

Studium Magisterskie

Kierunek Analiza Danych - Big Data

Karol Szerszeń  
 Nr albumu 71273

**Optymalizacja marketingu bezpośredniego z wykorzystaniem genetycznych sieci neuronowych**

Praca magisterska

napisana w Instytucie Ekonometrii

pod kierunkiem naukowym

Dr Michała Bernardellego

Warszawa 2017

Spis treści

[Wstęp 1](#_Toc506322360)

[1. Marketing bezpośredni 2](#_Toc506322361)

[1.1 Marketing bezpośredni 2](#_Toc506322362)

[1.2 Dotychczasowe podejścia 2](#_Toc506322363)

[2. Algorytmy genetyczne 4](#_Toc506322364)

[2.1 Idea algorytmów ewolucyjnych 4](#_Toc506322365)

[2.2 Klasyczny algorytm genetyczny 5](#_Toc506322366)

[2.3 Operatory 7](#_Toc506322367)

[2.4 Funkcja przystosowania 9](#_Toc506322368)

[2.5 Kodowanie 10](#_Toc506322369)

[3. Sieci Neuronowe 11](#_Toc506322370)

[3.1 Idea Sieci Neuronowych 11](#_Toc506322371)

[3.2 Topologia sieci 13](#_Toc506322372)

[3.3 Funkcje aktywacji 15](#_Toc506322373)

[3.4 Uczenie sieci 18](#_Toc506322374)

[4. Zbiór danych 20](#_Toc506322375)

[4.1 Źródło danych 20](#_Toc506322376)

[4.2 Charakterystyki zmiennych 20](#_Toc506322377)

[4.3 Przekształcenia zmiennych 25](#_Toc506322378)

[5. Modelowanie 27](#_Toc506322379)

[5.1 Optymalizowana topologia sieci neuronowej 27](#_Toc506322380)

[5.2 Implementacja algorytmu genetycznego 27](#_Toc506322381)

[5.3 Wyniki 30](#_Toc506322382)

[Zakończenie 35](#_Toc506322383)

## Wstęp

Celem pracy było porównanie genetycznych sieci neuronowych z obecnie stosowanym podejściem wykorzystującym klasyczną regresję logistyczną

## 1. Marketing bezpośredni

### 1.1 Pojęcie marketingu bezpośredniego

Współcześnie przeciętny konsument wystawiony jest dziennie na kilkaset do kilku tysięcy ofert marketingowych dziennie[[1]](#footnote-1), nie zdając sobie nawet z tego faktu sprawy a co za tym idzie ignorując je. Z tego powodu przyciągnięcie klienta za pomocą tradycyjnych, masowych kanałów promocji staje się coraz mniej popularne. Mianem nowej koncepcji marketingu został określony[[2]](#footnote-2) marketing bezpośredni, który w latach 90 XX. wieku zaczął się wyodrębniać jako autonomiczna dyscyplina teoretyczna i praktyczna spośród szerokiego grona technik promocji. Jak podaje Rzemieniak[[3]](#footnote-3) przez marketing bezpośredni określamy „system wzajemnego oddziaływania, który stosuje jeden lub wiele nośników reklamowych w celu skłonienia klienta do obserwowalnej reakcji, prowadzącej do zawarcia transakcji w dowolnym miejscu. Obejmuje całokształt działań marketingowych wykorzystujących wielofunkcyjne media reklamy bezpośredniej i bazę danych do zbudowania długotrwałej, indywidualnej i obopólnie korzystnej więzi z klientem”. Kanałami, które służą do kontaktu z klientem za pomocą tej techniki są coraz rzadziej poczta i ulotki a coraz częściej spersonalizowane wiadomości e-mail, powiadomienia *push* na stronach internetowych czy wiadomości sms. Wprost z powyższej definicji wynika, że główne cechy marketingu bezpośredniego to:[[4]](#footnote-4)

1. Zorientowanie na potrzeby indywidualnych klientów a nie ich grup
2. Znaczenie bazy danych o klientach
3. Budowanie relacji z najatrakcyjniejszymi klientami
4. Kierowanie nacisku na bezpośrednią komunikację z indywidualnym klientem

Staje się więc oczywiste, że ten rodzaj sposobu na docieranie do klienta jest preferowany jako przynoszący największą szansę na konwersję. Dodatkowo kanałem, który cieszy się największą popularnością jest wiadomość e-mail. Jednym z powodów tego stanu jest powszechny dostęp Internetu – jak podaje GUS[[5]](#footnote-5) w 2016 roku 80,4% polskich gospodarstw domowych posiadało dostęp do Internetu, a aż 97% internautów jest aktywnym użytkownikiem skrzynki pocztowej. Jeśli weźmiemy pod uwagę fakt, iż jest to też jeden z najtańszych sposobów na dotarcie klienta ze zwrotem z inwestycji około 4000%[[6]](#footnote-6) nie powinna dziwić jego popularność wśród marketingowców.

