

Jak piekarnia traci pieniądze na złym planowaniu wypieku – i jak można to ograniczyć dzięki prostym danym



Od zawsze piekarz mierzy się z tym samym pytaniem: ile i co upiec. To brzmi prosto, ale w praktyce jest to jeden z najtrudniejszych problemów w całej branży. Chleb, bułki, kajzerki, rogale, chałki, pieczywo krojone, pieczywo ciemne, jasne, z ziarnem, bez ziarna – każdy z tych wyrobów schodzi inaczej. Inaczej w poniedziałek, inaczej w piątek, inaczej przed świętem, inaczej po święcie. Inaczej rano, inaczej po południu. Inaczej w małym sklepie na osiedlu, inaczej w punkcie przy ruchliwej ulicy. A decyzję trzeba podjąć zanim klient wejdzie do sklepu. Często trzeba ją podjąć już nocą, gdy praca dopiero się zaczyna.

To właśnie tu leży sedno problemu. Pieczywa nie da się dorobić bez kosztu i bez czasu. Jeśli piekarnia wyprodukuje za dużo, część towaru zostanie na półce. Potem trzeba go przecenić, oddać, zutylizować albo wyrzucić. Każda z tych dróg oznacza stratę. Strata dotyczy tylko samego chleba. Strata dotyczy też mąki, prądu, gazu, pracy ludzi, czasu pracy pieca, pracy aut, miejsca na półce i całego wysiłku włożonego w produkcję. Jeśli zaś piekarnia wyprodukuje za mało, skutki są inne, ale wcale nie lepsze. Klient przyjdzie po chleb i go nie dostanie. Następnego dnia może pójść gdzie indziej. Po kilku takich sytuacjach zaczyna się psuć opinia o sklepie i o marce. Wtedy strata nie leży już tylko w jednej niesprzedanej sztuce albo w jednym braku. Wtedy strata dotyczy zaufania.

W prostym obrazie wygląda to tak. Gdy piekarz ma zrobić sto bochenków, a sprzedaje tylko siedemdziesiąt, to trzydzieści sztuk staje się ciężarem. Gdy zrobi siedemdziesiąt, a klienci chcą sto, to trzydziestu klientów może wyjść z pustą ręką. Z jednej strony jest nadmiar. Z drugiej strony brak. Jedno i drugie boli. Jedno i drugie kosztuje. Dlatego odwieczny problem piekarzy nie polega tylko na tym, jak dobrze piec. Polega także na tym, jak dobrze przewidzieć zapotrzebowanie rynku dzień wcześniej.

Doświadczenie kiedyś i dziś

Przez wiele lat ten problem rozwiązywano głównie doświadczeniem. Dobry piekarz i dobry właściciel sklepu „czuli”, ile mniej więcej trzeba zrobić. Znali swój rejon, znali ludzi, znali rytm tygodnia. Taka wiedza nadal ma wielką wartość. Nie da się jej lekceważyć. Jednak przy większej liczbie punktów sprzedaży i przy szerokim asortymencie samo wyczucie coraz częściej nie wystarcza. Gdy firma ma jeden sklep, można jeszcze wiele rzeczy



ocenić wzrokiem i pamięcią. Gdy sklepów jest kilka albo kilkanaście, zaczyna się chaos. W jednym punkcie schodzi więcej bułek rano. W drugim więcej chleba po południu. W trzecim duży ruch jest w piątek. W czwartym przed zamknięciem schodzą już tylko tańsze wyroby. Człowiek może znać część tych zjawisk, ale nie jest w stanie codziennie trafnie policzyć wszystkiego w głowie.

Dziś pojawił się jeszcze jeden ważny problem. Już nie wystarczy wiedzieć, ile wysłać do danego sklepu na cały dzień. Coraz częściej trzeba wiedzieć, ile dać na pierwszą zmianę, a ile na drugą. Innymi słowy: nie chodzi już tylko o pytanie „ile pieczywa ma trafić do sklepu”, lecz także o pytanie „o której godzinie ma tam trafić”. To zmienia skalę trudności. Bo jeśli rano wyśle się za dużo, część towaru może leżeć zbyt długo. Jeśli rano wyśle się za mało, półki będą puste jeszcze przed południem. A jeśli druga dostawa będzie źle policzona, problem wróci po południu. W efekcie piekarnia musi przewidzieć nie tylko wielkość popytu, ale też jego rozkład w czasie.

