

Czy Machine Learning może zwiększyć szansę zwierząt ze schroniska na adopcję?

Autor: Dawid Wiśniewski

O Autorze

Dawid Wiśniewski

Z wykształcenia jestem ekonomistą ze specjalizacją w Ekonomii Matematycznej. Wątek programistyczny towarzyszył mi przez całą moją edukację. Od Pascala i PHP w szkole średniej, które to zainspirowały mnie do pierwszych kroków w tworzeniu prostych programów, przez C, SASa oraz R na studiach, wspierających mnie w przetwarzaniu i analizie danych, po okazjonalne korzystanie z VBA, by ułatwić sobie żmudne zadania w pracy.

Udział w Bootcampie Kodołamacz Data Science był oczywistą decyzją w moim dążeniu, żeby zrozumieć czym „to całe ML i AI” różni się od mojej codziennej pracy w obszarze analityki. Pokazał mi, że merytoryczna wiedza statystyczna i ekonometryczna się nie dezaktualizuje oraz jak popularne języki programowania wykorzystywane w Data Science, w tym przede wszystkim Python, oraz najświeższe algorytmy uczenia nadzorowanego, otwierają nowe możliwości.

Wstęp

W projekcie, o którym przeczytacie poniżej, starałem się skupić na tym, co według mnie w Data Science jest najważniejsze, choć często spychane na drugi plan – jakie wnioski możemy wyciągać z zaawansowanych modeli matematycznych i statystycznych, żeby rzeczywistość, która nas otacza zmieniać na lepsze.

Na świecie jest wiele bezpańskich zwierząt, a dla wielu z nich proces adopcji jest czasochłonny i nie zawsze kończy się sukcesem. Technologia, w tym portale pośredniczące w adopcji, zapewne ją ułatwiły, jednak nadal wiele zwierząt długo oczekuje w schroniskach na nowych właścicieli.

Wspaniale byłoby, gdyby technologia była w stanie ten proces przyspieszyć. Data Science może pomóc w zrozumieniu prawidłowości, które wpływają na czas oczekiwania zwierząt na adopcję. Ich znajomość może zostać wykorzystana do tworzenia lepszych ogłoszeń adopcyjnych

Dane

[PetFinder.my](https://www.petfinder.my) jest największą platformą pośredniczącą w adopcji zwierząt w Malezji. Dysponuje ona bazą danych przeszło 15 000 zwierząt. Pod koniec 2018 roku udostępniła swoje dane w ramach konkursu na platformie Kaggle, w którym uczestnicy mieli stworzyć model przewidujący czas adopcji zwierząt.

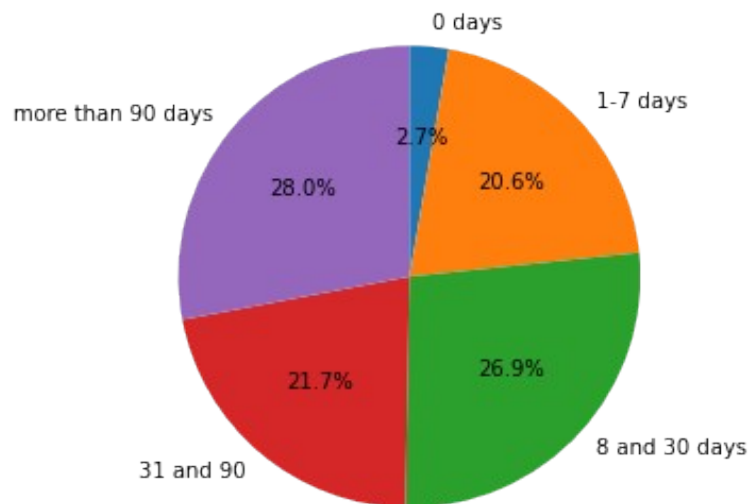
Postanowiłem przeanalizować te dane, jednak moim priorytetem było zrozumienie, co sprawia, że niektóre zwierzęta krótko oczekują na adopcję, a dla innych czas ten jest znacznie dłuższy. W przeciwieństwie do zadania konkursowego z Kaggle, najwyższa celność prognoz czasu adopcji nie była dla mnie aż tak ważna.

Czy zaadoptował(a)byś tego zwierzaka?