### 1.2 Analiza danych w marketingu

## 2. Algorytmy genetyczne

### 2.1 Idea algorytmów ewolucyjnych

Algorytmy ewolucyjne to techniki optymalizacyjne inspirowane analogiami biologicznymi ­– oparte są na idei doboru naturalnego i ewolucji osobników danej populacji pod wpływem zmieniających się czynników środowiska[[7]](#footnote-7). Presja tegoż środowiska stymuluje proces selekcji naturalnej prowadzący do promowania osobników posiadających lepsze predyspozycje i eliminacji tych z mniejszymi, tym samym zwiększając szanse przetrwania całej populacji. Idea algorytmów ewolucyjnych sięga lat 50 XX. wieku, ale matematyczne podstawy zostały opracowane przez Johna H. Hollanda[[8]](#footnote-8) a ich pierwszą wersję opublikował on w 1975 roku. W tej klasycznej już pracy przedstawił on modele pozwalające na opisanie nieliniowych interakcji z jakimi mamy do czynienia w procesie ewolucji i dostosowania do zmieniającego się środowiska. Techniki optymalizacyjne oparte na algorytmach ewolucyjnych używają podejścia opartego nie na pojedynczych rozwiązaniach ale na całych ich populacjach[[9]](#footnote-9) gdzie w pojedynczej iteracji bierze udział zbiór rozwiązań i „ewoluuje” w nowy zbiór rozwiązań w kolejnej iteracji. Gwiazda[[10]](#footnote-10) wyróżnia następujące zalety algorytmów ewolucyjnych:

* brak wymagań co do postaci rozwiązywanego problemu - nie jest wymagane podanie postaci funkcji celu a jedynie jej istnienie lub bardziej ogólnie - istnienie miary, która pozwala na wybranie lepszego z dwóch dostępnych rozwiązań, gdzie termin *lepsze* jest definiowany przez konkretne zagadnienie do którego algorytm jest wykorzystywany
* zdolność opuszczania lokalnych ekstremów - dzięki rozpatrywaniu populacji rozwiązań oraz mechanizmom samych algorytmów mimo znalezienia lokalnego ekstremum możliwe jest jego opuszczenie i kontynuowanie poszukiwania rozwiązania w innej przestrzeni
* uniwersalność stosowania - brak istnienia dedykowanego rozwiązania danego problemu nie jest przeszkodą do zastosowania algorytmu w jego klasycznej postaci

Dodatkowo algorytmy ewolucyjne operując na populacjach rozwiązań, a nie na pojedynczym, punktowym rozwiązaniu pozwalają na łatwe zastosowanie przetwarzania wielowątkowego co prowadzi do często lepszego czasu obliczeń w porównaniu do klasycznych metod optymalizacyjnych.

Natomiast do głównych wad algorytmów ewolucyjnych możemy zaliczyć[[11]](#footnote-11):

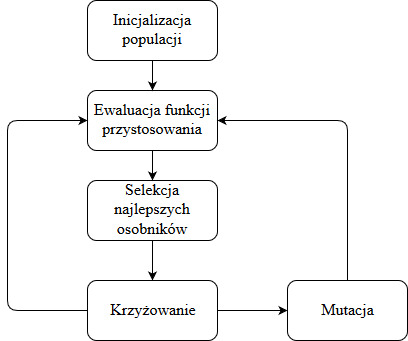
* uniwersalność prowadząca do mniejszej skuteczności niż w przypadku algorytmów dedykowanych
* większa złożoność obliczeniowa - algorytmy ewolucyjne są zwykle wolniejsze od metod zachłannych (obliczenia prowadzone na całych populacjach)
* niedeterministyczność - z założenia algorytm jest losowy, więc odtworzenie rozwiązania bywa niemożliwe
* postać funkcji celu ma duży wpływ na jakość rozwiązania

Algorytmy ewolucyjne są określeniem stosowanym do opisu klasy zagadnień bazujących na zbliżonej metodologii, obejmująca m.in.:

* algorytmy genetyczne
* programowanie genetyczne
* strategie ewolucyjne

W niniejszej pracy zastosowanie znajdzie głównie zagadnienie algorytmów genetycznych, głównie ze względu na ich możliwość przeszukiwania przestrzeni decyzyjnej w kilku punktach jednocześnie.

