



MAGICZNA 15,

CZYLI LISTA 15 PYTAŃ, KTÓRE WARTO
ZADAĆ PRZED ROZPOCZĘCIEM
PROJEKTU ZWIĄZANEGO Z UCZENIEM
MASZYNOWYM, ABY WYNIĘŚĆ GO NA
WYŻSZY POZIOM



MiroslawMamczur.pl

WSTĘP

Jeśli miałbym za każdym razem obciąć sobie palce, gdy zapomniałem o istotnej informacji podczas rozpoczynania projektów związanych z uczeniem maszynowym, to niestety nie miałbym czym pisać na klawiaturze. Większość błędów była niepotrzebna i można było ich uniknąć. Z czasem doszedłem do wniosku, że mogę podpisać się pod stwierdzeniem „*godzina planowania pozwala uniknąć dni programowania*”.



Gdyby od samego początku zadać sobie odpowiednie pytania łącząc sposób myślenia eksperta, menadżera lub prezesa, można byłoby wynieść projekt na wyższy poziom!

Postanowiłem podzielić się z Tobą listą 15 pytań, które warto sobie zadać. Jestem przekonany, że odpowiadając na te pytania przed rozpoczęciem projektu, osiągniesz znacznie lepsze wyniki!



A ZROZUM CEL PROJEKTU

1 CZY ROZUMIEM DOKŁADNIE CO JEST DO ZROBIENIA?

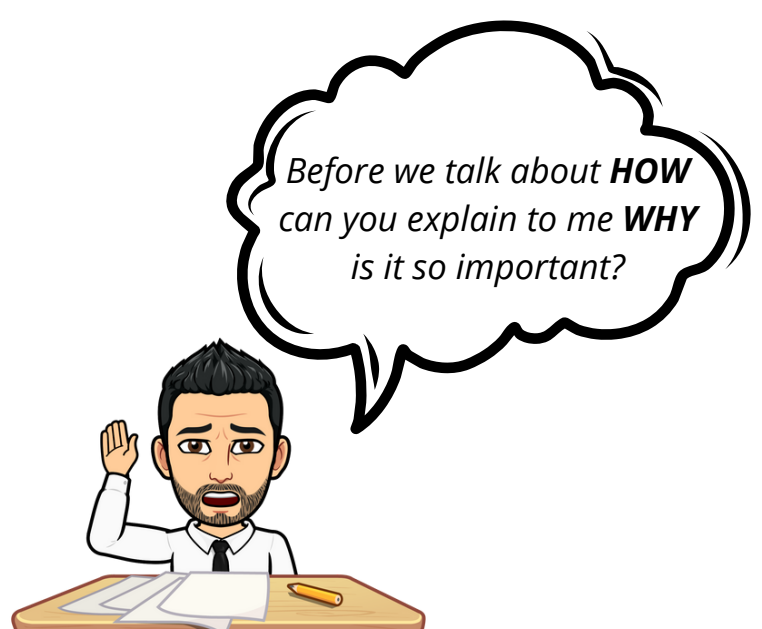
Jeśli chcesz dać z siebie 110%, najpierw musisz dokładnie zrozumieć, co masz zrobić. W tym miejscu warto doprecyzować, co dokładnie jest do zrobienia i jakie są oczekiwania. Polecam podczas rozmowy z osobą, która zleca nam dane zadanie używać **parafrazy**, czyli powtórzenia jak zrozumieliśmy zadanie naszymi własnymi słowami. Warto poprosić drugą stronę, by potwierdziła, czy dobrze zrozumieliśmy cel.

Super, gdybyś na tym etapie umiał swoimi słowami powiedzieć nowej osobie w projekcie, co musicie zrobić.

2 DLACZEGO TO JEST WAŻNE?

Zrozumienie dlaczego („WHY”) jest bardzo ważne. Pokazuje tak naprawdę dlaczego chcemy rozpocząć ten projekt i jakich rezultatów się spodziewamy. Wiele razy okazało się, że po zadaniu tego pytania można było zaproponować inne alternatywne podejście (np. zmiana celu z pozyskiwania więcej klientów na zwiększenie liczby zamawianych produktów przez tę samą ilość klientów, ale modyfikując sposób komunikacji z klientami).

Dodatkowo zadając pytanie „dlaczego to jest ważniejsze od innych rzeczy w tym momencie?” często wchodzimy w ciekawą dyskusję i możliwość zrozumienia strategii. Nie bój się kwestionować zdania odbiorców czy Project Managerów, jeśli masz wątpliwości (oczywiście o ile nie jest to dla ciebie kosztowne emocjonalnie).