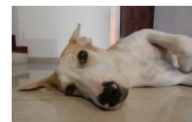
W oryginalnym zbiorze danych mamy informację, jak długo zwierzę oczekiwało na adopcję od momentu opublikowania jego profilu w portalu. Czas adopcji w oryginalnych danych został podzielony na pięć grup:

- adopcja w dniu publikacji ogłoszenia adopcyjnego (3%),
- adopcja poniżej jednego tygodnia od publikacji ogłoszenia (21%),
- czas oczekiwania między tygodniem a miesiącem (27%),
- adopcja między miesiącem a trzema miesiącami (22%),
- czas oczekiwania na adopcję powyżej trzech miesięcy (28%).

Adoption Speed frequencies



Dlaczego aż tak dużo zwierząt oczekiwało na adopcję powyżej trzech miesięcy? Patrząc na ich zdjęcia (poniżej) nie jestem w stanie zrozumieć, dlaczego te słodziaki nie zostały adoptowane od razu. Postanowiłem więc spróbować dowiedzieć się, jaka jest recepta na szybką adopcję zwierzaka.



W celu uogólnienia analizy, ostatecznie podzieliłem zwierzęta na dwie grupy: grupę zwierząt adoptowanej szybko (do 30 dni od publikacji profilu) oraz grupę zwierząt adoptowanych wolno (powyżej 30 dni od publikacji profilu).

Azor, Puszek, co możecie mi o sobie powiedzieć?

Żeby zrozumieć historię adopcji każdego z 15 tysięcy zwierząt, musiałem je lepiej poznać. W danych z [PetFinder.my](https://www.petfinder.com) poza samymi zdjęciami i informacją o czasie adopcji, znajduje się szereg dodatkowych

informacji z profilu każdego ze zwierząt. Oprócz podstawowych informacji osobowych, takich jak rodzaj zwierzęcia (w bazie były jedynie psy i koty), jego wiek, rasa czy kolor, znajdziemy też te „bardziej wrażliwe”, jak chociażby informacja o tym, czy zwierzę zostało wysterylizowane. Dodatkowo do dyspozycji mamy treść ogłoszeń adopcyjnych. Same zdjęcia zwierząt również dostarczają nam dodatkowe dane, m.in. o dominującej na nich kolorystyce czy o jakości i wymiarach zdjęcia.

Punktem wyjścia do dalszych analiz było około 100 zmiennych objaśniających, obejmujących informacje o zwierzęciu, informacje z profilu i informacje na podstawie zdjęć. W toku dalszego przetwarzania danych tworzone były kolejne zmienne, łącznie ponad 300, mające na celu zidentyfikowanie kluczowych aspektów profilu, które mogą wpływać na czas adopcji.

Modelowy profil

Dane zostały poddane szeregowi transformacji wynikających z najlepszych praktyk uczenia maszynowego, od grupowania kategorii zmiennych kategorycznych, przez tokenizację do standaryzacji zmiennych ilościowych.

W celu zidentyfikowania kluczowych czynników wpływających na czas adopcji, wypróbowałem szereg modeli z zakresu Machine Learning. Ostatecznie zdecydowałem się na używanie modelu Regresji Logistycznej jako kompromisu pomiędzy celnością modelu (63% vs. bazowe 50%, wynikające z rozkładu prawdopodobieństwa szybkości adopcji), a łatwością jego interpretacji. Jak okazało się w toku dalszych analiz, modelowy profil, który skutkuje szybką adopcją, nie jest prosty do stworzenia. Na czas adopcji wpływa wiele czynników i większość z nich tylko w niewielkim stopniu zwiększa szansę na szybką adopcję. Co więcej, dla wielu profili nie udało mi się zrozumieć ich recepty na sukces (szybką adopcję), o czym świadczy umiarkowana precyzja modelu.

Model Regresji Logistycznej pozwala na zidentyfikowanie czynników, które wpływają na analizowane zjawisko o charakterze kategorycznym – w naszym przypadku przedziały czasu adopcji stanowią analizowane kategorie. Jedną z zalet takiego modelu jest możliwość policzenia jak zmiana wartości jednego z czynników (przy pozostałych czynnikach pozostających na niezmiennym poziomie), wpłynie na szansę danego zwierzęcia na to, że zostanie zaadoptowane w ciągu 30 dni od dnia publikacji ogłoszenia, w stosunku do szans dłuższego oczekiwania na adopcję. W dalszej części artykułu będziemy często wykorzystywać tę własność modelu przytaczając szanse na szybką adopcję.