### 2.2 Klasyczny algorytm genetyczny

Algorytmy ewolucyjne charakteryzują się specyficzną nomenklaturą zaczerpniętą z pojęć stosowanych w dziedzinie biologii - genetyce. W klasycznym algorytmie genetycznym*[[12]](#footnote-12)* możemy wyróżnić struktury, zwane chromosomami (*ang. chromosomes)*, będące reprezentantami dziedziny optymalizowanej funkcji a składające się z mniejszych jednostek czyli genów. Zbiór struktur w poszczególnych etapach przetwarzania (w danej iteracji) tworzy populację. Określamy również funkcję przystosowania (*ang. fitness function)* która pozwala na mierzenie jakości dopasowania powyższych struktur podczas tworzenia nowej populacji. Najlepsze z nich „przeżywają” i uczestniczą w procesie reprodukcji przeprowadzanym przez mechanizm selekcji, natomiast pozostałe “wymierają” i nie są uwzględniane w przyszłych populacjach. Geny struktur wyselekcjonowanych do reprodukcji są wybierane przez operator krzyżowania   
i wymieniane pomiędzy rozwiązaniami rodzicielskimi. Z niewielkim prawdopodobieństwem poszczególne geny mogą zostać poddane mutacji - czyli wymienione na inne. Dzięki powyższym mechanizmom przeszukiwanie przestrzeni rozwiązań w kolejnych iteracjach zostaje zawężone do najbardziej obiecujących - potencjalnych - obszarów przestrzeni decyzyjnej a więc zbliżanie się do rozwiązania optymalnego. Zastosowanie algorytmu genetycznego wymaga również dążenia do zróżnicowania populacji, poprzez operator mutacji, co pozwala na wychodzenie z ekstremum lokalnego. Można to osiągnąć również poprzez maksymalne zróżnicowanie populacji początkowej - maksymalizacji entropii. 

Rysunek 1. Schemat klasycznego algorytmu genetycznego, Źródło: opracowanie własne

Powyższy diagram przedstawia działanie klasycznego algorytmu genetycznego - mutacja oznaczona jest jako opcjonalny krok w wyznaczaniu populacji potomków. Inicjalizacja populacji początkowej to tworzenie losowego zbioru na którym wykonywane będą dalsze kroki. Ważne jest, aby charakteryzowała się dużym zróżnicowaniem, gdyż efektywność początkowego algorytmu jest uzależniona właśnie od tego zbioru. Wielokrotne występowanie ciągów genów o wysokiej wartości funkcji przystosowania może prowadzić do całkowitego zdominowania rozwiązania przez właśnie te geny.

### 2.3 Operatory

#### 2.3.1 Selekcja

Operator selekcji pozwala na istnienie procesu określającego, które rozwiązania (chromosomy) przetrwają i będą miały szansę na przekazanie genów populacji potomnej, a które “wymrą”. Głównym celem tego operatora jest stawianie nacisku na wybór dobrze przystosowanych rozwiązań i eliminacja źle przystosowanych podczas utrzymywania wielkości populacji na stałym poziomie. Istnieje wiele rodzajów operatorów selekcji, ale ich wspólną charakterystyką jest próba zbalansowania przystosowania oraz różnorodności osobników w danej populacji[[13]](#footnote-13). Faworyzowanie przystosowania względem różnorodności - nacisku selektywnego (*ang. selective pressure*) - prowadzi do całkowitej dominacja jednego chromosomu nad innymi - zjawiska stłoczenia (*ang. crowding*). Metody eliminujące problem stłoczenia zaproponowali m.in. De Jong[[14]](#footnote-14) sugerując, że nowe osobniki powinny zastępować te najbardziej do nich podobne oraz Goldberg i Richardson[[15]](#footnote-15), którzy przyjęli funkcję podziału przystosowania, zmniejszającą wartość przystosowania danego chromosomu wprost proporcjonalnie do skali podobieństwa do innych chromosomów, tym samym nagradzając różnorodność.

