

2. Data Model Canvas. Data Science Workflow Canvas.
Definiowanie problemu odkrywania wiedzy.

Przetwarzanie danych i odkrywanie wiedzy

Tomasz Kajdanowicz, Kamil Tagowski

Plan wykładu

1. Kanwa danych (Data Canvas)
2. Kanwa modelowania danych (Data Model Canvas)
3. Kanwa modelu biznesowego (Business Model Canvas)
4. Szablon modelu biznesowego otwartych danych (The Open Data Business Model Canvas)
5. Kanwa uczenia maszynowego (The Machine Learning Canvas)
6. Szablon projektu AI (The AI Project Canvas)
7. Kanwa danologiczna (The Data Science Workflow Canvas)

Dane to surowiec

- dane to surowiec więcej warty niż ropa
- ilość danych rośnie szybciej niż technologia ich przetwarzania
- pracujący z danymi, zwykle nie są zaangażowani w opracowywanie nowych modeli biznesowych
- rozwijający biznes zwykle nie mają kompleksowej wiedzy na temat całego spektrum dostępnych danych

The
Economist

Regulating the internet giants

The world's most valuable resource is no longer oil, but data



Licencja zdjęcia: davidparkins.com

Źródło

<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

Dane a biznes

- potrzeba odkrywania potencjału danych w nowych modelach biznesowych

Pytania:

- jak przeglądać dostępne dane do zastosowań w nowych modelach biznesowych?
- jaki potencjał mają dane?
- jakie pytania zadawać by zrozumieć istotność danych w realizacji pomysłu biznesowego?

Data Canvas

Kanwa danych (The data canvas)

- formujemy różnorodną grupę (np. przedstawicieli różnych działów) na warsztaty
- od 3 do 9 osób (samodzielne wypełnianie Data Canvas jest mniej wnikliwe)
- prowadzimy ożywioną, ale zorganizowaną dyskusję
- jeden z uczestników przyjmuje rolę moderatora
- zawieszamy szablon na ścianie, aby wszyscy uczestnicy mogli nad nim równie dobrze pracować
- wypełniamy szablon Data Canvas
- umieszczamy karteczki samoprzylepne w różnych kształtach na których zapisujemy markerem treść

Przygotowanie

- przed warsztatem podsumuj jakimi danymi dysponujesz
 - dane wewnętrzne
 - dane zewnętrzne
 - dane publiczne (bez żadnych kosztów, za opłatą)
- ilość danych jest ogromna, więc wcześniej ustal kryteria:
 - czy jest jakaś specjalna grupa klientów, do której chcesz dotrzeć?
 - czy dane muszą mieć specjalny format?
 - czy posiadasz budżet na dostęp do dodatkowych źródeł danych?

W trakcie

- użyj jednej karteczki na każde źródło danych (napisz nazwę źródła danych)
 - można opisać hashtagiem odniesienie tematyczne do źródła danych
- karteczki samoprzylepne w różnych kształtach i kolorach doskonale nadają się do dodawania dodatkowych informacji
 - kształt:
 - okrągłe karteczki samoprzylepne w przypadku danych nieustrukturyzowanych
 - prostokątne w przypadku danych ustrukturyzowanych itd.
 - kolor:
 - zielone w przypadku wiarygodnych źródeł danych, których jakość została sprawdzona przez wiarygodny organ,
 - żółte w przypadku mniej wiarygodnych źródeł danych
 - czerwone w przypadku źródeł danych o wątpliwej jakości

W trakcie

- wprowadź symbole lub oznaczenia, np.
 - € w przypadku płatnych źródeł
 - § dla źródeł danych z ograniczeniami prawnymi lub etycznymi
- wyjaśnij symbole i oznaczenia w legendzie
- umieść karteczki samoprzylepne w jednym z czterech pól w zależności od tego, kto jest właścicielem danych i jak często są one aktualizowane

Kanwa

Kanwa danych (The Data Canvas)

Wewnętrzne dane cykliczne

Dane należące do Twojej organizacji
i aktualizowane w określonych momentach



Jakie dane są już dostępne w Twojej organizacji?
Jakie dane możesz pozyskać za pomocą produktu / usługi wykorzystywanej przez Twoich klientów?
Jakie dane mogą aktywnie dostarczać Twoi klienci?

