

Wszystko co trzeba wiedzieć na temat systemów rekomendacji dla sklepów branży e-commerce



Systemy rekomendacji służą do optymalizacji działalności ekonomicznej w obszarze sprzedaży oraz kosztów. W sprzedaży internetowej występują dwa podmioty, które poddane są systemom rekomendacji. To klienci i towary.

Mówiąc o tworzeniu rekomendacji możemy również mówić o tworzeniu narzędzi pomagających w decyzjach strategicznych. Mowa tu o analizach kanałów sprzedaży, kierunków rozwoju i eliminacji obszarów obciążających przedsiębiorstwo. Tego typu rekomendacje strategiczne nie pasują jednak do tematu, który chciałbym dzisiaj omówić.

Tak więc przedmiotem podstawowego systemu rekomendacji jest klient oraz towar.

Jaki jest cel systemów rekomendacji dla klientów platform e-commerce?

Zasadniczo można wyróżnić dwa cele, które powinny być omawiane oddzielnie.

Po pierwsze – chodzi o to, aby klient kupował więcej. Systemy rekomendacji w dość prosty i tani sposób mogą istotnie podnieść wielkość zakupów dokonywanych przez poszczególnych klientów.

Po drugie – chodzi o podniesienie wskaźnika retencji klientów. Dąży się tutaj do tego, aby klienci nie odchodzili ze sklepów. Oczywiście jest że dużo łatwiej jest stracić klienta niż pozyskać nowego. Z tego przekonania wynika troska o zrobienie wszystkiego co można, aby klientów zatrzymać.

Oba te cele: zwiększenie zdolności zakupowych oraz zatrzymanie klientów realizowane są przez różne, bardzo odmienne narzędzia. Wspomniałem, że cele te powinny być rozpatrywane rozłącznie. Jeżeli koncentrujemy się na

zatrzymaniu klientów, to nie powinniśmy dążyć do wzrostu ich efektywności przychodowej.

Z jednej strony to prawda, ponieważ takie podejście jest zgodne z żelazną zasadą optymalizacji, która mówi, że optymalizować można tylko jedną rzecz, a nie wiele rzeczy na raz. Z drugiej strony jest to niezgodne z intuicją i doświadczeniem. Jeżeli spowodowaliśmy za pomocą systemów rekomendacji, że klient zaczął kupować więcej niż wcześniej, to najprawdopodobniej związana było to z tym, że klient był bardzo zadowolony ze współpracy ze sklepem. Klient zadowolony ze współpracy zostanie w sklepie na dłużej. Oczywiście zawsze może pojawić się kontrargument – przecież można naciągnąć klientów na jakieś promocje i rzeczy, których nie potrzebują lub nie chcą. Klienci świadomie lub nieświadomie zmanipulowani zwykle tracą zaufanie i odchodzą. To dobry argument, który wskazuje, że system rekomendacji musi działać na uczciwych zasadach. Wtedy obok optymalizacji przychodowej również wpływa pozytywnie wzrost retencji klientów w sklepie.

System rekomendacji musi przewidywać przyszłość

System rekomendacji opiera się na zdarzeniach historycznych. Podstawą jego działania jest lista transakcji dokonywanych przez klientów w przeszłości. Jednak prawdziwym celem systemu rekomendacji jest przewidywanie co zrobi klient. Przewidywanie przyszłości jest podstawą optymalizacji. Optymalizacja ta dotyczy zarówno bieżącego procesu zakupowego, jak i przyszłego zaopatrzenia sklepu w produkty¹.

Obsługując pojedynczego klienta e-commerce, system rekomendacyjny może przewidzieć jakie produkty zaproponować, aby zwiększyć szanse dodatkowej sprzedaży. Jednocześnie system rekomendacyjny może pomóc przewidzieć, jakie będzie zapotrzebowanie na określone produkty w przyszłości.

Skuteczne przewidywanie przyszłości jest więc podstawą działania efektywnego systemu rekomendacji.

¹ Mowa tutaj o systemach replenishment, które z wyprzedzeniem przewidują popyt na poszczególne produkty i w porę uruchamiają proces ich zamawiania.

