



Uniwersytet Ekonomiczny
we Wrocławiu

KIERUNEK STUDIÓW
Zarządzanie

Krzysztof Marczyński
Nr albumu 181154

PRACA MAGISTERSKA

Analiza czynników wpływających na fluktuację pracowników IT z wykorzystaniem uczenia maszynowego

Promotor:
dr hab. Joanna Radomska, prof. UEW
Katedra Zarządzania Strategicznego

Pracę akceptuję i wnioskuję o jej dopuszczenie
do dalszych etapów postępowania.

.....
podpis Promotora

WROCŁAW 2022

Streszczenie

Celem pracy było sprawdzenie, czy uczenie maszynowe pozwala określić z zadowalającym stopniem pewności cechy wpływające na poziom zadowolenia i chęć zmiany pracy pracowników IT. Aby osiągnąć ten cel przeprowadzono przegląd literatury w zakresie problematyki fluktuacji i zadowolenia pracowników ze szczególnym zwróceniem uwagi na charakterystyki specyficzne dla branży informatycznej. Następnie wybrano źródło danych pozwalające przeprowadzić badania ilościowe dla społeczności pracowników zatrudnionych w branży informatycznej i przygotowano algorytm w języku Python pozwalający na zbudowanie modeli predykcji opartych o regresję i klasyfikację. W celu oceny skuteczności modelu predykcji wybrano zestawy odpowiednich metryk zarówno dla regresji, jak i klasyfikacji. Rezultaty badań mogą stanowić sugestie dla pracowników działów rekrutacyjnych i employer branding, jakie kwestie organizacyjne w firmie mogą ulec usprawnieniu, aby zwiększyć retencję i zadowolenie pracowników.

Abstract

The aim of this work was to check whether machine learning allows to determine with a satisfactory degree of certainty the features influencing the level of satisfaction and willingness to change the job of IT employees. To achieve this goal, a literature review was carried out in the field of employee fluctuation and satisfaction, paying particular attention to the characteristics specific to the IT industry. Then, a data source was selected to conduct quantitative research for the community of employees working in the IT industry, and an algorithm in the Python language was prepared to build prediction models based on regression and classification. In order to assess the effectiveness of the prediction models, sets of appropriate metrics were selected for both regression and classification. The results of the research may be a suggestion for employees of recruitment and employer branding departments which organizational issues in the company can be improved in order to increase retention and satisfaction of the employees.

Spis treści

Wstęp	4
Opis problemu	4
Cel i zakres pracy	5
Struktura pracy	6
1. Problem fluktuacji pracowników w literaturze	7
1.1. Zjawisko fluktuacji	7
1.2. Wpływ fluktuacji na firmę	9
1.3. Przyczyny fluktuacji pracowników	10
1.3.1. Zadowolenie i satysfakcja pracowników	11
1.3.2. Zaangażowanie organizacyjne pracowników	14
1.4. Koszt fluktuacji pracowników	15
1.5. Sposoby na retencję pracowników	16
2. Charakterystyka branży IT	18
2.1. Ogólna charakterystyka	18
2.2. Zadowolenie pracowników	19
2.3. Wpływ fluktuacji na utratę wiedzy w projektach informatycznych	20
2.4. Charakterystyka pracowników w branży IT	21
2.5. Trendy na rynku pracy	26
3. Optymalizacja procesu rekrutacji w branży IT z wykorzystaniem uczenia maszynowego	29
3.1. Metody badawcze	29
3.2. Wstępne przetworzenie danych	30
3.3. Selekcja cech z wykorzystaniem uczenia maszynowego	31
3.4. Prezentacja wyników	34
3.4.1. Wyniki regresji	34
3.4.2. Wyniki klasyfikacji	39
3.5. Wnioski i analiza możliwości praktycznego zastosowania zbudowanego modelu predykcji	44
Zakończenie	45

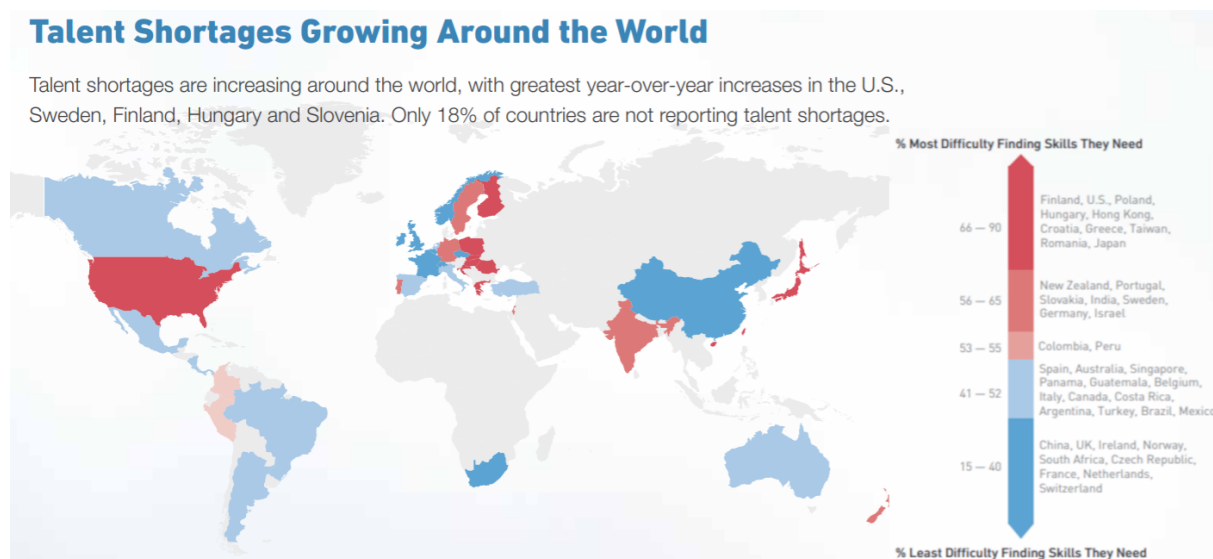
Bibliografia	47
Spis tabel	52
Spis rysunków	53
Spis kodów źródłowych	55
Dodatek A. Lista pytań w ankiecie StackOverflow	56
A.1. Pytania z roku 2017	56
A.2. Pytania z roku 2018	60
A.3. Pytania z roku 2019	63

Wstęp

Opis problemu

W ciągu ostatniego stulecia na świecie można zauważyć bezprecedensowe tempo rozwoju nowych technologii, w związku z czym nieustannie rośnie zapotrzebowanie na odpowiednio wykwalifikowanych pracowników technicznych. Na uwagę szczególnie zasługuje rozwój informatyzacji i zapotrzebowanie na pracowników zajmujących się wytwarzaniem oprogramowania. Według raportu Daxx, w 2021 roku na świecie jest zatrudnionych ok. 27 milionów programistów, natomiast globalne zapotrzebowanie jest o 40 milionów większe. Oszacowali też, że do 2030 roku globalnie może być zatrudnionych ok. 45 milionów programistów, a deficyt sięgać może nawet 85 milionów programistów [Daxx, 2022].

Na rysunku 1 przedstawiono problem niedoboru wykwalifikowanych pracowników w branży IT z podziałem na kraje.



Rysunek 1: Niedobór wykwalifikowanych pracowników IT na świecie

Źródło: [Daxx, 2020]

Mając na uwadze również, że całkowite koszty związane z odejściem pracownika na stanowisku programisty i zatrudnieniem na jego miejsce nowego wynoszą średnio ok. 50 tys. dolarów amerykańskich [Winter, 2016] (daje to kwotę ok. 200 tys. polskich złotych), wniosek nasuwa się prosty: firmom działającym w branży informatycznej zależy na zatrzymaniu (retencji) zatrudnionych programistów. Dzięki wysokim staraniom pracodawców związanych z retencją programistów, zawód programisty na początku 2022 roku znalazł się na 5 miejscu rankingu najlepszych zawodów według U.S.News [U.S.News, 2022].

W związku z omówionymi problemami związanymi z branżą informatyczną można sformułować następujące pytania badawcze:

1. Czy można wyłonić cechy (osoby lub przedsiębiorstwa) pozwalające oszacować zadowolenie i chęć zmiany pracy pracownika z branży IT?
2. Czy wyłonione cechy można użyć do poprawy procesu wstępnej selekcji kandydatów?
3. Czy wyłonione cechy można użyć do zwiększenia atrakcyjności przedsiębiorstwa dla pracowników i kandydatów?

Cel i zakres pracy

Cele pracy można podzielić na 3 kategorie:

1. Teoriopoznawcze:
 - przedstawienie zjawiska fluktuacji pracowników,
 - czynniki wpływające na fluktuację i zadowolenie pracowników w literaturze,
 - koszty związane z fluktuacją pracowników,
 - przedstawienie charakterystyki branży informatycznej i konsekwencji fluktuacji w tej branży;
2. Metodologiczne:
 - sprawdzenie, czy uczenie maszynowe pozwala określić z zadowalającym stopniem pewności (relatywnie niski znormalizowany błąd średniokwadratowy (ang. NRMSE)) cechy wpływające na poziom zadowolenia i chęć zmiany pracy pracowników IT;
3. Utylitarne:
 - możliwe usprawnienie pracy dla pracowników działów rekrutacyjnych i employer branding.

Struktura pracy

W pierwszym rozdziale pracy omawiane jest zjawisko fluktuacji pracowników z uwzględnieniem występującego w literaturze podziału na jej rodzaje, pokazania czym fluktuacja różni się od rotacji pracowników oraz w jaki sposób z fluktuacją związana jest retencja. Następnie przedstawione są czynniki prowadzące do wzrostu fluktuacji, pokazane jest czy wysoki wskaźnik fluktuacji zawsze oznacza problemy dla przedsiębiorstwa oraz jakie koszty ponosi firma w związku z fluktuacją.

W rozdziale drugim nakreślona zostaje charakterystyka zatrudnienia w branży informatycznej ze szczególnym zwróceniem uwagi na wskaźnik fluktuacji pracowników w tej branży. Szczegółowej analizie zostaje poddany problem utraty wiedzy w projektach informatycznych oraz jakości tworzonego oprogramowania w związku z odejściami pracowników.

W rozdziale trzecim przedstawiono źródło danych wybrane do przeprowadzenia badań ilościowych oraz uzasadniono zasadność przeprowadzenia badań z wykorzystaniem uczenia maszynowego dla omawianego w pracy problemu.

1. Problem fluktuacji pracowników w literaturze

1.1. Zjawisko fluktuacji

Fluktuacja pracowników (ang. employee turnover) w literaturze definiowana jest na wiele sposobów, jednak najczęstsze to "dobrowolne odejście z pracy pracowników dojrzałych" [Woźniak, 2016] lub szerzej jako "tempo w jakim pracownicy opuszczają firmę" [Cron and DeCarlo, 2006]. Czasami fluktuacja jest utożsamiana z rotacją pracowników (ang. job rotation), jednak częściej rotację definiuje się nie jako samo opuszczanie firmy przez pracowników, a raczej jako proces wymiany pracowników, który może służyć jako metoda rozwoju dla pracowników (np. poprzez przeniesienie pracownika na inne stanowisko wewnątrz organizacji lub awans) [Pocztowski, 2009]. Wysoka fluktuacja może nie być pożądana, choćby ze względu na wysokie koszty z nią związane [Philips and Edwards, 2009], więc odpowiedzią pracodawcy może być próba jej ograniczenia. Proces taki określany jest mianem retencji pracowników i jego głównym celem jest utrzymanie zatrudnienia na poziomie pozwalającym na "sprawną realizację procesów biznesowych" [Pocztowski, 2007].

Fluktuacja może być skategoryzowana na kilka różnych sposobów, co zostało przedstawione w tabeli 1.1.

Tabela 1.1: Rodzaje fluktuacji

Cecha fluktuacji	Opis
pożądana	<ul style="list-style-type: none"> • pożądana - opuszczenie pracownika pozytywnie wpływa na działanie firmy (np. pracownik o niskich kwalifikacjach), • niepożądana - pracownik jest trudny do zastąpienia i jego pracy przynosi firmie korzyści;
dobrowolna	<ul style="list-style-type: none"> • dobrowolna - pracownik sam odchodzi z firmy, • niedobrowolna - pracodawca zwalnia pracownika lub pracownik w inny sposób zostaje zmuszony do odejścia;
zapobiegalna	<ul style="list-style-type: none"> • zapobiegalna - firma jest jej w stanie zapobiec, • niezapobiegalna - zależy od czynników zewnętrznych, na które firma nie ma wpływu (np. wprowadzenie nowych regulacji prawnych);
nadmierna	<ul style="list-style-type: none"> • związana ściśle z wewnętrznym działaniem firmy - cechy takie jak złe warunki pracy i nieadekwatne wynagrodzenie mogą wpłynąć na zwiększenie fluktuacji pracowników w danym przedsiębiorstwie;

Źródło: opracowanie własne na podstawie: [Spychała et al., 2019]

Do mierzenia poziomu fluktuacji wykorzystuje się wskaźnik fluktuacji definiowany jako stosunek osób opuszczającej w firmę w danym roku do średniej liczby pracowników zatrudnionych w danym roku. Do wyliczenia fluktuacji można wykorzystać liczbę wszystkich pracowników, którzy opuścili organizację - niezależnie od powodu opuszczenia tejże organizacji - ale można też obliczyć ten wskaźnik uwzględniając jedynie fluktuację dobrowolną. Badanie fluktuacji dobrowolnej jest o tyle istotne, że może pozwolić na wykrycie problemów w firmie i opracowanie sposobu na ich przeciwdziałanie [Spychała et al., 2019]. Co więcej, badania pokazują, że większość odejść w organizacjach stanowią właśnie odejścia dobrowolne [Dalton et al., 1982]. Z tego względu, dalsze rozważania będą dotyczyły fluktuacji dobrowolnej, o ile wprost nie będzie napisane inaczej.

1.2. Wpływ fluktuacji na firmę

Fluktuacja pracowników niekoniecznie musi oznaczać problem dla przedsiębiorstwa. Z analizy badań Human Capital Index przeprowadzonych przez firmę Watson Wyatt w 2005 r. wynika, że zarówno bardzo niska i bardzo wysoka fluktuacja nie są korzystne dla przedsiębiorstw. Organizacje z umiarkowanym poziomem wskaźnika fluktuacji wynoszącym ok. 15% miały zwrot z inwestycji akcjonariuszy (ang. total shareholders return - TSR) na poziomie 43%, co stanowiło rezultat średnio o 9 punktów procentowych lepszy niż firmy o niższym lub o wyższym wskaźniku fluktuacji [Król and Ludwiciński, 2006].

Nie zawsze jednak można jednoznacznie stwierdzić, że wysoki poziom fluktuacji będzie negatywny w skutkach dla danego biznesu. Decyzję o tym należy poprzedzić analizą sytuacji w przedsiębiorstwach konkurencyjnych oraz specyfiki rynku pracy [Taylor, 2006]. Heurystyka pozwalająca określić czy wysoki poziom fluktuacji jest akceptowalny w przypadku danej branży została przedstawiona w tabeli 1.2.

Tabela 1.2: Heurystyka do oceny wysokiego poziomu fluktuacji

Wysoki poziom fluktuacji nieakceptowalny	Wysoki poziom fluktuacji akceptowalny
<ul style="list-style-type: none">• niedobór kandydatów o odpowiednich kompetencjach na rynku pracy,• fluktuacja jest wyższa niż u bezpośredniej konkurencji,• wysokie koszty rekrutacji;	<ul style="list-style-type: none">• wielu kandydatów o odpowiednich kompetencjach na rynku pracy,• niskie koszty rekrutacji,• niski koszt wdrożenia nowego pracownika,• niskie ryzyko utraty wiedzy w wyniku odejścia pracownika,• przewidywanie redukcji etatów w niedalekiej przyszłości;

Źródło: opracowanie własne na podstawie: [Taylor, 2006]

1.3. Przyczyny fluktuacji pracowników

Najstarszy model opisujący dobrowolną fluktuację pracowników został opracowany przez Marcha i Simona w 1958 roku. Model ten wyróżnia 2 czynniki wpływające istotnie na poziom fluktuacji [Woźniak, 2012]:

- to jak pracownik ocenia swoją chęć zmiany pracy - związane jest to głównie z niską satysfakcją z wykonywanej pracy oraz niskim zaangażowaniem organizacyjnym,
- to jak pracownik ocenia łatwość zmiany pracy - związane jest to z dostępnością na rynku pracy posad interesujących pracownika.

W ciągu ostatnich 50 lat powstało wiele modeli prognozujących fluktuację i większość z nich korzysta mniej lub bardziej bezpośrednio z cech wyróżnionych przez Marcha i Simona. Badania empiryczne prowadzone w tym czasie potwierdziły, że te cechy mają znaczący wpływ na fluktuację - jednak nie oddają w pełni istoty fluktuacji. Przy pracy nad tymi modelami wyszczególniono wiele drugorzędnych cech wpływających na fluktuację [Steel and Lounsbury, 2009]. Cechy te zostały przedstawione w tabelach 1.3 i 1.4.

Tabela 1.3: Najczęściej pojawiające się cechy w modelach fluktuacji

Kategoria	Cechy
cechy osobiste	<ul style="list-style-type: none">• osobowość,• wyznawane wartości,• wiek,• staż pracy,• wiedza,• doświadczenie,• profesjonalizm,• odpowiedzialność wobec rodziny;

Źródło: opracowanie własne na podstawie: [Steel and Lounsbury, 2009]

Tabela 1.4: Najczęściej pojawiające się cechy w modelach fluktuacji (ciąg dalszy)

Kategoria	Cechy
cechy stanowiska pracy	<ul style="list-style-type: none"> • postrzeganie pracy, • skomplikowanie pracy, • oczekiwania wobec wykonywanej pracy, • wynagrodzenie i benefity, • koszt zmiany pracy, • stres, • dopasowanie wykonywanej pracy do oczekiwań pracownika, • rozmiar firmy;
mechanizmy zmiany stanowiska	<ul style="list-style-type: none"> • chęć zmiany, • oczekiwania względem przyszłej pracy, • wysiłek potrzebny do zmiany bieżącej sytuacji, • możliwość przejścia do firmy powiązanej lub innego oddziału firmy, • możliwość awansu lub degradacji, • alternatywne sposoby opuszczenia pracy;
konsekwencje opuszczenia lub pozostania w firmie	<ul style="list-style-type: none"> • konsekwencje pozapracowe, • wydajność pracy;
mechanizmy wpływające na proces decyzyjny	<ul style="list-style-type: none"> • zdarzenia nieprzewidziane, • myśli o odejściu z firmy;

Źródło: opracowanie własne na podstawie: [Steel and Lounsbury, 2009]

1.3.1. Zadowolenie i satysfakcja pracowników

Satysfakcję z pracy najprościej można zdefiniować jako "pozytywne i negatywne uczucia oraz postawy wobec wykonywanej pracy"[Schultz and Schultz, 2002]. Często w literaturze zamiennie do określenia "satysfakcja z pracy" występuje "zadowolenie z pracy", jednak czasem wskazywane jest rozróżnienie ze względu na czas trwania [Sowińska, 2014]:

- zadowolenie jest uczuciem chwilowym,
- satysfakcja jest odczuwana po długim czasie odczuwania zadowolenia.