Podstawowym i najczęściej stosowanymi operatorami selekcji są:

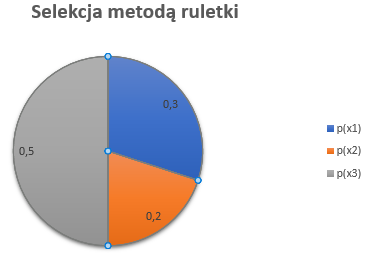
* metoda ruletki
* selekcja rankingowa
* selekcja turniejowa

W poniższej pracy zastosowana została metoda ruletki, która swą nazwę zawdzięcza analogii do losowania za pomocą koła kasynowej ruletki. Ogólny schemat tej metody wygląda następująco:

1. Obliczenie sumy wartości funkcji celu (przystosowania):

f(x) =

1. Obliczenie wkładu każdego osobnika w sumę: p() = f() / f(x)
2. Traktujemy wartości p() jako rozkład prawdopodobieństwa i n-krotnie losujemy osobniki zgodnie z tym rozkładem



Rysunek 2 Selekcja metodą ruletki, Źródło: opracowanie własne

Selekcja chromosomu może być postrzegana jako obrót koła ruletki, dzięki czemu zostaje wybrany chromosom, należący do wybranego fragmentu koła. Im większy jest fragment koła, tym prawdopodobieństwo wyboru jest większe a tym samym wprost proporcjonalne do wartości funkcji przystosowania danego chromosomu - im lepsze dane rozwiązanie tym częściej będzie wybierane.

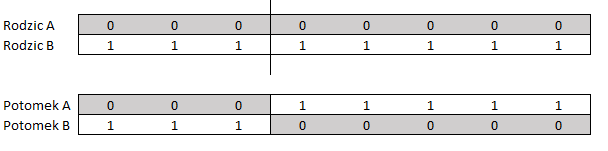
#### 2.3.2 Krzyżowanie

Krzyżowanie to proces wymiany informacji zapisanej w genach poszczególnych genotypów pochodzących od różnych rozwiązań rodzicielskich. Pozwala on na otrzymywanie nowych kombinacji cech - zarówno łącznie lepszych niż oba rozwiązania rodzicielskie jak i z możliwością bycia gorszym od każdego z nich. Krzyżowanie jest dominującym operatorem w algorytmach genetycznych - decyduje o konieczności utrzymania różnorodnej oraz licznej populacji. Istnieje wiele metod krzyżowania jednak w poniższej pracy zastosowany zostanie operator krzyżowania prostego, którego schemat został przedstawiony poniżej.

Krzyżowanie proste[[16]](#footnote-16) polega na wybraniu losowej liczby *n* - punktu krzyżowania. Następnie zastosowana jest następująca reguła:

* pierwszy potomek otrzymuje pierwsze *n* genów od pierwszego rozwiązania rodzicielskiego, natomiast pozostałe od drugiego
* drugi potomek otrzymuje pozostałe geny obu rozwiązań rodzicielskich czyli pierwsze *n* od drugiego oraz pozostałe od pierwszego

Poniższy rysunek obrazuje podziały chromosomów rozwiązań rodzicielskich oraz tworzenie nowych chromosomów przy pomocy operatora krzyżowania, z wartością   
*n* = 3:



Rysunek 3 Tworzenie nowych chromosomów w wyniku operatora krzyżowania, Żródło: opracowanie własne

#### 2.3.3 Mutacja

Operator mutacji polega na zmianie wartości jednego z genów danego chromosomu na inny, np. dla chromosomu [11001] mutacji może ulec gen trzeci w wyniku czego powstanie chromosom [11101]. Mutacja zachodzi z niewielkim prawdopodobieństwem *p*, ustalanym przy inicjalizacji algorytmu, najczęściej nieprzekraczającym kilku procent. Prawdopodobieństwo mutacji może również być modyfikowane w czasie - podobnie jak operator temperatury w innym algorytmie optymalizacyjnym - wyżarzaniu symulowanym[[17]](#footnote-17). Jedną z metod wyboru genów do mutacji jest wylosowanie liczby w przedziale [0, 1] dla każdego genu i zastosowanie mutacji dla tych, gdzie wylosowana liczba jest mniejsza od prawdopodobieństwa p.