To nie jest wyłącznie problem logistyki. To jest przede wszystkim problem produkcji. Produkcji nie planuje się

"Za dużo chleba źle, za mało też źle"



o trzynastej, gdy już widać, że czegoś brakuje. Produkcję trzeba zaplanować wcześniej. Często już dzień przed sprzedażą trzeba wiedzieć, ile sztuk danego wyrobu będzie potrzebne rano w sklepie A, ile po południu w sklepie B, ile pieczywa jasnego, ile ciemnego, ile drobnych bułek, ile chleba rodzinnego. To oznacza, że dobra decyzja sprzedażowa musi powstać zanim zacznie się nocna praca piekarni. Najpierw trzeba przewidzieć popyt. Potem trzeba przełożyć tę prognozę na plan wypieku. Potem trzeba ten plan wykonać. Dopiero na końcu klient widzi gotowy towar na półce.

Można to opisać prostym przykładem. Sklep ma być otwarty od szóstej do dziewiętnastej. Rano przychodzą ludzie po świeży chleb do pracy i do szkoły. Po południu przychodzą inni klienci, często po drobne zakupy na wieczór. Jeśli ktoś policzy cały dzień jako jedną liczbę, np. „sprzeda się sto dwadzieścia sztuk”, to nadal nie wie, czy osiemdziesiąt z nich ma być gotowe rano, a czterdzieści po południu, czy odwrotnie. A to robi wielką różnicę. Bo od tego zależy, ile trzeba upiec w nocy, ile zostawić na później, jak załadować auto i jak rozdzielić towar między sklepy.

Czy trzeba prognozować?

W tym miejscu widać jasno, że brak dobrej prognozy nie jest drobną niedogodnością. To realny koszt. Gdy firma działa na wyczucie, łatwo o błędne decyzje, które powtarzają się codziennie. Jeden dzień z nadmiarem to pojedyncza strata. Dziesięć dni z nadmiarem to już stały wyciek pieniędzy. Jeden dzień z brakami to kilku niezadowolonych klientów. Wiele dni z brakami to utrata części rynku. Dochodzi do tego niepotrzebna praca ludzi. Ktoś robi ciasto, ktoś obsługuje piec, ktoś ładuje auto, ktoś rozkłada towar, a potem okazuje się, że część tej pracy nie dała wyniku, bo towar nie znalazł nabywcy. Marnują się też surowce.

Z drugiej strony dobrze policzona produkcja daje zysk nie dlatego, że dzieje się coś nadzwyczajnego. Daje zysk dlatego, że znika chaos. Towar trafia tam, gdzie ma największą szansę się sprzedać. Sklepy nie świecą z pustą ladą. Nie trzeba w pośpiechu łątać braków.

Dlatego pytanie: „ile pieczywa upiec” nie jest pytaniem prostym. To jedno z głównych pytań zarządczych w piekarni. Łączy w sobie sprzedaż, produkcję, transport, pracę ludzi i relację z klientem. Kto odpowiada na nie źle, ten płaci codziennie, czasem nie widząc pełnej skali strat. Kto odpowiada na nie dobrze, ten zyskuje nie tylko lepszy wynik finansowy, ale też większy porządek w całej firmie. Właśnie dlatego przewidywanie popytu nie jest modą ani ozdobą. Jest narzędziem do ograniczania strat i do lepszego sterowania piekarnią.

Ten artykuł dotyczy właśnie tego problemu. Nie chodzi w nim o teorię dla teorii. Chodzi o praktyczne pytanie: jak przewidzieć, ile i jakiego pieczywa będzie potrzebne w danym sklepie, o danej porze, w danym dniu, tak aby ograniczyć straty i nie tracić klienta. Bo w piekarni błąd prognozy nie kończy się w tabeli. Błąd prognozy kończy się na półce, w koszu albo w opinii klienta.

Czy prognozować średnią, a może średnią warunkową?

Pierwszy odruch bywa prosty. Skoro dany sklep sprzedaje zwykle 40 bochenków chleba dziennie, to można uznać, że jutro też sprzeda 40. Taki sposób myślenia jest łatwy i wygodny. Daje szybki wynik. Problem w tym, że życie sklepu nie działa pod linijkę. Sprzedaż nie jest co dzień taka sama. Raz jest większa, raz mniejsza. Sama średnia zbyt mocno wygładza świat. Pokazuje środek, ale chowa różnice, które w piekarni mają duże znaczenie.