Ludzie lubią zdjęcia

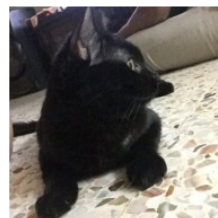
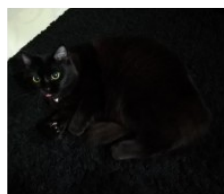
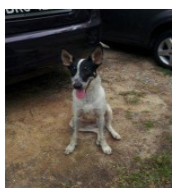
Obecnie wiele ludzi chętnie dzieli się swoimi zdjęciami, w różnych social mediach i spędza dużo czasu przeglądając zdjęcia innych. Prawdopodobnie dlatego też oczekują, że profil zwierzęcia do adopcji będzie miał zdjęcie. Umieszczenie zdjęcia zwierzęcia na jego profilu adopcyjnym zwiększa szanse szybkiej adopcji niemal siedmiokrotnie!

Wybór zdjęcia również może mieć wpływ na czas adopcji. Zwierzęta, których profil wizualizuje małe zdjęcie, mają mniejsze szanse na szybką adopcję.

Co takie zdjęcie powinno przedstawiać? Przede wszystkim powinno podkreślać najbardziej pożądane cechy zwierzęcia, o których napiszę poniżej. Dodatkowo pozytywny wpływ na przyspieszenie adopcji będzie miało, jeśli zwierzę na zdjęciu wygląda podobnie do najbardziej pożądanych ras.

Natomiast osoba oddająca zwierzę do adopcji powinna unikać zdjęć, które źle eksponują zwierzę. Zdjęcia te zostały sklasyfikowane przez narzędzie z zakresu Computer Vision (wykorzystany przez

[PetFinder.my](https://petfinder.my) do przetwarzania zdjęć na dane ilościowe) odpowiednio jako *bezpieczny pies* oraz *czarny kot*, a więc nawet algorytm AI służący do rozpoznawania obiektów na zdjęciach nie był w stanie ich sklasyfikować precyzyjnie. Przykłady takich zdjęć znajdziecie poniżej.



Z rodziną najlepiej wychodzi się na zdjęciach.... Chyba że chcesz być szybko zaadoptowanym zwierzęciem

W serwisie [PetFinder.my](https://petfinder.my) jeden profil może dotyczyć więcej niż jednego zwierzęcia. Jak pokazują dane, tworzenie profilu dla więcej niż dwóch zwierząt wpływa niekorzystnie na czas adopcji – jedno lub dwa zwierzęta zwykle są adoptowane szybciej, więc z dużą rodziną na zdjęciu nie najlepiej wychodzi się w kwestii czasu adopcji.

W ramach rzeczy, których tak jak rodziny się nie wybiera, wpływ na czas adopcji mają również:

- płeć – męskie osobniki mają o około 30% wyższą szansę na szybką adopcję,
- rozmiar (teoretyczny rozmiar dojrzałego osobnika) – duże osobniki częściej są adoptowane szybko,
- długość sierści – zwierzęta o długiej sierści są szybciej adoptowane,
- stan zdrowia – zwierzęta, które miały kontuzję, niezależnie czy bardziej, czy mniej poważną, mają mniejsze szanse na adopcję,
- to czy zwierzę jest psem czy kotem, oraz jaką ma rasę i kolor, ale te czynniki szczegółowo omówimy w dalszej części artykułu.

Pies czy kot?

Często uczestniczę w dyskusjach pomiędzy zagorzałymi zwolennikami psów, wychwalającymi ich przywiązanie, oraz właścicielami kotów, próbującymi przekonać tych pierwszych o wyższości indywidualności kocich charakterów. Dane z Malezji pokazują, że pies ma o 33% większe szanse na szybką adopcję niż kot, więc to psy są na uprzywilejowanej pozycji w tym kraju.

Rasa ma również ogromne znaczenie. Wśród kotów najszybciej adoptowane mogą zostać persy i koty syjamskie. Wśród psów największymi szczęściarzami są pudle, golden retrievery i shih tzu. Niestety, nierasowe psy i koty, które stanowią największą grupę wśród wszystkich zwierząt poddawanych pod adopcję w Malezji, mają znacząco niższe szanse na szybką adopcję niż zwierzęta rasowe.

	Szanse na adopcję w 30 dni (vs. pozostałe rasy)
Breed1_Poodle	44.33%
Breed1_Golden Retriever	35.33%
Breed1_Persian	32.21%
Breed1_Siamese	30.55%
Breed1_Shih Tzu	21.73%
Breed1_Domestic Short Hair	-20.58%
Breed1_Labrador Retriever	-23.41%
Breed1_other	-24.31%
Breed1_Terrier	-54.79%
Breed1_Mixed Breed	-63.62%

Co więcej, kolor sierści również ma wpływ na czas adopcji. Preferowane są pod tym względem jasne zwierzęta (kremowe, białe i złote), choć żółty kolor sierści stanowi wyłom w tej regule – żółtej maści zwierzęta mają mniejsze szanse na szybką adopcję.