### 2.4 Funkcja przystosowania

Funkcja przystosowania służy do oceny jakości przystosowania danego chromosomu - określa jak dobrze dany chromosom rozwiązuje dany problem. Pozwala ona na porównywanie poszczególnych rozwiązań pomiędzy sobą - jej wartość jest konieczna w przypadku zastosowania operatora selekcji. Funkcja przystosowania jest wybierana na podstawie problemu do którego jest stosowany algorytm - będzie ona inna przy problemie komiwojażera oraz inna przy optymalizacji wag sieci neuronowej. Funkcja przystosowania w problemie zastosowania algorytmu genetycznego do optymalizacji topologii sieci neuronowej będzie odzwierciedlać poprawność klasyfikacji przez sieć o danej topologii. Zostanie ona omówiona w dalszej części pracy.

### 2.5 Kodowanie

Jednym z najistotniejszych problemów zastosowania algorytmów genetycznych jest sposób kodowania chromosomów oraz pojedynczych genów. Zależy on głównie od optymalizowanego zagadnienia, co sprawia, że istnieje wiele sposobów kodowania chromosomów. Najpopularniejszym a jednocześnie najprostszym sposobem kodowania jest kodowanie binarne, polegające na zapisaniu chromosomu za pomocą ciągu zer i jedynek[[18]](#footnote-18), jak na przykład dla dwóch poniższych chromosomów A i B:

A = 1 0 0 1 0 1 0 0 1 1 1

B = 1 0 0 0 1 1 1 0 1 0 1

Taki sposób kodowania został opracowany jako pierwszy przez Hollanda[[19]](#footnote-19) oraz jego współpracowników, zatem i dla niego zostały przewidziane wszystkie operatory – selekcji, krzyżowania i mutacji.

Jednym z częściej stosowanych a podobnym do binarnego jest kodowanie za pomocą liczb rzeczywistych lub całkowitych[[20]](#footnote-20). W tym sposobie chromosomy również są ciągami liczb o równej długości, ale zakres liczbowy obejmuje więcej niż tylko dwie wartości:

A = 2 0 14 5 7 9 230

B = 3 11 9 5 83 1 39

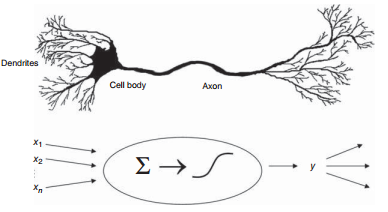
Dla wielu zastosowań jest to bardziej naturalny sposób reprezentacji rozwiązań niż kodowanie binarne. Wydajność algorytmu genetycznego może zależeć od sposobu reprezentacji genów[[21]](#footnote-21) jednak jest ona w dużej mierze zależna od danego problemu oraz pozostałych składowych algorytmu.

Poza tymi dwoma sposobami kodowań istnieje jeszcze wiele innych, jednak nie są one użyte w poniższej pracy, więc zostają pominięte.

## 3. Sieci Neuronowe

### 3.1 Idea Sieci Neuronowych

Sztuczne sieci neuronowe zostały stworzone aby odzwierciedlać strukturę ludzkiego mózgu w celu stworzenia sztucznej inteligencji. Ludzki umysł składa się z około 1011 neuronów, każdy połączony do średnio 10,000 innych neuronów, co daje około 1015 połączeń. Sieci neuronowe są próbą odtworzenia nieliniowego sposobu uczenia obecnych w sieciach neuronowych występujących w przyrodzie. Poniższy rysunek[[22]](#footnote-22) przedstawia prawdziwy neuron, który używa dendrytów do zbierania bodźców od innych neuronów i łączy informacje wejściowe, generując nieliniową odpowiedź i przesyłając ją przez akson do innych neuronów.



Rysunek 4 Schemat ludzkiego neuronu oraz prostej sieci neuronowej, Źródło: Data Mining and Predictive Analytics, Larose, D. T. , s.340

Rysunek przedstawia również postać sztucznego neuronu, który obecny jest w większości sieci neuronowych. Informacje wejściowe (xi) są przechwytywane z nadrzędnych warstw neuronów (lub zbioru danych w przypadku pierwszej warstwy), następnie zbierane przez funkcję kombinacji oraz przekształcane przez nieliniową funkcję aktywacji w celu stworzenia informacji wyjściowej, która jest przekazywana do kolejnej warstwy neuronów. Pierwszym prosty model neuronu został zaproponowany przez McCullocha i Pittsa w 1943 roku [[23]](#footnote-23)- udało im się matematycznie opisać pojedynczy neuron w sposób, który pozwalał na zrozumienie podstawowych procesów uczenia zachodzących w mózgu. Sukces tego podejścia polegał na prostocie rozwiązania, które wprawdzie nie było specjalnie potężne, ale dało narzędzie, które w połączeniu w sieć stały się osobną dziedziną nauki a ich możliwości przerosły oczekiwania twórców.