Jaki jest cel tworzenia systemów rekomendacji dla towarów?

Tutaj należy wskazać na pewną nieścisłość. System rekomendacji, jak sama nazwa wskazuje, służy do rekomendowania czegoś, podpowiadania czegoś. Czy można coś podpowiadać towarom? Można jednak optymalizować strukturę tych towarów, można obserwować tendencje i procesy używając narzędzi statystycznych. Dzięki temu możliwe jest zoptymalizowanie struktury towarów w sklepie, wysyłania odpowiednich asortymentów w miejsca, gdzie są one bardzo potrzebne i zabierania ich z miejsc, w których nie cieszą się zbyt dużą popularnością. Dzięki systemowi rekomendacji towarów możliwe jest zoptymalizowanie procesu sprzedaży i uzyskanie wysokich zysków lub zmniejszenie kosztów.

Podstawowym podmiotem, do którego skierowany jest system rekomendacji jest klient, ponieważ to on podejmuje decyzje i jest elementem w jakimś stopniu niestabilnym, który należy obserwować w ujęciu predykcyjnym.

System rekomendacji analizuje zderzenia zachowań klientów ze swego rodzaju „zachowaniem” towarów. Każdy klient ma swój niepowtarzalny wzór zachowań, który bardzo rzadko się zmienia. Towary mają pewne okresy popularności, mają też między sobą pewne powiązania. Klienci sami walidują towary i powiązania między nimi. Rolą nauki o danych jest wychwycenie tych więzi między klientami a towarami.

Najważniejszym narzędziem w systemach rekomendacji jest grupowanie podmiotów w klastrach. Tak grupowani są zarówno klienci, jak i towary. W klastrach grupuje się obiekty podobne do siebie pod względem określonych cech. Niezwykle ważne jest, aby wskazywać te cechy, które są ważne dla procesu sprzedaży. Powyższy opis może wydawać się nieco zagmatwany i niezrozumiały, dlatego poniżej będę starał się sprecyzować krok po kroku sposób tworzenia systemów rekomendacji.

Wewnętrzne źródła danych do systemów rekomendacji e-commerce

Podstawowym źródłem danych w systemach sprzedażowych e-commerce jest rejestr sprzedaży z zawartymi w nim logowaniami klientów. Historia zakupów poszczególnych klientów jest śledzona i modelowana przez analityków danych.

Podobne rejestry występują w sprzedaży detalicznej, w której klienci używają kart lojalnościowych. Zgodnie z badaniami, które prowadziłem rok temu, około 70% klientów robiących zakupy w największych polskich sieciach detalicznych wprowadzało podczas transakcji kody kart lojalnościowych. Klienci są skutecznie nakłaniani do rejestrowania

się w systemie sklepu. Sieci detaliczne, takie jak Biedronka, stosują szereg promocji. Część z nich jest możliwa wyłącznie dla osób, które zarejestrowały swoją kartę klienta. Wróćmy jednak do systemów rekomendacji dla platform e-commerce.

Kolejnym źródłem danych dla systemów rekomendacji e-commerce są statyczne dane klienta. Dane te pozyskuje się w trakcie zakładania konta klienta i nie ma ich w rejestrze sprzedaży.

Klienci anonimowi

Część klientów dokonujących zakupów nie rejestruje się w sklepach internetowych. Czasem klienci z jakichś powodów nie ujawniają się w systemie.

Pojawiają się więc transakcje bez informacji ID klienta. Transakcje te zawierają jednak szereg informacji, które pozwalają systemom z łatwością dopasować dane klienta do jego numeru ID. Mowa tutaj o numerze karty kredytowej, numerze telefonu, adresie, na który klient życzy sobie otrzymać kupowany produkt. Możemy też wspomnieć o unikatowym adresie urządzenia, z którego dokonywane są zakupy. Klient identyfikowany według metadanych, który nie chciał ujawnić, może być przyporządkowywany do transakcji za pomocą prostych aplikacji działających w ramach systemu e-commerce.