Dlatego ktoś może pójść krok dalej i powiedzieć: nie bierzmy jednej średniej dla wszystkiego, tylko liczymy średnią warunkową. To znaczy średnią dla określonych warunków. Np. osobno dla poniedziałku, osobno dla piątku. Osobno dla rana, osobno dla popołudnia. Osobno dla dni suchych, osobno dla deszczowych. To już jest lepsze podejście, bo uwzględnia fakt, że nie każdy dzień jest taki sam.

Warunki mogą być bardzo proste. Dzień tygodnia. Pora dnia. Pogoda. Czas przed świętem. Czas po święcie. Szkoła otwarta albo ferie. To wszystko wpływa na ruch klientów. Gdy więc ktoś mówi o średniej warunkowej, to w praktyce mówi o średniej policzonej dla konkretnej sytuacji. Nie dla całego roku naraz, lecz dla wycinka rzeczywistości.

Przykład jest prosty. W czwartki sklep może sprzedawać średnio 50 bochenków. Ale gdy rozdzieli się te czwartki na dwa typy, obraz może się zmienić. W czwartki bez deszczu sprzedaż może wynosić średnio 58 sztuk. W czwartki z deszczem tylko 41. Jedna zwykła średnia daje liczbę 50, ale ta liczba nie pasuje dobrze ani do dnia suchego, ani do dnia mokrego. Jest pośrodku. A sklep nie działa pośrodku. Sklep działa w konkretnej pogodzie. No dobrze, a czy podobnie to działa w zimę i w lecie? A co z wiosną?

To samo widać w ciągu dnia. Jeśli ktoś policzy jedną średnią dla całej doby, może uznać, że dany punkt sprzedaje 120 bułek. Ale z tego nic jeszcze nie wynika dla planu pracy. Może być tak, że rano schodzi 80, a po południu 40. A może być odwrotnie. Jedna średnia dzienna nie mówi, kiedy klient przyjdzie. A dla piekarni to rzecz kluczowa, bo pieczywo trzeba upiec wcześniej i ustalić na której zmianie będzie jaki wypiek? Może lepiej było kiedyś jak był jeden sklep i 3 asortymenty?

Średnia warunkowa jest więc krokiem naprzód, ale nadal ma swoje granice. Największy problem polega na tym, że świat nie składa się z prostych, osobnych szuflad. Czynniki łączą się ze sobą. Dzień tygodnia działa razem z pogodą. Pogoda działa razem z porą dnia. Pora dnia działa razem z typem sklepu. A to prowadzi do zależności nieliniowych.

Nieszczęsne zależności nieliniowe

Zależność nieliniowa oznacza tyle, że zmiana jednego czynnika nie daje zawsze tej samej zmiany wyniku. To nie jest układ typu: jeden krok w lewo, minus pięć sztuk; jeden krok w prawo, plus pięć sztuk. W realnym handlu wpływ bywa nierówny, skokowy albo zależny od innych warunków.

Pierwszy prosty przykład: deszcz. Lekki deszcz może prawie nie zmienić sprzedaży w sklepie osiedlowym. Ludzie i tak wyjdą po chleb. Ale mocny deszcz po południu może już wyraźnie zmniejszyć ruch. Czyli wpływ pogody nie rośnie równo. Mały opad i duży opad nie działają tak samo.

Drugi przykład: pora dnia. Rano brak kilku bochenków może oznaczać dużą stratę, bo wtedy ruch jest szybki i klient chce kupić od razu. Po południu ten sam brak kilku sztuk może mieć mniejsze znaczenie, bo ruch jest słabszy albo klienci kupują inny towar. Ta sama liczba braków nie daje więc tego samego skutku o każdej porze. To też jest nieliniowość.

W praktyce oznacza to jedno: nawet dobra średnia warunkowa bywa zbyt uboga. Ona dobrze działa wtedy, gdy świat jest prosty i spokojny. Gdy jednak sklep żyje rytmem dnia, tygodnia, pogody i lokalnych zwyczajów, sama średnia zaczyna przegrywać. Jest za sztywna. Za mało widzi. Nie umie dobrze łączyć wielu wpływów naraz.