Walk the walk, talk the talk

Założmy, że zwierzak zrobił wszystko, co mógł pozując do zdjęcia, by pokazać swoje walory. Tutaj pałeczkę zwiększania szans na szybką adopcję powinien przejąć autor profilu. Opis zwierzęcia jest nie bez znaczenia i odpowiednia treść może sprzyjać przyspieszeniu adopcji, dlatego poddałem ją analizie tekstu. Chciałem zidentyfikować kluczowe słowa sprzyjające przyspieszeniu adopcji oraz sprawdzić, czy długość opisu ma istotne znaczenie.

Poza tym, że profil zwierzęcia posiada podsumowanie podstawowych informacji o nim, takich jak wiek czy stan zdrowia, dłuższe opisy, rozwijające te informacje służą zwiększeniu szans na szybką adopcję (każde dodatkowe zdanie w opisie zwiększa szanse o niemal 2%). Jeśli zwierzę jest szczeniakiem lub młodym kotkiem, warto to podkreślić. Podobnie jak jego dobry stan zdrowia.

Pozytywny wpływ na przyspieszenie adopcji, wydaje się mieć również podanie powodu oddania zwierzęcia do adopcji. W szczegółowych opisach zwierząt, które szybko zostały zaadoptowane, pojawiają się powody osobiste ze strony poprzedniego właściciela, czy historie jak zwierzę zostało znalezione.

Time is Money, Money is Money

Jeden z wyników analizy, który smuci mnie niezmiernie, jest to, że czas, a dokładniej wiek zwierzęcia, ma ogromny wpływ na szansę na szybką adopcję. Każde 18 miesięcy dodane do wieku zwierzęcia zmniejsza szansę na szybką adopcję aż o 27%.

Osoba oddająca zwierzę do adopcji, może uczynić jeszcze jedno, by ją przyspieszyć – nie oczekiwać zapłaty. Profile, gdzie za zwierzę oczekiwana jest opłata, mają o około 25% mniejsze szanse niż te, gdzie opłata nie jest oczekiwana.

Czy Malezyjczycy mają podobne preferencje jak Polacy?

W czasie analizy często zapominałem, że mam do czynienia z danymi z portalu adopcyjnego zwierząt w Malezji ponieważ mam poczucie, że w Polsce wiele preferencji w procesie adopcyjnym zwierzęcia byłoby podobnych. Z mojego doświadczenia wiek, rasa zwierząt czy ich stan zdrowia wpływają na decyzję o adopcji.

Wydaje mi się również, że przy adopcji kotów w Polsce preferowane są koty, które zostały wysterylizowane. Natomiast z danych dla Malezji wynika, że tam fakt wysterylizowania zwierzęcia zmniejsza szanse na jego szybką adopcję i fakt ten dotyczy zarówno psów, jak i kotów. Być może w Malezji zdolność do reprodukcji stanowi zaletę. Niestety nie jestem w stanie na podstawie dostępnych danych zweryfikować z czego to wynika. Nie jestem też w stanie ocenić, które wnioski bazujące na danych z Malezji miałyby również zastosowanie na gruncie polskim. Pozostawiam to więc Waszej subiektywnej ocenie.

Podsumowanie

Przeprowadzona przeze mnie analiza, moim zdaniem pokazuje szereg pożytecznych wniosków przy tworzeniu profili adopcyjnych zwierząt. O ile bazuje na danych z Malezji, jestem przekonany, że takie rady jak umieszczanie zdjęć dobrej jakości, dobrze eksponujących zwierzę czy stworzenie szczegółowego opisu, który potencjalnemu nowemu właścicielowi przybliży jego nowego kompana, są użyteczne również w Polsce.

Oczywiście rady te mogą jedynie pomóc przyspieszyć adopcję zwierzęcia, ale nie dają gwarancji sukcesu. Wszakże mówimy o sytuacji, gdzie jedna osobowość ma się związać z drugą i niemierzalna danymi „chemia” pomiędzy nimi musi zrobić swoje.

Źródła:

- Dane, na których bazuje analiza: <https://www.kaggle.com/c/petfinder-adoption-prediction>
- Kod użyty do analizy wraz z wynikami możecie zobaczyć na moim koncie w GitHubie: https://github.com/dawidjwisniewski/pet_adoption/blob/main/article.ipynb