Należy pamiętać o tym, że klienci, którzy zauważą, że są identyfikowalni przez inteligentne systemy rekomendacyjne, mimo że nie ujawnili swoich danych, mogą utracić zaufanie do sklepu, może ich to skłonić do odejścia.

Klienci wieloracy

Z kontami wielorakimi mamy do czynienia wtedy, gdy jedno konto klienta służy wielu osobom. Na przykład firma produkcyjna ma jeden kod ID w sklepie internetowym i tam dokonywane są zakupy do wszystkich działów firmy. Kupowane są rury i zawory hydrauliczne, materace, elektronarzędzia i farby. Co gorsze, często zakupów dokonują różne osoby mające różne wzorce zachowań. Z konta firmowego mogą też być dokonywane zakupy prywatne pracowników.

Generalnie konta wielorakie są elementem zakłócającym efektywność systemów e-commerce. Powinny być eliminowane, podobnie jak robi się z wartościami odstającymi w modelowaniu. Niestety zazwyczaj konta wielorakie to konta o wielkich obrotach. Jednym ze sposobów radzenia sobie z takimi kontami jest namawianie firmy do tworzenia wielu kont w zamian za otrzymywane rabaty i promocje za specjalizację.

Zewnętrzne źródła danych

Obok informacji zgromadzonych w bazach sklepu internetowego ważne są również dane zewnętrzne. Niestety klient nie podejmuje decyzji wyłącznie na podstawie tego co widzi w sklepie. Dużą część swoich decyzji klienci opierają na informacjach zewnętrznych. Jeżeli ktoś widział wcześniej w innym sklepie internetowym płyn za 5 zł, to nie skusi się na zakup płynu do mycia naczyń za 8 zł. Reklama

rekomendacyjna, która pojawi się w dole okna zakupowego z płynem do mycia naczyń za 10 zł bardziej zniechęci klienta, niż skłoni do dalszego kupowania.

Ten sam klient skusi się na zakup płynu do mycia naczyń za 8 zł, jeżeli dostanie coś w zamian. Żeby zaproponować atrakcyjną ofertę system musi być doskonale zorientowany sytuacyjnie. Musi znać klienta, ale również musi znać rynek i ofertę konkurencji.

Informacja o cenach i dostępności towarów konkurencyjnych powinna więc znaleźć się jako zmienna w modelach rekomendacji. Modele predykcyjne powinny również zawierać bardzo wiele regularnych informacji zewnętrznych, takich dzień tygodnia, miesiąc, sezon oraz wiele ważnych zmiennych sytuacyjnych takich jak prognoza pogody, wskaźnik optymizmu konsumentów. System powinien znać trendy i mody na rynku. Wszystkie te informacje są przetwarzane jako zmienne w świecie cyfrowych modeli matematycznych.

Klastrowanie populacji klientów

Technicznie możliwe jest względnie precyzyjne przewidywanie zachowań każdego klienta. Kupujący zachowują się w sposób powtarzalny. Jednak nikt nie próbuje traktować klientów indywidualnie, ponieważ rozwiązanie to byłoby nieefektywne. Celem procesu klastrowania, czyli bardziej efektywnej formy do prostego grupowania, jest znalezienie klientów, którzy mają bardzo podobne zachowania. Klient ma indywidualne zwyczaje i nawyki, jego zachowanie jest podobne do zachowania innych osób. System stara się przyporządkować do klastra ludzi bardzo podobnych pod względem określonych cech. Stają się oni swego rodzaju cyfrowymi bliźniakami, działają bardzo podobnie, mają te same fobie, preferencje, częstotliwości dokonywania zakupów i wrażliwość na reklamy. Populacja składająca się ze 100 tys. klientów może być więc podzielona na 5000 klastrow. Ludzie przyporządkowani do poszczególnych klastrów będą traktowani przez system sprzedażowy inaczej. Dla każdej z tych grup stosowana będzie inna formy zachęcania i perswazji. Doskonałe rezultaty daje znajdowanie w ramach jednego klastra osób, które odstają od innych i które, dzięki znajomości cech pozostałych osób w klastrze, łatwo namówić na większe lub efektywniejsze zakupy.