Dlatego w pewnym momencie warto przejść dalej. Nie odrzucać doświadczenia. Nie odrzucać prostych statystyk, ale potraktować je jako punkt startu, a nie jako cel końcowy. Średnia i średnia warunkowa mogą dać pierwsze przybliżenie. Mogą pokazać bazę. Jednak tam, gdzie ważna jest dokładność, mniejsze straty i lepszy plan wypieku, potrzebne są modele, które widzą więcej niż prosty przeciętny dzień. To właśnie dlatego nowoczesne prognozy sprzedaży nie kończą się na średniej. Zaczynają się od niej, ale idą dalej.

Regresja liniowa jako pierwszy krok do lepszej prognozy

Liczenie samej średniej warunkowej brzmi rozsądnie tylko do chwili, gdy ktoś spróbuje zrobić to naprawdę. Jeśli piekarnia ma 30 produktów i 15 sklepów, sytuacja szybko staje się trudna. Trzeba by osobno liczyć średnie dla chleba, osobno dla bułek, osobno dla rogalików. Do tego osobno dla każdego sklepu. Potem jeszcze osobno dla poniedziałku, wtorku, środy. Potem dla rana i dla popołudnia. Potem jeszcze dla dni z deszczem i bez deszczu. W pewnym momencie robi się z tego gąszcz liczb, z którego trudno wyciągnąć prostą decyzję. To jest ciężka, katorżnicza praca. Męcząca nie tylko dla właściciela piekarni; to „droga krzyżowa” nawet dla doświadczonego analityka. A co gorsza, efekt nadal bywa zbyt słaby jak na dzisiejsze potrzeby.

Dlatego warto przejść krok dalej. Takim krokiem jest regresja liniowa. Tu poziom pracochłonności istotnie spada. Nie trzeba od razu budować skomplikowanego systemu. Na poziomie podstawowym można to policzyć nawet w Excelu. I to jest jej duża zaleta. Daje prostszy, szybszy i bardziej uporządkowany sposób myślenia o sprzedaży.

Najprościej mówiąc, regresja liniowa to wzór, który łączy kilka ważnych informacji i na tej podstawie wylicza przewidywaną sprzedaż. Do tego wzoru wstawia się zmienne, czyli takie dane, które mają wpływ na popyt. Np. temperaturę, opady, siłę wiatru, dzień tygodnia, numer sklepu albo porę dnia. Potem wzór wylicza, ile sztuk danego wyrobu powinno się sprzedać.

To można sobie wyobrazić bardzo prosto. Średnia mówi: „zwykle sprzedaje się 40 sztuk”. Regresja liniowa mówi coś więcej: „jeśli jutro będzie czwartek, będzie chłodno, będzie padać, a sprzedaż dotyczy sklepu przy osiedlu i pory porannej, to przewidywana sprzedaż wynosi tyle i tyle”. To już nie jest patrzenie na jeden przeciętny dzień. To jest policzenie konkretnej sytuacji.

Właśnie tu widać przewagę takiego podejścia. Wiele rzeczy wiemy wcześniej. Wiemy, jaki jutro będzie dzień tygodnia. Możemy też sprawdzić prognozę pogody. Możemy znać przewidywaną temperaturę, opady i siłę wiatru. Wiemy też, do którego sklepu ma trafić towar. Wiemy również, czy mówimy o sprzedaży przed południem, czy po południu. Skoro te informacje są dostępne wcześniej, można ich użyć do przewidywania sprzedaży dzień przed wypiekami. Tak, właśnie dzięki temu będziemy mogli zaplanować odpowiedni poziom produkcji dzień, dwa dni wcześniej.

To jest ważne także dlatego, że sklepy nie są takie same. Sklep naprzeciwko stacji działa inaczej niż sklep na osiedlu. W jednym punkcie ruch może być szybki i poranny. W innym większa sprzedaż może być rozłożona w ciągu dnia. W jednym miejscu klienci wpadają po drodze. W innym kupują spokojnie, przy okazji codziennych zakupów. Dlatego numer sklepu, a szerzej sam sklep

jako miejsce sprzedaży, też jest ważną zmienną. To nie jest detal. To jedna z podstawowych informacji.

Podobnie jest z porą dnia. Sprzedaż przed południem i po południu często wygląda inaczej. Rano klienci kupują świeże pieczywo na start dnia. Po południu ruch może być mniejszy albo inny. Kto policzy tylko sprzedaż dzienną, ten nadal nie wie, ile towaru powinno być w sklepie rano, a ile po południu. Dlatego warto dodać do wzoru jeszcze jedną zmienną: porę dnia. To bardzo prosty ruch, a może wiele zmienić.