Na satysfakcję z pracy wpływają głównie 3 grupy czynników [Schultz and Schultz, 2002]:

- czynniki związane z pracą - np. zakres obowiązków, lokalizacja biura, relacje z współpracownikami,

- czynniki indywidualne - np. staż pracy, wiek, zdrowie, zależności rodzinne,
- motywacja i aspiracje.

W tabelach 1.5 i 1.6 omówiono poszczególne czynniki wpływające na zadowolenie z pracy.

Tabela 1.5: Czynniki organizacyjne wpływające na zadowolenie pracowników

Czynnik	Opis
doświadczenia zawodowe	Satysfakcja spada u pracowników z kilkuletnim doświadczeniem, jeśli w swojej pracy nie są odpowiednio nagradzani i motywowani.
wykorzystanie umiejętności	Pracownicy, którzy w pracy mogą korzystać z wcześniej zdobytych umiejętności są zwykle bardziej zadowoleni.
odpowiedniość pracy	Pracownicy odczuwają większą satysfakcję jeśli wymagania stanowiska pracy pokrywają się z ich zdolnościami i aspiracjami.
status pracy	Status społeczny stanowiska istotnie wpływa na poziom zadowolenia pracownika.

Źródło: opracowanie własne na podstawie: [Schultz and Schultz, 2002]

Tabela 1.6: Czynniki osobowe wpływające na zadowolenie pracowników

Czynnik	Opis
wiek	Statystycznie wraz z wiekiem rośnie zadowolenie z pracy.
płeć	Nie wykazano bezpośredniej korelacji płci z zadowoleniem z pracy, natomiast pośrednio na zadowolenie przedstawicieli danej płci może wpływać różnica w poziomach wynagrodzeń kobiet i mężczyzn na takich samych stanowiskach.
rasa	Statystycznie przedstawiciele mniejszości etnicznych częściej są przyjmowani na stanowiska gorsze lub gorzej płatne niż przedstawiciele lokalnej większości etnicznej.
zdolności poznawcze	Osoby o wysokich zdolnościach poznawczych częściej odczuwają brak satysfakcji jeśli praca nie angażuje ich intelektualnie.
cechy osobowości	Badania wskazują na korelację pomiędzy stabilnością emocjonalną, a zadowoleniem z pracy.

Źródło: opracowanie własne na podstawie: [Schultz and Schultz, 2002]

W pierwszych próbach badań nad zagadnieniem zadowolenia z pracy, zadowolenie definiowano jako przestrzeń jednowymiarową rozpinającą się od niezadowolenia do zadowolenia [Sowińska, 2014]. Oznaczało to, że występowanie określonego czynnika mogło skutkować zadowoleniem, a jego brak - niezadowoleniem. Nowe spojrzenie na tę kwestię wprowadził F. Herzberg definiując teorię dwuczynnikową, która zakłada, że czynniki wpływające na zadowolenie mogą być rozpatrywane niezależnie od czynników wpływających na niezadowolenie. Teoria ta jednak została zdefiniowana w sposób uznany za nieprecyzyjny, przez co spotkała się z szeroką falą krytyki [Sowińska, 2014].

Rozwinięciem teorii Herzberga jest teoria trychotomii czynników motywacji, którą zaproponował L. Koziół [Koziół, 2011]. Według tej teorii można wyszczególnić 3 kategorie czynników wpływających na zadowolenie: motywatory, demotywatory i czynniki higieny. Czynniki te zostały omówione w tabeli 1.7.

Tabela 1.7: Wpływ czynników motywacji na zadowolenie pracowników

Kategoria	Wpływ	Przykłady
motywatory	ich występowanie wpływa na zadowolenie	<ul style="list-style-type: none"> • odpowiedzialność, • uznanie, • możliwość rozwoju osobistego, • osiągnięcia, • awans;
demotywatory	ich występowanie wpływa na niezadowolenie	<ul style="list-style-type: none"> • niejasne oczekiwania, • zmuszanie pracowników do działań sprzecznych z etyką zawodową, • praca ponad siły, • nierówny podział obowiązków, • brak szkoleń, • niedoceniać pracowników z wieloletnim stażem, • nadmierna presja, • strach przed utratą pracy;
czynniki higieny	ich niewystępowanie wpływa na niezadowolenie	<ul style="list-style-type: none"> • wynagrodzenia, • zajmowane stanowisko, • warunki pracy, • nadzór techniczny, • bezpieczeństwo pracy, • polityka firmy, • świadczenia socjalne;

Źródło: opracowanie własne na podstawie: [Koziół, 2011]

1.3.2. Zaangażowanie organizacyjne pracowników

Zaangażowanie organizacyjne jest w literaturze definiowane na wiele różnych sposobów. Istniejące definicje można podzielić na 3 koncepcje [Juchnowicz, 2010]:

1. zaangażowanie jako postawa pracownika,
2. zaangażowanie przejawiane poprzez zachowanie pracownika,
3. zaangażowanie jako wymiana świadczeń na poziomie relacji firma – pracownik.

M. Juchnowicz [Juchnowicz, 2010] pokazuje, że pierwsza koncepcja traktująca zaangażowanie organizacyjne jako postawę pracownika najszerzej opisuje zjawisko zaangażowania, gdyż zawiera 3 czynniki kształtujące zaangażowanie: poznawczy, emocjonalny i behawioralny. Czynniki te zostały omówione w tabeli 1.8.

Tabela 1.8: Czynniki kształtujące zaangażowanie

Czynnik	Akcja	Opis
poznawczy	myślenie	Wiedza o organizacji jest podstawą zaangażowania.
emocjonalny	odczuwanie	Konieczny jest stosunek emocjonalny w stosunku do sposobu działania organizacji oraz jej wartości i celów.
behawioralny	działanie	Do zaangażowania potrzebna jest chęć podjęcia działań względem organizacji.

Źródło: opracowanie własne na podstawie: [Juchnowicz, 2010]

Na tej podstawie zaangażowanie można zdefiniować jako "intelektualne i emocjonalne oddanie organizacji" [Juchnowicz, 2010]. Dla tak zdefiniowanego zaangażowania można wyszczególnić 4 podstawowe cechy postawy pracownika zaangażowanego [Juchnowicz, 2010]:

- stabilizacja - pracownikowi zależy żeby należeć do firmy,
- identyfikacja - pracownik wierzy w misję, wartości i cele organizacji i chce uczestniczyć w ich realizacji,
- pasja - wykonywana praca jest zgodna z cechami, aspiracjami i zainteresowaniami pracownika,
- efektywne działanie na rzecz pracodawcy - pracownik dąży do wykorzystania pełni swojego potencjału, aktywnie dzieli się wiedzą.

1.4. Koszt fluktuacji pracowników

Z perspektywy menadżerskiej wydawać by się mogło, że głównym kosztem związanym z fluktuacją pracowników jest koszt prowadzenia rekrutacji przez dział HR. Problem jest jednak zdecydowanie bardziej złożony. Edwards i Philips [Philips and Edwards, 2009] pokazują, że - w zależności od stanowiska i wymaganych na nim kompetencji - całkowity koszt związany z odejściem pracownika i zatrudnieniem w jego miejsce nowego oscyluje od 30 do nawet 400 procent rocznego wynagrodzenia na danym stanowisku. Wyszczególnione przez nich koszty przedstawiono w tabeli 1.9.

Tabela 1.9: Typy kosztów związanych z fluktuacją

Typ	Cechy
koszty związane z odejściem starego pracownika	<ul style="list-style-type: none">• przekazanie wiedzy innym pracownikom,• po podjęciu decyzji o odejściu z pracy, odchodzący pracownik może być mniej zaangażowany w wykonywane obowiązki;
koszt prowadzenia rekrutacji	<ul style="list-style-type: none">• koszt związany z publikowaniem ogłoszeń o pracę,• selekcja aplikantów,• prowadzenie rozmów rekrutacyjnych,• koszt operacyjny zakontraktowania nowego pracownika - związany między innymi z procesowaniem umowy czy skierowaniem na badania lekarskie;
koszty związane z wdrożeniem nowego pracownika	<ul style="list-style-type: none">• czas nowego pracownika potrzebny na zapoznanie się z obowiązkami i wdrożenie na nowe stanowisko pracy (w zależności od stanowiska może to trwać nawet kilka miesięcy),• czas doświadczonych pracowników potrzebny na wdrażanie nowego pracownika,• w zależności od specyfiki stanowiska zakup odpowiedniego sprzętu dla pracownika, np. ubrań roboczych czy laptopa;
szacowane utracone korzyści	<ul style="list-style-type: none">• potencjalnie większe obciążenie pracowników którzy pozostali w firmie (przy brakach kadrowych),• utrata części wiedzy odchodzącego pracownika;
inne koszty powiązane	<ul style="list-style-type: none">• możliwe pogorszenie relacji z klientem, co może negatywnie wpłynąć na sprzedaż,• odejście pracownika może zachęcić innych do rozważenia zmiany pracy;

Źródło: opracowanie własne na podstawie: [Philips and Edwards, 2009]

1.5. Sposoby na retencję pracowników

Przy rozważaniach na temat zwiększenia retencji pracowników warto zwrócić uwagę czynniki fluktuacji z modelu Marcha i Simona, które zostały opisane w sekcji 1.3. Wynika z nich, że przedsiębiorstwa powinny dążyć do zwiększenia zadowolenia pracowników oraz do zwiększenia ich zaangażowania organizacyjnego.

Postulaty te są ujęte w założeniach koncepcji "employer branding" [Spychała et al., 2019]. W literaturze koncepcja ta jest definiowana jako całokształt działań danej organizacji ukierunkowanych do pracowników obecnych, byłych oraz potencjalnych, które służą kreowaniu wizerunku firmy jako atrakcyjnego miejsca pracy, jednocześnie tworząc odpowiednie środowisko do realizacji jej celów biznesowych [Kozłowski, 2012].

Akcje związane z koncepcją employer branding są podejmowane na 2 poziomach relacji firmy [Spychała et al., 2019]:

1. z obecnymi pracownikami,
2. z byłymi pracownikami oraz potencjalnymi przyszłymi pracownikami.

W celu zarządzania tymi relacjami firmy podejmują akcje marketingowe - w przypadku relacji z obecnymi pracownikami jest to marketing wewnętrzny, natomiast w przypadku byłych i potencjalnych pracowników marketing zewnętrzny. Marketing wewnętrzny opiera się na relacjach pracowników z współpracownikami i przełożonymi. Aby zwiększyć przywiązanie organizacyjne i utożsamianie się pracowników z celami przedsiębiorstwa przełożeni powinni podejmować następujące akcje [Spychała et al., 2019]:

- transparentne przedstawianie pracownikom powodów kryjących się za działaniami firmy,
- wykazywanie empatii wobec pracowników,
- promowanie indywidualnego rozwoju pracowników,
- umożliwianie pracownikom pracy dającej poczucie dawania istotnego wkładu w działanie firmy.

Marketing zewnętrzny nakierowany jest na zwiększenie rozpoznawalności przedsiębiorstwa i budowaniu pozytywnych skojarzeń z nim związanych. Z tego powodu wiele firm angażuje się w akcje związane z tzw. "społeczną odpowiedzialnością biznesu". Takie akcje mogą być związane np. z przekazywaniem środków na cele charytatywne czy promowaniem ograniczania emisji spalin. Inną formą akcji marketingowych zwiększających zasięg firmy wśród potencjalnych pracowników jest prowadzenie prelekcji dla studentów kierunków związanych z działalnością firmy. Należy podkreślić, że marketing zewnętrzny opiera się w dużej mierze na opiniach pracowników, dlatego ważne żeby wpierw dopracowany był marketing wewnętrzny [Spychała et al., 2019].

Skuteczny employer branding opiera się na następujących założeniach [Spychała et al., 2019]:

- pracownicy traktowani są jako klienci przedsiębiorstwa na równi z konsumentami, czy ogólnie - klientami zewnętrznymi,
- pracownik i pracodawca są równie istotni żeby możliwe było zrealizowanie celów organizacji,
- pracownicy tworzą wewnętrzny rynek pracy - firma powinna umożliwiać pracownikom objęcie nowych wakatów w celu umożliwienia pracownikom rozwoju i zaspokojenia ich ambicji,
- postrzeganie firmy zależy przede wszystkim od pracowników.

2. Charakterystyka branży IT

2.1. Ogólna charakterystyka

Branża IT jest stosunkowo młoda. Koncepcja stanowiąca podstawę architektury wszystkich współczesnych komputerów - a ogólniej w zasadzie niemal wszystkich urządzeń korzystających z procesora - została opracowana w 1945 roku przez Johna von Neumanna, Johna Mauchly'ego i Johna Eckerta [Riya, 2020]. Model komputera implementującego tę architekturę nazwali "przykładową maszyną cyfrową". Na przestrzeni lat wraz z postępem technologicznym następował stopniowy wzrost znaczenia technologii informatycznych, jednak prawdziwy rozkwit branży nastąpił po uruchomieniu w 1993 roku ogólnosiwiatowej sieci komputerowej (ang. World Wide Web) [Rangarajan and Tiwari, 2014], którą potocznie nazywamy "Internetem".

Rok 1993 można uznać za symboliczną datę kiedy rynek pracy IT zaczął przybierać znaną obecnie postać, jednak warto wyszczególnić jeszcze kilka ważnych dat z historii rozwoju branży:

- 2002 - Amazon Web Services - pierwsza implementacja chmury obliczeniowej [Alabbas, 2017],
- 2004 - Web 2.0 - ewolucja sposobu kreowania treści w internecie w sposób bardziej angażujący użytkowników [Yakovlev, 2007],
- 2007 - iPhone - początek ery urządzeń mobilnych [D. Pierce, 2018].

W Polsce branża IT stanowi bardzo istotny sektor gospodarki. W 2021 roku działalność w branży IT stanowiła w przybliżeniu 8% PKB Polski [PAIH, 2021]. Według raportu HackerRank Polscy programiści znajdują się w światowej czołówce pod względem umiejętności, plasując się na trzecim miejscu, zaraz za Chinami i Rosją [Tripathi, 2021]. Wysokie kwalifikacje polskich programistów są jedną z przyczyn dlaczego wiele zagranicznych firm decyduje się na outsourcing projektów do polskich firm informatycznych. Wśród innych powodów wybierania Polski do procesu outsourcingu wyszczególnić można [Software Mind, 2021]:

- wysoki stopień znajomości języka angielskiego,
- przynależność Polski do Unii Europejskiej zapewnia zgodność z dyrektywami RODO dotyczącymi nie przetwarzania danych użytkowników poza terenem Unii Europejskiej,

- Polska jest zbliżona kulturowo i lokalizacyjnie do zachodnich zagranicznych inwestorów z krajów takich jak Niemcy czy Wielka Brytania, a nawet Stany Zjednoczone (gdyż różnica czasu między Polską, a wschodnim wybrzeżem Stanów Zjednoczonych wynosi zaledwie 6 godzin).

Duże zainteresowanie zagranicznych inwestorów polskim rynkiem IT oraz ogólny światowy niedobór wykwalifikowanych programistów [Daxx, 2022] kreują sytuację, w której pracodawcy prześcigają się w przygotowywaniu coraz atrakcyjniejszych ofert pracy zawierających szereg benefitów pracowniczych [Rojewska, 2020]. Tak powstała subiektywnie postrzegana łatwość zmiany pracy może skutkować zwiększeniem fluktuacji pracowników. Według raportu LinkedIn [Booz, 2018] w 2017 roku branża IT odnotowała najwyższy wskaźnik fluktuacji wynoszący 13.2%, gdzie globalna średnia dla innych zawodów wynosiła zaledwie 10.9%. Szczególnie wysokie wskaźniki fluktuacji odnotowano dla następujących specjalizacji w branży IT:

- 23.3% - projektant doświadczeń użytkownika,
- 21.7% - analityk danych,
- 21.7% - programista systemów wbudowanych.

Jak wspomniano w rozdziale 1.2, wysoki poziom fluktuacji niekoniecznie musi oznaczać problemy dla firmy, jeśli dostępnych jest wielu potencjalnych pracowników, a koszty rekrutacji są niskie. Te warunki nie są jednak spełnione w przypadku branży IT, gdzie wydatki związane z zatrudnieniem nowego pracownika po odejściu starego mogą sięgać nawet rzędu 200 tys. polskich złotych [Winter, 2016].

2.2. Zadowolenie pracowników

Jak pokazano w rozdziale 1.3 wiele z klasycznych teorii badających czynniki wpływające na fluktuację wyróżnia zadowolenie z pracy jako jeden z kluczowych czynników. W przypadku branży IT zadowolenie z pracy wpływa również znacząco na produktywność programistów i przede wszystkim na jakość tworzonego oprogramowania [Graziotin et al., 2018]. Według Roberta C. Martina - autorytetu w dziedzinie stosowania dobrych wzorców w programowaniu - jakość kodu ma kluczowe znaczenie dla przyszłości tworzonego oprogramowania, a także dla przyszłości firmy. Martin wyszczególnił następujące skutki zaniechania dbałości o jakość oprogramowania [Martin, 2014]:

- wydłużenie czasu dostarczania nowych funkcjonalności oprogramowania,
- częstsze pojawianie się awarii oprogramowania,
- spadek produktywności programistów,
- spadek motywacji programistów,
- konieczność poświęcania większej ilości czasu na analizowanie istniejącego kodu,

- zwiększenie kosztów tworzenia oprogramowania,
- wydłużenie czasu trwania projektu,
- możliwa konieczność napisania całego kodu od nowa.

Często stosowanym sposobem zwiększenia zadowolenia pracowników jest ograniczanie niezadowolenia. Czynniki wpływające na niezadowolenie pracowników w branży IT można podzielić na 3 podstawowe kategorie [Graziotin et al., 2017]:

1. wewnętrzne związane z osobą pracownika
2. zewnętrzne związane z organizacją firmy
3. zewnętrzne związane z oprogramowaniem

W tabeli 2.1 przedstawiono częste przyczyny niezadowolenia pracowników z uwzględnieniem powyższych kategorii.

Tabela 2.1: Przyczyny niezadowolenia pracowników w branży IT

Kategoria	Przyczyny
czynniki wewnętrzne związane z osobą pracownika	<ul style="list-style-type: none"> • fiksacja na punkcie rozwiązywania problemów, • syndrom impostora - poczucie nie posiadania wystarczających kompetencji do pełnionej roli, • problemy osobiste;
czynniki zewnętrzne związane z organizacją firmy	<ul style="list-style-type: none"> • presja czasu, • monotonne lub powtarzające się zadania, • podejmowanie złych decyzji (zwłaszcza technicznych) na wyższych szczeblach w organizacji;
czynniki zewnętrzne związane z oprogramowaniem	<ul style="list-style-type: none"> • niska jakość kodu i złe praktyki programistyczne, • praca z kodem, który nie działa z niewiadomych przyczyn, • narzucone ograniczenia dotyczące narzędzi wykorzystywanych do programowania (takiej jak język programowania czy środowisko);

Źródło: opracowanie własne na podstawie: [Graziotin et al., 2017]

2.3. Wpływ fluktuacji na utratę wiedzy w projektach informatycznych

Jednym z aspektów związanych z fluktuacją pracowników w branży IT jest utrata wiedzy w projektach informatycznych. Można wyszczególnić kilka podstawowych przyczyn utraty wiedzy [Nesh, 2021]:

- nieudokumentowane dane - wiele decyzji projektowych podejmowanych jest w oparciu o doświadczenie pracownika, testowanie wielu rozwiązań i wyciąganie wniosków z sukcesów i porażek - powstaje w ten sposób rodzaj "wiedzy plemiennej", która może być w pełni zrozumiała dla obecnych członków zespołu, ale nie jest w całości przekazywana nowym pracownikom,
- udokumentowane, ale nieuporządkowane dane - wiedza projektowa często jest spisywana w formie różnorodnych dokumentów, które funkcjonują w obiegu między pracownikami, jednak w przypadku braku uporządkowanej struktury bazy wiedzy projektowej dokumenty takie mogą łatwo zaginąć w przypadku odejścia pracownika,
- relacje międzyludzkie - decyzje projektowe mogą być podejmowane na podstawie wymagań klienta; w wielu projektach informatycznych wymagania klienta nie są udokumentowane w formalnych dokumentach przygotowanych przez analityków biznesowych, ale są przekazywane w formie ustnej lub mailowej, z tego powodu w przypadku odejścia pracownika wiedza wynikająca z relacji z klientem jest często trudna do przekazania.