Transpozycje podziałów

Przyporządkowywanie klientów do klastrow odbywać się może według rozmaitych zmiennych. Wstępnie można zastosować podejście RFM² (*Recency, Frequency and Monetary*), gdzie klienci są grupowani według częstotliwości i wielkości dokonywanych zakupów. W ten sposób dokonywana jest wstępna analiza klientów. Następnie klienci mogą być grupowani według asortymentu, jaki wybierali. W ten sposób może powstać pierwszy zbiór klastrow

² „Mastering E-commerce Product Recommendations in Python. Using RFM Scores and TF-IDF Scores for Product Recommendations”, Sadrach Pierre, Ph.D. (<https://medium.com/datafabrika/mastering-e-commerce-product-recommendations-in-python-7c12a4bf0c2c>)



klientów. Teraz można zrobić transpozycję, czyli używając grupowania RFM możemy nałożyć klastry asortymentowe i przeanalizować jakie są skupiska klientów poszczególnych transpozycjach. W ten sposób tworzy się macierze transpozycji, które wskazują na nisze rynkowe, anomalie i miejsca, w których proces sprzedaży jest nienaturalnie nadmierny.

Klienci mogą być przyporządkowani do rozmaitych zbiorów klastrow i w każdym z tych zbiorów można analizować pod kontem korelacji lub związków nieliniowych³ z innymi podziałami prostymi takimi jak sezonowość, wielorakość, miejsce zamieszkania, płeć czy dzień tygodnia. Zrozumiemy, że możliwość analizy zachowań klientów jest praktycznie nieskończona.

Podobnie grupować można towary według częstotliwości sprzedaży, rotacji lub sezonowości. Obok prostych grupowań można tworzyć złożone plastrowanie populacji towarów. Klastry towarów i klientów mogą być ze sobą zestawiane, tworząc nieskończone ilości wzorców zachowań. Najważniejsze jest, żeby znaleźć powtarzalność wzorów zachowań klientów w odniesieniu do powiązanych ze sobą towarów. Bardzo ważna jest tu umiejętność likwidowania nadmiaru informacji.

³ Związek pomiędzy grupowaniem RFM i zbiorem klastrow asortymentowych nie jest związkiem liniowym, ponieważ RFM rozpatruje się w kategoriach danych dyskretnych, czyli klasowych.

Prosty system e-commerce w praktyce

Założmy, że na platformie e-commerce pojawia się klient. Klient ten jest jednym z 140 klientów zgrupowanych w jednym klastrze. W klastrze tym klienci zachowują się mniej więcej w zbliżony sposób. Klient ten włożył do koszyka młotek. System momentalnie znajduje powiązany z nim drugi przedmiot. Jest nim śrubokręt. Okazało się, że w tym klastrze klienci, którzy wybrali śrubokręt zwykle kupowali też młotek oraz jeżeli wybrali młotek, to kupowali również śrubokręt. Spośród 140 klientów 38 dokonało w przeszłości właśnie takiego wyboru. System szacuje więc, że klient, który wybrał na platformie e-commerce młotek z prawdopodobieństwem 24% za chwilę włoży do koszyka śrubokręt. Jednocześnie system wie, że w 76% klient ten nie włoży śrubokręta do koszyka⁴. Duże prawdopodobieństwo kupienie śrubokręta powoduje, że system wyświetla rekomendację. Klient na pasku platformy zakupowej znajduje kilka śrubokrętów. Nie wiemy czy bez tej informacji reklamowej klient kupiłby śrubokręt. Należy zakładać, że klienci z tego klastra jeżeli nie otrzymają reklamy ze śrubokrętami, to prawdopodobnie go nie kupią. Jednak zaaplikowana rekomendacja zwiększa drastycznie prawdopodobieństwo zakupu tego narzędzia. Tak działa system oparty o dwa ważne narzędzia, a mianowicie klastrowanie klientów i tworzenie tzw. par rekomendacji⁵.