Co to jest ta regresja?

To po prostu sposób liczenia, wzór, jak receptura na wypiek makowca: 0,5 kg maku, 0,02 kg drożdży oraz 10 kg mąki pszennej typu 500. Dokładnie tak samo działa regresja. Wiatr wpływa na sprzedaż obniżając ją o 4%, słońce podnosi sprzedaż o 12%, w czwartki sprzedaż jest wyższa o 20%. Regresja to po prostu receptura ma obliczenie wielkości sprzedaży. W praktyce taki wzór działa tak, że każda zmienna wnosi swoją część informacji. Temperatura mówi coś o zachowaniu klientów. Opady mówią coś o skłonności do wyjścia z domu. Siła wiatru też może mieć wpływ, zwłaszcza przy gorszej pogodzie. Dzień tygodnia pokazuje rytm życia sklepu. Numer sklepu pokazuje różnicę między lokalizacjami. Pora dnia pokazuje, kiedy ruch jest większy, a kiedy mniejszy. Regresja liniowa zbiera to w jedno i daje końcową liczbę.

To jest jak przepis. Do miski trafia kilka składników: dzień tygodnia, pogoda, sklep, pora dnia. Wymieszane razem dają prognozę. Nie trzeba już ręcznie liczyć setek średnich dla każdej możliwej kombinacji. Jeden wzór robi to w sposób uporządkowany.

To właśnie czyni regresję liniową dobrym punktem startu. Jest prostsza niż żmudne liczenie średnich warunkowych. Jest też bardziej praktyczna, bo pozwala uwzględnić wiele rzeczy naraz. A przy tym nie odstrasza. Daje się wdrożyć. Daje się policzyć. Daje się wyjaśnić. Właściciel piekarni nie musi od razu znać wszystkich szczegółów technicznych. Wystarczy, że zrozumie prostą rzecz: jeśli da się wcześniej zebrać ważne informacje o jutrzejszym dniu, to da się też lepiej przewidzieć jutrzejszą sprzedaż.

Regresja liniowa nie rozwiąże od razu każdego problemu. Ale daje pierwszy prawdziwy krok naprzód. Zamiast zgadywać, zaczyna się liczyć. Zamiast ręcznie budować dziesiątki tabel, buduje się jeden spójny wzór. I właśnie dlatego dla wielu piekarni może to być rozsądny początek drogi do lepszej produkcji, mniejszych strat i spokojniejszego planowania.

I kto ma to robić? Tę regresję...

W tym miejscu wielu właścicieli piekarni ma tę samą myśl: to brzmi sensownie, ale kto ma to zrobić. Znowu liczby, tabele, wzory, obliczenia. Nie każdy musi to lubić.

Nie każdy musi to umieć. I nie ma w tym nic złego. Piekarz ma znać piec, ciasto, garowanie, rytm produkcji i sprzedaży. Od obliczeń są ludzie, którzy się tego uczyli.

Warto powiedzieć to wprost. Zbudowanie prostej regresji liniowej nie jest zadaniem dla profesora matematyki. To jest jedna z podstawowych czynności w pracy osób, które zajmują się analizą finansową, raportowaniem, controllingiem albo prognozami. Dla osoby po ekonomii, finansach, statystyce albo nawet dla juniora, który pracował w Excelu i miał kontakt z prostą analizą danych, jest to rzecz zwyczajna. Nie ręczne liczenie na kartce, tylko normalna praca w arkuszu.

To ważna wiadomość, bo wielu przedsiębiorców od razu zakłada, że skoro wchodzi temat modelu, to trzeba zatrudniać drogiego specjalistę od sztucznej inteligencji. Nie. Regresja liniowa to najprostszy model prognostyczny. Właśnie dlatego jest dobrym startem. Jest dość prosta, by ją wdrożyć szybko, i dość użyteczna, by dać realną poprawę.

Dane też zwykle już są. I to jest druga dobra wiadomość. Danych nie trzeba zbierać od zera. Są w punktach sprzedaży, czyli w POS-ach. Każda sprzedaż zostawia ślad. Te dane odkładają się w rejestrze sprzedaży. Księgowość dobrze wie, co to jest. W praktyce są to dane cyfrowe, w formie arkusza Excel lub txt gotowe do eksportu z systemu sprzedażowego. Widać tam, co się sprzedało, gdzie, kiedy i w jakiej liczbie.