Donadelli [Donadelli, 2015] wyszczególnia dodatkowe czynniki związane z wysoką fluktuacją w projektach informatycznych:

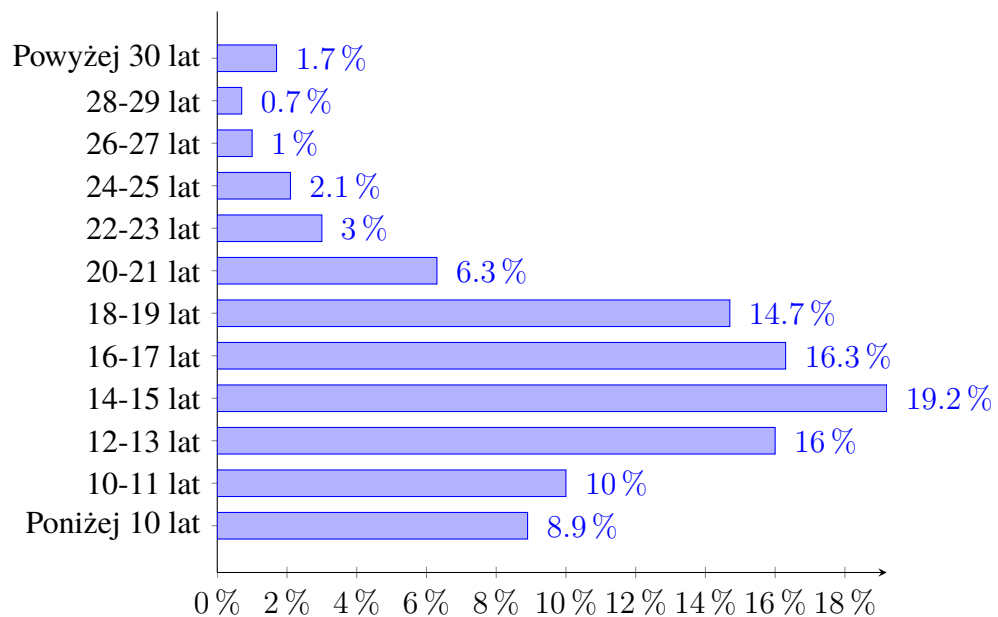
- spadek produktywności firmy,
- obniżenie jakości wytwarzanego kodu,
- zwiększenie liczby "porzuconych" plików (tj. takich, które od dawna nie były edytowane przez żadnego z członków zespołu),
- ograniczenie rozprzestrzeniania wiedzy w projekcie.

Na szczególną uwagę zasługuje ostatni wymieniony punkt. W zwinnych metodykach zarządzania projektami stosuje się "wskaźnik autobusu", który dosłownie oznacza wskazanie ilu członków zespołu musiałoby zostać potrąconych przez autobus, aby w znaczący sposób wpłynęło to negatywnie na przyszłość prowadzonego projektu [Berteig, 2005]. Wskaźnik autobusu zwykle jest szczególnie istotny w projektach prowadzonych przez małe zespoły, gdzie każdy z członków zespołu zajmuje się pracą nad odrębną częścią systemu. W takich projektach odejście pojedynczego pracownika może doprowadzić do utraty znaczącej części wiedzy. Sytuacja zwykle inaczej się prezentuje w przypadku dużych projektów, gdzie wiedza jest znacząco bardziej rozproszona i odejście pracownika nie wpływa istotnie na dalsze prowadzenie projektu.

2.4. Charakterystyka pracowników w branży IT

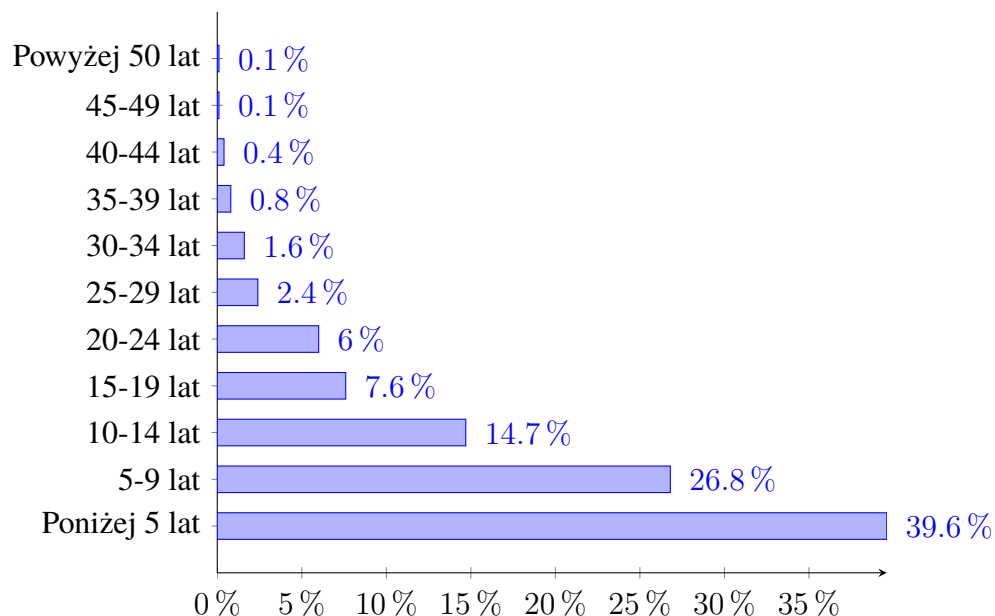
W niniejszym podrozdziale przedstawiona zostanie demografia pracowników z branży informatycznej. Opracowanie zostało przygotowane na podstawie danych z ankiety deweloperskiej przeprowadzonej przez StackOverflow w 2020 roku, w której wzięło udział około 65 tysięcy programistów z całego świata [Stack Overflow, 2020].

Na rysunku 2.1 pokazano w jakim wieku respondenci napisali swoją pierwszą linijkę kodu. Na wykresie widać, że przeważająca większość respondentów - aż 54% - miała pierwszy kontakt z programowaniem już przed 16 rokiem życia. Jednakże, wczesny kontakt z programowaniem nie przekłada się wśród respondentów na długi staż zawodowy, gdyż jak widać na rysunku 2.2 około 65% badanych osób ma mniej niż 10 lat doświadczenia zawodowego. Można domniemywać, że w związku ze wzrostem znaczenia branży IT w ciągu ostatnich kilkudziesięciu lat oraz wzrostem zapotrzebowania na specjalistów IT na rynku pracy wzrasta zainteresowanie branżą IT wśród młodszych pokoleń. Obrazować może to fakt, że aż 70% respondentów to osoby w wieku poniżej 35 lat, co pokazuje rysunek 2.3.



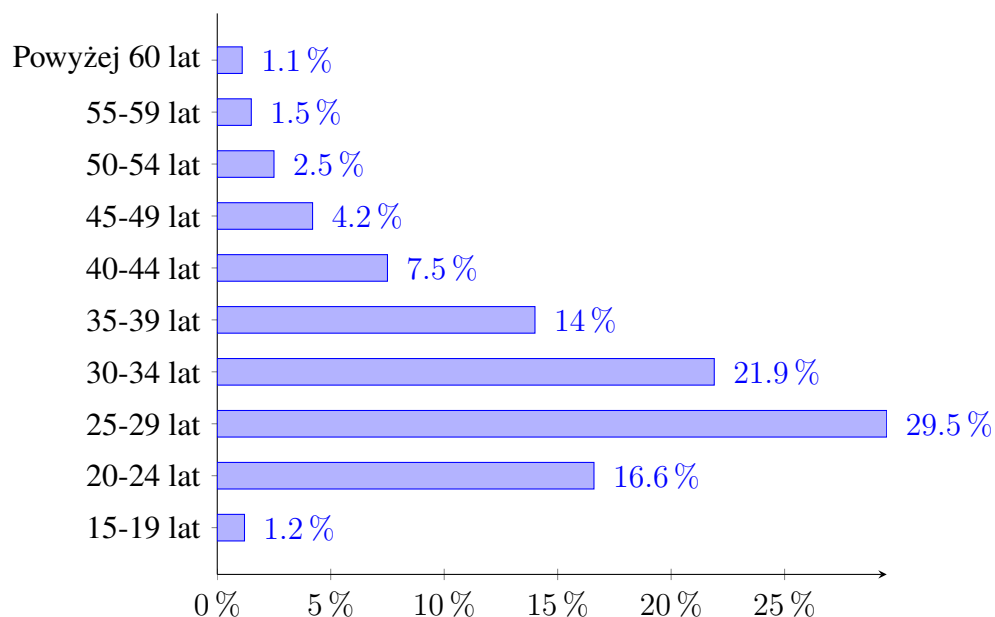
Rysunek 2.1: Rozkład wieku, w którym respondenci ankiety StackOverflow napisali pierwszą linijkę kodu

Źródło: opracowanie własne na podstawie: [Stack Overflow, 2020]



Rysunek 2.2: Rozkład stażu pracy respondentów ankiety StackOverflow w latach

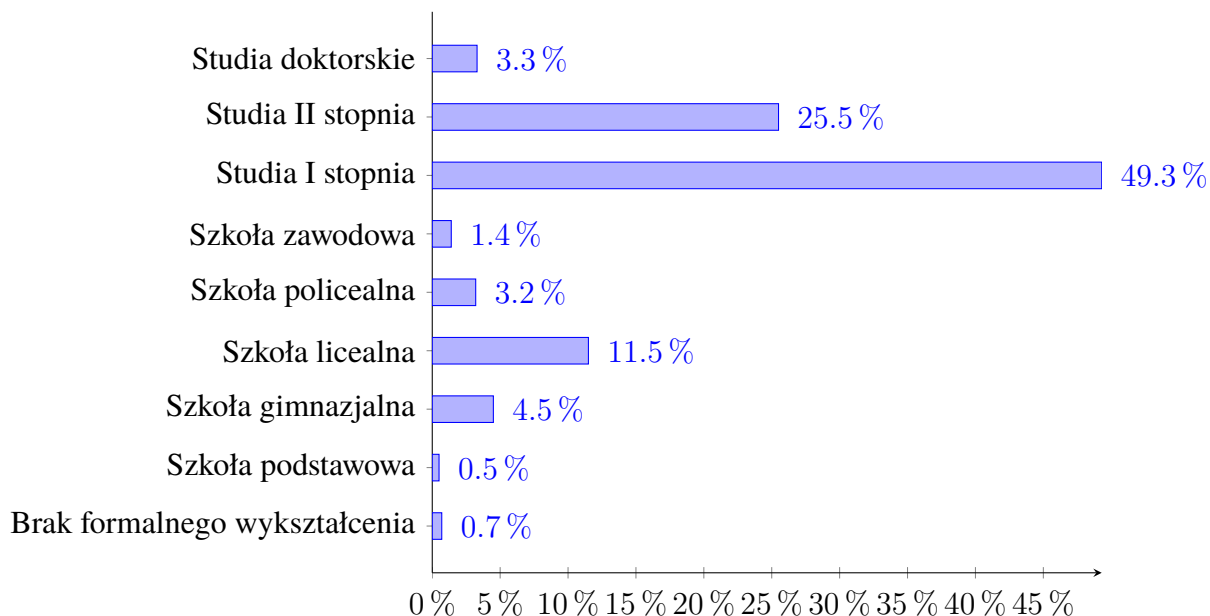
Źródło: opracowanie własne na podstawie: [Stack Overflow, 2020]



Rysunek 2.3: Rozkład wieku respondentów ankiety StackOverflow

Źródło: opracowanie własne na podstawie: [Stack Overflow, 2020]

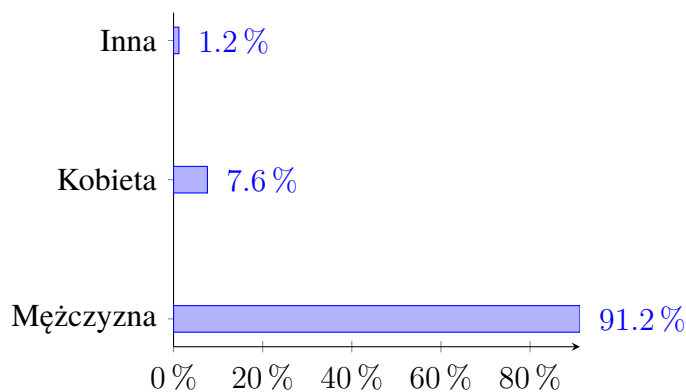
Na rysunku 2.4 pokazano jakie najwyższe wykształcenie uzyskali respondenci ankiety. Ciekawe jest, że mimo iż firmy z branży IT zwykle nie wymagają od pracowników formalnego wykształcenia, aż 75% badanych osób ma ukończone co najmniej studia I stopnia.



Rysunek 2.4: Rozkład wykształcenia respondentów ankiety StackOverflow

Źródło: opracowanie własne na podstawie: [Stack Overflow, 2020]

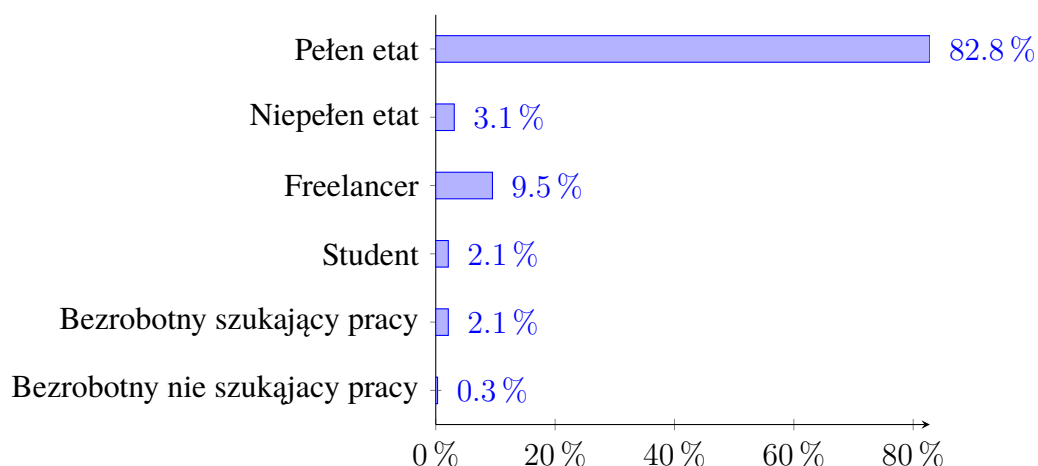
Na rysunku 2.5 przedstawiono rozkład płci respondentów i jak widać branża IT jest silnie zdominowana przez mężczyzn (ponad 90% respondentów).



Rysunek 2.5: Rozkład płci respondentów ankiety StackOverflow

Źródło: opracowanie własne na podstawie: [Stack Overflow, 2020]

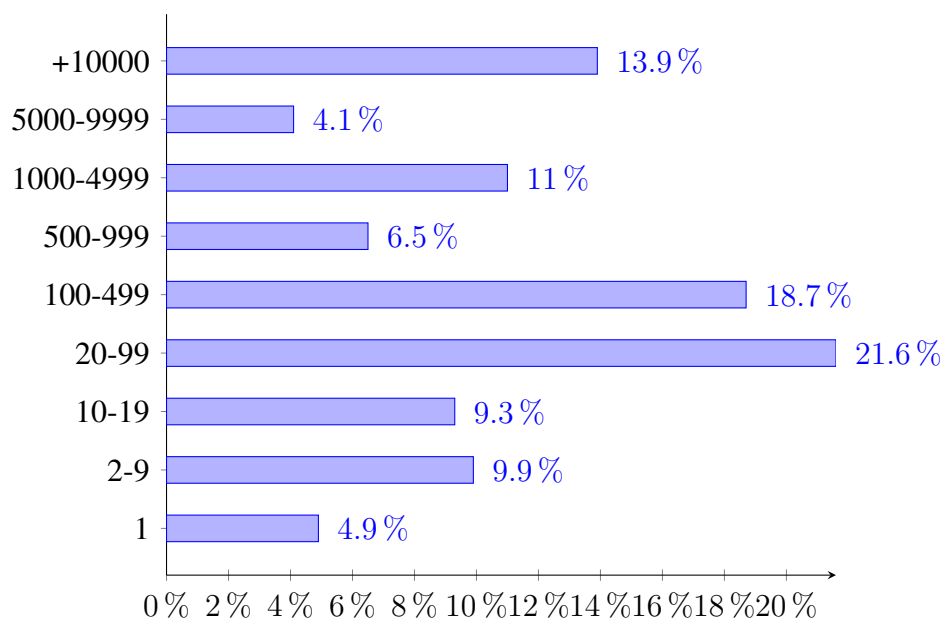
Na rysunku 2.6 przedstawiono rozkład form zatrudnienia w oparciu o które pracują respondenci. Niemal 80% respondentów zdecydowało się na podjęcie stosunku pracy w pełnym wymiarze godzin ze swoimi pracodawcami.



Rysunek 2.6: Rozkład rodzaju zatrudnienia respondentów ankiety StackOverflow

Źródło: opracowanie własne na podstawie: [Stack Overflow, 2020]

Rozkład rozmiaru firm w których pracują respondenci został przedstawiony na rysunku 2.7. Najwięcej jest średnich firm (od 20 do 500 pracowników) i dużych korporacji powyżej 10 tysięcy pracowników.