*Before we talk about **HOW**
can you explain to me **WHY**
is it so important?*



B DEFINICJA PROBLEMU

3 CO KONKRETNIE CHCESZ PRZEWIDZIEĆ?

W przypadku projektów związanych z uczeniem maszynowym bardzo ważne jest konkretne zdefiniowanie problemu. Musisz więc **być tak precyzyjny, jak to tylko możliwe**. Samo powiedzenie, że potrzebny jest model na „przewidzenie odejścia klienta” to za mało. Należy doprecyzować, czym jest odejście klienta. Czy przypadkiem nie chodzi tak naprawdę o utratę aktywności (pознаłeś „dlaczego”, więc wiesz czego chce biznes)? W jakim horyzoncie czasowym?

Zadawaj pytania i doprecyzuj tak długo, jak będziesz mieć pewność, że wszystko jest jasne. Pamiętaj o doprecyzowaniu, w którym momencie model będzie przeliczany na produkcji, gdyż będzie to wpływać na odpowiedni dobór danych, które będziesz wybierać.

4 JAKIE MASZ DANE WEJŚCIOWE DLA ALGORYTMÓW?

Jedynym sposobem, w jaki model może przewidzieć wyniki, jest czerpanie ich z danych, którymi karmisz model. Bardzo ważne jest powiedzenie „*trash in trash out*”. Tak więc, jeśli chcesz, aby model dobrze wykonywał swoje zadanie, musisz mieć dobrej jakości dane. Dodatkowo warto pomyśleć o wielu niestandardowych źródłach danych.

Pamiętaj, że w dzisiejszym świecie nawet informacja z logów aplikacji jest bardzo cenna. To jak użytkownik używa aplikacji, w jakim tempie klika czy nawet trzyma telefon może w naszym zadaniu okazać się istotne. Bardzo polecam myślenie „**Out of the Box**”. Co przez to rozumiem? Testowanie metod uczenia maszynowego, które poznasz w inny sposób niż standardowy. Dla przykładu:

- mając kilka danych w formie szeregów czasowych (np. saldo karty kredytowej, dni przeterminowania, % wykorzystania karty) potraktować je jak zdjęcie i wykorzystać algorytmy sieci konwulcyjnych,
- mając mnóstwo danych z dodatkowego źródła, możesz przepuścić je przez sieć i do standardowego modelowania włożyć charakterystykę jako prawdopodobieństwo z tego modelu,
- przeglądając pytania z ankiety (np. o zarobkach) możesz wrzucić odpowiedzi oraz pytania i zbudować prosty model. Dzięki temu z danych zobaczysz, które pytanie i odpowiedź wpływają na najwyższe zarobki.



5 ILE MAMY DANYCH?

Wiedząc, ile masz dostępnych danych, będziesz mógł lepiej dobrać algorytm uczenia maszynowego. Dzięki tej wiedzy od razu będzie wiadomo jakim wachlarzem algorytmów możesz dysponować. Dla małej liczby obserwacji (np. mniej niż 100) najprawdopodobniej najlepszym rozwiązaniem będzie zaproponowanie modelu eksperckiego z domenowymi ekspertami. Oczywiście im więcej danych tym lepiej.

A co jeśli w tym momencie nawet nie wiadomo, ile jest danych (np. rząd wielkości)? Wówczas, zanim przystąpisz do projektu, dokładniej rozpoznaj temat lub dopytaj biznes czy na pewno im się to opłaci, skoro nie zna skali.



6 JAKIE SĄ NAJISTOTNIEJSZE CZYNNIKI DO MODELU?

Tak naprawdę algorytm nie rozumie naszego świata. Algorytmy uczenia maszynowego to po prostu metody, które na podstawie danych przez nas dostarczonych próbują znaleźć reguły rozwiązujące dany problem. Zatem bardzo ważne jest, abyś dostał kilka cennych wskazówek na temat zagadnienia, które rozwiązujesz:

- jakie dane według eksperta domenowego są istotne,
- na podstawie czego sam podjąłby decyzję, aby przewidzieć zjawisko,
- na co warto uważać,
- o czym nie warto zapominać,
- kogo należy wykluczyć biznesowo z próbki (np. dane historyczne dotyczyły sprzedaży dla mężczyzn i kobiet, a obecnie w sprzedaży są tylko sukienki).



C POMIAR EFEKTYWNOŚCI

7 CO BĘDZIE TWOJĄ METRYKĄ SUKCESU?

Jest to jedno z ważniejszych pytań. Warto wprost zdefiniować, co jest naszą metryką sukcesu. Najważniejsze, by była ona zgodna z tym, jak model będzie używany. Jeśli dla przykładu będzie to model służący do podjęcia decyzji czy pożyczysz komuś własne pieniądze, to pewnie nie będziesz patrzył na standardowe metryki jak Accuracy, AUC czy F1. Ważniejszą informacją będzie ile zarobisz lub ewentualnie stracisz pieniędzy.