Wewnętrzne dane ciągłe

Dane należące do Twojej organizacji
i dostępne w ciągłym strumieniu



Zewnętrzne dane cykliczne

Dane będące własnością strony trzeciej
i aktualizowane w określonych momentach

Jakie dane mogą dostarczyć Twoi partnerzy?
Jakie partnerstwa możesz zawrzeć w celu uzyskania danych?
Z jakiego rodzaju otwartych danych można korzystać?
Jakie dane możesz uzyskać z internetu?
Jakie dane można kupić?

Zewnętrzne dane ciągłe

Dane będące własnością strony trzeciej
i dostępne w ciągłym strumieniu

Po wypełnieniu - oceń kanwę

- nie poświęcaj zbyt wiele czasu na wypełnianie pierwszej wersji

Pytania:

- w jaki sposób źródła danych są rozmieszczone w czterech kwadratach?
- gdzie są główne punkty zainteresowania i które pola zawierają niewiele danych?
- jakie dodatkowe źródła danych pozwoliłyby Ci wzmocnić Twoje mocne strony i wypełnić luki?
- w których przypadkach masz dużo danych lub szczególnie cenne dane?
- z jakich danych możesz korzystać przy minimalnej inwestycji?
- które dane są drogie w zbieraniu lub ulepszaniu?
- jakie źródła danych można łączyć?

Data Model Canvas

Kanwa modelowania danych (Data Model Canvas)

Narzędzie do planowania projektu opartego o dane:

- Perspektywa twórcy vs perspektywa odbiorcy
- Kompetencje i interesariusze
- Scenariusz wykorzystania danych
- Korzyści i koszty
- Prezentacja i dotarcie

Perspektywy

Odbiorca

Kogo problemy adresujemy?

W jakim cierpieniu przynosimy ulgę?

Kto będzie oceniał czy nam się udało?

Jaką wartość dostarcza projekt?

Jak odbiorca wejdzie w interakcję z wynikami modelu?

Twórca

Jakie zyski przynosi projekt twórcy?

Dlaczego się go podejmuje?

Jakie będą tego koszty?

Kompetencje i kontekst

Kompetencje

Jakich potrzebuję kompetencji, żeby zrealizować projekt?

Skąd mogę uzyskać kompetencje jeśli ich brakuje?

Kontekst

Czy ktoś się tym już zajmuje?

Czy w realizacji tego projektu powinienem z kimś współpracować?

Scenariusz wykorzystania danych

Dane

Jakich danych potrzebuję i czy mam do nich dostęp? Skąd mogę je wziąć?

Czy są kompatybilne, dobrze reprezentowane, czy muszę je przetwarzać, czyścić?

Kwestie etyczne, prawne, prywatność.

Jaki jest rozmiar danych? Czy jest ich wystarczająco?

Scenariusz wykorzystania danych

Model

Jakich metod, narzędzi, technik mogę użyć do rozwiązania problemu na których danych?

Czy buduję jeden model czy strukturę zależnych od siebie modeli?

Dlaczego tak?

Czy ktoś już robił coś podobnego i w jaki sposób?

Scenariusz wykorzystania danych

Ewaluacja wyników

Jakich wyników się spodziewam?

Jak będę oceniał model i jego wyniki? Jakie przyjmę miary?

Jak umiejscowię wiedzę ekspercką?

Które kluczowe elementy projektu powiedzą mi, że robię coś nie tak?

Jakie liczby, rezultaty, intuicje - będą testem zdroworozsądkowym?

