

Cześć I

„O co w tym chodzi?” Najważniejsze zasady Data Science

Przy całej uwadze poświęconej najwyraźniej nieskończonemu potencjałowi technologii i różnorodnym możliwościom, jakie oferuje ona sprytnym przedsiębiorcom, ktoś mógłby zapytać po co w ogóle zajmować się Data Science. Czemu po prostu nie opanować technologii. W końcu to one napędzają świat i nie widać żadnych oznak spowolnienia ich rozwoju. Czytelnik zainteresowany rozwojem swojej kariery zawodowej może uważać, że wiedza o tym, jak opanować nową technologię, zagwarantuje mu powodzenie.

Łatwo jest uznać technologię za siłę, która zmienia świat – zawdzięczamy jej komputery osobiste, internet, sztuczne organy, autonomiczne pojazdy, system GPS, ... Niewielu ludzi myśli o Data Science jako o sile napędowej wielu z tych wynalazków. Właśnie dlatego powinieneś przeczytać tę książkę zamiast książki poświęconej technologii – żeby coś zmienić, musisz poznać sposób, w jaki to coś działa.

Nie powinniśmy uważać danych za nudnego, ale przydatnego przodka, a technologii za szykownego nastolatka. Znaczenie Data Science nie sprowadza się do wyjaśnienia, że technologia potrzebuje danych jako jednego z wielu swoich składników. W ten sposób zaprzeczylibyśmy pięknu danych i wielu ich interesującym zastosowaniom, zarówno w pracy, jak i zabawie. W skrócie, rozwój technologii byłby niemożliwy bez Data Science i *vice versa*. To znaczy, że poznanie podstaw Data Science otwiera wiele drzwi, bo specjaliści danych potrzebni są w wielu dziedzinach. Data Science jest wyjątkowo obiecującym obszarem prac i badań.

Część I wprowadzi Cię w świat wszechobecnych danych i najważniejszych zasad Data Science, które pomogą Ci pogłębić posiadaną wiedzę. Pojęcia przedstawione w trzech pierwszych rozdziałach dadzą Ci zarys obrazu, w jaki sposób dane dotyczą Ciebie, i pozwolą Ci się zastanowić nie tylko nad tym, jak Ty i Twoja firma mogą z nich skorzystać, ale również nad tym, jak możesz użyć danych do rozwoju swojej kariery i nie tylko.

Dokąd zmierzamy

W rozdziale 1 zrobimy pierwsze kroki naszej wędrówki po Data Science. Wyjaśnię w nim, jak my wszyscy, żyjąc w erze komputerów, przyczyniamy się do gigantycznego przyrostu danych. Następnie przejdę do odpowiedzi na pytania, jak ludzie gromadzą i pracują z danymi oraz – przede wszystkim – jak można użyć danych z korzyścią dla różnorodnych projektów prowadzonych w ramach Data Science i poza tą dziedziną.

Przekonamy się, że jednym z problemów dotyczących Data Science nie jest jej trudność, ale to, że dla wielu osób obszar Data Science pozostaje niejasny. Dopiero gdy zdamy sobie sprawę z ogromnej ilości danych i poznamy metody ich zbierania, będziemy mogli zastanowić się nad różnorodnymi sposobami pracy z danymi. Osiągnęliśmy poziom rozwoju technologicznego pozwalający gromadzić i przechowywać dane na potrzeby najróżniejszych branż – wystarczy spojrzeć na liczbę ogólnie dostępnych zbiorów danych i projektów rządowych mających na celu zestawienie danych zbieranych przez różne instytucje kulturalne i polityczne. Jednak wciąż relatywnie mało osób wie, jak dostać się do tych danych i jak je przeanalizować. Bez specjalistów rozumiejących ich przydatność, te wspaniałe zbiory danych będą się tylko kurzyły. Ten rozdział odpowiada na pytania, dlaczego właśnie teraz Data Science stała się ważna, dlaczego nie jest tylko chwilową modą i dlaczego warto rozważyć jej użycie w codziennej pracy.

