

Kim jest Data Scientist. Jak im pomóc, aby zyskała firma



Łukasz
Kobyliński

Pracują z różnorodnymi językami programowania, znajomość statystyki mają w małym palcu, a rozwiązywanie kwestii związanych z prowadzeniem firmy przy pomocy technik opartych na danych to dla nich bułka z masłem. Data Scientist to obecnie najbardziej pożądany zawód na rynku IT. Ogromny wzrost popularności tej specjalizacji pokazuje, jak na przestrzeni lat w dobie Big Data zmieniło się podejście przedsiębiorstw do tematu zarządzania danymi. Zarówno większe, jak i mniejsze firmy przekonały się, że wszelkie dostępne dane - te wewnętrzne, zewnętrzne, ustrukturyzowane i nieustrukturyzowane - trzeba brać pod uwagę w codziennej działalności przedsiębiorstw. Kim jest, z jakich narzędzi korzysta i ile zarabia specjalista Data Science?

Specjaliści Data Science robią zawrotną karierę na rynku IT. Poszukuje ich coraz więcej firm, a zapotrzebowanie na nich nieustannie rośnie. Data Scienciści, mówiąc najprościej, to osoby analizujące dane w taki sposób, aby wyciągnąć z nich wnioski niosące realną wartość dla biznesu. To matematycy, informatycy i badacze trendów w jednym. Specjaliści w tej dziedzinie posiadają unikalne umiejętności wsparcia zarówno zagad-

nień technologicznych, jak i biznesowych. Charakteryzują się ukierunkowaną na wykrywanie i rozwiązywanie istotnych dla przedsiębiorstw kwestii. Z jakich narzędzi korzysta Data Scientist? Są to przede wszystkim: wizualizacja danych, uczenie maszynowe, uczenie pogłębione, rozpoznawanie wzorców, przygotowywanie danych oraz analityka tekstu. Raport na temat tej specjalizacji udostępniła w Polsce platforma Kaggle. Wynika z niego, że

zostanie “badaczem danych” to najlepszy kierunek zawodowy, na jaki mogą zdecydować się młodzi ludzie. To nie tylko ciekawa i pełna wyzwań praca, ale także możliwość zarobienia naprawdę dobrych pieniędzy. Mediana zarobków w 2018 roku wynosi bowiem minimum 7 tysięcy złotych miesięcznie.

Czym Data Scientist różni się od standardowego analityka danych? Przede wszystkim tym, że umie programować. Bardzo istotne jest tu również zagadnienie Big Data, ponieważ przy tak ogromnych ilościach danych jakie gromadzą dziś firmy każde rozwiązanie powinno być projektowane w kontekście wydajności i planowania wdrożenia produkcyjnego, a tego w oderwaniu od programowania

niez znajomość koncepcji przetwarzania języka naturalnego, czy analizy obrazu. Dla organizacji, w które pracuje Data Scientist ważne jest, aby tego rodzaju specjalista potrafił znajdować w danych reguły i wzorce, a także wykrywać trendy mogące poprawić pewien aspekt działania firmy, na przykład sprzedaż, czy skuteczność marketingu.

Każdy dobry Data Scientist powinien umieć świetnie się komunikować. Mowa tu o komunikacji zarówno pisemnej, jak i mówionej. Przeważnie jest tak, że po wykonaniu wszystkich analiz, przetestowaniu wielu różnych modeli statystycznych osiągnięte rezultaty trzeba przedstawić w formie prezentacji, naukowego artykułu albo po prostu



zrobić się nie da. Językami, które w dziedzinie Data Science są najważniejsze to R i Python (razem z bibliotekami Numpy i Pandas). Ze względu na częstą pracę z bazami danych także język zapytań SQL jest niezbędną umiejętnością Data Scientista. Specjalista w tej dziedzinie musi także idealnie poruszać się w materii inżynierii oprogramowania oraz algorytmów i struktur danych. Kolejnym, bardzo ważnym aspektem jego pracy jest obszar uczenia maszynowego: tworzenia modeli predycyjnych na podstawie historycznych danych. Dane takie mogą mieć charakter liczbowy, ale mogą być też tekstem, czy obrazem. Stąd ważna jest rów-

raportu. Stąd tak ważne jest, aby potrafić formułować swoje myśli w taki sposób, by przekazać wszystkie ważne informacje i problemy związane z analizą w sposób zrozumiały dla osób zajmujących się rozwojem biznesu, a nie zgłębianiem danych.

