

Jak prognozować? – praktyczne porady

O NAS

Od ponad 15 lat wspieramy największe przedsiębiorstwa w uzyskaniu i utrzymaniu przewagi konkurencyjnej poprzez wykorzystanie dedykowanych systemów informatycznych. Projektujemy i uruchamiamy platformy informatyczne pozwalające gromadzić, przetwarzać i analizować duże zbiory danych. Specjalizujemy się w dostarczaniu rozwiązań dla klientów z branży finansowej i retail.

Dzięki wiedzy, doświadczeniu i pasji ponad 95-osobowego zespołu realizujemy zaawansowane technologicznie projekty, wydobywając z dużych zbiorów danych wartość biznesową. Nasz sukces opieramy na wsłuchiwaniu się w potrzeby klientów, podążaniu za mega trendami, ciągłym rozwoju i wykorzystaniu nowych technologii. Pozwala nam to utrzymać pozycję lidera w branży.

Pomagamy przedsiębiorstwom wprowadzać podejście
Data-Driven Business.



17

lat doświadczenia

95

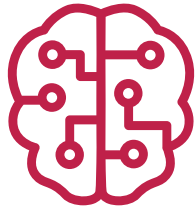
osobowy zespół

200

zrealizowanych projektów

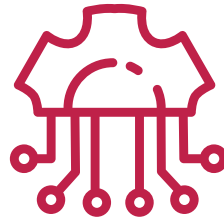
ROZWIĄZANIA

Projektujemy i uruchamiamy platformy informatyczne pozwalające gromadzić, przetwarzać i analizować duże zbiory danych.



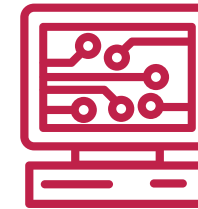
Artificial Intelligence

Zaawansowana analityka
biznesowa oparta
o **Machine Learning** i **Data Science**.



Data Management

Nadawanie danym wartości
biznesowej w oparciu
o **technologie Big Data**.



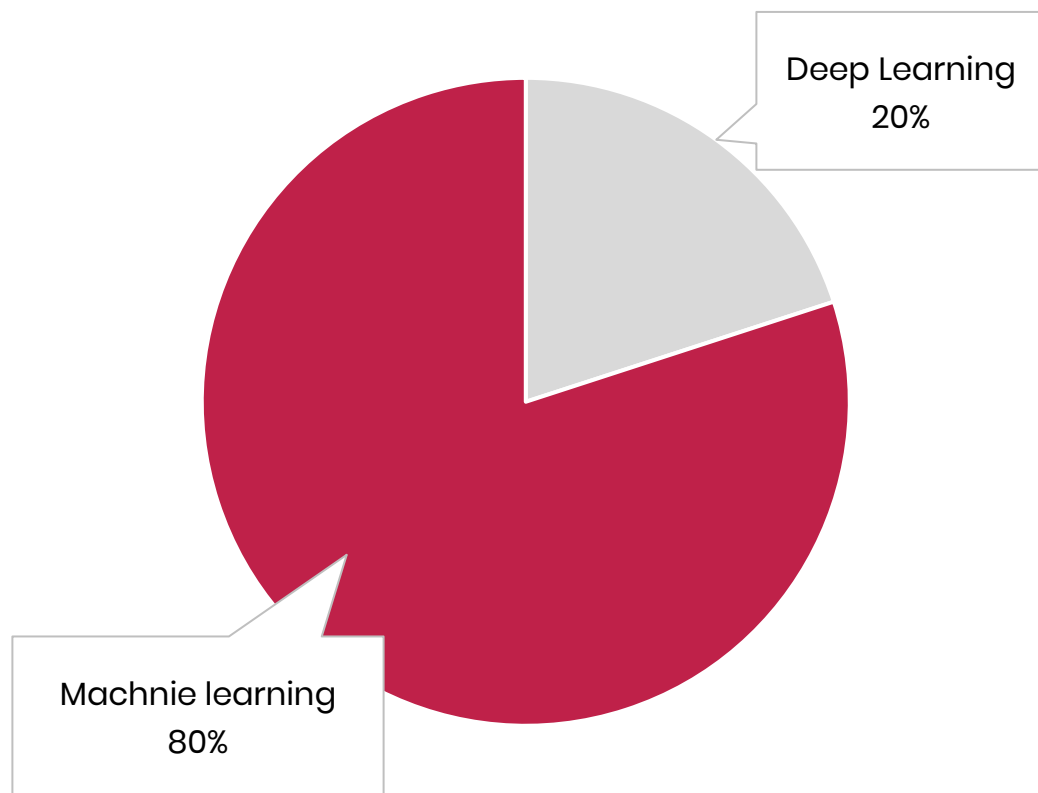
Business Automation

Rozwój dedykowanych
rozwiązań IT
na poziomie **enterprise**.