Czy jest lista rzeczy, które powinienem sprawdzać w modelu zanim będę go używał?

Scenariusz wykorzystania danych

Ewaluacja wyników

Jakich wyników się spodziewam?

Jak będę oceniał model i jego wyniki? Jakie przyjmę miary?

Jak umiejscowię wiedzę ekspercką?

Które kluczowe elementy projektu powiedzą mi, że robię coś nie tak?

Jakie liczby, rezultaty, intuicje - będą testem zdroworozsądkowym?

Czy jest lista rzeczy, które powinienem sprawdzać w modelu zanim będę go używał?

Kanwa Modelu Danych (Data Model Canvas)

Problem biznesowy/pytanie badawcze

Jak brzmi specyficzne pytanie biznesowe na które ma odpowiedzieć projekt? Jakie pytanie zadajemy, w imieniu kogo?



Wartość biznesowa

Jakie potrzeby Klientów zostaną zaadresowane?



Zespół

Jakie kompetencje są wymagane od zespołu realizującego projekt? Czy masz wystarczająco specjalistów? Jak ich pozyskać?



Dane

Jakie dane są nam potrzebne? Gdzie są dane? Skąd pozyskać? Rozmiar danych? Czy wystarczy danych? Jakie operacje na przetwarzaniu danych? Czyszczenie, transformacje? Etyka, prywatność, interpretowalność, transparentność, rozliczalność. Model dojrzałości danych.



Model

Jaki model zastosujesz do rozwiązania problemu? Które modele wpgóle są adekwatne do tego problemu? Czy istnieją podobne projekty lub papiery które rozwiązywały taki problem i muszę się ich nauczyć?

Ewaluacja

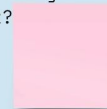
Spodziewany efekt prac? Jak zostanie oceniony? Skąd będziesz wiedział, że projekt jest prawidłowo wykonany? Jakie części projektu są kluczowe i od razu po nich widać, że coś jest nie tak? W jakiej kolejności, kiedy i jakie wyniki sprawdzę w projekcie, by być pewnym że działa.

Uruchomienie/Wdrożenie/UX

Jak zaprojektować końcowy efekt metody by transformował do wartości biznesowej? Jaki będzie końcowy efekt działania? W jaki sposób pokażesz go światu? Rozwiązanie dopasowane czy ogólnego zastosowania?

Użytkownicy

Kto będzie użytkownikiem rozwiązania? Jakie mają wymagania? Jak o je zweryfikujesz?



Ewaluacja

W jaki sposób zostanie oceniony efekt prac?



Przewidywane benefity

Co nam da zaproponowanie rozwiązania problemu?



Data Model Canvas 3.0

Business Problem/Question What is the specific business question the project is intended to answer? What question are you asking/answering and for whom?		Business Value How this project powers business value?	
Team/Collaborations Do you have the necessary data science capability? How can you build this capability?	Data Which data should we use and where is it located? Ethics, privacy, interpretability, transparency, accountability. Data maturity model. Where am I going to get the data? What is the potential size of the data? Is it enough data? What data cleaning, munging, etc will I need to do to this data?		Deployment/UX How do you design the model's output such that it can be efficiently consumed by the business? How do you find a balance between solutions tailored specifically for one application to a more general solution that can be used in other applications? How do you promote the operationalization of the solution in the enterprise? What is the right support model once the solution is deployed in production?
	Model What methods/techniques should I use? Why do I think these are the correct methods/techniques to use for this type of problem and data set? Are there similar projects / papers that have already done this that I can learn from before I get started?		
	Evaluation What do you expect the result to be? How will you evaluate your methods and results? How will I know I did the analysis and project correctly? What are key parts of the project that will tell me that I am doing things incorrectly? What numbers / results / insights will I sense check? What are simple logical chronological checkpoints I can put into my project to ensure I check to see what if what I am doing is working?		
Partners/Collaborations Do you need to collaborate with external partners?	Users Who will be the user/customers of your data product? What are their requirements? How will you teach them?		
Expected Costs Is this a quick-win, low-complexity project or a long-term engagement?		Expected Benefits	