Pary rekomendacji

Produkty sprzedawane parami lub trójkami można wyszukiwać bez oparcia o klastry klientów. Po prostu jeżeli ktoś kupuje gąbkę do zmywania naczyń, to prawdopodobnie też będzie kupował płyn do mycia naczyń. Z jakichś powodów jednak część klientów tak robi, część osób kupuje same gąbki. Może to wynikać z tego, że gąbki nie będą przeznaczone do zmywania naczyń lecz na przykład do czyszczenia felg w samochodzie.

A może jest tak, że gąbki i płyn do mycia naczyń kupują częściej kobiety, a gąbki bez płynu do mycia naczyń kupują częściej mężczyźni. To typowy wzór zachowań, w którym istotną rolę odgrywa przynależność klienta do określonej grupy, klastra. Ten prosty przykład wskazuje na to, że ważne jest, aby systemy par rekomendacji używać w oparciu o klastry. Takich pojedynczych wzorców spotykamy tysiące. Na szczęście wyszukiują je i selekcionują masowo narzędzia.

⁴ Generalnie nikt nie wnika dlaczego tak się dzieje, to czysta statystyka, nie jest ważne, żeby dowiedzieć się jakie są prawdziwe powody takich wyborów.

⁵ „Understanding the Apriori Algorithm in Market Basket Analysis”, Ajeeth. Medium <https://medium.com/@ajeethaj99/understanding-the-apriori-algorithm-in-market-basket-analysis-bbeceba4d348>

Analiza niekompletności zakupów w klastrze

Jak wspomniałem wcześniej klienci przyporządkowani do jednego klastra cechują się określonym zbiorem specyficznych zachowań. Weźmy na przykład klaster osób, które kupują artykuły związane z pokryciami dachowymi. Algorytm wziął pod uwagę częstotliwość oraz wartość kupowanych przedmiotów oraz typ klienta (na przykład small business). Zebrana spora grupa klientów została sklasyfikowana jako drobni wykonawcy dekarze. Teraz możliwe jest dokonanie kompletności prowadzonych przez nich zakupów. Część klientów kupuje w obszarze określonych kategorii oferowanych przez sklep, takich jak materiały izolacyjne, rynny, dachówki, pokrycia bitumiczne, lepiki.

Założmy, że analiza kompletności wykazała, iż 27% klientów z tego klastra kupuje pokrycia bitumiczne lecz nie kupuje lepików. Jednak technologia tworzenia pokryć bitumicznych zmusza do wykorzystywania lepików. Wynika z tego, że spora część klientów nie kupuje lepików w naszym sklepie lecz kupuje je gdzie indziej⁶. To przykład wykrycia niszy rynkowej.

Nie pozostaje nic innego jak zaproponowanie wyłącznie tym 27% klientów z klastra pokrycia dachowe doskonałej oferty na lepiki. Ważne jest, żeby klienci zmienili swoje zwyczaje i zaczęli kupować lepiki w naszym sklepie.

Analiza przerwanych zwyczajów

Grupowanie klientów w klastry zakupowe sprzyja możliwości wykrycia zmiany ich zwyczajów.

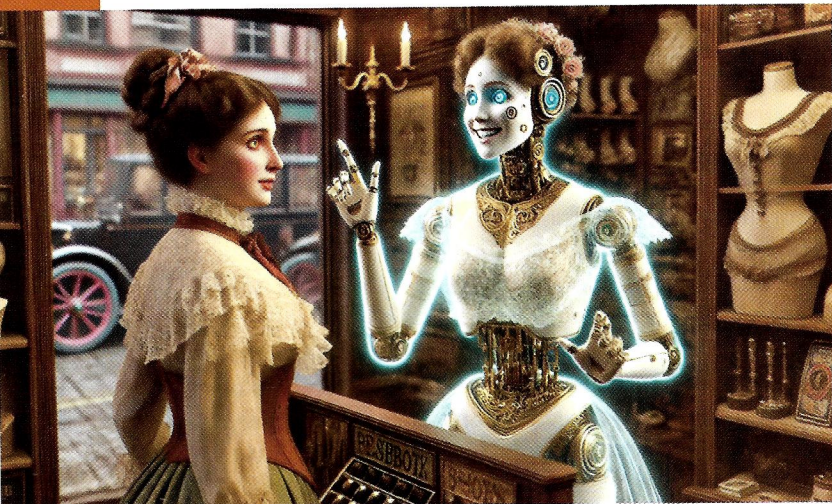
Założmy, że mamy klaster hydraulików. Sezonowo w klastrze tym kupowane były małe hydrofory. Sezonowość wynikała z tego, że na wiosnę często okazywało się, że małe hydrofory zamontowane na działkach rekreacyjnych trzeba było wymieniać. Jednak w pewnym momencie okazało się, że hydrofory nie były sprzedawane na wiosnę w klastrze hydraulików.