Data-Driven Business

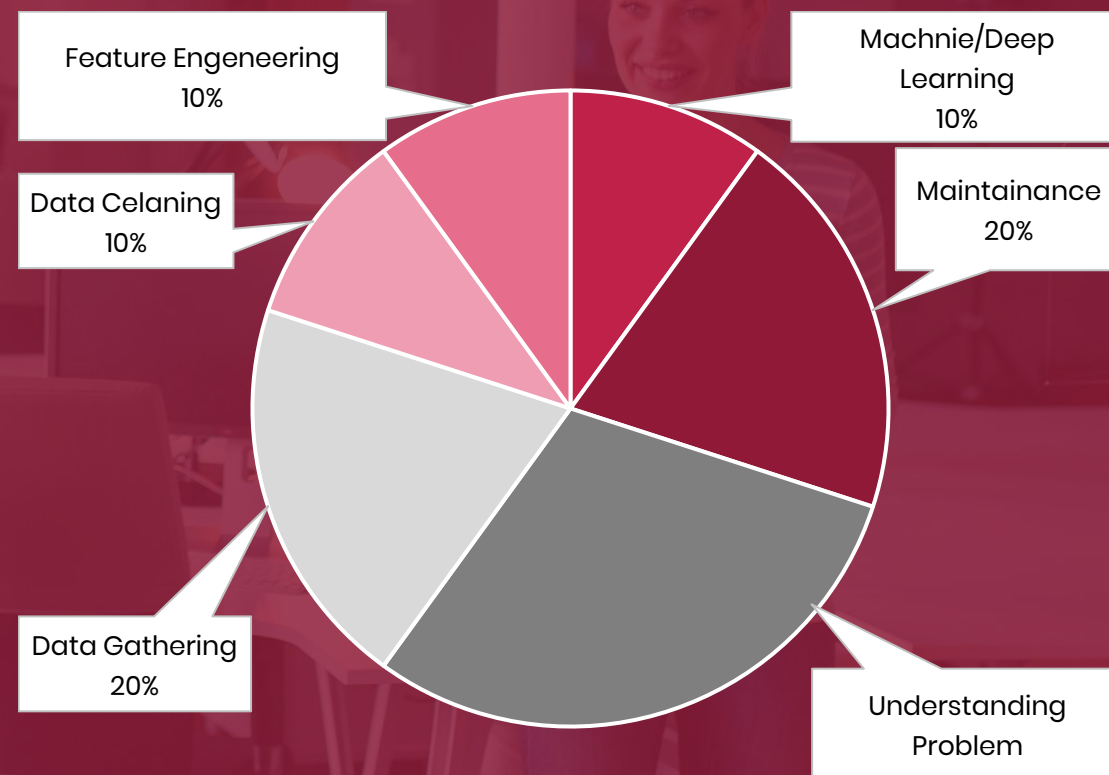
Data Science

Oczekiwania



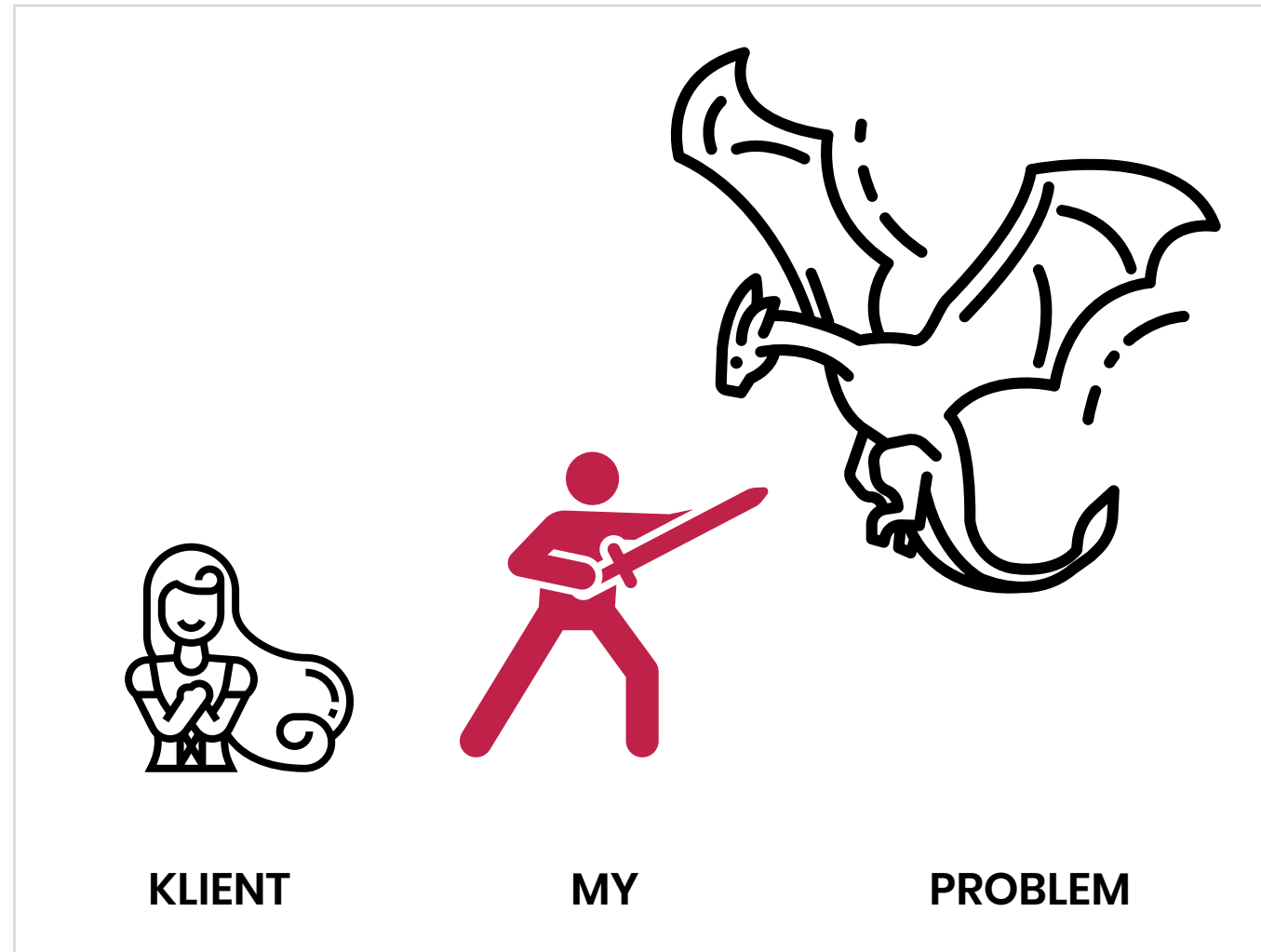
VS

Rzeczywistość



PoC – walka o klienta

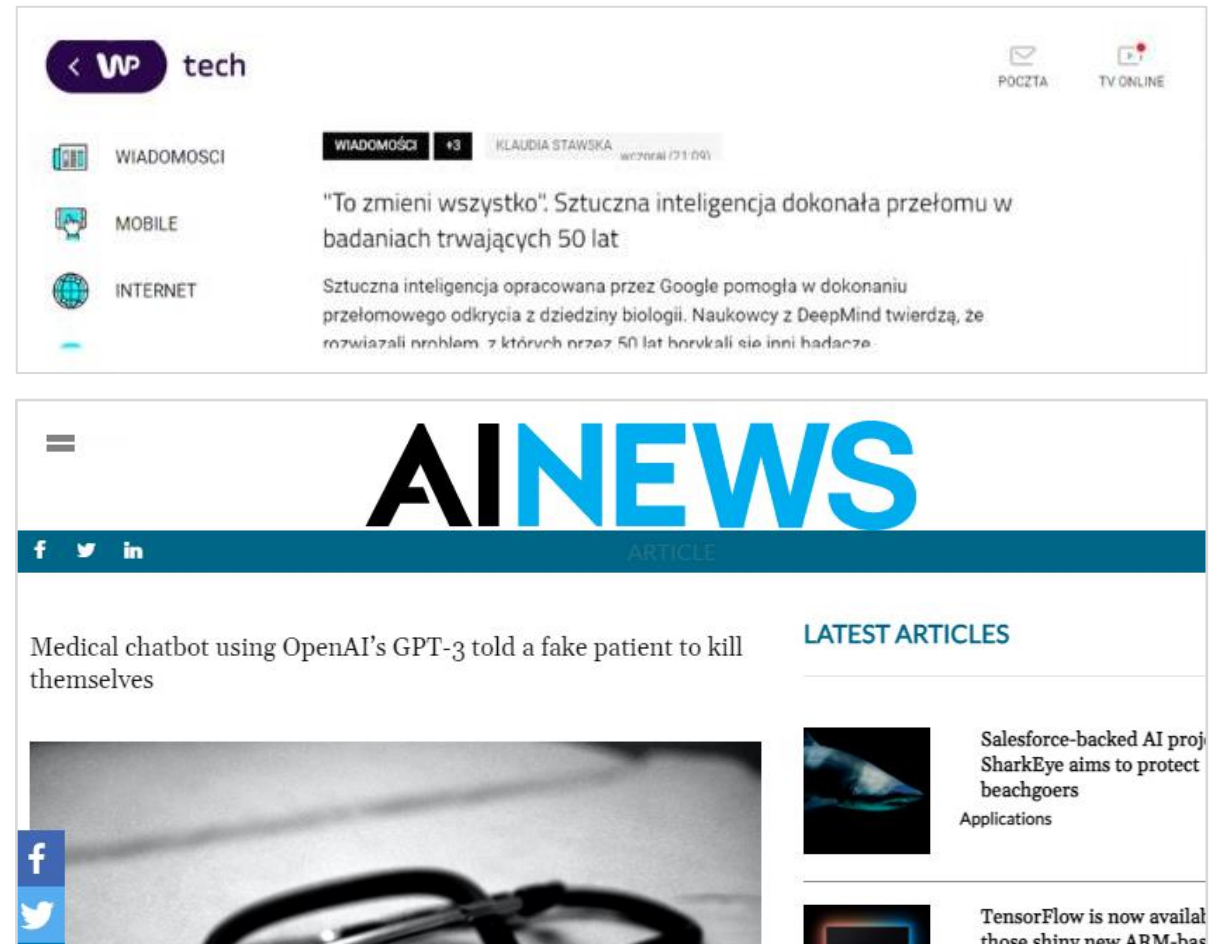
- Pewną „symulacją” projektów typu PoC są serwisy typu **Kaggle**
- Takie projekty kuszą możliwością skorzystania z metod, których wykorzystanie produkcyjne w praktyce:
 - albo powoduje **znaczący przyrost pracy** w celu utrzymania takiego rozwiązania
 - albo jest wręcz **niemożliwe** np.: ręczne uzupełnianie danych, dostęp do danych na określonych licencjach.