Zdecydowana większość specjalistów Data Science zaczynała swoją karierę na stanowisku analityka danych lub statystyka. Jednak razem z rozwojem Big Data zmieniły się ich zadania. Wszystko dlatego, ponieważ dane nie są już dla informatyków poboczną kwestią. Są to kluczowe, wymagające analitycznego i logicznego podejścia informacje,

które potrzebują też, aby podejść do nich z dociekliwością i, przede wszystkim, umieć przełożyć pomysł z zakresu zaawansowanych technologii na sposoby generowania przychodów.

Sprawdźmy jeszcze, jak wyglądają zarobki specjalistów Data Science. Według raportu Kaggle, w Polsce mediana rocznych wynagrodzeń wynosi 23.894\$, co w przeliczeniu na złotówki daje około 85 tysięcy złotych. Miesięcznie jest to 7 tysięcy złotych. Jak na polskie warunki to naprawdę dobrze. Wyniki ankiety udowadniają również, że większość zarobków polskich analityków danych mieści się poniżej dwukrotnie wyższej kwoty, czyli 50.000\$ rocznie.

Kilka lat temu rynek IT zorientował się, że Data Scientistów jest po prostu za mało. Dało to szansę wypełnienia kolejnej niszy w polskim systemie edukacji i wpłynęło na powstawanie nowych kierunków studiów lub intensywnych kursów zawodowych (tzw. bootcampów) w obszarze analizy i przetwarzania dużych zbiorów danych.

Bootcampy systematycznie zyskują na popularności i przyciągają rzesze osób, które albo chcą się przebranżowić i rozpocząć od podstaw karierę w dziedzinie Data Science albo pragną rozwinąć swoje aktualne, programistyczne umiejętności. Koszt takiego specjalistycznego kursu, choć na pierwszy rzut oka wysoki, jest szybko zwracającą się inwestycją. Jego wysokość wynika właśnie z zaangażowania mentorów i trenerów, czyli do-

świadczonych specjalistów z danego obszaru, którzy poświęcają ponad 8 godzin dziennie na przekazywanie swojej wiedzy kursantom.

Czym dokładnie jest bootcamp? To intensywny, kilkutygodniowy kurs pozwalający zdobyć kompetencje w poszukiwanych na rynku zawodach w czasie o wiele krótszym, niż na przykład w przypadku studiów. Uczestnicy bootcampu uczą się pracując przynajmniej 8 godzin dziennie i są nadzorowani przez dedykowanego mentora. Statystyki ze Stanów Zjednoczonych pokazują, że 30% absolwentów znajduje pracę w ciągu 30 dni od ukończenia kursu, a 89% w okresie ponad 120 dni. Również na polskim rynku IT potrzeby są coraz większe. Szacuje się bowiem, że brakuje u nas kilkadziesiąt tysięcy programistów.

Absolwenci bootcampów mają dużą przewagę na rynku pracy, ponieważ po ukończeniu kursu posiadają dokładnie te kompetencje, jakie są poszukiwane i pożądane przez pracodawców. Stąd też oferta bootcampów jest po prostu zoptymalizowaną formą kształcenia – najbardziej dopasowaną i efektywną czasowo oraz kosztowo skierowaną do osób, które chcą podjąć pierwszą pracę w nowym zawodzie.



O autorze:

Łukasz Kobylński, dyrektor zarządzający w Sages i Kodołamacz

Chief Science Officer w Sages oraz adiunkt w Instytucie Podstaw Informatyki PAN, gdzie prowadzi projekty w Zespole Inżynierii Lingwistycznej. Od wielu lat zajmuje się analizą danych i uczeniem maszynowym, początkowo w odniesieniu do obrazów - w roku 2012 obronił doktorat na Politechnice Warszawskiej z tego zakresu - a obecnie w zastosowaniu do przetwarzania języka naturalnego. Szczególnie zainteresowany lingwistyką korpusową, analizą tekstu na poziomie morfoskładniowym i semantycznym, a także efektywnym przetwarzaniem dużych zbiorów danych. Posiada duże doświadczenie trenerskie - przeprowadził ponad 1000 godzin szkoleń i warsztatów.