Business Model Canvas

Kanwa modelu biznesowego (Business Model Canvas)

Segmenty klientów

Propozycja wartości

Kanały

Relacje z klientami

Strumienie przychodów

Kluczowe zasoby

Kluczowe działania

Kluczowi partnerzy

Struktura kosztów

Szablon modelu biznesowego (The Business Model Canvas)

Kluczowi partnerzy

Kim są Twoi kluczowi partnerzy, aby uzyskać przewagę konkurencyjną?



Kluczowe działania

Jakie są kluczowe kroki, aby dotrzeć do klientów?



Propozycja wartości

Jak sprawisz, że życie Twoich klientów będzie szczęśliwsze?



Relacje z klientami

Jak często będziesz wchodzić w interakcje z klientami?



Segmenty klientów

Kim są Twoi klienci? W kilku słowach opisz grupę docelową.



Kluczowe zasoby

Jakich zasobów potrzebujesz, aby Twój pomysł zadziałał?



Kanały

Jak zamierzasz dotrzeć do swoich klientów?



Struktura kosztów

Ile planujesz wydać na rozwój produktu i marketing w określonym okresie?



Strumienie przychodów

Ile planujesz zarobić w określonym okresie? Porównaj swoje koszty i przychody.



Kanwa modelu biznesowego (Business Model Canvas)

Segmenty klientów

- pierwszy obszar modelu biznesowego
- zawiera różne grupy ludzi i organizacji do których chcemy skierować nasze działania
- podział względem obsługiwanych segmentów, tzn. klienci o podobnej charakterystyce powinni się znaleźć w jednym segmencie

Pytania pomocnicze: Kto jest naszym klientem? Dla kogo budujemy produkt/usługę? Komu oferujemy wartość? Kto będzie płacił?

Propozycja wartości

- chcemy zaoferować pewną wartość
 - zbiór produktów/usług, które mają dla klienta znaczenie
- podstawowe rozróżnienie między tym co oferujemy my, a nasza konkurencja
- agreguje korzyści, które otrzyma nasz klient

Pytania pomocnicze: Jaką wartość generujemy dla naszych klientów? Za co będą płacić? Co ich boli, a co ma dla nich kluczowe znaczenie? Jakie problemy klientów rozwiązujemy? Jakie produkty i usługi będziemy oferować?

Kanwa modelu biznesowego (Business Model Canvas)

Kanały

- sposób w jakie dostarczamy naszą propozycję wartości klientowi
 - kanały komunikacji,
 - dystrybucji,
 - sprzedaży.
- punkty, w których dochodzi do kontaktu między klientem a naszą firmą

Pytania pomocnicze: Gdzie bywają nasi klienci? Gdzie chcemy spotkać naszych klientów? Z jakich kanałów będziemy korzystać przy nawiązywaniu kontaktu z klientem?

Relacje z klientami

- charakterystyka relacji, jakie nawiązujemy z klientami podczas przekazywania propozycji wartości
 - osobista relacja,
 - w pełni automatyczna obsługa.
- w jakiego rodzaju relacje chcemy wchodzić ze swoimi klientami?

Pytania pomocnicze: Jakich relacji oczekują od nas nasi klienci? Czy oczekują osobistego wsparcia, a może szybkiej i automatycznej obsługi? Czy sposób nawiązywania relacji z klientami jest zintegrowany z pozostałymi obszarami modelu biznesowego?