PoC czyli walka o zainteresowanie klienta
– jak najbardziej, byle pozostawać otwartym na komunikację z klientem!

Wysokie oczekiwania klientów

- Dziedziny takie jak **Big Data, Machine Learning, Data Science, Artificial Intelligence, Advanced Analytics**, itd. często w oczach osób decyzyjnych są „lekiem na całe zło”
- Do powszechnej świadomości dziedziny te przenikają dzięki **newsom** takim jak np. „Uczenie maszynowe sposobem na wynalezienie leku na raka”

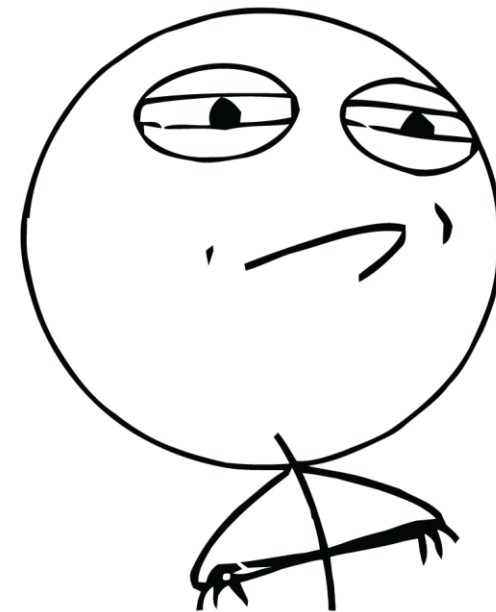


Tak powierzchowna wiedza powoduje, że klienci do projektów podchodzą z bardzo wysokimi oczekiwaniami – „sky is the limit!”

Techniczne wyzwania w procesie modelowania

- Odpowiedni dobór **atrybutów**
- Wybór modelu przy **strukturze kaskadowej** lub **rownoległej**
 - Walidacja wykonywana na treningu: walidacja kryzysowa czy walidacja szeregu czasowego
- Ocena **jakości** modelu
 - Metody statystyczne np: Maximum likelihood estimation (MLE), AIC, BIC, MAE, MAPE, rozkłady błędów
 - Metody empiryczne (**wypadkowa kilku metod**)
 - Ocena cząstkowa i całościowa
- **Trening modelu** a generowana prognoza
- **Częstotliwość** trenowania modelu
- Odpowiedni dobór **algorytmu** do danych
- **Hiperparametryzacja** algorytmu
- **Przedziały** w jakim prognozy powinny się znajdować