Gdybyśmy analizowali poziom sprzedaży hydroforów bez klastrów moglibyśmy nie wykryć tej anomalii. Hydrofory mogą być kupowane przez różne grupy klientów. Jednakże na hydraulikach zależy nam najbardziej. Dzięki wykrytej anomalii możliwe jest zbadanie sytuacji i zaproponowanie nowych lepszych warunków tak, żeby hydraulicy nie kupowali hydroforów u konkurencji. Działanie to nie będzie wprowadzone automatycznie wykonane przez system rekomendacji lecz system rekomendacji pozwala na przeprowadzenie takiego działania przez menedżerów sprzedaży.

Jak zatrzymać klientów w sklepie?

Dotąd moje rozważania koncentrowały się na wzroście efektywności zakupowej klientów. Jeden klient kupujący młotek dzięki trafnie wyświetlanej reklamie mógł kupić dodatkowo śrubokręt. Kobieta kupująca płyn do mycia naczyń dostawała rekomendację zakupu gąbek.

⁶ Niszę tę można wykryć przez znalezienie par rekomendacji lecz może okazać się, że nikt nie kupuje lepiku w naszym sklepie bo: „każdy dekarz wie, że najtaniej jest zamówić lepik u producenta” i wtedy algorytm nie wykryje tej zależności.



Dzięki tym i tysiącom podobnie efektywnie wykorzystywanym wzorcom łączna średnia efektywność zakupowa klientów może wzrosnąć o 30-40%.

Odczują oni też poprawę komfortu zakupów, ponieważ dobrze dopasowana podpowiedź zwykle jest mile widziana u klientów.

Wzrost efektywności zakupowej klientów może jednak nie zrekompensować spadku sprzedaży wynikającego z odejść części klientów. Klienci odchodzą zwykle z jakiegos powodu, zwykle przyczynia się do tego jakaś źle dopasowana format traktowania lub oferta innych sklepów. Powód może też nie być związany z działalnością sklepu. Aby uniknąć odejść wynikających z działalności naszego sklepu należy stworzyć kolejny system rekomendacji, który będzie skoncentrowany wyłącznie na zatrzymywaniu klientów.

Tu również musimy przeprowadzić klastrowanie i podzielić klientów na grupy zrzeszające bliźniaków cyfrowych. Następnie konieczne jest uruchomienie testów przetrwania.

Survival test⁷ to metoda opracowana dawno temu dla placówek medycznych, która miała na celu statystyczną analizę czasu przeżywalności pacjentów z określonym prawdopodobieństwem w stosunku do czasu w oparciu o obecnie stosowane zabiegi medyczne. W ten sposób testowano różne kuracje i w oparciu o to narzędzia obliczano prawdopodobieństwo w jakim nastąpią przyszłe zgony pacjentów. Tak więc stosując kurację dla określonych grup pacjentów, przewidywano, że mają oni na przykład 80% szans na przeżycie pierwszego roku, 52% na przeżycie drugiego roku i 10% na przeżycie trzeciego roku. Narzędzie to szybko zostało zaadaptowane przez firmy komercyjne. Pojęcie zgonu pacjenta zostało zamienione na pojęcie odejścia klienta ze sklepu.

