logistycznych. Prowadzi badania w obszarze rozwoju i zastosowania sztucznej inteligencji. Od lat zaangażowany w popularyzację ekonometrii oraz *data science* w środowiskach biznesowych