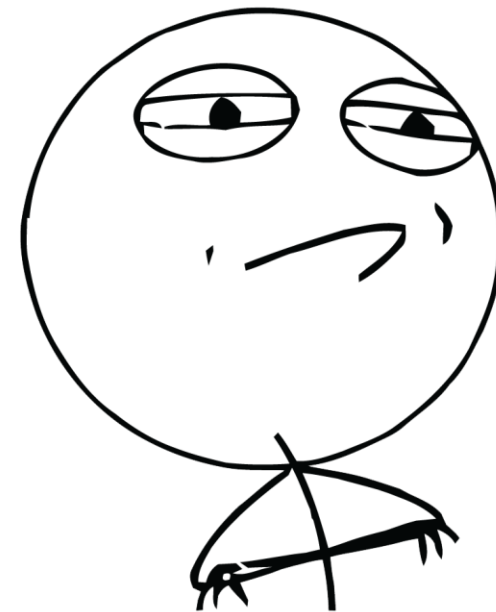
CHALLENGE ACCEPTED



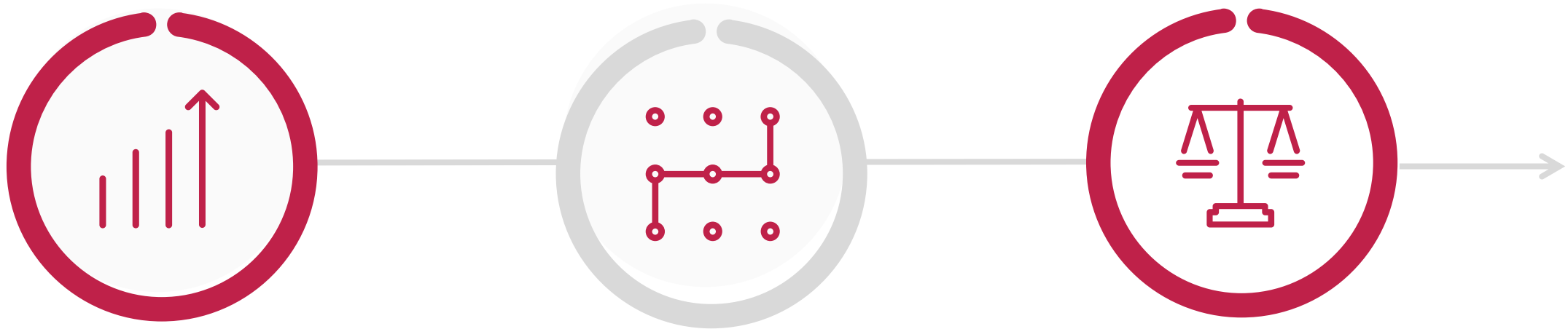
Biznesowe wyzwania w procesie modelowania

- Overstock, understock, write-off
(niedoszacowania i przeszacowania)
- Produkty nowe
- Produkty o krótkiej historii sprzedaży
- Produkty rzadko rotujące
- Sklepy sezonowe
- Akcje promocyjne
- Kanibalizacja i asocjacja produktów
- Ruchome święta, wakacje
- Horyzont prognozy
- Logistyka

CHALLENGE ACCEPTED



Implementacja rozwiązania



Optymalizacja procesu

Architektura workflow

Weryfikacja implemtacji

porównanie wdrożonego
rozwiązania z koncepcją
wypracowaną w Lab

Proces Modelowania w produkcji

1. Pre-processing

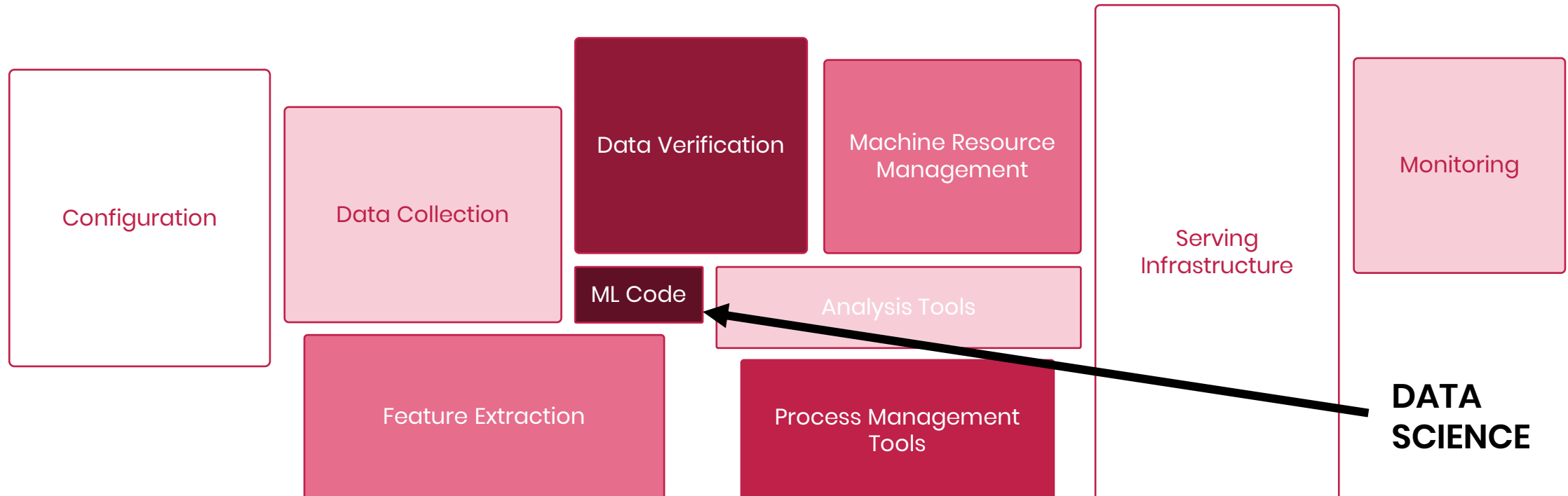
- Import danych
- Tworzenie zbiorów danych
- Przygotowanie danych (agregacja)
- Walidacja danych (NA, NULL, itp.)
- Outliers
- Przeliczanie atrybutów