Kanwa modelu biznesowego (Business Model Canvas)

Strumienie przychodów

- celem biznesu jest zarabianie pieniędzy
- dostarczanie klientom odpowiedniej propozycji wartości, powinno generować strumienie przychodów
- wskazujemy w jaki sposób nasz produkt lub usługa będzie zarabiać
 - dla każdego segmentu klientów możemy mieć inny mechanizm cenowy

Pytania pomocnicze: Za co klienci są w stanie zapłacić? Za co i ile będą płacić? Które elementy naszego produktu/usługi będą darmowe, a które płatne?

Kluczowe zasoby

- wartości materialne i niematerialna, które potrzebujemy by dać klientom odpowiednią wartość
 - ludzie,
 - maszyny i materiały produkcyjne
 - licencje

Pytania pomocnicze: Jakich kluczowych zasobów potrzebujemy, żeby zaoferować naszą propozycję wartości? Jakich zasobów wymagają nasze kanały dotarcia do klienta i nawiązywane z nim relacje?

Kanwa modelu biznesowego (Business Model Canvas)

Kluczowe działania

- odpowiednie działania, które są kluczowe do stworzenia i zaoferowania klientom propozycji wartości
- rodzaj kluczowych działań ściśle zależy od charakteru naszego biznesu

Pytania pomocnicze: Jakie działania musimy podejmować, by dostarczyć naszym klientom propozycję wartości? Jakich działań wymagają nasze kanały dotarcia do klienta i nawiązywane z nim relacje?

Kluczowi partnerzy

- biznes opiera się na współpracy z kluczowymi partnerami
- firmy lub organizacje, których usługi lub produkty są konieczne do zaoferowania propozycji wartości
- niezbędni dostawcy lub podwykonawcy

Pytania pomocnicze: Kto jest naszym kluczowym partnerem? Jakie zewnętrzne firmy lub organizacje są nam niezbędne do działania? Jakie kluczowe zasoby i działania realizują nasi partnerzy?

Kanwa modelu biznesowego (Business Model Canvas)

Struktura kosztów

- po stronie klienta mamy generowane strumienie przychodów
- po stronie naszych wewnętrznych zasobów i działań mamy strukturę kosztów
- wszystkie wydatki ponoszone w związku z funkcjonowaniem modelu biznesowego
- wyliczyć, gdy znamy kluczowe zasoby, działania i partnerów naszego modelu biznesowego

Pytania pomocnicze: Jakie koszty generuje nasz model biznesowy? Jakie nakłady finansowe generują kluczowe zasoby, działania, partnerzy?

Open Data — Business Model Canvas

Szablon modelu biznesowego otwartych danych (The Open Data Business Model Canvas)

Kluczowi partnerzy

lokalny biznes
obywatel
agencje rządowe
developerzy aplikacji
dostawcy map
entuzjaści otwartych danych
startupy



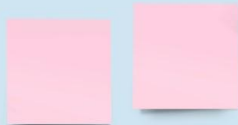
Kluczowe działania

zaangażowanie interesariuszy
publikacja danych otwartych
utrzymywanie aktualności danych
utrzymanie portalu
dostępowego



Propozycja wartości

prosty i intuicyjny
interfejs do otwartych danych
prostota obsługi
zapotrzebowanie na dane
darmowy dostęp do danych
darmowa licencja



Relacje z klientami

samoobsługa na portalu



Segmenty klientów

developerzy aplikacji
dostawcy map
dziennikarze



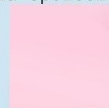
Kluczowe zasoby

społeczność
marka
platforma/portal



Kanały

portal z danymi otwartymi
hackatonu
media społecznościowe



Struktura kosztów

portal - rozwój i utrzymanie
marketing i zarządzanie relacjami ze społecznością













Strumienie przychodów

brak bezpośrednich strumieni przychodu
wkład dla dobra wspólnego



Machine Learning Model Canvas

Decisions  How are predictions used to make decisions that provide the proposed value to the end-user?	ML task  Input, output to predict, type of problem.	Value Propositions  What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?	Data Sources  Which raw data sources can we use (internal and external)?	Collecting Data  How do we get new data to learn from (inputs and outputs)?
Making Predictions  When do we make predictions on new inputs? How long do we have to <u>featurize</u> a new input and make a prediction?	Offline Evaluation  Methods and metrics to evaluate the system before deployment.		Features  Input representations extracted from raw data sources.	Building Models  When do we create/update models with new training data? How long do we have to <u>featurize</u> training inputs and create a model?
	Live Evaluation and Monitoring Methods and metrics to evaluate the system after deployment, and to quantify value creation.			