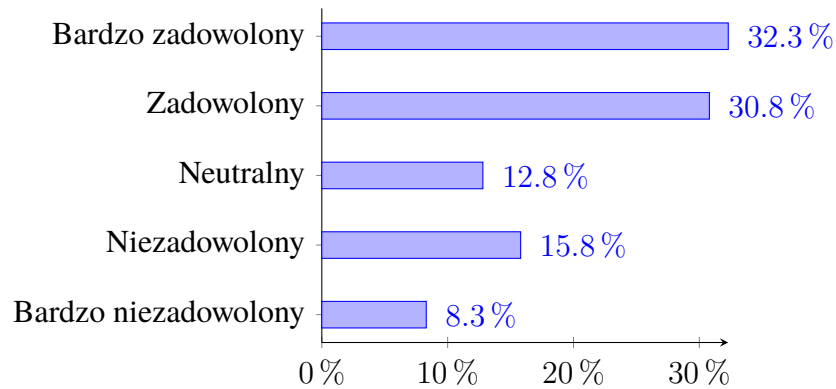


Rysunek 2.7: Rozkład rozmiaru firm, w których pracują respondenci ankiety StackOverflow

Źródło: opracowanie własne na podstawie: [Stack Overflow, 2020]

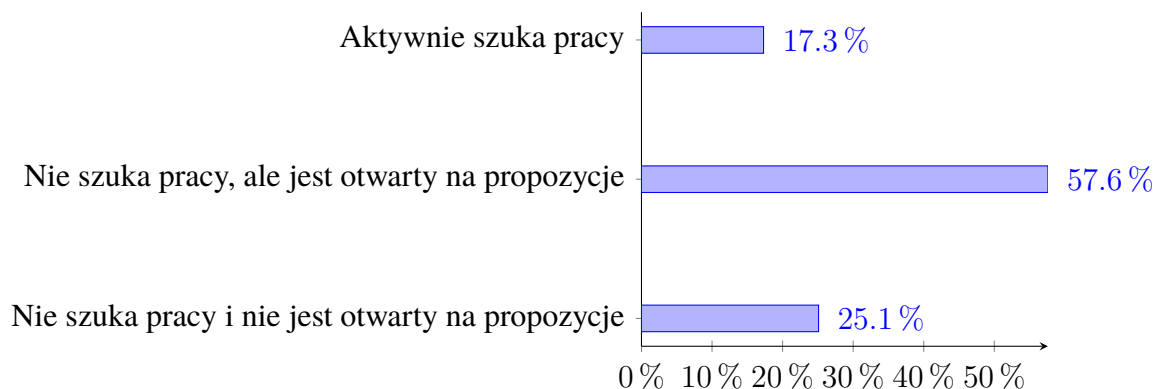
Rysunek 2.8 pokazuje, że około 63% respondentów jest zadowolona lub bardzo zadowolona z wykonywanej pracy, a jedynie około 24% respondentów jest niezadowolona lub bardzo niezadowolona z pracy. Porównując to z rysunkiem 2.9 obrazującym rozkład statusu szukania pracy widać korelację odsetka niezadowolonych respondentów z odsetkiem respondentów

aktywnie poszukujących pracy. Ciekawe natomiast jest, że aż 57,6% respondentów deklaruje, że mimo iż nie szuka aktywnie pracy, to jest otwarta na interesujące propozycje. Pokazuje to, że nawet zadowoleni z pracy pracownicy mogą być otwarci na zmianę pracy jeśli nowa oferta spotka się z ich zainteresowaniem.



Rysunek 2.8: Rozkład zadowolenia z pracy respondentów ankiety StackOverflow

Źródło: opracowanie własne na podstawie: [Stack Overflow, 2020]



Rysunek 2.9: Rozkład statusu szukania pracy wśród respondentów ankiety StackOverflow

Źródło: opracowanie własne na podstawie: [Stack Overflow, 2020]

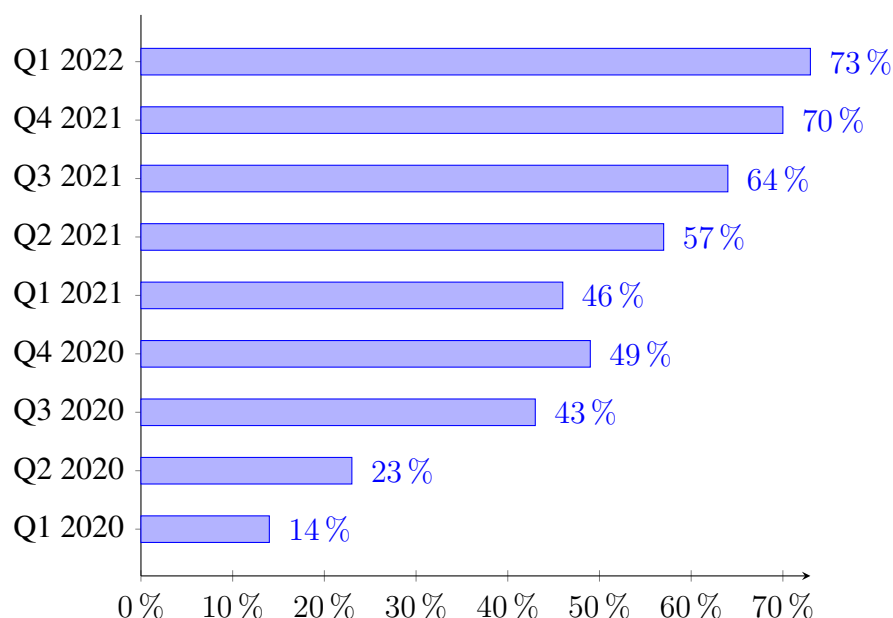
2.5. Trendy na rynku pracy

Branża informatyczna jest bardzo dynamicznie rozwijającą się branżą. Wiele trendów obserwowanych na rynku pracy związanych jest z rozwojem technologii, w tym przede wszystkim sztucznej inteligencji, urządzeń mobilnych czy wirtualnej rzeczywistości. Wraz z nowymi technologiami pojawia się większa potrzeba specjalizacji, w efekcie czego stanowisko "inżyniera oprogramowania" jest zastępowane bardziej konkretnymi stanowiskami takimi jak "programista aplikacji mobilnych na system iOS", czy "programista aplikacji webowych w technologii Java" [Sonmez, 2015].

Od 2020 roku sytuacja związana z pandemią koronawirusa miała istotny wpływ na wyłonienie się nowych trendów w branży IT. Na początku pandemii wiele firm zdecydowało się ograniczyć lub wstrzymać rekrutację, a czasem nawet zwolnić niektórych pracowników, jednak takie działania okazały się nierozsądne, gdyż wbrew obawom zapotrzebowanie na usługi informatyczne nie zmalało, a wręcz wzrosło w trakcie pandemii. Przez nałożone przez rząd ograniczenia firmy zostały zmuszone do oferowania w większym zakresie usług zdalnie, co przyczyniło się do przyspieszenia transformacji cyfrowej [Kalbarczyk, 2021].

Sytuacja pandemiczna pokazała również, że branża IT jest w stanie bardzo dobrze dostosować się do pracy zdalnej i nawet przeprowadzanie rekrutacji czy wdrażania nowych pracowników w formie zdalnej stało się branżową normą. Mniejsze znaczenie zaczęła mieć lokalizacja siedziby firmy, stąd też pracownicy coraz chętniej decydują się na pracę dla zagranicznego pracodawcy, gdzie często mogą liczyć na korzystniejsze warunki finansowe niż w przypadku pracy dla polskiej firmy [Kalbarczyk, 2021].

Na rysunku 2.10 pokazano udział ofert pracy zdalnej w stosunku do wszystkich ofert pracy w kolejnych kwartałach od 2020 do 2022 roku. W ciągu zaledwie 2 lat liczba ta wzrosła od 14% aż do 73%. Jednak liczby te nie są jedynie odzwierciedleniem sytuacji pandemicznej. Według badań przeprowadzonych we wrześniu 2021 roku przez firmę DataArt, aż 64% ankietowanych programistów byłoby gotowych złożyć wypowiedzenie, jeśli pracodawca postanowi narzucić im powrót do pracy stacjonarnej z biura [Pasterczyk, 2021].



Rysunek 2.10: Udział ofert pracy zdalnej w stosunku do wszystkich ofert pracy

Źródło: opracowanie własne na podstawie: [Ułan, 2022]

Pandemia koronawirusa nie jest jedynym istotnym w skali świata wydarzeniem mającym wpływ na sytuację rynku pracy branży IT. Z powodu rosyjskiej inwazji na Ukrainę w 2022 roku wiele firm postanowiło lub zostało zmuszonych do zawieszenia działalności swoich oddziałów w Ukrainie, Rosji i Białorusi [Błaszczak, 2022]/ Kraje te były w ostatnich latach popularnym kierunkiem outsourcingu usług ze względu na stosunkowo niski koszt zatrudnienia pracowników w porównaniu do krajów zachodnich oraz dużą liczbę wysoko wykwalifikowanych specjalistów. Ekspert z firmy Everest Group oceniają, że ograniczenia działalności we wspomnianych krajach może spowodować wzrost zainteresowania Polską, Węgrami i Rumunią jako kierunkiem outsourcingu [Overby, 2022]. W Polsce częściowo może to znaleźć odzwierciedlenie w liczbie ofert pracy. Według analizy Adecco Poland w lutym 2022 roku liczba ogłoszeń była prawie o połowę większa w stosunku do grudnia 2021 roku [Błaszczak, 2022].

3. Optymalizacja procesu rekrutacji w branży IT z wykorzystaniem uczenia maszynowego

3.1. Metody badawcze

Biorąc pod uwagę rosnące zapotrzebowanie na usługi informatyczne, deficyt wykwalifikowanych pracowników [Daxx, 2022] oraz wysoki koszt rekrutacji nowych pracowników [Winter, 2016] istotne jest poszukiwanie sposobów na zwiększenie retencji pracowników IT. Badania będące przedmiotem niniejszej pracy mają na celu analizę cech skorelowanych z zadowoleniem pracowników branży IT i z ich chęcią do zmiany pracy.

Celem badania jest określenie czy i jeśli tak to jakie cechy firmy i pracownika mogą mieć istotny wpływ na zadowolenie pracownika oraz czy można wyszczególnić cechy organizacyjne firmy, których modyfikacja mogłaby pozwolić na zwiększenie zadowolenia pracowników. Określenie takich cech miałoby szczególne znaczenie już na etapie rekrutacji, gdyż mogłoby umożliwić selekcjonowanie kandydatów o profilu "kompatybilnym" z profilem firmy.

Badania zostaną przeprowadzone w oparciu o ankietę developerską StackOverflow [Stack Overflow, 2011]. Ankieta ta jest przeprowadzana corocznie od 2011 roku, jednak popularność zyskała dopiero w 2016 roku, kiedy to liczba respondentów biorących udział w badaniu przekroczyła 50 tysięcy. W latach 2016-2022 liczba respondentów utrzymywała się na poziomie od 50 do 100 tysięcy. Odpowiedzi w ankiecie udzielali specjaliści z całego świata reprezentując różne grupy wiekowe i kulturowe. Tak wysoka liczba odpowiedzi i różnorodność respondentów może pozwolić na przeprowadzenie badań ilościowych i umożliwić próbę generalizacji wniosków na całą branżę IT.

Do analizy danych zostaną wykorzystane metody uczenia maszynowego, a w szczególności regresja z wykorzystaniem regresora XGBoost opartego na algorytmie wzmocnienia gradientowego. Zdecydowano się na modelowanie problemu jako regresji, a nie klasyfikacji, gdyż zadowolenie pracowników oraz chęć zmiany pracy można przedstawić na skali liczbowej. Punnoose i Ajit [Punnoose and Ajit, 2016] przeprowadzili badania związane ze skutecznością przewidywania retencji pracowników w firmie z wykorzystaniem uczenia maszynowego.

Klasyfikator XGBoost osiągnął aż 86% skuteczności klasyfikacji co stanowiło ponad 50% lepszy wynik niż pozostałe testowane metody klasyfikacji tradycyjnie używane podczas badania problemu fluktuacji i retencji pracowników. Do celów badań będących przedmiotem niniejszej pracy implementacja algorytmu regresji i klasyfikacji zostanie przygotowana w języku Python.

3.2. Wstępne przetworzenie danych

W celu przeprowadzenia badań konieczne jest wyznaczenie zmiennej zależnej, której wartość będzie próbował oszacować model predykcji. Ze względu na omawiany w pracy problem retencji pracowników, jako zmienną zależną w badaniach można przyjąć następujące aspekty z ankiet StackOverflow:

1. ocena zadowolenia z obecnie wykonywanej pracy w 10 stopniowej skali,
2. status poszukiwania pracy w 3 stopniowej skali.

Badania zdecydowano przeprowadzić dla 3 edycji ankiety, aby zwiększyć szansę na zbudowanie modelu o satysfakcjonującym poziomie predykcji. Ankiety z lat 2021 i 2022 zostały odrzucone ze względu na brak pytań dotyczących zadowolenia z pracy i statusu poszukiwania pracy. Natomiast ankieta z roku 2020 została odrzucona, ponieważ była przeprowadzana w ciągu pierwszych 2 miesięcy trwania pandemii koronawirusa i istnieje obawa, że mogło to negatywnie wpłynąć na spójność zbioru danych w kontekście modelowania predykcji. Biorąc powyższe pod uwagę, do przeprowadzenia badania wybrano ankiety deweloperskie StackOverflow z lat 2017, 2018, 2019 (dostępne do pobrania pod adresem <https://insights.stackoverflow.com/survey>) ze względu na to, że są to 3 najnowsze edycje ankiety w których zadano pytania o wymienione powyżej aspekty.

Powyższe aspekty zostaną wykorzystane w badaniu jako zmienne zależne, czyli zmienne których wartość chcemy wyznaczać wykorzystując model predykcji. Dla badań z wykorzystaniem algorytmu regresji zmienne zależne pozostawiono w oryginalnej skali, natomiast dla algorytmu klasyfikacji odpowiedzi sprowadzono do klasyfikacji binarnej przy następujących założeniach:

1. pracownik jest zadowolony z pracy jeśli ocenił poziom zadowolenia na 5 lub więcej,
2. pracownik ma status osoby poszukującej jeśli szuka pracy aktywnie lub jest otwarty na atrakcyjne oferty pracy.

W celu przygotowania danych do badania konieczne było oczyszczenie i znormalizowanie danych. Proces ten został przeprowadzony w następujący sposób:

1. Usunięto pytania, które nie mają istotnego wpływu na badany problem (na przykład pytanie dotyczące preferencji związanej z wymawianiem słowa "GIF").
2. Uwspólniono nazwy etykiet dotyczących tych samych pytań w kolejnych edycjach ankiety.

3. Usunięto odpowiedzi respondentów, którzy nie mają doświadczenia zawodowego lub nie udzielili odpowiedzi na pytania dotyczące zadowolenia z pracy i statusu poszukiwania pracy.
4. Dla pytań z odpowiedziami w skali Likerta (typu od "bardzo się nie zgadzam" do "bardzo się zgadzam") zastosowano przekształcenie na skalę liczbową.
5. Dla pytań z odpowiedziami reprezentowanymi przez kategorie, które można przedstawić w skali (na przykład edukacja) zastosowano przekształcenie na skalę liczbową.
6. Dla pytań wielokrotnego wyboru oraz dla pytań z odpowiedziami reprezentowanymi przez kategorie, których nie można przedstawić w skali (na przykład technologie) potraktowano każdą kategorię jak osobne pytanie i zastosowano kodowanie binarne (0 - niewystępuje i 1 - występuje).
7. Kraj pochodzenia respondenta po zastąpieniu kodowaniem binarnym daje w rezultacie macierz rzadką z zaledwie 1 elementem niezerowym. Z tego powodu kraj pochodzenia respondenta zastąpiono wartością wskaźnika rozwoju społecznego HDI (ang. Human Development Index) [UNDP, 2021] odpowiednią dla danego kraju i roku badania.
8. Wynagrodzenie respondentów przeliczono na wartość w dolarach amerykańskich (USD) według kursu z 1 dnia roku w którym przeprowadzono badanie.
9. W pytaniach na które respondent nie udzielił odpowiedzi, wartości uzupełniono średnią wartością z pozostałych odpowiedzi, żeby wartości te nie wpływały negatywnie na działanie algorytmu regresji.
10. Na koniec wszystkie wartości przeskalowano aby otrzymać wartości z przedziału 0 - 1.

3.3. Selekcja cech z wykorzystaniem uczenia maszynowego

Do trenowania modelu uczenia maszynowego wykorzystano algorytm przedstawiony na listingu 3.1.

Listing 3.1: Algorytm uczenia modelu regresji

```
1 import xgboost as xgb
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import math
5 from sklearn.model_selection import train_test_split
6 from sklearn.metrics import mean_squared_error, r2_score
7 from sklearn.feature_selection import SelectFromModel
8
9 def fit_model(data, dependent_variable, name):
10     def fit_estimator(x_set, y_set):
11         estimator = xgb.XGBRegressor(**estimator_params,
12                                     ↪ eval_metric='rmsle')
13         estimator.fit(x_set, y_set)
14         return estimator
15
16     x_tilde = data.drop(regression_dependent_variables +
17                       ↪ class_dependent_variables, axis=1)
18     y_tilde = data[dependent_variable]
19
20     x_train, x_test, y_train, y_test = train_test_split(x, y,
21                                                         ↪ test_size=0.3, stratify=y)
22
23     selection = SelectFromModel(fit_estimator(x_train, y_train),
24                               ↪ threshold=.015, prefit=True)
25
26     selection_estimator =
27     ↪ fit_estimator(selection.transform(x_train), y_train)
28     preds = selection_estimator.predict(selection.transform(x_test))
29     print("RMSE: %.2f" % math.sqrt(abs(mean_squared_error(y_test,
30                                                         ↪ preds))))
31     print("R2: %.2f" % r2_score(y_test, preds))
32
33     xgb.plot_importance(selection_estimator, max_num_features=10)
34     plt.tight_layout()
35     plt.savefig(f'data/feat_importance_{name}.png')
```

Źródło: opracowanie własne

Hiperparametry estymatora wyznaczono w oparciu o projekt Harathi [Harathi, 2018] oraz z wykorzystaniem narzędzia GridSearchCV. Dodatkowo skorzystano z metody histogramowej do budowania drzewa decyzyjnego w celu skrócenia czasu uczenia modelu [Galarnyk, 2021]. Wyznaczone parametry przedstawiono na listingu 3.2.

Listing 3.2: Parametry estmatora

```
1 estimator_params = {
2     'tree_method': "hist",
3     'single_precision_histogram': True,
4     'n_jobs': -1,
5     'n_estimators': 1500,
6     'importance_type': 'weight',
7     'use_label_encoder': False,
8     'booster': 'gbtree',
9     'scale_pos_weight': 1,
10    'reg_alpha': 5,
11    'colsample_bytree': .8,
12    'learning_rate': .1,
13    'min_child_weight': 7,
14    'subsample': .5,
15    'max_depth': 6,
16    'gamma': 0.1,
17    'reg_lambda': 1
18 }
```

Źródło: opracowanie własne

Zbiór danych podzielono losowo w stosunku 7:3 w następujący sposób:

- z 70% danych utworzono zbiór uczący (ang. training set)
- z 30% danych utworzono zbiór testowy (ang. test set)

Zbiór uczący został wykorzystany do trenowania modelu predykcji, natomiast zbiór testowy do sprawdzenia skuteczności predykcji przy pomocy odpowiednich miar.

Do sprawdzenia skuteczności predykcji modelu regresji wykorzystano następujące miary:

- Pierwiastek logarytmu błędu średniokwadratowego RMSLE (ang. Root Mean Squared Logarytmic Error) - pokazuje jak bardzo predykcja jest oddalona od oczekiwanej wartości. Im niższa wartość tym lepiej.
- Współczynnik determinacji R² - pokazuje w jakim stopniu predykcja jest dopasowana do oczekiwanej wartości. Im wyższa wartość tym lepiej.