2. Processing

- Modelowanie

3. Post-processing

- Weryfikacja finalnych prognoz (nakładanie współczynników)
- Dostarczenie rezultatów do klienta



Zasada KISS

Kee**p** **I**t **S**imple, **S**tupid

Kee**p** **I**t **S**uper **S**imple

Inżynieria cech › Algorytmy

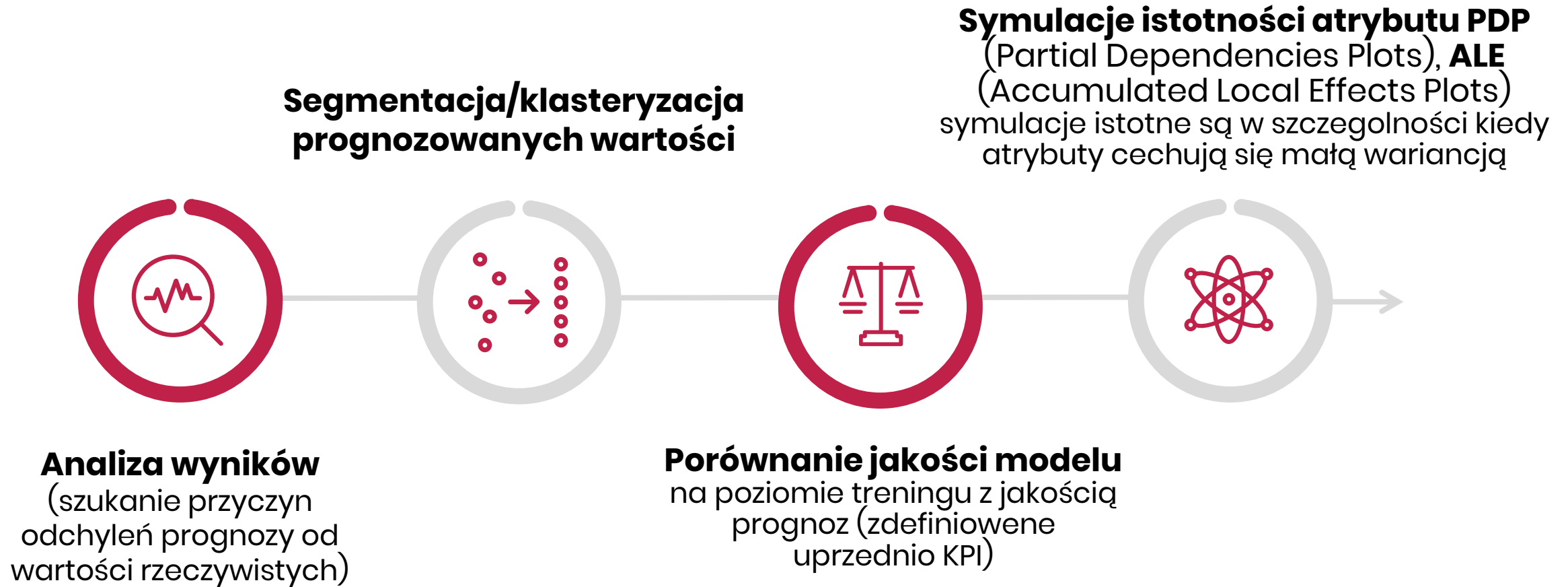
Proste interpretowalne atrybuty › Złożone atrybuty

Raportowanie i dokumentacja

Raport	Dokumentacja
<ul style="list-style-type: none">▪ Często najważniejsza składowa projektu▪ Końcowy raport to przysłowiowa “kropka nad i”▪ Często decyduje o powodzeniu całego przedsięwzięcia▪ Raport powinien weryfikować zasadność hipotezy▪ Raport powinien być:<ul style="list-style-type: none">▪ Logiczny▪ Spójny▪ Łatwy w odbiorze, to kategoria odbiorcy decyduje do kogo raport ma być skierowany (zarząd, business, dział techniczny)	<ul style="list-style-type: none">▪ Z reguły powstaje po akceptacji merytorycznej projektu▪ Update dokumentacji jest niezbędny▪ Dokumentacja powinna być<ul style="list-style-type: none">▪ Spójna▪ Przejrzysta▪ Kompletna

Nawet najlepsze rozwiązanie, jeśli nie jest wydajne, a co za tym idzie oszczędne w implementacji, nie ma szans na sukces.