Dlatego nie patrz jedynie na podstawowe metryki, tylko popatrz jak model będzie zarabiał na produkcji.



8 CZY SĄ BENCHMARKI

Bardzo ciężko w dzisiejszym świecie znaleźć problem, którego nikt jeszcze nie próbował rozwiązać lub po prostu nie rozwiązał przed Tobą. Warto poszukać inspiracji, czy ktoś już danym zagadnieniem się zajmował.

Najpierw wewnątrz organizacji, dla której robisz projekt, a następnie poza nią. Jeśli budujesz kolejną wersję już istniejącego modelu, to masz punkt odniesienia. Pomaga także duża sieć kontaktów. Możesz poprosić znajomych o informacje jak to wygląda u nich (jeśli oczywiście mogą je podać).

9 JAKIEGO MINIMALNEGO WYNIKU SIĘ SPODZIEWASZ?

Naprawdę cenną informacją jest świadomość, jakiego minimalnego wyniku się spodziewasz. Warto pamiętać o zasadzie Pareto: 20% pracy daje 80% wyniku.

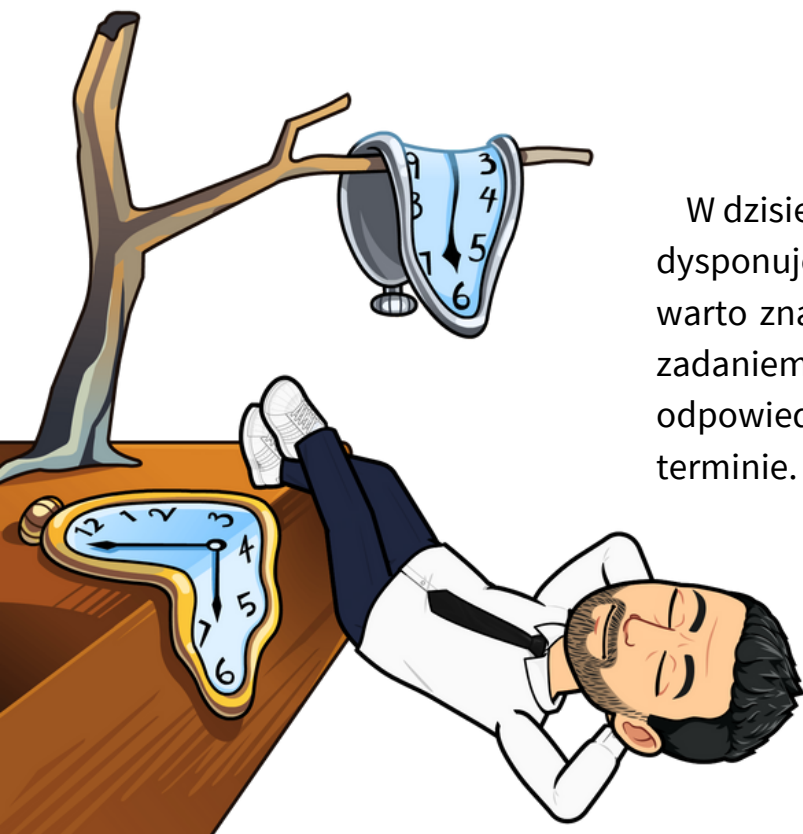
Jeśli spełnimy minimalny poziom, czasami warto się zatrzymać i rozpocząć przygodę z kolejnym projektem niż dopieszczać model do granic możliwości.

Czasami walka o odrobinę lepszy model może trwać bardzo długo nie dając aż tak dużego wyniku, jaki już osiągnęliśmy mniejszym nakładem pracy.



D CZAS

10 ILE MASZ CZASU NA DANY PROJEKT?



W dzisiejszym świecie najcenniejszą rzeczą, jaką dysponujemy, jest czas. Od samego początku warto znać wszystkie terminy związane z danym zadaniem. Dzięki temu będzie można dobrać odpowiednie metody i zakończyć projekt w terminie.

11 KIEDY SPODZIEWASZ SIĘ PIERWSZYCH WYNIKÓW?

Mając projekt rozpisany na dwa miesiące możesz zaplanować, ile każdy etap powinien zająć czasu. Oczywiście wraz z doświadczeniem będziesz robić to coraz lepiej. Zawsze warto mieć bufor na nieprzewidziane zdarzenia. Jeszcze żaden projekt nie przeszedł idealnie i zawsze pojawia się jakaś niespodzianka :)

Jeśli od początku będziesz wiedzieć, kiedy spodziewasz się pierwszych wyników, lepiej zapanujesz nad całym projektem. Dla przykładu przygotowując model i posiadając 5 różnych źródeł danych, w pierwszym kroku możesz zbudować prosty model tylko na jednym źródle. Dzięki temu może się okazać, że już od początku będziesz spełniać minimum, które było zaplanowane i znacznie szybciej zakończysz projekt. A co jeśli wynik okaże się gorszy od oczekiwanego? Wówczas na spokojnie sprawdź, czy nie popełniłeś gdzieś błędu.