Celem modelu jest maksymalizacja R² i minimalizacja RMSLE. Dobrze dopasowany model powinien mieć wartość R² na poziomie powyżej 70% [Fernando, 2021], jednak nawet przy niskiej wartości R² model może być użyteczny jeżeli RMSLE ma wartość poniżej 10% [Alexander et al., 2015]. Natomiast przy akceptowalnym poziomie współczynnika R², model

może być użyteczny jeśli RMSLE nie przekracza 50% [ResearchGate, 2017]. Ma to szczególne znaczenie w przypadku badań prowadzonych w obszarze nauk o społeczeństwie, gdyż nawet model nie oferujący bardzo wysokiej skuteczności predykcji może pozwolić na pogłębienie wglądu w kwestię zachowań społecznych.

Do sprawdzenia skuteczności predykcji modelu klasyfikacji wykorzystano następujące miary:

- Dokładność (ang. accuracy) - pokazuje w ilu przypadkach klasyfikacja oszacowała poprawną wartość,
- Precyzja (ang. precision) - "proporcja prawidłowych pozytywnych klasyfikacji względem wszystkich pozytywnych klasyfikacji" [Szeliga, 2019],
- Czułość (ang. recall) - "proporcja poprawnych klasyfikacji przykładów pozytywnych względem wszystkich przykładów należących do klasy pozytywnej" [Szeliga, 2019].

Na podstawie opracowania Brownlee [Brownlee, 2020], w badaniach przyjęto wartości miar klasyfikacji za dobre, jeśli wynoszą powyżej 85% i za złe jeśli wynoszą poniżej 75%.

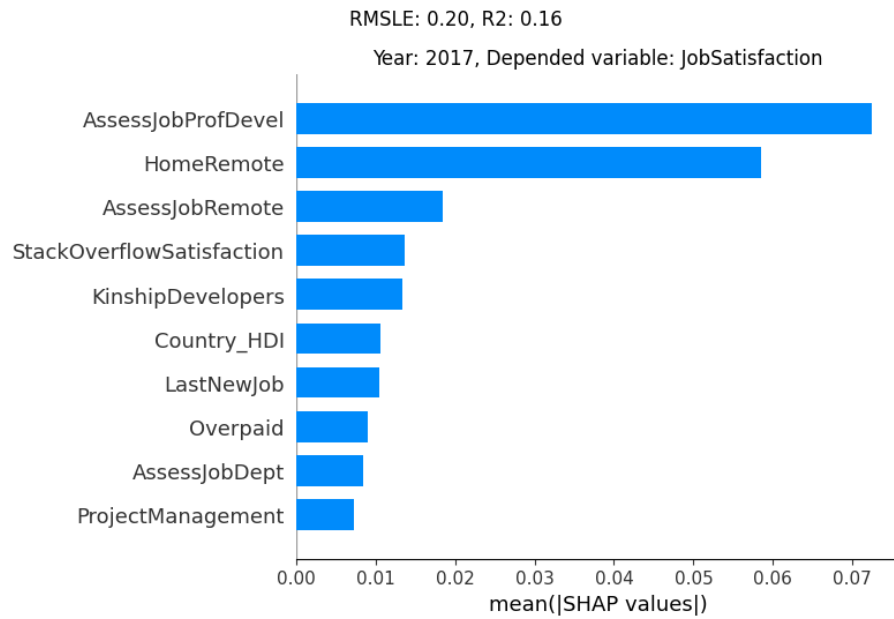
Selekcję pytań najsilniej wpływających na predykcję wykonano z wykorzystaniem metody SHAP (ang. SHapley Additive exPlanations). Jest to metoda oparta o kooperacyjną teorię gier i jest wykorzystywana do zwiększenia możliwości interpretacyjnych modeli uczenia maszynowego. Pozwala na selekcję cech silnie skorelowanych z przewidzianym wynikiem, jednak nie pozwala na ocenę skuteczności samej predykcji [Trevisan, 2017].

3.4. Prezentacja wyników

Algorytm przedstawiony na listingu 3.1 został uruchomiony dla znormalizowanych danych z lat 2017-2019. Lista atrybutów biorących udział w badaniu i odpowiadających im pytań została przedstawiona w dodatku A. Wyniki zostały podzielone ze względu na modelowanie problemu jako regresję i jako klasyfikację.

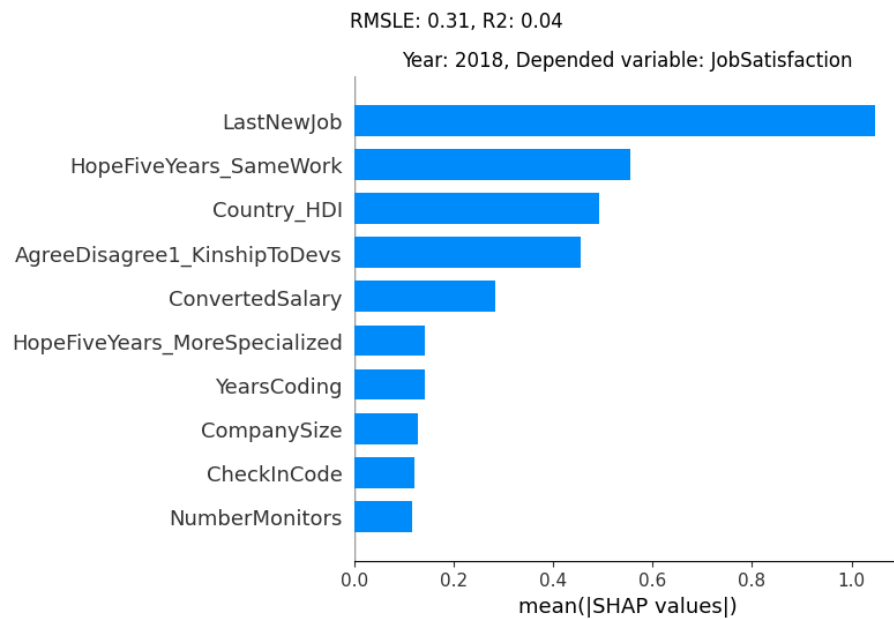
3.4.1. Wyniki regresji

Na rysunkach 3.1, 3.2 i 3.3 przedstawiono wyniki regresji dla zmiennej zależnej "JobSatisfaction", natomiast na rysunkach 3.4, 3.5 i 3.6 przedstawiono wyniki regresji dla zmiennej zależnej "JobSeekingStatus".



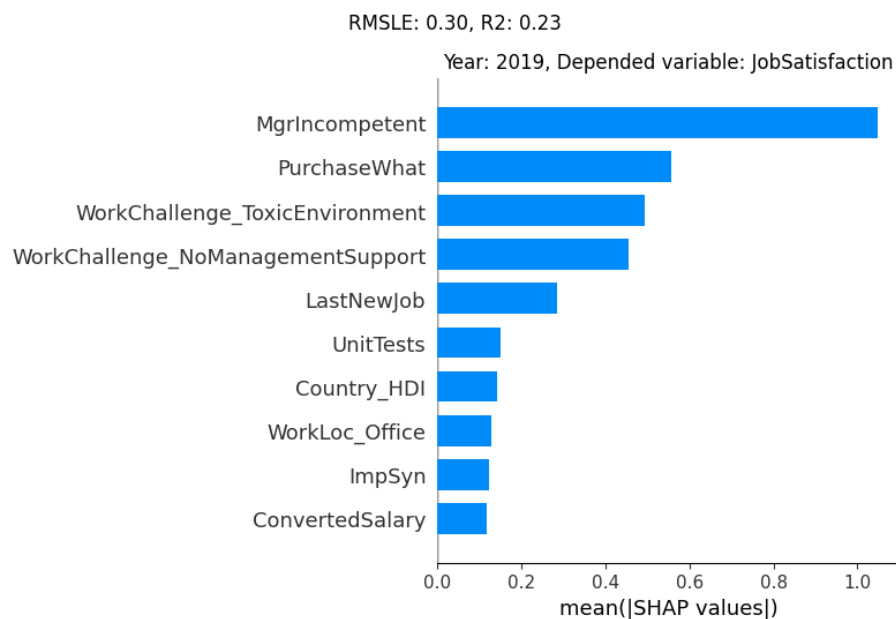
Rysunek 3.1: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2017

Źródło: opracowanie własne



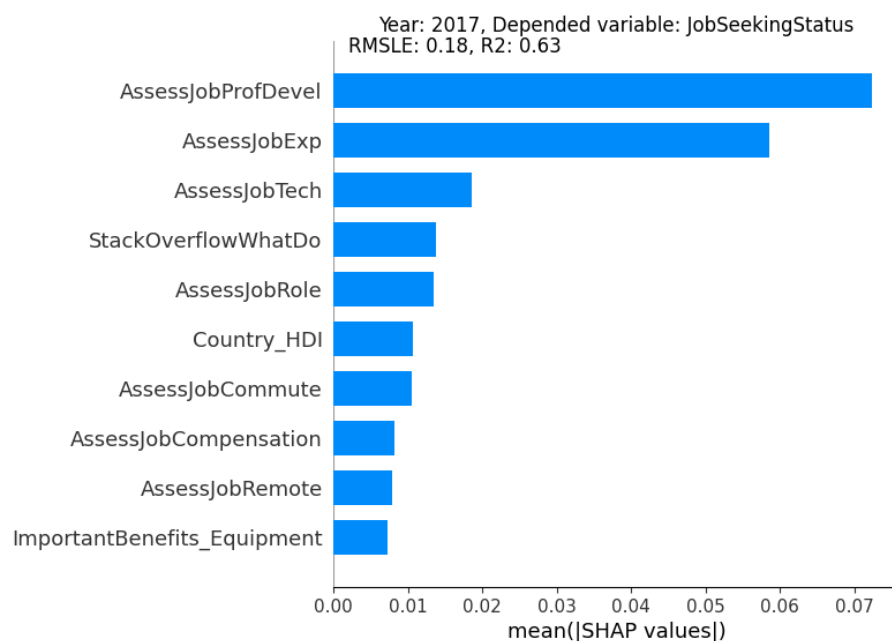
Rysunek 3.2: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2018

Źródło: opracowanie własne



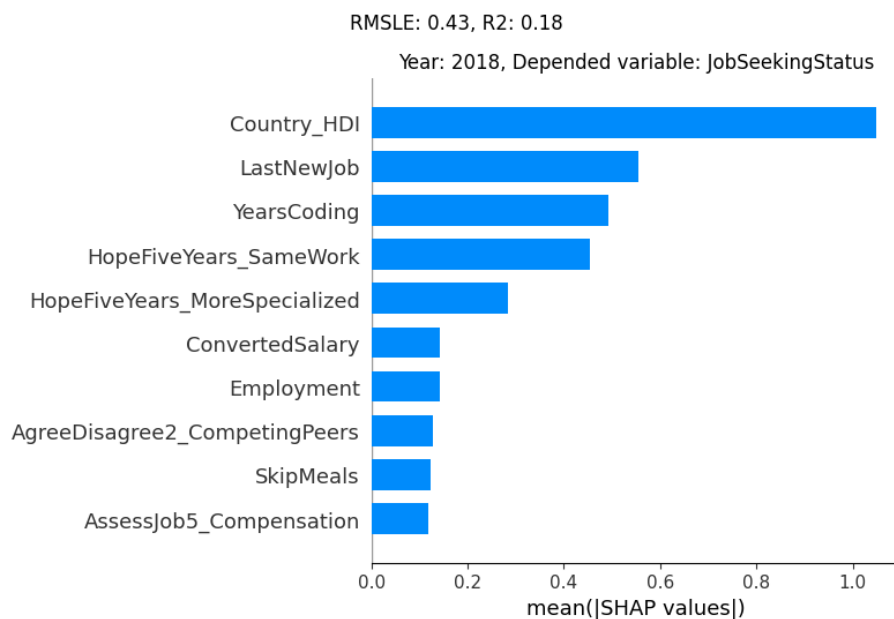
Rysunek 3.3: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2019

Źródło: opracowanie własne



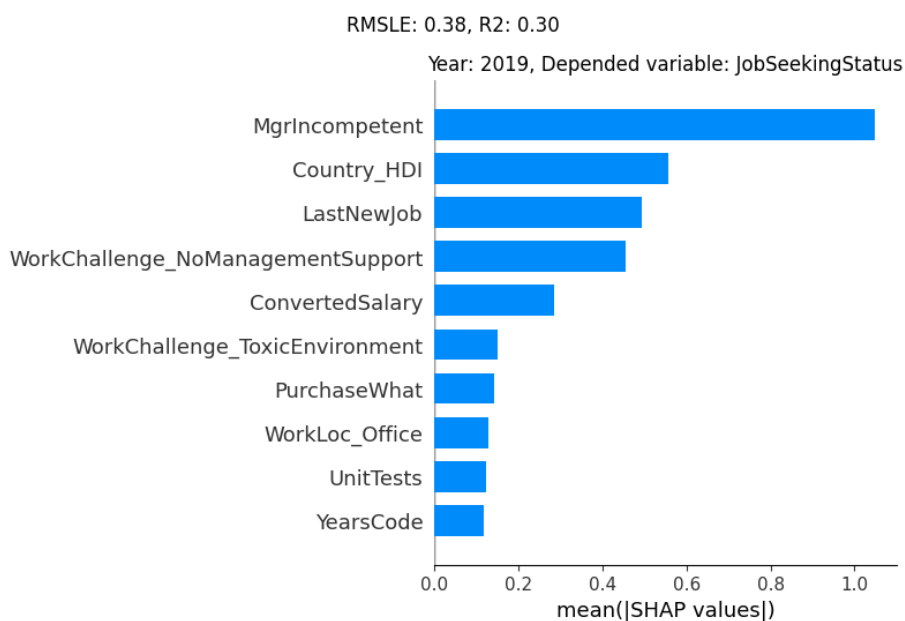
Rysunek 3.4: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2017

Źródło: opracowanie własne



Rysunek 3.5: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2018

Źródło: opracowanie własne



Rysunek 3.6: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2019

Źródło: opracowanie własne

W tabelach 3.1 i 3.2 przedstawiono podsumowanie skuteczności modeli regresji, odpowiednio dla zmiennych zależnych JobSatisfaction i JobSeekingStatus.

Tabela 3.1: Porównanie wyników regresji dla zmiennej zależnej JobSatisfaction

Rok	2017	2018	2019
RMSLE	0.20	0.31	0.30
R2	0.16	0.04	0.23
Pytania z siłą wpływu większą od 0.2	<ul style="list-style-type: none"> AssessJobProfDevel HomeRemote AssessJobRemote 	<ul style="list-style-type: none"> LastNewJob HopeFiveYears_SameWork Country_HDI AgreeDisagree1_KinshipToDevs ConvertedSalary 	<ul style="list-style-type: none"> MgrIncompetent PurchaseWant WorkChallenge_ToxicEnvironment WorkChallenge_NoManagementSupport LastNewJob

Źródło: opracowanie własne

Tabela 3.2: Porównanie wyników regresji dla zmiennej zależnej JobSeekingStatus

Rok	2017	2018	2019
RMSLE	0.18	0.43	0.38
R2	0.63	0.18	0.30
Pytania z siłą wpływu większą od 0.2	<ul style="list-style-type: none"> AssessJobProfDevel AssessJobExp AssessJobTech 	<ul style="list-style-type: none"> Country_HDI LastNewJob YearsCoding HopeFiveYears_SameWork HopeFiveYears_MoreSpecialized 	<ul style="list-style-type: none"> MgrIncompetent Country_HDI LastNewJob WorkChallenge_NoManagementSupport ConvertedSalary

Źródło: opracowanie własne

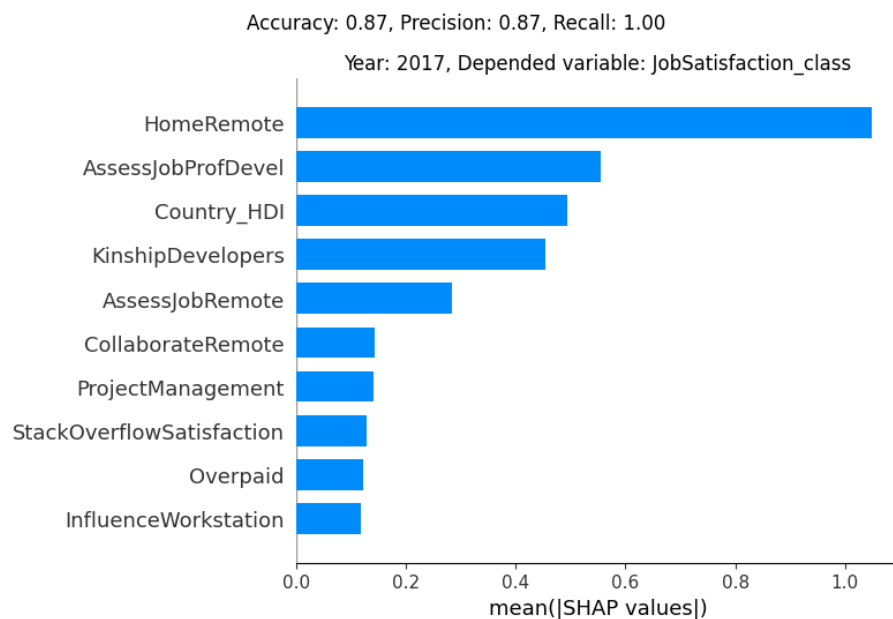
Mimo stosunkowo dobrych wartości RMSLE, w większości badanych przypadków osiągnięto niską wartość R2, co oznacza, że wyselekcjonowane atrybuty są zasadniczo słabo skorelowane ze zmienną zależną. Satysfakcjonujący rezultat osiągnięto dla danych z 2017 roku i zmiennej zależnej JobSeekingStatus. Na tej podstawie można wnioskować, że na chęć poszukiwania pracy mogą mieć wpływ następujące cechy:

- Czy respondent uważa, że możliwość rozwoju osobistego w firmie jest ważna,

- Czy respondent uważa, że wymagane na stanowisku doświadczenie jest ważne,
- Czy respondent uważa, że technologie wykorzystywane w pracy są ważne.

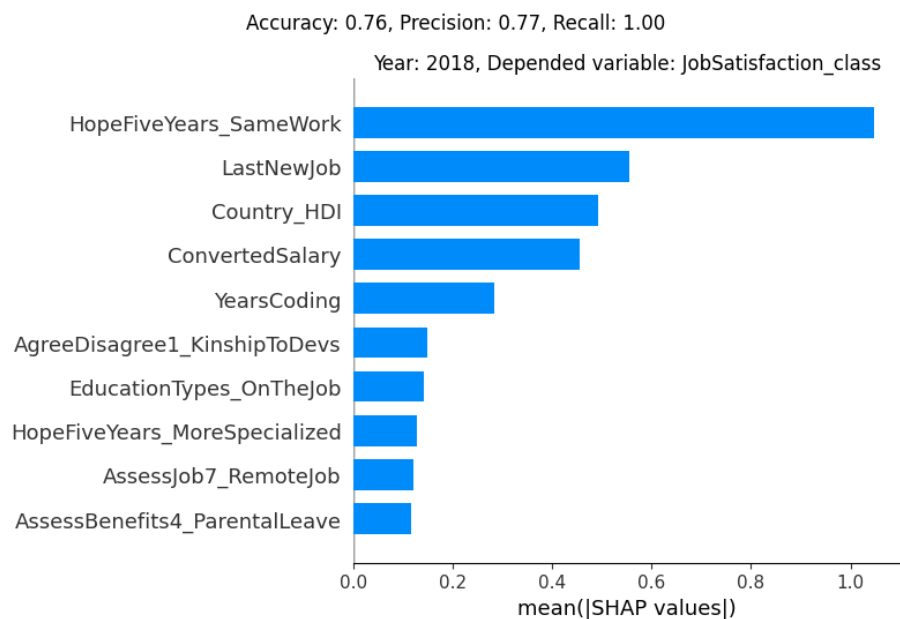
3.4.2. Wyniki klasyfikacji

Na rysunkach 3.7, 3.8 i 3.9 przedstawiono wyniki klasyfikacji dla zmiennej zależnej "JobSatisfaction", natomiast na rysunkach 3.10, 3.11 i 3.12 przedstawiono wyniki klasyfikacji dla zmiennej zależnej "JobSeekingStatus".