AI Project Canvas

AI Project Canvas

Kolejność:








1. Dane, umiejętności i wynik
2. Propozycja wartości
3. Integracja, interesariusze, klienci
4. Koszty i przychody

Szablon projektu AI (The AI Project Canvas)

<div>Dane Które dane są nam potrzebne?</div> <div><div></div><div></div></div>	<div>Umiejętności Jakie umiejętności są wymagane w celu zrealizowania projektu?</div> <div><div></div></div> <div>Wynik Jakie metryki optymalizujemy?</div> <div><div></div></div>	<div>Propozycja wartości Jak projekt AI podniesie wartość?</div> <div><div></div><div></div></div>	<div>Integracja Jak projekt będzie się integrował?</div> <div><div></div></div> <div>Interesariusze Kim są zainteresowani projektem AI?</div> <div><div></div></div>	<div>Klienci Kim są klienci?</div> <div><div></div></div>
<div>Struktura kosztów Jakie koszty w projekcie poniesiemy?</div> <div><div></div></div>		<div>Strumienie przychodów Jak projekt wygeneruje przychód?</div> <div><div></div></div>		

AI Project Canvas

Title: Anonymization

Data  <i>Which data do you need?</i> 10'000 labeled images from car perspective Dataset not available yet JPEG or PNG	Skills  <i>Which skills do you need for development?</i> Computer Vision Python & Data Engineering	Value Proposition  <i>What is the value added by your project?</i> Protect privacy Enable data collection for function development	Integration  <i>How will the project be integrated?</i> Microservice offering anonymization endpoint	Customers  <i>Who are the end customers?</i> Driver assistance function developers
Cost  <i>What costs will the project incur?</i> 1 FTE for 3 months Image labeling Cloud Computing			Revenue  <i>How will the project generate revenue?</i> SaaS	

Data Science Workflow Canvas

Kanwa danologiczna (The Data Science Workflow Canvas)

Definicja problemu

Jaki problem staramy się rozwiązać?
Jakie zagadnienia dotyczą problemu?



Wyniki działania rozwiązania

Jakiego typu wyniki chcemy uzyskać? Np. dla predykcji zidentyfikuj dane predykcyjne X i predykowane Y.



Pozyskanie danych

Jakie źródła danych będziemy pozyskiwać? Czy jest wystarczająco danych? Czy możemy ich użyć?

Modelowanie

Jakie modele są adekwatne dla zakładanych wyników?

Ewaluacja modeli

Jak możesz ewaluować jakość swojego modelu?

Przygotowanie danych

Jak musisz przygotować dane by działał model i realizował zakładane wyniki?

Podsumowanie

Aby dobrze wykorzystać szablony pamiętaj o tych trzech rzeczach:

- Nie bój się popełniać błędów
 - użyj canvas jako przestrzeni do burzy mózgów
 - wracaj do szablonów, aby udoskonalić swój proces
 - zachowaj to, co działa, i usuń to, co nie działa
- Skoncentruj się na tym, co chcesz osiągnąć
 - przy zmianie początkowych celów, skup się na tym, co chcesz osiągnąć
 - myśl o celach niezależnie od tego jak często musisz wracać i je aktualizować
- AI oraz nauka o danych jest procesem nieliniowym i iteracyjnym
 - nie ma poprawnego ani liniowego sposobu realizacji projektu AI
 - canvas to zasób ułatwiający rozpoczęcie pracy nad projektem