Do zalet sieci neuronowych możemy zaliczyć przede wszystkim ich ogólny charakter – nie ma dużego znaczenia jaki problem próbujemy rozwiązać, sieć będzie się sprawdzać tak samo dobrze. Dzięki tej ogólności można je stosować do rozwiązywania problemów, dla których nie istnieją wyspecjalizowane sposoby radzenia sobie z nimi lub są one zbyt trudne do rozwiązania metodami analitycznymi, np. zagadnienia modelowania cen na rynkach papierów wartościowych czy też zachowania silników indukcyjnych[[24]](#footnote-24). Dodatkowo sieci pozwalają na pominięcie długiego i często bardzo skomplikowanego procesu programowania danego rozwiązania zastępując je procesem uczenia sieci.

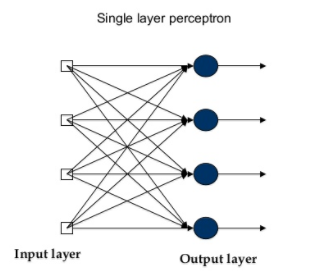
Podstawową wadą sieci neuronowych jest niewielka interpretowalność wyników. W przeciwieństwie do klasycznych metod regresji nie jesteśmy w stanie powiedzieć o ile zmieni się odpowiedź klasyfikacji, jeśli zmienimy wartość jednego z atrybutów. Nieco pomaga w interpretacji analiza wrażliwości, jednak daje ona tylko informacje o tym, który z parametrów wejściowych jest najbardziej istotny, natomiast nie pozwala na określenie jak bardzo zmieni się wartość wyjściowa.

Sieci neuronowe, tak jak wcześniej wspomniano, znalazły zastosowania w wielu dziedzinach nauki. Potharst, Kaymak i Pijls[[25]](#footnote-25) zastosowali sieci neuronowe do doboru grupy docelowej dla akcji charytatywnej. Natomiast Yang, Liu i Coid[[26]](#footnote-26) użyli ich jako jednej z technik do klasyfikacji poważnych przestępstw oraz określania prawdopodobieństwa recydywy wśród skazanych. Jednym z najpopularniejszych zastosowań sieci neuronowych jest przewidywanie cen instrumentów finansowych, ze względu na nieprzewidywalność rynków oraz ciężkie do określenia zależności, gdzie jako przykład można podać pracę Dunisa[[27]](#footnote-27), w której zastosował kilka hybrydowych rozwiązań łączących sieci neuronowe z algorytmami genetycznymi, uzyskując dobre rezultaty.

### 3.2 Topologia sieci

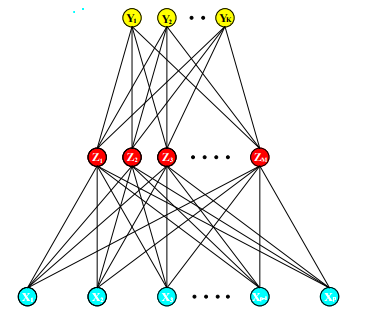
Topologia sieci neuronowych (architektura) to określenie obrazujące strukturę danej sieci czyli sposób połączenia poszczególnych neuronów oraz miejsc, w których wykonywane są obliczenia. Istnieją trzy podstawowe klasy architektury sieci[[28]](#footnote-28).

Najprostszą architekturą jest jednowarstwowa jednokierunkowa sieć (*ang.* single-layer feedforward network) gdzie neurony ułożone są w warstwy. W tej najprostszej formie możemy wyróżnić warstwę wejściową, która przesyła informacje bezpośrednio do warstwy wyjściowej, ale nigdy na odwrót. Nazwa „jednowarstwowa” odnosi się do jednej tylko warstwy wykonującej obliczenia – wyjściowej.