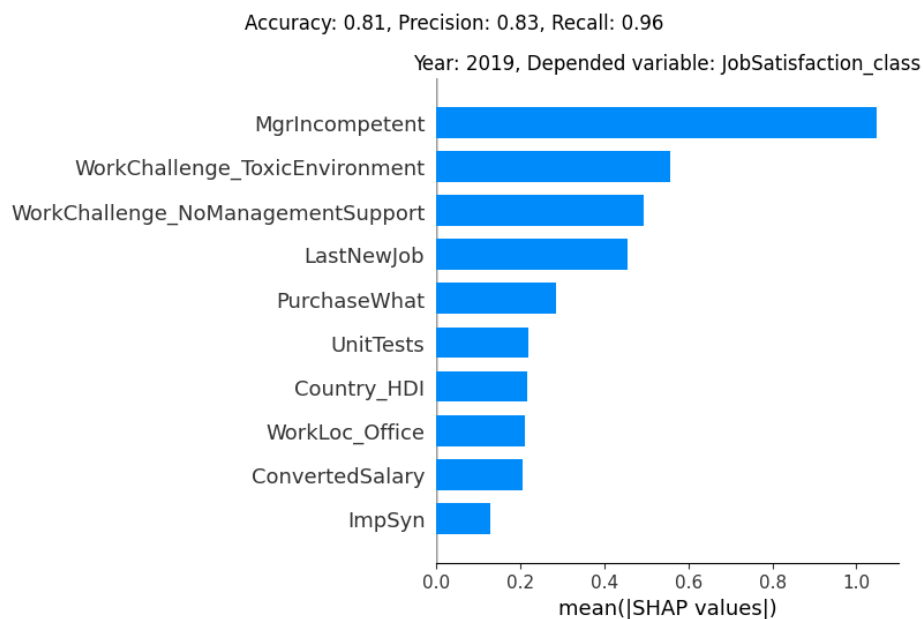
Rysunek 3.7: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2017

Źródło: opracowanie własne



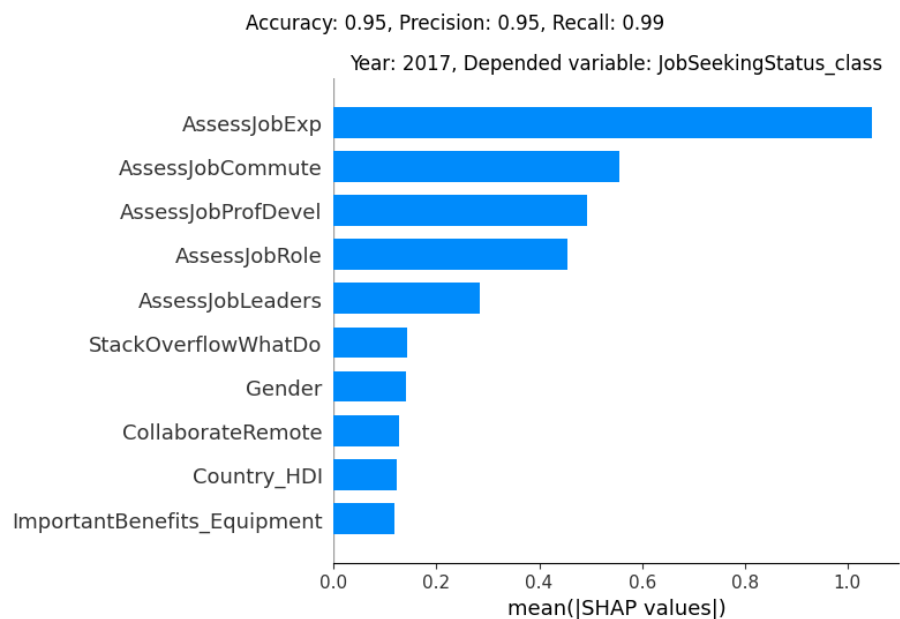
Rysunek 3.8: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2018

Źródło: opracowanie własne



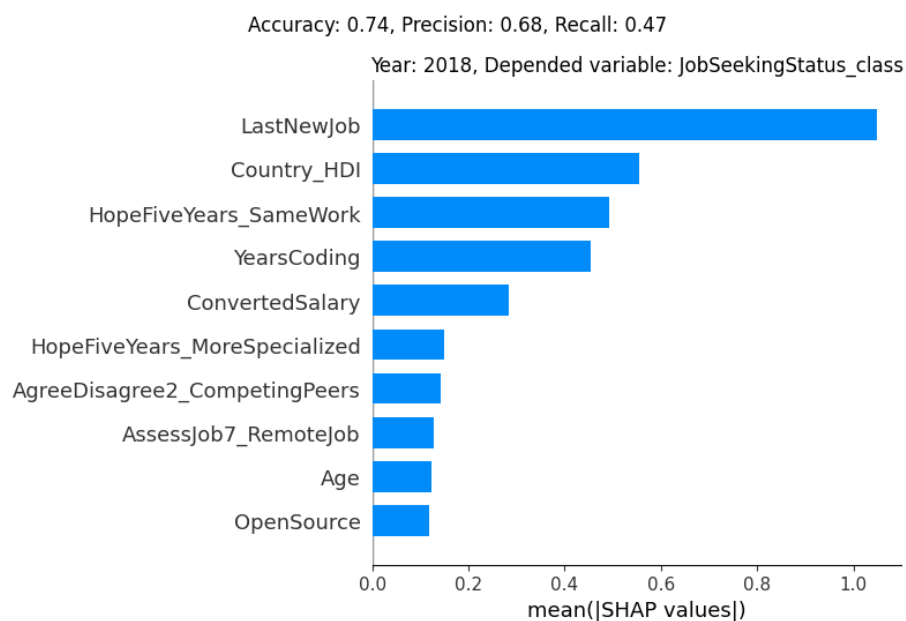
Rysunek 3.9: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2019

Źródło: opracowanie własne



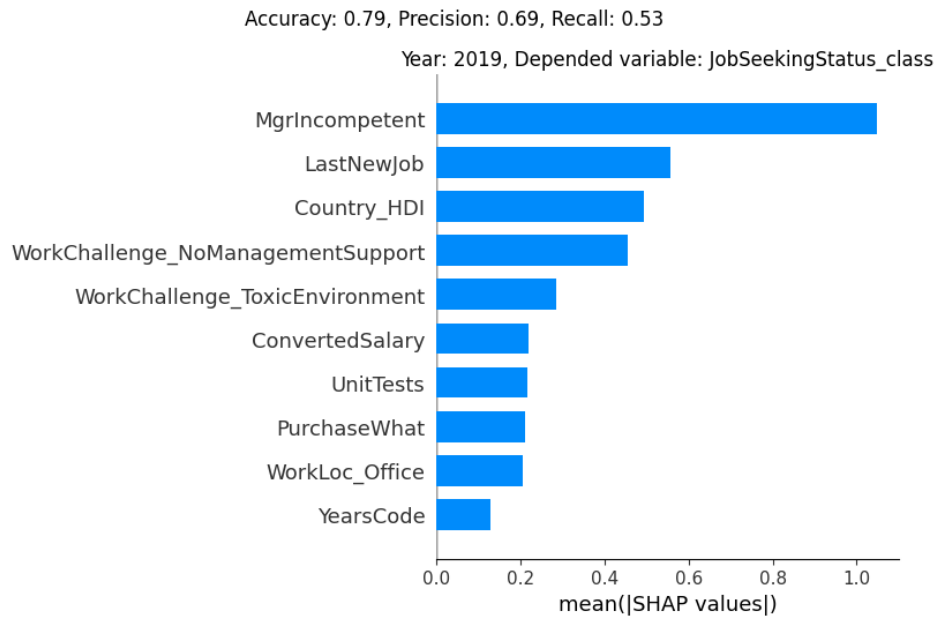
Rysunek 3.10: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2017

Źródło: opracowanie własne



Rysunek 3.11: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2018

Źródło: opracowanie własne



Rysunek 3.12: Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2019

Źródło: opracowanie własne

W tabelach 3.3 i 3.4 przedstawiono podsumowanie skuteczności modeli klasyfikacji, odpowiednio dla zmiennych zależnych JobSatisfaction i JobSeekingStatus.

Tabela 3.3: Porównanie wyników klasyfikacji dla zmiennej zależnej JobSatisfaction

Rok	2017	2018	2019
Dokładność	0.87	0.76	0.81
Precyzja	0.87	0.77	0.83
Czułość	1.00	1.00	0.96
Pytania z siłą wpływu większą od 0.2	<ul style="list-style-type: none"> HomeRemote AssessJobProfDevel Country_HDI KinshipDevelopers AssessJobRemote 	<ul style="list-style-type: none"> HopeFiveYears_SameWork LastNewJob Country_HDI ConvertedSalary YearsCoding 	<ul style="list-style-type: none"> MgrIncompetent WorkChallenge_ToxicEnvironment WorkChallenge_NoManagementSupport LastNewJob PurchaseWhat

Źródło: opracowanie własne

Tabela 3.4: Porównanie wyników klasyfikacji dla zmiennej zależnej JobSeekingStatus

Rok	2017	2018	2019
Dokładność	0.95	0.74	0.79
Precyzja	0.95	0.68	0.69
Czułość	0.99	0.47	0.53
Pytania z siłą wpływu większą od 0.2	<ul style="list-style-type: none"> AssessJobExp AssessJobCommute AssessJobProfDevel AssessJobRole AssessJobLeaders 	<ul style="list-style-type: none"> LastNewJob Country_HDI HopeFiveYears_SameWork YearsCoding ConvertedSalary 	<ul style="list-style-type: none"> MgrIncompetent LastNewJob Country_HDI WorkChallenge_NoManagementSupport WorkChallenge_ToxicEnvironment

Źródło: opracowanie własne

Satysfakcjonujący rezultat osiągnięto dla danych z 2017 roku i obu badanych zmiennych zależnych. Na tej podstawie można wnioskować, że na zadowolenie z pracy mogą mieć wpływ następujące cechy:

- Jak często respondent obecnie pracuje zdalnie,
- Czy respondent uważa, że możliwość rozwoju osobistego w firmie jest ważna,
- Indeks rozwoju społecznego kraju, w którym mieszka respondent,
- W jakim stopniu respondent czuje solidarność ze swoimi współpracownikami,
- Czy respondent uważa, że możliwość pracy zdalnej jest ważna.

Natomiast na chęć poszukiwania pracy mogą mieć wpływ następujące cechy:

- Czy respondent uważa, że doświadczenie wymagane na stanowisku pracy jest ważne,
- Czy respondent uważa, że czas dojazdu do pracy jest ważny,
- Czy respondent uważa, że możliwość rozwoju osobistego w firmie jest ważna,
- Czy respondent uważa, że nazwa roli lub stanowiska pracy jest ważna,
- Czy respondent uważa, że reputacja seniorów pracujących w firmie jest ważna.

3.5. Wnioski i analiza możliwości praktycznego zastosowania zbudowanego modelu predykcji

Przeprowadzone badania wykazały, że dla zbioru danych opartego o wyniki ankiety deweloperskiej StackOverflow [Stack Overflow, 2011] można zbudować skuteczny model predykcji zarówno modelując problem jako regresję, jak i jako klasyfikację.

Opracowane modele pozwoliły na wytypowanie cech silnie wpływających na możliwość predykcji zadowolenia z pracy i chęci szukania nowej pracy.

Szczególnie wysoką jakość modelu predykcji osiągnięto dla zbioru danych z 2017 roku.

Osiągnięte rezultaty mogą być uwzględnione podczas rekrutacji nowych pracowników oraz mogą być sugestią dla pracowników działów zasobów ludzkich, w jakich aspektach można poprawić funkcjonowanie firmy, aby zwiększyć retencję pracowników.

W przyszłości badania można rozszerzyć uwzględniając podział pracowników ze względu na takie cechy jak:

- wiek,
- płeć,
- kraj zamieszkania,
- obszar specjalizacji;

oraz uwzględniają podział firm ze względu na takie cechy jak:

- rozmiar firmy,
- gałąź przemysłu, w obrębie której działa firma.

Przypuszczać można, że stosując taki granularny podział możliwe będzie wyszczególnienie cech szczególnie pożądanых wśród kandydatów zatrudnianych do określonej firmy lub cech przedsiębiorstwa, na które konkretny programista poszukujący pracy powinien zwracać uwagę.

Zarówno dla bieżących modeli predykcji, jak i potencjalnie możliwych do opracowania modeli szczegółowych, możliwe byłoby opracowanie kwestionariusza, w którym po uzupełnieniu danych dotyczących kandydata i przedsiębiorstwa algorytm korzystający z opracowanych modeli predykcji byłby w stanie oszacować, w jakim stopniu pracownik będzie zadowolony z pracy.

Zakończenie

Celem pracy było sprawdzenie, czy uczenie maszynowe pozwala określić z zadowalającym stopniem pewności cechy wpływające na poziom zadowolenia i chęć zmiany pracy pracowników IT.

Aby osiągnąć ten cel dla danych z ankiet deweloperskich StackOverflow [Stack Overflow, 2011] z lat 2017, 2018 i 2019 opracowano algorytm w języku Python pozwalający na zbudowanie modeli predykcji opartych o regresję i klasyfikację.

W wyniku badań wyłoniono zestaw cech silnie wpływających na zadowolenie pracowników:

- Jak często respondent obecnie pracuje zdalnie,
- Czy respondent uważa, że możliwość rozwoju osobistego w firmie jest ważna,
- Indeks rozwoju społecznego kraju, w którym mieszka respondent,
- W jakim stopniu respondent czuje solidarność ze swoimi współpracownikami,
- Czy respondent uważa, że możliwość pracy zdalnej jest ważna;

oraz na deklarowaną przez pracowników chęć szukania nowej pracy:

- Czy respondent uważa, że możliwość rozwoju osobistego w firmie jest ważna,
- Czy respondent uważa, że wymagane na stanowisku doświadczenie jest ważne,
- Czy respondent uważa, że technologie wykorzystywane w pracy są ważne,
- Czy respondent uważa, że czas dojazdu do pracy jest ważny,
- Czy respondent uważa, że nazwa roli lub stanowiska pracy jest ważna,
- Czy respondent uważa, że reputacja seniorów pracujących w firmie jest ważna.

Na podstawie analizy wyników badań można podjąć się odpowiedzi na postawione w pracy pytania badawcze:

1. Można wyłonić cechy (osoby lub przedsiębiorstwa) pozwalające oszacować zadowolenie i chęć zmiany pracy pracownika z branży IT.
2. Wyłonione cechy są raczej związane z preferencjami dotyczącymi stanowiska, które firma może dostosować, aby zwiększyć atrakcyjność miejsca pracy dla swoich pracowników. W wyniku badania nie wyłoniono cech, które silnie związane byłyby z profilem kandydata, więc nie można stwierdzić, że wyłonione cechy można użyć do poprawy procesu wstępnej selekcji kandydatów.

3. Wyłonione cechy można użyć do zwiększenia atrakcyjności przedsiębiorstwa dla pracowników i kandydatów, na przykład poprzez:

- oferowanie pracownikom możliwości rozwoju osobistego,
- oferowanie pracownikom możliwości pracy zdalnej,
- oferowanie pracownikom atrakcyjnej nazwy stanowiska.

Ze względu na dostępność danych, badania wykonano dla ankiet przeprowadzonych przed wybuchem epidemii koronawirusa. Jak wspomniano w rozdziale 2.5, pandemia miała istotny wpływ na trendy w branży IT, można więc przypuszczać, że wpłynęła również na zmianę priorytetów w zakresie cech wpływających na zadowolenie pracowników. Szczególnie konieczność pracy zdalnej w czasie pandemii mogła znacząco wpłynąć na oczekiwania w tym zakresie w okresie postpandemicznym.

Przeprowadzone badania mogą być potraktowane jako wstęp do problematyki predykcji zadowolenia i retencji pracowników. Uzyskana skuteczność predykcji pokazuje, że przy odpowiednio postawionych pytaniach możliwe jest poprzez przeprowadzenie ankiety wyłonienie cech, które odpowiadają za zadowolenie i chęć zmiany pracy pracowników. Badania przeprowadzone w ramach niniejszej pracy traktowały jednorodnie całą branżę informatyczną. Przyszłe badania mogłyby skupić się na podziale pracowników (np. ze względu na specjalizację) i firm (np. ze względu na organizację) oraz na przygotowaniu modeli predykcji dla innych zbiorów danych. Uzyskane specjalistyczne modele dla wielu zbiorów danych mogą następnie być połączone, aby dalej zwiększać skuteczność predykcji. Następnie modele predykcji mogą zostać wykorzystane w formie aplikacji komputerowej, w której po wprowadzeniu danych firmy i kandydata, możliwe mogłoby być określenie dopasowania kandydata do firmy.

Bibliografia

Spis literatury

- [Cron and DeCarlo, 2006] Cron, W. L. and DeCarlo, T. E. (2006). *Darlymple's Sales Management*. John Wiley Sons, Inc, New Jersey.
- [Dalton et al., 1982] Dalton, D. R., Todor, W. D., and Krackhardt, D. M. (1982). Turnover overstated: A functional Taxonomy. *Academy of Management Review*, 7.
- [Donadelli, 2015] Donadelli, S. M. (2015). The Impact of Knowledge Loss on Software Projects: Turnover, Customer Found Defects, and Dormant Files.
- [Graziotin et al., 2017] Graziotin, D., de Fabian Fagerholm, Fagerholm, F., and Wang, X. (2017). On the Unhappiness of Software Developers.
- [Graziotin et al., 2018] Graziotin, D., Fagerholm, F., Wang, X., and Abrahamsson, P. (2018). What happens when software developers are (un)happy. *ArXiv*, 432.
- [Juchnowicz, 2010] Juchnowicz, M. (2010). *Zarządzanie przez zaangażowanie. Koncepcja, Kontrowersje, Aplikacje*. Polskie Wydawnictwo Ekonomiczne, Warszawa.
- [Kozioł, 2011] Kozioł, L. (2011). Trychotomia czynników motywacji w pracy – zarys koncepcji. *Zeszyty Naukowe MWSE*, 19.
- [Kozłowski, 2012] Kozłowski, M. (2012). *Employer branding. Budowanie wizerunku pracodawcy krok po kroku*. Oficyna Wolter Kluwer, Warszawa.
- [Król and Ludwiciński, 2006] Król, H. and Ludwiciński, A. (2006). *Zarządzanie zasobami ludzkimi. Podręcznik*. Wydawnictwo Naukowe PWN, Warszawa.
- [Martin, 2014] Martin, R. C. (2014). *Czysty kod*. Helion, Gliwice.
- [Philips and Edwards, 2009] Philips, J. and Edwards, L. (2009). *Managing Talent Retention: An ROI Approach*. Pfeiffer, San Francisco.
- [Pocztowski, 2007] Pocztowski, A. (2007). *Zarządzanie zasobami ludzkimi*. PWE, Warszawa.