Dzięki tej metodzie łatwo jest określić prawdopodobieństwo utrzymania klienta przez najbliższy lata. Można dodawać do algorytmu różne bodźce, takie jak karty lojalnościowe, rabaty, okresowe indywidualne zniżki i uzyskiwać informację jaki będzie efekt zastosowania tych technik w utrzymaniu klienta. Tym stosunkowo łatwym i prostym

⁷ „Data Science concept — Time-to-event Analysis (Survival Analysis)”. Kiel Dang. Medium <https://medium.com/@kirudang/data-science-concept-time-to-event-analysis-survival-analysis-b2b4a4c1828a>

sposobem można w sposób drastyczny zmniejszyć prawdopodobieństwo odejścia klienta. Co ciekawe, taki system może również służyć do bieżącej obsługi klientów. Może być podręcznym doradcą przedstawicieli handlowych rozmawiających z szczególnie klientami.

Analiza rezygnacji zakupowych

Ważnym elementem poprzedzającym decyzję o odejściu klienta z platformy e-commerce jest wskaźnik rezygnacji zakupowych. Klient może zdecydować się na odejście ze sklepu, ponieważ w ciągu dłuższego czasu nie mógł znaleźć właściwej dla siebie oferty. A takiej decyzji również nie wykryje test przetrwania.

Aby możliwe było utworzenie wskaźnika rezygnacji zakupowej należy zrekonfigurować system zbierania danych. Jak wspomniałem powyżej system rekomendacyjny tworzony jest na bazie rejestru sprzedaży. Rejestr sprzedaży składa się z transakcji, które zostały zawarte z klientem w określonym czasie i na określone towary. Jeżeli klient wejdzie do sklepu internetowego i kręci się po różnych stronach tego zachowania nie znajdziemy w rejestrze sprzedaży.

Jedynym sposobem na analizę zachowań klienta jest historia przeglądania stron sklepu internetowego. Wymaga to zastosowania specjalistycznych narzędzi śledzących, jednak przedsięwzięcie to jest opłacalne, ponieważ efektywnie wskazuje na wahania klienta oraz jego determinację zakupową⁸.

Jak zbudować system rekomendacji?

Najważniejsze w budowie systemu rekomendacji jest wskazanie celu. Na szczęście zwykle wszystkie firmy chcą tego samego, czyli wzrostu efektywności zakupowej klientów. Jednocześnie sklepy e-commerce najczęściej chcą wskazania czynników, które przyczyniają się do odejść klientów. Oba te cele można łatwo pogodzić opierając się na prostych zasadach budowania systemu rekomendacji.

Najważniejsze są dane, które muszą być kompletne, wiarygodne i obszerne. Na podstawie małej ilości danych nie jest możliwe stworzenie efektywnego systemu rekomendacyjnego. Rejestr transakcji sprzedażowych zawierający minimum kilkaset tysięcy operacji powinien być połączony z innymi bazami uzyskanymi ze źródeł zewnętrznych.

Kolejnym krokiem jest właściwe wykorzystanie metod klastrowania danych. Często przy bardzo wielu zmiennych stosuje się algorytm PCA (*Principal Component Analysis*). To metoda łączenia wielu

⁸ „Everything You Need to Know About Tracking Consumer Behavior Online”, Illia Lahunou, Verfacto <https://www.verfacto.com/blog/data-driven-marketing/tracking-consumer-behavior-online/>

cech w 3, 4 skonsolidowane cechy składowe. Cechy cząstkowe łączone są tam w cechy główne na podstawie siły wzajemnych kowariancji dla cech⁹.

Wyobraźmy sobie informacje o kliencie i spróbujmy policzyć ile może on mieć cech. To może być adres zamieszkania, ilość transakcji w miesiącu, ilość różnych kategorii produktów, forma płatności oraz dziesiątki innych zmiennych. Przy pomocy takich chmur danych nie można efektywnie przyporządkować klienta do klastra. Dlatego właśnie stosuje się różne narzędzia, które powodują zmniejszenie ilości zmiennych przez ich redukcję lub grupowanie w duże zmienne za pomocą algorytmu PCA.