E WSPÓŁPRACA

12 CZY WIEM KTO BĘDZIE MÓGŁ MI POMÓC?

Warto zastanowić się, czy jak ruszysz z projektem i będziesz mieć wątpliwości, to będziesz wiedział, z kim możesz się skontaktować:

- kto pomoże mi uzyskać dostęp do danych,
- kto może pomóc zrozumieć dane,
- kto prowadzi projekt,
- z kim można konsultować projekt w zakresie wdrożenia na produkcję (kilka razy to uratowało mi życie – dowiedziałem się, że niektóre rozwiązania są niemożliwe. Dzięki temu przygotowałem odpowiednią alternatywę, którą szybko można było wdrożyć na produkcji),
- **SAM** dla siebie też możesz być wsparciem wykorzystując kody z wcześniejszych swoich projektów.



13 JAK BĘDZIE WYGLĄDAĆ WSPÓŁPRACA Z ODBIORCAMI?

Najgorszą rzeczą, jaką można zrobić, to wziąć na barki przygotowanie projektu, a następnie pokazać odbiorcy ostateczną wersję. Ważne byś od samego początku ustalił, co jaki czas będziesz pokazywał wyniki i omawiał je z odbiorcami.

Dzięki temu już na wczesnych etapach będzie można dokonać korekt lub poprawek. Jeśli przygotujesz coś nie do końca poprawnie (choć dzięki wcześniejszym pytaniom mam nadzieję, że do tego nie dojdzie!) szybko będziesz mógł dokonać zmian. Dyskusja z odbiorcami w trakcie trwania projektu potrafi znacznie przybliżyć Cię do osiągnięcia sukcesu.



F TAK, LET'S GO! ALE CZY NA 100%?

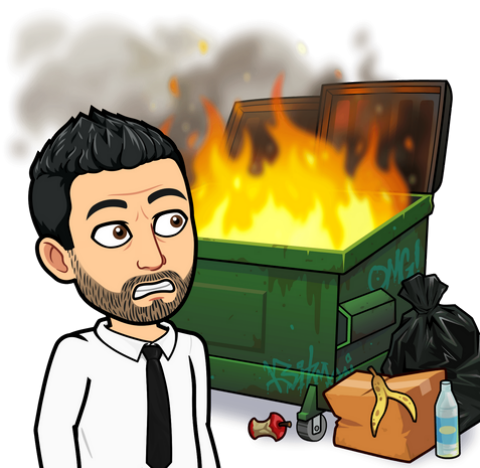
14 CZY MOŻNA ZREZYGNOWAĆ NA RZECZ "QUICK WIN"?

Czasami, zamiast budować model do rozwiązania jakiegoś zagadnienia, warto się zastanowić, czy nie ma prostszej ścieżki. Jeżeli zaczniesz myśleć w ten sposób, to będziesz jednym z najcenniejszych pracowników.



Jeśli wchodzisz w nowy projekt, to spróbuj spojrzeć na niego świeżym okiem. Czasami nie trzeba budować modelu, który będzie przewidywał „*kiedy będzie awaria windy?*”, jeśli zauważysz, że nigdy nie były przeprowadzane przeglądy i prace serwisowe. Może warto wówczas najpierw zacząć od tych kroków, niż przewidywać od razu awarie :)

15 DLACZEGO PROJEKT SIĘ NIE UDA?



To moje ulubione pytanie. Próbuąc wcielić się w rolę krytyka próbujemy znaleźć wszystkie powody, dlaczego może nam się nie udać. Dzięki temu będziesz mógł przygotować się na różne ewentualności, które przyjdą Ci do głowy.

Zawsze można zadać to pytanie w całym zespole projektowym – będzie ciekawa zabawa :). A jak już udzielisz odpowiedzi na te pytania, to ruszaj z projektem! Powodzenia!



PODSUMOWANIE

Odpowiedz sobie na wszystkie powyższe pytania, a projekt, w którym bierzesz udział, będzie miał znacznie większe szanse na sukces!

Dodatkowo zadając niektóre pytania od razu w oczach innych zostaniesz uznany za eksperta w swojej dziedzinie. Z wyprzedzeniem będziesz mógł przewidzieć, co może się nie udać. Ponadto pokażesz swoim współpracownikom swoje podejście do projektów Data Science i monetyzacji danych z perspektywy biznesowej. Bo najważniejszy jest klient, a nie metody, jakich użyjesz.

**Powodzenia w przyszłych projektach!
I zacznij korzystać z tych pytań już dziś!**

Pozdrawiam serdecznie z całego serducha,

Miroslaw Mamcew