- [Pocztowski, 2009] Pocztowski, A. (2009). *Procesy płynności i retencji zatrudnienia w organizacji*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków.
- [Punnoose and Ajit, 2016] Punnoose, R. and Ajit, P. (2016). Prediction of Employee Turnover in Organizations using Machine Learning Algorithms. *International Journal of Advanced Research in Artificial Intelligence*, 5.
- [Rangarajan and Tiwari, 2014] Rangarajan, K. and Tiwari, S. (2014). Evolution of Global IT Services Industry: Impact of Technological advancements and Business needs. *IOSR Journal of Business and Management*, 16.
- [Riya, 2020] Riya, J. (2020). Computer Fundamentals - Von Neumann Model. Dostęp 18.03.2022.
- [Schultz and Schultz, 2002] Schultz, D. and Schultz, S. (2002). *Psychologia a wyzwania dzisiejszej pracy*. Wydawnictwo Naukowe PWN, Warszawa.
- [Sowińska, 2014] Sowińska, A. (2014). Zadowolenie z pracy - problemy definicyjne. *Studia Ekonomiczne / Uniwersytet Ekonomiczny w Katowicach*, 197.
- [Spychała et al., 2019] Spychała, M., Bartecki, T., and Brzózka, M. (2019). Employer branding – nowoczesne rozwiązanie problemu rotacji pracowników. *Zeszyty Naukowe Politechniki Poznańskiej. Organizacja i Zarządzanie*.
- [Steel and Lounsbury, 2009] Steel, R. P. and Lounsbury, J. W. (2009). Turnover process models: Review and synthesis of a conceptual literature. *Human Resource Management Review*, 19.
- [Taylor, 2006] Taylor, S. (2006). *Płynność zatrudnienia. Jak zatrzymać pracowników w firmie*. Oficyna Wolters Kluwer, Kraków.
- [Woźniak, 2012] Woźniak, J. (2012). *Współczesne systemy motywacyjne*. Wydawnictwo Naukowe PWN, Warszawa.
- [Woźniak, 2016] Woźniak, J. (2016). *Fluktuacja informatyków i jej przyczyny w świetle wyników badań*. Wyższa Szkoła Finansów i Zarządzania, Warszawa.
- [Yakovlev, 2007] Yakovlev, I. (2007). Web 2.0: Is It Evolutionary or Revolutionary? *IT Professional*, 9:43–45.

Wykaz innych źródeł

- [Alabbas, 2017] Alabbas, A. (2017). A Brief History of AWS. <https://mediatemple.net/blog/cloud-hosting/brief-history-aws/>. Dostęp 18.03.2022.

- [Alexander et al., 2015] Alexander, D., Tropsha, A., and Winkler, D. (2015). Beware of R2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4530125/>. Dostęp 06.06.2022.
- [Błaszczak, 2022] Błaszczak, A. (2022). Wojna nasili niedobór kadr w IT. Napływ nowych zleceń do Polski. <https://www.rp.pl/rynek-pracy/art35856021-wojna-nasili-niedobor-kadr-w-it-naplyw-nowych-zleceń-do-polski>. Dostęp 26.05.2022.
- [Berteig, 2005] Berteig, M. (2005). Truck Factor. <https://www.agileadvice.com/2005/05/15/agilemanagement/truck-factor/>. Dostęp 26.05.2022.
- [Booz, 2018] Booz, M. (2018). These 3 Industries Have the Highest Talent Turnover Rates. <https://www.linkedin.com/business/talent/blog/talent-strategy/industries-with-the-highest-turnover-rates>. Dostęp 18.03.2022.
- [Brownlee, 2020] Brownlee, J. (2020). Failure of Classification Accuracy for Imbalanced Class Distributions. <https://machinelearningmastery.com/failure-of-accuracy-for-imbalanced-class-distributions/>. Dostęp 26.09.2022.
- [D. Pierce, 2018] D. Pierce, L. G. (2018). The WIRED Guide to the iPhone. <https://www.wired.com/story/guide-iphone/>. Dostęp 18.03.2022.
- [Daxx, 2020] Daxx (2020). Talent Shortage in the Nordics: How Norway, Sweden, Finland, and Denmark Can Solve the Software Developer Shortage in 2021. <https://www.daxx.com/blog/development-trends/nordics-tech-talent-shortage>. Dostęp 17.01.2022.
- [Daxx, 2022] Daxx (2022). The Software Developer Shortage in the US and the Global Tech Talent Shortage in 2022. <https://www.daxx.com/blog/development-trends/software-developer-shortage-us>. Dostęp 17.01.2022.
- [Fernando, 2021] Fernando, J. (2021). R-Squared. <https://www.investopedia.com/terms/r/r-squared.asp>. Dostęp 06.06.2022.
- [Galarnyk, 2021] Galarnyk, M. (2021). How to Speed Up XGBoost Model Training. <https://www.anyscale.com/blog/how-to-speed-up-xgboost-model-training>. Dostęp 06.06.2022.
- [Harathi, 2018] Harathi, P. (2018). Stack Overflow Survey Analysis. <https://medium.com/@s.pranav.harathi/stack-overflow-survey-analysis-ed45127691b>. Dostęp 06.06.2022.

- [Kalbarczyk, 2021] Kalbarczyk, A. (2021). Sytuacja na rynku pracy IT w dobie koronawirusa. <https://itwiz.pl/sytuacja-na-ryнку-pracy-it-w-dobie-koronawirusa>. Dostęp 26.05.2022.
- [Nesh, 2021] Nesh (2021). Employee Turnover = Knowledge Loss? Let's change the equation. <https://hellonesh.io/employee-turnover-knowledge-loss-lets-change-the-equation>. Dostęp 26.05.2022.
- [Overby, 2022] Overby, S. (2022). Wojna w Ukrainie stawia globalny outsourcing IT w trudnej sytuacji. <https://www.computerworld.pl/news/Wojna-w-Ukrainie-stawia-globalny-outsourcing-IT-w-trudnej-sytuacji,437151.html>. Dostęp 26.05.2022.
- [PAIH, 2021] PAIH (2021). Polska Izba Informatyki i Telekomunikacji i Polska Agencja Inwestycji i Handlu łączą siły by promować branżę teleinformatyczną. https://www.paih.gov.pl/20210322/wspolpraca_piit_i_paih. Dostęp 18.03.2022.
- [Pasterczyk, 2021] Pasterczyk, J. (2021). Przymusowy powrót do biura z home office? W branży IT 64 proc. wróciłoby z wypowiedzeniem. <https://geek.justjoin.it/przymusowy-powrot-do-biura-z-home-office-w-branzy-it-64-proc-wrociloby-z-wypowiedzeniem>. Dostęp 26.05.2022.
- [ResearchGate, 2017] ResearchGate (2017). What's the acceptable value of Root Mean Square Error (RMSE), Sum of Squares due to error (SSE) and Adjusted R-square? <https://www.researchgate.net/post/Whats-the-acceptable-value-of-Root-Mean-Square-Error-RMSE-Sum-of-Squares-due-to-error-SSE-and-Adjusted-R-square>. Dostęp 06.06.2022.
- [Rojewska, 2020] Rojewska, K. (2020). Najciekawsze benefity oferowane przez firmy IT. <https://nofluffjobs.com/blog/najciekawsze-benefity-oferowane-przez-firmy-it/>. Dostęp 18.03.2022.
- [Software Mind, 2021] Software Mind (2021). Why choose Poland for IT Outsourcing in 2021? <https://softwaremind.com/why-choose-poland-for-it-outsourcing-top-factors/>. Dostęp 18.03.2022.
- [Sonmez, 2015] Sonmez, J. Z. (2015). Specialists vs. polyglots – What kind of software developer are you? <https://jaxenter.com/specialists-vs-polyglots-what-kind-of-software-developer-are-you-120333.html>. Dostęp 26.05.2022.

- [Stack Overflow, 2011] Stack Overflow (2011). Stack Overflow Annual Developer Survey. <https://insights.stackoverflow.com/survey>. Dostęp 06.06.2022.
- [Stack Overflow, 2020] Stack Overflow (2020). 2020 Developer Survey. <https://insights.stackoverflow.com/survey/2020>. Dostęp 23.05.2022.
- [Szeliga, 2019] Szeliga, M. (2019). Uczenie maszynowe – ocena, interpretacja i wdrożenie modeli. <http://www.it-professional.pl/archiwum/art,8537,uczenie-maszynowe-ocena-interpretacja-i-wdrozenie-modeli.html>. Dostęp 26.09.2022.
- [Trevisan, 2017] Trevisan, V. (2017). Using SHAP Values to Explain How Your Machine Learning Model Works. <https://towardsdatascience.com/using-shap-values-to-explain-how-your-machine-learning-model-works-732b3f40e137>. Dostęp 26.09.2022.
- [Trikha, 2021] Trikha, R. (2021). Which Country Would Win in the Programming Olympics? <https://blog.hackerrank.com/which-country-would-win-in-the-programming-olympics/>. Dostęp 18.03.2022.
- [Ułan, 2022] Ułan, G. (2022). Złoty okres na specjalistów IT - wygląda na to, że praca zdalna pozostanie w branży IT na stałe. <https://antyweb.pl/praca-zdalna-w-it>. Dostęp 26.05.2022.
- [UNDP, 2021] UNDP (2021). Human Development Index. <https://hdr.undp.org/en/content/human-development-index-hdi>. Dostęp 06.06.2022.
- [U.S.News, 2022] U.S.News (2022). Software Developer. <https://money.usnews.com/careers/best-jobs/software-developer>. Dostęp 17.01.2022.
- [Winter, 2016] Winter, T. (2016). Have you ever wondered how much it costs to recruit a new programmer? The numbers may surprise you. <https://devskiller.com/true-cost-of-recruiting-a-developer-infographic>. Dostęp 17.01.2022.

Spis tabel

1.1	Rodzaje fluktuacji	8
1.2	Heurystyka do oceny wysokiego poziomu fluktuacji	9
1.3	Najczęściej pojawiające się cechy w modelach fluktuacji	10
1.4	Najczęściej pojawiające się cechy w modelach fluktuacji (ciąg dalszy)	11
1.5	Czynniki organizacyjne wpływające na zadowolenie pracowników	12
1.6	Czynniki osobowe wpływające na zadowolenie pracowników	12
1.7	Wpływ czynników motywacji na zadowolenie pracowników	13
1.8	Czynniki kształtujące zaangażowanie	14
1.9	Typy kosztów związanych z fluktuacją	15
2.1	Przyczyny niezadowolenia pracowników w branży IT	20
3.1	Porównanie wyników regresji dla zmiennej zależnej JobSatisfaction	38
3.2	Porównanie wyników regresji dla zmiennej zależnej JobSeekingStatus	38
3.3	Porównanie wyników klasyfikacji dla zmiennej zależnej JobSatisfaction	42
3.4	Porównanie wyników klasyfikacji dla zmiennej zależnej JobSeekingStatus	43

Spis rysunków

1	Niedobór wykwalifikowanych pracowników IT na świecie	4
2.1	Rozkład wieku, w którym respondenci ankiety StackOverflow napisali pierwszą linię kodu	22
2.2	Rozkład stażu pracy respondentów ankiety StackOverflow w latach	23
2.3	Rozkład wieku respondentów ankiety StackOverflow	23
2.4	Rozkład wykształcenia respondentów ankiety StackOverflow	24
2.5	Rozkład płci respondentów ankiety StackOverflow	24
2.6	Rozkład rodzaju zatrudnienia respondentów ankiety StackOverflow	25
2.7	Rozkład rozmiaru firm, w których pracują respondenci ankiety StackOverflow .	25
2.8	Rozkład zadowolenia z pracy respondentów ankiety StackOverflow	26
2.9	Rozkład statusu szukania pracy wśród respondentów ankiety StackOverflow . .	26
2.10	Udział ofert pracy zdalnej w stosunku do wszystkich ofert pracy	27
3.1	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2017	35
3.2	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2018	35
3.3	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2019	36
3.4	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2017	36
3.5	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2018	37
3.6	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu regresji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2019	37

3.7	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2017	39
3.8	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2018	40
3.9	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSatisfaction dla ankiety StackOverflow z roku 2019	40
3.10	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2017	41
3.11	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2018	41
3.12	Średnia siła wpływu 10 pytań najsilniej wpływających na predykcję w modelu klasyfikacji dla zmiennej zależnej JobSeekingStatus dla ankiety StackOverflow z roku 2019	42

Spis kodów źródłowych

3.1	Algorytm uczenia modelu regresji	32
3.2	Parametry estmatora	33

A. Lista pytań w ankiecie StackOverflow

W sekcjach A.1, A.2, A.3 przedstawiono przetłumaczoną listę pytań wykorzystanych w przeprowadzonych w ramach niniejszej pracy badaniach. Przedstawione pytania pochodzą z ankiety deweloperskiej StackOverflow [Stack Overflow, 2011], pominięte jednak zostały pytania nieistotne z punktu widzenia prowadzonego badania.

A.1. Pytania z roku 2017

- **"AnnoyingUI"** - Czy denerwuje cię, jeśli aplikacja ma zły interfejs użytkownika?
- **"AssessJobCommute"** - Rozważając ofertę pracy, jak ważny dla ciebie jest: czas dojazdu do pracy?
- **"AssessJobCompensation"** - Rozważając ofertę pracy, jak ważne dla ciebie są: benefity pracownicze?
- **"AssessJobDept"** - Rozważając ofertę pracy, jak ważny dla ciebie jest: konkretny dział lub zespół, w którym będziesz pracować?
- **"AssessJobDiversity"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: różnorodność (ang. diversity) w firmie?
- **"AssessJobExp"** - Rozważając ofertę pracy, jak ważne dla ciebie jest: wymagane w ogłoszeniu doświadczenie?
- **"AssessJobFinances"** - Rozważając ofertę pracy, jak ważne dla ciebie jest: powodzenie finansowe firmy?
- **"AssessJobIndustry"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: gałąź przemysłu, w której operuje firma?
- **"AssessJobLeaders"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: reputacja seniorów pracujących w firmie?
- **"AssessJobOffice"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: organizacja biura?
- **"AssessJobProduct"** - Rozważając ofertę pracy, jak ważne dla ciebie jest: jak szeroko wykorzystywany będzie produkt wytwarzany przez firmę?
- **"AssessJobProfDevel"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: możliwość rozwoju osobistego?
- **"AssessJobProjects"** - Rozważając ofertę pracy, jak ważne dla ciebie jest: jak projekty są zarządzane?

- **"AssessJobRemote"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: możliwość pracy z domu lub zdalnie?
- **"AssessJobRole"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: rola lub nazwa stanowiska?
- **"AssessJobTech"** - Rozważając ofertę pracy, jak ważny dla ciebie jest: język lub technologia, w której będziesz pracować?
- **"BoringDetails"** - Czy nudzą cię szczegóły implementacyjne?
- **"BuildingThings"** - Czy budowanie rzeczy daje ci satysfakcję?
- **"CareerSatisfaction"** - Czy jesteś zadowolony ze swojej kariery?
- **"ChallengeMyself"** - Czy lubisz stawiać sobie wyzwania?
- **"ChangeWorld"** - Czy chcesz zmienić świat?
- **"CheckInCode"** - Jak często committujesz kod do systemu kontroli wersji?
- **"CollaborateRemote"** - Czy uważasz, że trudniej jest współpracować zdalnie ze swoimi współpracownikami?
- **"CompanySize"** - Ilu pracowników zatrudnia firma, w której pracujesz?
- **"CompanyType"** - Jaki jest typ firmy, w której pracujesz?
- **"CompetePeers"** - Czy lubisz rywalizować ze swoimi współpracownikami?
- **"ConvertedSalary"** - Jakie jest twoje wynagrodzenie w przeliczeniu na USD?
- **"Country"** - W jakim kraju mieszkasz?
- **"CousinEducation"** - Co polecisz osobie chcącej przebranżowić się na programistę?
- **"DatabaseWorkedWith"** - Z jakich baz danych korzystałeś w ciągu ostatniego roku?
- **"DevType"** - Jaki typ programisty najlepiej do ciebie pasuje?
- **"DifficultCommunication"** - Czy masz problem z komunikacją ze swoimi współpracownikami?
- **"DiversityImportant"** - Czy różnorodność (ang. diversity) w miejscu pracy jest dla ciebie ważna?
- **"EducationImportant"** - Czy uważasz, że edukacja formalna znacząco wpływa na sukces zawodowy?
- **"EducationTypes"** - Poza edukacją formalną, z jakich form edukacji korzystałeś?
- **"Employment"** - Jaki jest twój status zatrudnienia?
- **"EnjoyDebugging"** - Czy lubisz zajmować się debugowaniem kodu?
- **"EquipmentSatisfiedCPU"** - Biorąc pod uwagę twoje obecne stanowisko do pracy, w jakim stopniu jesteś zadowolony z: mocy obliczeniowej (CPU/GPU)?
- **"EquipmentSatisfiedMonitors"** - Biorąc pod uwagę twoje obecne stanowisko do pracy, w jakim stopniu jesteś zadowolony z: monitorów (liczba, wielkość, rozdzielczość)?
- **"EquipmentSatisfiedRAM"** - Biorąc pod uwagę twoje obecne stanowisko do pracy, w jakim stopniu jesteś zadowolony z: ilości pamięci RAM?
- **"EquipmentSatisfiedRW"** - Biorąc pod uwagę twoje obecne stanowisko do pracy, w jakim stopniu jesteś zadowolony z: prędkości zapisu/odczytu komputera?