Monitoring



Analiza biznesowa w projektach Data Science

- Nierzadko sam klient **nie zdaje sobie sprawy z poziomu skomplikowania zagadnienia** – szczególnie w przypadku, w którym realizowany projekt jest pierwszą okazją do zetknięcia się z tym zagadnieniem
- Odpowiednia komunikacja z klientem pozwala na:
 - Uświadomienie klientowi **komplikacji** bądź **ograniczeń** na drodze do sukcesu (urealnienie oczekiwań)
 - Zebranie **spostrzeżeń i doświadczeń klienta** kiedy postawiony problem jest mu znany z autopsji, a jego wiedza domenowa może mieć ogromne znaczenie i znaczący wkład w realizacji projektu z sukcesem

*Nie bój się zadawać pytań
jeżeli czegoś nie rozumiesz.*

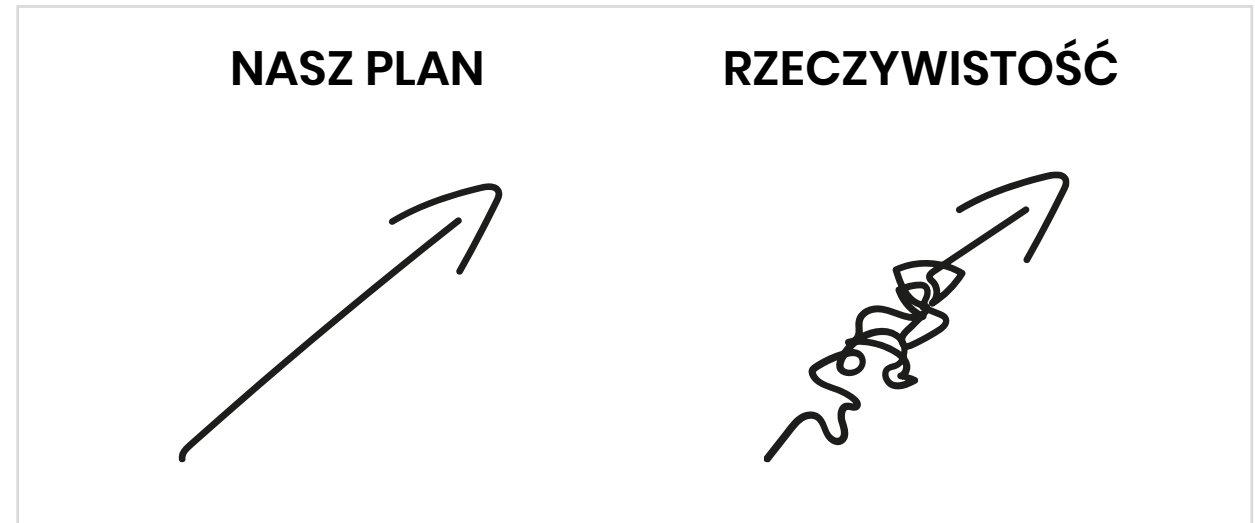
*Nie bój się zadawać pytań
wtedy kiedy myślisz, że
wszystko już rozumiesz!*

Kluczem do sukcesu jest prawidłowe zrozumienie problemu.
Nawet najlepsza odpowiedź na nieprawidłowo sformułowane pytanie jest błędna!

Kompleksowość i spójność



W pierwszej kolejności
„sięgaj po nisko wiszące owoce”



**Unikaj wyjątków i sytuacji szczególnych,
ale nie zapominaj o nich!**

To często one w długiej perspektywie
decydują o końcowym sukcesie.