Do tego dochodzą dane pogodowe. One są dostępne od ręki z internetu. Można pobrać prognozę temperatury, opadów i siły wiatru. To już wystarcza na prosty model. Następna rzecz to dzień tygodnia. Tę informację daje sama data. Excel potrafi wyznaczyć ją prostą funkcją. Czy jutro jest czwartek, piątek czy sobota, wiadomo z góry. Nie ma tu żadnej przeszkody.

Warto też dodać sklep jako osobną informację. To jest bardzo ważne. Inaczej działa punkt przy stacji, inaczej punkt na osiedlu, inaczej sklep przy szkole, a inaczej przy ruchliwej ulicy. Dobrze jest też rozdzielić sprzedaż na czas przed południem i po południu, bo ruch bywa różny. W efekcie model dostaje kilka prostych danych: datę, dzień tygodnia, pogodę, sklep i porę dnia. To już daje sensowną bazę do prognozy. Co ważne, im więcej danych tym lepiej. Jak mamy dane z 5 lat tym lepiej, dla Excela nie ma różnicy czy baza składa się z 200 linii czy z 200 tys.

Analizę regresji robi nisko wykwalifikowany specjalista, robi raz na rok, zajmuje mu to 2 dni

Tu trzeba podkreślić jedną rzecz. Nie chodzi o to, by właściciel piekarni sam siedział nad Excelem i uczył się statystyki po nocach. Chodzi o to, by wiedział, że to można komuś zlecić. Taki model może przygotować pracownik biurowy, analityk, ekonomista, osoba od raportów albo ktoś z zewnątrz na krótkie zlecenie. Nie musi to być etat. Można zlecić budowę modelu raz, a potem tylko co pewien czas sprawdzać, czy nadal działa dobrze.

“Mam zaufanie i się nie stresuje”

I właśnie tak warto na to patrzeć. Nie jak na wielki projekt technologiczny, tylko jak na zwykłą usługę. Tak jak serwisuje się piec, samochód albo kasę fiskalną, tak samo można co jakiś czas sprawdzić model prognozy. Sam model nie musi być tworzony co tydzień od nowa. Jeśli jest zrobiony poprawnie i oparty na porządnym danych, może działać stabilnie przez wiele miesięcy. Potem wymaga tylko przeglądu, korekty i odświeżenia.

Krótko mówiąc: danych zwykle nie brakuje, narzędzia są powszechne, a sama umiejętność budowy prostej regresji liniowej jest szeroko znana wśród ekonomistów i analityków. Dla branży piekarskiej to dobra wiadomość. Bo oznacza, że nie trzeba bać się matematyki. Trzeba tylko wiedzieć, komu zlecić pracę i czego od tej osoby wymagać.

To jest wiedza podstawowa. Dla analityka tak podstawowa, jak dla piekarza wiedza o tym, jak pracują drożdże i kiedy ciasto rośnie dobrze, a kiedy źle. Dlatego właściciel piekarni nie powinien pytać: czy dam radę to sam policzyć. Lepsze pytanie brzmi: kto ma to dla mnie sprawnie zbudować i jak szybko mogę zacząć z tego korzystać.

Następny krok jest prosty. Najpierw bierze się dane z rejestru sprzedaży. Potem dodaje pogodę, dzień tygodnia, sklep i porę dnia. A potem osoba, która zna Excela i podstawy prognoz, buduje pierwszy model. Bez wielkiej teorii. Bez strachu. Po prostu po to, by ograniczyć straty i lepiej planować pracę.

Gdy Excel już nie wystarcza

W tym miejscu warto powiedzieć rzecz prostą. Właściciel piekarni nie musi rozumieć całej tej mechaniki. Nie musi znać nazw modeli, nie musi znać wzorów, nie musi umieć tego liczyć. To nie jest jego rola. Jego rolą jest wiedzieć, że da się to zrobić lepiej niż w Excelu i że da się to zlecić. Dobrze zrobiony system może działać w tle, sam pobierać dzień tygodnia, sam pobierać prognozę pogody, sam liczyć przewidywaną sprzedaż, a wynik pokazywać na prywatnej stronie firmowej, zwykłej stronie internetowej dostępnej tylko dla właściciela i osób uprawnionych. Rano albo dzień wcześniej można tam

po prostu wejść i zobaczyć: ile przewiduje się sprzedaży w danym sklepie, dla danego wyrobu, przed południem i po południu następnego dnia, za dwa dni.