- **"EquipmentSatisfiedStorage"** - Biorąc pod uwagę twoje obecne stanowisko do pracy, w jakim stopniu jesteś zadowolony z: pamięci komputera?
- **"FormalEducation"** - Jaki jest twój najwyższy ukończony poziom edukacji formalnej?
- **"FrameworkWorkedWith"** - Z jakich narzędzi deweloperskich, bibliotek i frameworków korzystałeś w ciągu ostatniego roku?
- **"FriendsDevelopers"** - Czy większość twoich znajomych to programiści, naukowcy albo inżynierowie?
- **"Gender"** - Z jaką płcią się identyfikujesz?
- **"HighestEducationParents"** - Jaki jest najwyższy poziom edukacji osiągnięty przez kogośkolwiek z twoich rodziców?
- **"HomeRemote"** - Jak często pracujesz z domu lub zdalnie?
- **"ImportantBenefits"** - Jakie benefity pozapłacowe są dla ciebie ważne?
- **"ImportantHiringAlgorithms"** - Według ciebie jak ważna w procesie rekrutacji jest: znajomość algorytmów i struktur danych?
- **"ImportantHiringCommunication"** - Według ciebie jak ważna w procesie rekrutacji jest: umiejętność komunikacji?
- **"ImportantHiringCompanies"** - Według ciebie jak ważne w procesie rekrutacji jest: w jakich firmach pracował kandydat?
- **"ImportantHiringEducation"** - Według ciebie jak ważne w procesie rekrutacji jest: formalne wykształcenie?
- **"ImportantHiringGettingThingsDone"** - Według ciebie jak ważna w procesie rekrutacji jest: historia dostarczania zobowiązań w terminie?
- **"ImportantHiringOpenSource"** - Według ciebie jak ważna w procesie rekrutacji jest: kontrybucja do projektów typu Open Source?
- **"ImportantHiringPMExp"** - Według ciebie jak ważne w procesie rekrutacji jest: doświadczenie w pracy z metodykami zarządzania projektem?
- **"ImportantHiringRep"** - Według ciebie jak ważna w procesie rekrutacji jest: reputacja na forum Stack Overflow?
- **"ImportantHiringTechExp"** - Według ciebie jak ważne w procesie rekrutacji jest: doświadczenie w pracy z narzędziami deweloperskimi wykorzystywanymi w firmie?
- **"ImportantHiringTitles"** - Według ciebie jak ważne w procesie rekrutacji jest: jakie kandydat zajmował wcześniej stanowiska?
- **"InfluenceCloud"** - Jak duży wpływ w swojej firmie masz na zakup: systemów chmurowych?
- **"InfluenceCommunication"** - Jak duży wpływ w swojej firmie masz na zakup: narzędzi do współpracy i komunikacji?
- **"InfluenceConsultants"** - Jak duży wpływ w swojej firmie masz na zakup: konsultacji?
- **"InfluenceDatabase"** - Jak duży wpływ w swojej firmie masz na zakup: systemów bazodanowych?

- **"InfluenceDeptTech"** - Jak duży wpływ w swojej firmie masz na zakup: rozwiązań technologicznych w swoim departamencie?
- **"InfluenceHardware"** - Jak duży wpływ w swojej firmie masz na zakup: sprzętu komputerowego dla innych do pracy?
- **"InfluenceInternet"** - Jak duży wpływ w swojej firmie masz na zakup: łącza internetowego?
- **"InfluenceRecruitment"** - Jak duży wpływ w swojej firmie masz na zakup: narzędzi dla rekruterów?
- **"InfluenceServers"** - Jak duży wpływ w swojej firmie masz na zakup: serwerów?
- **"InfluenceTechStack"** - Jak duży wpływ w swojej firmie masz na zakup: rozwiązań technologicznych w swojej firmie?
- **"InfluenceVizTools"** - "Jak duży wpływ w swojej firmie masz na zakup: narzędzi do analizy i wizualizacji danych?"
- **"InfluenceWorkstation"** - Jak duży wpływ w swojej firmie masz na zakup: sprzętu komputerowego dla siebie do pracy?
- **"InTheZone"** - Czy programując jesteś "w transie"?
- **"InvestTimeTools"** - Czy inwestujesz dużo czasu w dostosowanie narzędzi, których używasz?
- **"JobSatisfaction"** - Czy jesteś zadowolony ze swojej pracy?
- **"JobSecurity"** - Czy stabilność zatrudnienia jest dla ciebie ważna?
- **"JobSeekingStatus"** - Czy szukasz obecnie nowej pracy?
- **"KinshipDevelopers"** - Czy czujesz solidarność ze swoimi współpracownikami?
- **"LanguageWorkedWith"** - W jakich językach programowania pisałeś kod w ciągu ostatniego roku?
- **"LastNewJob"** - Kiedy ostatni raz zmieniałeś pracę?
- **"LearningNewTech"** - Czy nauka nowych rzeczy sprawia ci przyjemność?
- **"Methodology"** - Z jakimi metodykami zarządzania projektem miałeś okazję pracować?
- **"MetricAssess"** - Jakie metryki powinny być użyte do oceny pracy senior dewelopera?
- **"MobileDevType"** - Jaki typ programisty aplikacji mobilnych najlepiej do ciebie pasuje?
- **"OtherPeoplesCode"** - Czy uważasz, że utrzymywanie czyjegoś kodu jest formą tortury?
- **"Overpaid"** - Czy uważasz, że dostajesz lepsze wynagrodzenie niż zasługujesz?
- **"PlatformWorkedWith"** - Na jakie platformy tworzyłeś oprogramowanie w ciągu ostatniego roku?
- **"ProblemSolving"** - Czy lubisz rozwiązywać problemy?
- **"Professional"** - Czy zawodowo zajmujesz się programowaniem?
- **"ProgramHobby"** - Czy programujesz jako hobby lub kontrybuujesz kod do projektów typ Open Source?
- **"ProjectManagement"** - Czy uważasz, że większość technik zarządzania projektami jest bezużyteczna?
- **"SeriousWork"** - Czy poważnie podchodzisz do swojej pracy?

- **"ShipIt"** - Czy uważasz, że lepiej szybko dostarczyć kod, nawet jeśli nie jest zoptymalizowany?
- **"StackOverflowSatisfaction"** - Czy lubisz korzystać forum Stack Overflow?
- **"StackOverflowWhatDo"** - Czy uważasz, że nie byłbyś w stanie pracować bez forum Stack Overflow?
- **"UndergradMajor"** - W jakiej specjalizacji ukończyłeś edukację?
- **"University"** - Czy jesteś obecnie studentem na wyższej uczelni?
- **"VersionControl"** - Jakich systemów kontroli wersji używasz?
- **"WantWorkedDatabase"** - Z jakich baz danych chciałbyś korzystać w ciągu następnego roku?
- **"WantWorkedFramework"** - Z jakich narzędzi deweloperskich, bibliotek i frameworków chciałbyś korzystać w ciągu następnego roku?
- **"WantWorkedLanguage"** - W jakich językach programowania chciałbyś pisać kod w ciągu następnego roku?
- **"WantWorkedPlatform"** - Na jakie platformy chciałbyś tworzyć oprogramowanie w ciągu następnego roku?
- **"WebDevType"** - Jaki typ programisty aplikacji webowych najlepiej do ciebie pasuje?
- **"WorkPayCare"** - Czy nie przejmujesz się, na czym polega twoja praca pod warunkiem, że dostajesz odpowiednie wynagrodzenie?
- **"YearsCodingProf"** - Ile lat programujesz zawodowo?
- **"YearsProgram"** - Ile lat temu nauczyłeś się programować?

A.2. Pytania z roku 2018

- **"Age"** - Ile masz lat?
- **"AgreeDisagree1_kinshipToDevs"** - Czy czujesz solidarność ze swoimi współpracownikami?
- **"AgreeDisagree2_competingPeers"** - Czy lubisz rywalizować ze swoimi współpracownikami?
- **"AgreeDisagree3_worseThanPeers"** - Czy uważasz, że jesteś gorszym programistą niż twoi współpracownicy?
- **"AssessBenefits1_salary"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: wynagrodzenie i premie?
- **"AssessBenefits10_transport"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: zniżki na transport?
- **"AssessBenefits11_eduBudget"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: budżet edukacyjny?
- **"AssessBenefits2_stockOrShares"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: udziały lub akcje firmy?

- **"AssessBenefits3_healthInsurance"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: ubezpieczenie zdrowotne?
- **"AssessBenefits4_parentalLeave"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: urlop macierzyński/tacierzyński?
- **"AssessBenefits5_multisport"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: karta sportowa?
- **"AssessBenefits6_retirementPlan"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: oszczędności emerytalne?
- **"AssessBenefits7_mealsAndSnacks"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: posiłki zapewniane przez firmę?
- **"AssessBenefits8_computerAllowance"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: "kieszonkowe" na wydatki biurowe?
- **"AssessBenefits9_childcare"** - Spośród benefitów pracowniczych, jak ważne dla ciebie jest: opieka nad dzieckiem?
- **"AssessJob1_industry"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: gałąź przemysłu, w której operuje firma?
- **"AssessJob10_projectImpact"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: jak szeroko wykorzystywany będzie produkt wytwarzany przez firmę?
- **"AssessJob2_companyFinancialStatus"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: powodzenie finansowe firmy?
- **"AssessJob3_department"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: konkretny dział lub zespół, w którym będziesz pracować?
- **"AssessJob4_languages"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: język lub technologia, w której będziesz pracować?
- **"AssessJob5_compensation"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: benefity pracownicze?
- **"AssessJob6_companyCulture"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: organizacja biura lub kultura firmy?
- **"AssessJob7_remoteJob"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: możliwość pracy z domu lub zdalnie?
- **"AssessJob8_development"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: możliwość rozwoju osobistego?
- **"AssessJob9_diversity"** - Rozważając ofertę pracy, jak ważna dla ciebie jest: różnorodność (ang. diversity) w firmie?
- **"CareerSatisfaction"** - Czy jesteś zadowolony ze swojej kariery?
- **"CheckInCode"** - Jak często committujesz kod do systemu kontroli wersji?
- **"CompanySize"** - Ilu pracowników zatrudnia firma, w której pracujesz?
- **"ConvertedSalary"** - Jakie jest twoje wynagrodzenie w przeliczeniu na USD?
- **"Country"** - W jakim kraju mieszkasz?

- **"DatabaseDesireNextYear"** - Z jakich baz danych chciałbyś korzystać w ciągu następnego roku?
- **"DatabaseWorkedWith"** - Z jakich baz danych korzystałeś w ciągu ostatniego roku?
- **"Dependents"** - Czy masz dzieci albo inne osoby od ciebie zależne?
- **"DevType"** - Jaki typ programisty najlepiej do ciebie pasuje?
- **"EducationParents"** - Jaki jest najwyższy poziom edukacji osiągnięty przez któregoś z twoich rodziców?
- **"EducationTypes"** - Poza edukacją formalną, z jakich form edukacji korzystałeś?
- **"Employment"** - Jaki jest twój status zatrudnienia?
- **"ErgonomicDevices"** - Jakich urządzeń używasz do zwiększenia ergonomii pracy?
- **"EthicalImplications"** - Czy uważasz, że jako programista masz obowiązek rozważyć etyczne konsekwencje kodu, który piszesz?
- **"EthicsChoice"** - Jeśli dostaniesz zlecenie napisania nieetycznego kodu, to czy napiszesz go mimo wszystko?
- **"EthicsReport"** - Jeśli dostaniesz zlecenie napisania nieetycznego kodu, czy zgłosisz to komuś?
- **"EthicsResponsible"** - Na kim spoczywa odpowiedzialność za nieetyczny kod?
- **"Exercise"** - Ile razy w tygodniu wykonujesz ćwiczenia fizyczne?
- **"FormalEducation"** - Jaki jest twój najwyższy ukończony poziom edukacji formalnej?
- **"FrameworkDesireNextYear"** - Z jakich narzędzi deweloperskich, bibliotek i frameworków chciałbyś korzystać w ciągu następnego roku?
- **"FrameworkWorkedWith"** - Z jakich narzędzi deweloperskich, bibliotek i frameworków korzystałeś w ciągu ostatniego roku?
- **"Gender"** - Z jaką płcią się identyfikujesz?
- **"Hobby"** - Czy programujesz jako hobby?
- **"HopeFiveYears"** - Co planujesz robić za 5 lat?
- **"HoursComputer"** - Ile godzin dziennie spędzasz przy komputerze?
- **"HoursOutside"** - Ile godzin dziennie spędzasz na dworze?
- **"JobSatisfaction"** - Czy jesteś zadowolony ze swojej pracy?
- **"JobSeekingStatus"** - Czy szukasz obecnie nowej pracy?
- **"LanguageDesireNextYear"** - W jakich językach programowania chciałbyś pisać kod w ciągu następnego roku?
- **"LanguageWorkedWith"** - W jakich językach programowania pisałeś kod w ciągu ostatniego roku?
- **"LastNewJob"** - Kiedy ostatni raz zmieniałeś pracę?
- **"Methodology"** - Z jakimi metodykami zarządzania projektem miałeś okazję pracować?
- **"NumberMonitors"** - Ilu monitorów używasz w swojej stacji roboczej?
- **"OpenSource"** - Czy kontrybuujesz do projektów typu Open Source?
- **"OperatingSystem"** - Jakiego systemu operacyjnego używasz do pracy?

- **"PlatformDesireNextYear"** - Na jakie platformy chciałbyś tworzyć oprogramowanie w ciągu następnego roku?
- **"PlatformWorkedWith"** - Na jakie platformy tworzyłeś oprogramowanie w ciągu ostatniego roku?
- **"Salary"** - Jakie jest twoje wynagrodzenie?
- **"SkipMeals"** - Ile razy w tygodniu pomijasz posiłek, żeby nie przerywać pracy?
- **"StackOverflowVisit"** - Jak często odwiedzasz forum Stack Overflow?
- **"Student"** - Czy jesteś obecnie studentem na wyższej uczelni?
- **"TimeFullyProductive"** - Ile czasu potrzebowałby nowy członek zespołu, aby w pełni wdrożyć się do projektu, przy którym pracujesz?
- **"UndergradMajor"** - W jakiej specjalizacji ukończyłeś edukację?
- **"VersionControl"** - Jakich systemów kontroli wersji używasz?
- **"WakeTime"** - O której godzinie zwykle wstajesz w dni robocze?
- **"YearsCoding"** - Ile lat temu nauczyłeś się programować?
- **"YearsCodingProf"** - Ile lat programujesz zawodowo?

A.3. Pytania z roku 2019

- **"Age"** - Ile masz lat?
- **"Age1stCode"** - W jakim wieku napisałeś swoją pierwszą linię kodu?
- **"BetterLife"** - Czy uważasz, że współcześnie ludzie mają lepsze życie niż ich rodzice?
- **"CareerSatisfaction"** - Czy jesteś zadowolony ze swojej kariery?
- **"CodeRev"** - Czy wykonujesz review kodu swoich współpracowników?
- **"CodeRevHrs"** - Ile godzin tygodniowo poświęcasz na wykonywanie review kodu?
- **"CompanySize"** - Ilu pracowników zatrudnia firma, w której pracujesz?
- **"ConvertedSalary"** - Jakie jest twoje wynagrodzenie w przeliczeniu na USD?
- **"Country"** - W jakim kraju mieszkasz?
- **"DatabaseDesireNextYear"** - Z jakich baz danych chciałbyś korzystać w ciągu następnego roku?
- **"DatabaseWorkedWith"** - Z jakich baz danych korzystałeś w ciągu ostatniego roku?
- **"Dependents"** - Czy masz dzieci albo inne osoby od ciebie zależne?
- **"DevEnviron"** - Z jakich środowisk programistycznych korzystasz regularnie?
- **"DevType"** - Jaki typ programisty najlepiej do ciebie pasuje?
- **"EduOther"** - Poza edukacją formalną, z jakich form edukacji korzystałeś?
- **"Employment"** - Jaki jest twój status zatrudnienia?
- **"Extraversion"** - Preferujesz rozmowy na żywo czy poprzez czat online?
- **"FormalEducation"** - Jaki jest twój najwyższy ukończony poziom edukacji formalnej?
- **"Gender"** - Z jaką płcią się identyfikujesz?
- **"Hobbyist"** - Czy programujesz jako hobby?

- **"ImpSyn"** - Jak oceniasz swój poziom kompetencji zawodowych?
- **"JobFactors"** - Jakie czynniki związane z pracą są dla ciebie najważniejsze?
- **"JobSatisfaction"** - Czy jesteś zadowolony ze swojej pracy?
- **"JobSeekingStatus"** - Czy szukasz obecnie nowej pracy?
- **"LanguageDesireNextYear"** - W jakich językach programowania chciałbyś pisać kod w ciągu następnego roku?
- **"LanguageWorkedWith"** - W jakich językach programowania pisałeś kod w ciągu ostatniego roku?
- **"LastNewJob"** - Kiedy ostatni raz zmieniałeś pracę?
- **"MgrIncompetent"** - Czy uważasz, że twój menadżer wie co robi / jest kompetentny?
- **"MgrMoney"** - Czy uważasz, że musisz być menadżerem, żeby zarabiać więcej pieniędzy?
- **"MgrWant"** - Czy chciałbyś w przyszłości pracować na stanowisku menadżerskim?
- **"MiscTechDesireNextYear"** - Z jakich narzędzi deweloperskich, bibliotek i frameworków chciałbyś korzystać w ciągu następnego roku?
- **"MiscTechWorkedWith"** - Z jakich narzędzi deweloperskich, bibliotek i frameworków korzystałeś w ciągu ostatniego roku?
- **"OpenSourceQuality"** - Co sądzisz o jakości oprogramowania typu Open Source?
- **"OpenSourcer"** - Jak często kontrybuujesz do oprogramowania typu Open Source?
- **"OpSys"** - Jakiego systemu operacyjnego używasz do pracy?
- **"PlatformDesireNextYear"** - Na jakie platformy chciałbyś tworzyć oprogramowanie w ciągu następnego roku?
- **"PlatformWorkedWith"** - Na jakie platformy tworzyłeś oprogramowanie w ciągu ostatniego roku?
- **"PurchaseHow"** - Jak twoja firma podejmuje decyzje zakupowe związane z nowymi technologiami?
- **"PurchaseWhat"** - Jak duży wpływ masz w twojej firmie na decyzje zakupowe związane z nowymi technologiami?
- **"SOComm"** - Czy czujesz się jak członek społeczności forum Stack Overflow?
- **"SOFindAnswer"** - Ile razy w tygodniu znajdujesz odpowiedź na swój problem na forum Stack Overflow?
- **"SOHowMuchTime"** - Ile czasu zaoszczędziłeś dzięki forum Stack Overflow?
- **"SOPartFreq"** - Jak często zadawałeś pytania lub udzielałeś odpowiedzi na forum Stack Overflow?
- **"SOTimeSaved"** - Szybciej jest znaleźć odpowiedź na Stack Overflow czy w innych źródłach?
- **"SOVisitFreq"** - Jak często odwiedzasz forum Stack Overflow?
- **"SOVisitTo"** - W jakim celu odwiedzasz forum Stack Overflow?
- **"Student"** - Czy jesteś obecnie studentem na wyższej uczelni?
- **"Trans"** - Czy identyfikujesz się jako osoba transseksualna?
- **"UndergradMajor"** - W jakiej specjalizacji ukończyłeś edukację?

- **"UnitTests"** - Czy w twojej firmie regularnie pisane są testy jednostkowe?
- **"WebFrameDesireNextYear"** - Z jakich frameworków webowych chciałbyś korzystać w ciągu następnego roku?
- **"WebFrameWorkedWith"** - Z jakich frameworków webowych korzystałeś w ciągu ostatniego roku?
- **"WorkChallenge"** - Jakie są twoje największe przeszkody w osiągnięciu lepszej produktywności jako programista?
- **"WorkLoc"** - Gdzie chciałbyś pracować?
- **"WorkPlan"** - Jak ustrukturyzowana lub zaplanowana jest twoja praca?
- **"WorkRemote"** - Jak często pracujesz zdalnie?
- **"WorkWeekHrs"** - Ile godzin tygodniowo pracujesz?
- **"YearsCode"** - Ile lat temu nauczyłeś się programować?
- **"YearsCodingProf"** - Ile lat programujesz zawodowo?