Analiza przetrwania, podobnie jak narzędzia zwiększające efektywność zakupową klientów, jest narzędziem prostym w tworzeniu i obsłudze. Jednocześnie metodologia jest uważana za żmudną z uwagi na to, że pojawia się ogromna ilość interakcji wątków i zmian zewnętrznych, które wymagają od badacza doświadczenia. Bardzo ważną umiejętnością jest utrzymywanie właściwych priorytetów badań, polegających na eliminacji nieefektywnych wzorców i rozbudowaniu wzorców efektywnych. Wymaga to wysokiej samodyscypliny i cierpliwości.

Najpierw prototyp potem wdrożenie

W tworzeniu systemu rekomendacji najpierw tworzy się prototyp, zazwyczaj pisany w języku Python i oparty o biblioteki tego języka. Prototyp jest testowany i doskonalony w środowisku programistycznym. Niestety prototyp nie nadaje się do wykorzystania w procesie sprzedażowym. Rozwiązanie to musi być wdrożone do środowiska produkcyjnego. Tym zajmuje się inżynier danych, który wdraża program do środowiska chmurowego lub do lokalnego środowiska obsługującego systemy sprzedażowe.

Systemy wsadowe i strumieniowe

Systemy rekomendacji mogą bazować na rejestrach historycznych. Prototyp jest zbudowany na bazie pewnej historii, która z biegiem czasu się zmienia¹⁰. System więc musi być aktualizowany. Ktoś kto do tej pory kupował jogurty nagle zaczął



kupować kefiry. Ktoś kto raz kupił młotek i śrubokręt raczej już nie będzie chciał powtórzyć swoich decyzji. Wiemy jednak, że klient, który kiedyś kupił młotek i śrubokręt po pewnym czasie chce kupić wiertarkę. Te przykłady pokazują, że konieczne jest okresowe aktualizowanie systemu rekomendacji. Im ta aktualizacja jest dokonywana częściej, tym lepiej. A najlepiej w pewnych określonych sytuacjach jest, gdy aktualizacja prowadzona jest na okrągło. Możliwe jest więc budowanie systemu rekomendacji opartego na stałym dostarczaniu informacji w formie strumienia danych i przetwarzania ich do systemu rekomendacyjnego¹¹.

Podsumowanie

Każdy wie, że właściwie sformułowana, szyta na miarę oferta powoduje wzrost efektywności zakupowej klienta i dodatkowo powoduje wzrost przywiązania klienta do sklepu. Problem polega na tym, że aby dostosować ofertę dla klienta musimy bardzo dużo wiedzieć na jego temat i użyć narzędzi, które spowodują, że ta wiedza będzie efektywnie wykorzystana.

Żeby możliwe było takie przedsięwzięcie trzeba doskonale zorganizować dane oraz proces ich przetwarzania i wykorzystywania w analizie.

Proces jest trudny lecz potrzebny, ponieważ jeżeli my tego nie zrobimy, to robi to konkurencja. System rekomendacji na platformie e-commerce może być doskonałym źródłem przewagi konkurencyjnej, w której bez ponoszenia większych nakładów na ogólną reklamę lub nadmierne, szerokie stosowanie promocji, uzyskujemy większą efektywność sprzedaży, a dzięki temu wyższe zyski zabezpieczające trwałość i zwiększającą wartość prowadzonej firmy e-commerce.

Wojciech Moszczyński

⁹ „Segmentation of a population containing very many features. Pca analysis and clustering by k-means metod”, Wojciech Moszczyński. Medium, <https://wtm695450085.medium.com/segmentation-of-a-population-containing-very-many-features-928ca8ac5112>

¹⁰ Raczej nie mówimy tutaj o wytrenowanych modelach prognozytycznych ponieważ systemy rekomendacji zazwyczaj mają charakter systemów tworzonych bez nadzoru, czyli nie testowanych na próbkach testowych.

¹¹ „Microservices in Python: Kafka and Django”, Manish Sharma, Medium, <https://medium.com/@mansha99/microservices-using-django-and-kafka-3776e8592ef3>

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.