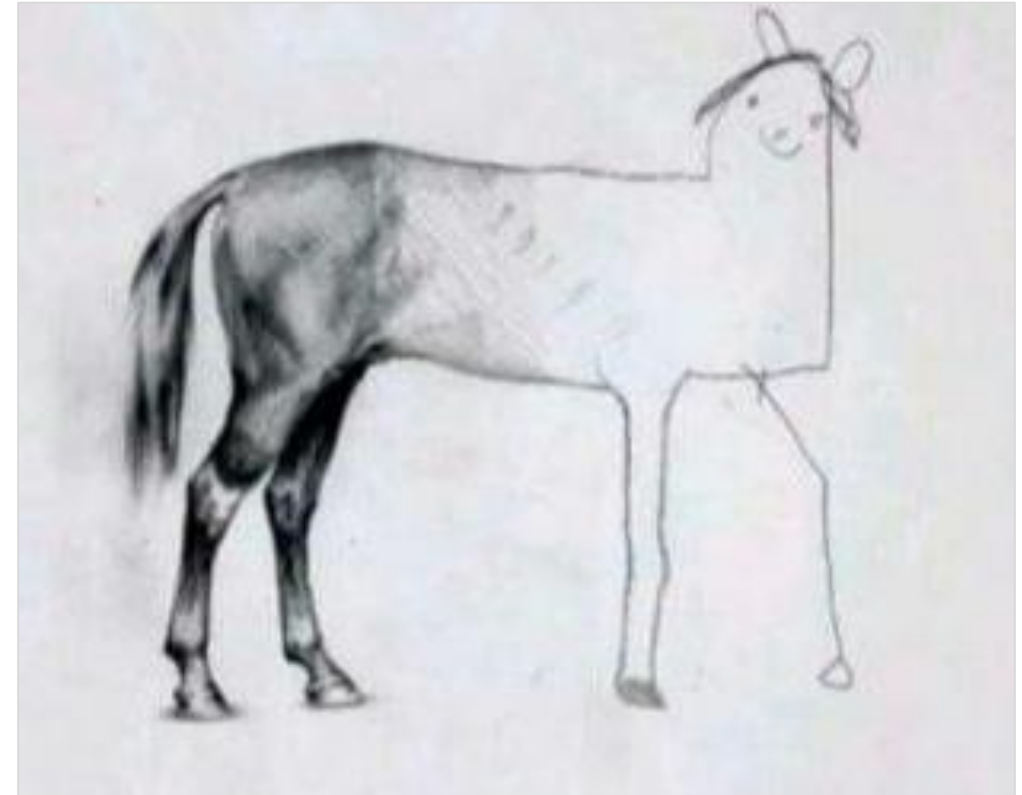
Wyraźnie sygnalizuj sytuacje,
w których realizacja pewnych
obszarów jest zagrożona bądź
wręcz niemożliwa!

Ustal kryteria sukcesu, które możliwie najszerszej opisują zakres realizacji.

Budżet projektu

- Rozwiązywanie problemów na potrzeby zaliczeń, czy wręcz dla „zabawy” na własną rękę, raczej nie jest powiązana z ograniczeniami budżetowymi. Z takimi ograniczeniami zawsze mamy jednak do czynienia kiedy rozwiązujemy rzeczywiste problemy biznesowe klientów.
- W praktyce biznesowej natknąć można się na sytuację w której dostarczymy:
 - **Niekompletne rozwiązanie** – ten przypadek szybko do nas wróci wraz z niezadowolonym klientem
 - **Kompletne rozwiązanie, ale ukończone „na prędcie”** – ten przypadek pewnie do nas nie wróci aż tak szybko, ale pewnym jest, że po ujrzeniu światła dziennego klient nie będzie zadowolony
 - **Kompletne rozwiązanie ukończone na „koszt własny”** – klient będzie zadowolony co pewnie ciężko powiedzieć o PM

POCZĄTEK vs. KONIEC projektu



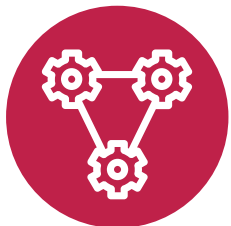
Zbytnie poświęcenie uwagi szczegółom na początku projektu może spowodować problemy z dostarczeniem ostatecznego rozwiązania zakładanej (wysokiej) jakości.

Budżet klienta (koszty klienta)



Na początkowym etapie współpracy

Kosztom klienta jest **zaangażowanie zespołu Data Science** w opracowanie rozwiązania na zadany problem.



Na dalszym etapie współpracy

Pojawiają się jednak inne rodzaje kosztów, które z czasem decydują o **rentowności** (z punktu widzenia klienta) całego projektu. Są to np. koszty:

- serwerów i infrastruktury
- chmury obliczeniowej
- wszelkich licencji
- odpłatnych dostępów i źródeł danych
- zespołu utrzymaniowego

Niekiedy realizowane projekty rozwiązują problemy klientów w przypadku, których szeroki zakres korzyści może mieć **charakter niemierzalny**, albo trudny do skwantyfikowania.

Współpraca i implementacja

- Na uwadze należy mieć fakt, że sukces projektów realizowanych w ramach szeroko rozumianej dziedziny Data Science zależy od wielu czynników. A najczęściej również od wielu osób zaangażowanych w różnym charakterze.
- Skuteczna współpraca, jasna i klarowna komunikacja są czynnikiem kluczowym.

*Sukces jest wspólny,
ale wspólna jest też
porażka.*



Należy zwracać uwagę na zgodność wyników uzyskanych w fazie rozwoju projektu z wynikami otrzymywanymi po ich implementacji produkcyjnej.

10m

<https://forms.gle/ga8BSg5GM3nVnPb39>



3Soft S.A.

Porcelanowa 23
40-246 Katowice, Polska

www.3soft.pl



www.facebook.com/3soft



www.linkedin.com/company/3soft



Jarek Tkocz

Member of the Board
Chief Data Science Officer

jarek.tkocz@3soft.pl



Adam Staszczyk

Head of Data Science

adam.staszczyk@3soft.pl

Dziękujemy za uwagę