W drugiej części rozdziału szczegółowo wyjaśnię, czemu gwałtowny rozwój technologii nie pozwala nawet na chwilę wstrzymać badań w zakresie Data Science. Jakikolwiek byłyby nasze obawy względem nadchodzącego świata, nie możemy przestać zbierać, przechowywać i używać danych. Niemniej nie możemy zignorować faktu, że dane jako takie nie są etyczne lub nieetyczne i że otwiera to pole do nadużyć. Ci z Was, którym kwestie etyczne leżą na sercu, mogą wziąć sprawę w swoje ręce i dołączyć do dyskusji prowadzonych przez międzynarodowe instytucje powołane w celu rozwiązywania etycznych problemów wynikających z używania danych. Temat ten jest tak wciągający, że poświęciłem mu osobny punkt rozdziału 3 – Ciemna strona sztucznej inteligencji.

Przyszłość to dane

Wszystko, każdy proces, każde urządzenie, wkrótce będzie sterowane danymi. To całkowicie zmieni sposób prowadzenia biznesu. Przewiduję, że w ciągu 10 lat od każdego pracownika, we wszystkich firmach, będzie się oczekiwało umiejętności pracy z danymi oraz zdolności analitycznych, dzięki którym wypracuje on dodatkową wartość dla firmy. Nie tak bardzo zwariowany pomysł, jeśli wziąć pod uwagę, że teraz od wielu osób spodziewamy się umiejętności posługiwania się systemem płatności mobilnych i cyfrowym portfelem Apple Pay, który zadebiutował na rynku w roku 2014.

W rozdziale 2 wyjaśnię, że każdy aspekt naszego życia wiąże się z danymi. Dane nami kierują, a ich siła rośnie wraz z ich ilością. Chociaż technologia dopiero niedawno, jak na całą historię ludzkości, stała się ważną częścią naszego życia, dane zawsze odgrywały w nim zasadniczą rolę. Nasze DNA zawiera najbardziej podstawowe dane na nasz temat. To one nami sterują: odpowiadają za sposób, w jaki patrzymy, za kształt naszych kończyn, za budowę i działanie naszych mózgów i za zakres emocji, których doświadczamy. Jesteśmy pojemnikami na dane, chodzącymi pamięciami przenośnymi biochemicznych informacji, przekazującymi je naszym dzieciom i programującymi je mieszanką danych naszych i naszego partnera. Brak zainteresowania danymi oznacza brak zainteresowania najważniejszymi zasadami naszego istnienia.

Z tego rozdziału dowiesz się, jak dane są używane w różnych obszarach. Żeby to zilustrować, posłużyłem się przykładami bezpośrednio nawiązującymi do teorii hierarchii potrzeb Masłowa, teorii, która powinna być znana wielu studentom, przedsiębiorcom i menedżerom. Jeżeli nie znasz teorii Masłowa, nie martw się – wyjaśnię ją i jej związek z danymi w rozdziale 2.

Hamowanie postępu

W ostatnim rozdziale części I przyjrzę się bieżącemu stanowi badań nad sztuczną inteligencją, jej potencjalnym zastosowaniom i związanymi z nimi zagrożeniami. Rozwój sztucznej inteligencji pośrednio wpłynął na różne obszary naszego życia. Pojawiły się pytania o przyszłość specjalistów danych, jak również badaczy i praktyków, wykraczające poza obszar Data Science. Jeżeli planujesz karierę specjalisty danych, ten rozdział może dostarczyć inspiracji na zagospodarowanie nisz tak bardzo potrzebujących wykwalifikowanych fachowców.

Żeby uzupełnić przykłady z rozdziału 2, zawierającego przekonujące przykłady użycia danych w codziennym życiu, w rozdziale 3 zwracam uwagę na pięć najbardziej obiecujących zastosowań sztucznej inteligencji w biznesie. Różnorodność zastosowań sztucznej inteligencji może utrudniać ich poznanie. Ten rozdział dostarczy

Ci podstawowych informacji o najważniejszych trendach użycia SI i zachodzących w nich zmianach.

Pozytywny wpływ sztucznej inteligencji jest oczywisty, ale nie powinien nas zaślepić. Dlatego w rozdziale 3 znajdziemy też informacje o zagrożeniach bezpieczeństwa wynikających z danych i ich użyciu przez sztuczną inteligencję oraz o tym, jak specjaliści danych mogą rozwiązywać takie, bieżące i przyszłe, problemy. Etyka jest fascynująca ze względu na jej wpływ na kierunki rozwoju Data Science. Na podstawie tego, co wiemy na temat metod zbierania danych i sposobów ich użycia przez maszyny i usługi online, etyka danych wyznacza obszar dialogu między ludźmi a technologią.