Kto to ma zrobić?

To już nie jest poziom Excela. To jest poziom osoby, która umie pracować w Pythonie i budować modele prognostyczne. Taki specjalista nie musi codziennie siedzieć w piekarni. Może przygotować rozwiązanie raz, uruchomić je, a potem tylko co jakiś czas sprawdzać i odświeżać. Jeśli dane są porządne, a rynek wokół sklepu nie zmienia się gwałtownie, taki model może działać stabilnie przez wiele miesięcy. Zwykle nie psuje się z dnia na dzień. Częściej problem pojawia się wtedy, gdy zmienia się otoczenie biznesowe: obok otwiera się supermarket, rusza nowa piekarnia, zamyka się konkurent albo zmienia się ruch w okolicy.

Czas zbudowania takiej aplikacji na stronie internetowej to około jednego tygodnia. Można dać sobie 2 tygodnie. To już jest maksymalny czas na zbudowanie takiego rozwiązania. Cena? Załóżmy, że ekspert weźmie 200 zł za godzinę. To bardzo wysoka cena, ale takich specjalistów nie ma za dużo, wszyscy o nich walczą, więc się cenią. Policzmy 80 godzin po 200 zł, co daje 16 tys. zł + vat. Teraz – ile na tym zaoszczędzimy? Czy to się opłaca? Ten system ma chodzić bez obsługi przez co najmniej 12 miesięcy, w praktyce 24 miesiące. Tu nie ma co się zepsuć. Nawet jak zacznie słabiej prognozować, zawsze będzie lepiej prognozował niż regresji. W skali 24 miesięcy wychodzi 25 zł dziennie. Jeżeli model zaoszczędził więcej niż 24-30 zł dziennie to znaczy, że inwestycja się opłaciła. Nie liczymy już, że odpadają nam trudy samodzielnego liczenia oraz stres odpowiedzialności za straty, którego gorący oddech czujemy każdego dnia na swoich plecach. Idziemy z duchem czasu, to przecież przewaga konkurencyjna na cyfrowym polu walki.

Lepsze od regresji

Warto też uczciwie powiedzieć, że regresja liniowa jest dobrym początkiem, ale nie końcem drogi. To model prosty i pożyteczny. Jednak są metody lepiej dopasowane do tego typu zadania. W piekarni nie przewiduje się klasy typu „mało”, „średnio”, „dużo”, tylko konkretną liczbę sztuk. Do takich danych dobrze pasują modele liczące wielkości nieujemne.

Teraz będzie wiadomość dla ewentualnego wykonawcy „cyfrowej wróżki”.

W scikit-learn jest do tego m.in. PoissonRegressor, czyli model Poissona, a krok wyżej stoi HistGradientBoostingRegressor z ustawieniem loss=“poisson”. Ten drugi lepiej radzi sobie z nieliniowością, a dokumentacja scikit-learn podaje też, że dla większych zbiorów jest dużo szybszy od klasycznego gradient boostingu. W przykładowym porównaniu scikit-learn zarówno

PoissonRegressor, jak i HistGradientBoostingRegressor dawały lepszą zgodność prognoz z obserwacjami niż prostszy model liniowy typu Ridge (odmiana regresji liniowej).

Jeśli więc junior chce się rozwijać, to właśnie tu zaczyna się następny krok. Najpierw buduje prostą regresję liniową. Potem uczy się modeli licznikowych, które jeszcze znajdzie w Excelu. A potem przechodzi do modeli drzewiastych, które lepiej wychwytyją zależności ukryte w danych. Dla piekarza najważniejszy wniosek jest prosty: da się mieć prognozę lepszą niż ta z Excela, ale do tego trzeba już człowieka, który zna te narzędzia.

Komu to dać?

Tu jednak pojawia się ważna sprawa. Właściciel piekarni nie powinien oddawać tematu w ciemno. Nie musi rozumieć wszystkiego, ale powinien wiedzieć, czego wymagać. Po pierwsze, model ma przewidywać liczbę sztuk. Po drugie, ma być oceniany nie abstrakcyjnym wskaźnikiem, tylko błędem w sztukach. Nie interesuje go, że wskaźnik wygląda ładnie w raporcie. Interesuje go, o ile średnio model pomylił się w liczbie bochenków, bułek czy rogali. Właśnie dlatego na start dużo lepsza jest miara MAE, czyli średni błąd bezwzględny. W scikit-learn MAE jest opisany jako miara błędu regresji, której najlepsza wartość to 0. Dla biznesu brzmi to jasno: „mylimy się średnio o 4 sztuki”. Z kolei R^2 może być przydatny dla analityka, ale dla właściciela bywa mało czytelny, a dokumentacja wprost podaje, że może być nawet ujemny.

Po trzecie, taki model trzeba sprawdzać tak, jak działa czas. Nie wolno mieszać w nauce wczoraj z jutrem. Dokumentacja scikit-learn zaleca do danych uporządkowanych w czasie TimeSeriesSplit, bo zwykłe metody podziału mogłyby prowadzić do uczenia na przyszłości i oceniania na przeszłości. To jest bardzo ważne. W piekarni model ma udawać prawdziwe życie: najpierw zna stare dane, a dopiero potem ma przewidzieć nowy dzień. Jeżeli specjalista nie rozumiał: mówię o podziale na zmienne treningowe i testowe w metodzie nauki z nadzorem.

Dla specjalisty wskazówka jest więc krótka. Jeśli ma zrobić model lepszy niż Excel, powinien zacząć od celu w postaci liczby sprzedanych sztuk, potraktować dane jako problem regresyjny z porządkiem czasowym, nie robić losowego mieszania train i test, mierzyć błąd w sztukach, a jako sensownych kandydatów rozważyć PoissonRegressor albo – częściej – HistGradientBoostingRegressor(loss="poisson"). PoissonRegressor

to uogólniony model liniowy z rozkładem Poissona, a HistGradientBoostingRegressor wprost dopuszcza strategię poisson.

Cyfrowy piekarz

Dla właściciela piekarni ważne jest coś jeszcze. Taki model nie ma być czarną skrzynką zamkniętą na zawsze. Powinien codziennie pokazywać nie tylko prognozę, ale też własny błąd. Najprościej: model podał prognozę na jutro, piekarnia według niej zaplanowała wypiek, a dzień później widać już sprzedaż i zwroty. Wtedy można sprawdzić, o ile model się pomylił. Takie dane o wielkościach zwrotów może wprowadzać do Excela kierownik produkcji. Jeśli przez wiele tygodni myli się mało, działa dobrze. Jeśli nagle zaczyna mylić się wyraźnie bardziej, to znak, że trzeba go poprawić albo odświeżyć. W ten sposób właściciel nie musi znać matematyki. Wystarczy, że codziennie patrzy na jedną prostą rzecz: jaka była prognoza i jaki był błąd. Można też jako opcję poprosić wykonawcę, aby model sam się douczał co miesiąc na nowych danych. To normalna funkcjonalność takich systemów. Wtedy jego trwałość i wysoka jakość prognoz będzie liczona w latach.

To właśnie daje największy spokój. Nie trzeba wierzyć modelowi na słowo. Można go kontrolować. Można go pilnować. Można wymagać, by raportował własne nie-trafienia. Dzięki temu model nie żyje obok firmy, tylko pracuje dla firmy. Model ma się sam douczać. Ma być bezproblemowy jak aplikacja w telefonie czy piec do wypieku pieczywa.

Zakończenie, happy end

I to jest dobre zakończenie całego tematu. Najpierw była intuicja. Potem średnia. Potem średnia warunkowa. Potem regresja liniowa w Excelu. A na końcu pojawia się rozwiązanie bardziej dojrzałe: model zrobiony przez specjalistę, działający automatycznie, pokazujący prognozę na prywatnej stronie i regularnie raportujący własny błąd. Bez wielkiej filozofii. Bez modnych haseł. Model bezobsługowy, który co miesiąc poprawia swoje zdolności, bo automatycznie się doucza. Po prostu jako narzędzie do lepszego planu wypieku, mniejszych strat i spokojniejszego prowadzenia piekarni. Oczywiście najpierw pojawi się stres, że musimy coś zmienić, potem trud znalezienia kogoś, kto to robi. Trochę stresu, trochę wstydu i nerwów. Idziemy do przodu, warto, naprawdę warto.

Wojciech Moszczyński

Wojciech Moszczyński – absolwent Katedry Ekonometrii i Statystyki Uniwersytetu Mikołaja Kopernika w Toruniu, specjalista z zakresu ekonometrii, finansów, *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 *machine learning* oraz *data science* w środowiskach biznesowych.