Rysunek 5. Jednowarstwowa jednokierunkowa sieć neuronowa, Źródło: https://www.slideshare.net/MohammedBennamoun/artificial-neural-network-lect4-single-layer-perceptron-classifiers

Drugą klasą sieci neuronowych, która odróżnia się od poprzedniej przez obecność warstwy *ukrytej* jest najczęściej spotykana wielowarstwowa jednokierunkowa sieć (ang*. multilayer feedforward network*, lub równoznacznie *Multi Layer Perceptron - MLP*), która zostanie użyta również w tej pracy. W tej architekturze warstwa ukryta, czy też neurony ukryte, odpowiada za większość obliczeń. Określenie *warstwa ukryta* odnosi się do faktu, iż neurony tej warstwy nie są widziane ani od strony wejściowej sieci, ani od warstwy wyjściowej. Główną funkcją jaką spełniają neurony warstwy ukrytej to przechwytywanie informacji z warstwy wejściowej oraz przekształcanie ich w sygnał wyjściowy za pomocą nieliniowych funkcji, zwanych funkcjami aktywacji. Sygnały wejściowe w warstwie wejściowej sieci dostarczają informacji o wektorze wejściowym, który jest przekształcany przez funkcję aktywacji i przekazywany do drugiej warstwy, którą w tym przypadku jest pierwsza warstwa ukryta. Sygnały wyjściowe z tej warstwy przekazywane są do kolejnej, przekształcane przez funkcję aktywacji i tak dalej, dla całej struktury sieci, aż do końcowej warstwy wyjściowej. Zestaw danych wyjściowych z ostatniej warstwy – warstwy wyjściowej – to ostateczny wynik, odpowiedź, sieci na problem zadany przez dane wejściowe. W ogólnym przypadku sieć posiadająca *m* źródeł danych, *h1* neuronów pierwszej warstwy ukrytej, *h2* neuronów drugiej warstwy ukrytej oraz *q* neuronów w warstwie wyjściowej definiujemy jako sieć *m - h1 - h2 - q*. Z reguły takie sieci są w pełni połączone, tzn. każdy neuron jest połączony z każdym innym neuronem warstwy poprzedzającej oraz następnej. Poniższy rysunek prezentuje właśnie taką architekturę, gdzie przez Xp oznaczone zostały neurony warstwy wejściowej, Zm neurony warstwy ukrytej oraz Yk to neurony warstwy wyjściowej.



Rysunek 6 Jednowarstwowa jednokierunkowa sieć neuronowa z warstwą ukrytą, Źródło:[[29]](#footnote-29)

Sieci rekurencyjne (ang. *recurrent networks*) różnią się od pozostałych architektur sposobem przepływu informacji – w tej sieci istnieje co najmniej jedna pętla przekazująca informacje między warstwą wyjściową a wejściową. Jeśli sieć posiada jedną warstwę wejściową i jedną wyjściową to sygnał z każdego neuronu warstwy wyjściowej jest przekazywany do pozostałych neuronów warstwy wejściowej. Pętla informacyjna (ang. *feedback loop)* może zawierać również przepływy informacji wyjściowej danego neuronu do niego samego (ang. *self-feedback loop)*. Obecność pętli informacyjnych ma znaczący wpływ na sposób uczenia się sieci oraz na jej wydajność i dokładność. Co więcej sieci rekurencyjne korzystają z gałęzi złożonych z elementów opóźnionych w czasie, co skutkuje nieliniowym zachowaniem, zakładając, że sieć zawiera nieliniowe elementy.

W związku z ogromnym zainteresowaniem sieciami neuronowymi w latach 90-tych XX wieku powstało wiele innych topologii, jednakże ich zastosowanie z reguły ograniczało się do rozwiązywania jednego, konkretnego problemu. Większość sieci neuronowych współcześnie stosowanych opiera się na powyższych trzech schematach, różniąc się nieznacznie od pierwowzoru takimi hiper-parametrami jak rodzaj funkcji aktywacji (sieci o radialnej funkcji aktywacji – *ang.* radial basis function networks[[30]](#footnote-30)), liczbą neuronów w warstwie ukrytej czy też liczbą warstw ukrytych.

Topologia sieci neuronowej z reguły jest wynikiem manualnego dostosowywania parametrów, określania liczebności warstw ukrytych oraz neuronów w każdej z nich. W związku z ogromną liczbą możliwych kombinacji połączenia już najprostszej sieci MLP proces ten jest żmudny i pochłaniający dużą ilość czasu. Jednym z popularniejszych i często używanych rozwiązań tego problemu jest zastosowanie algorytmu genetycznego do optymalizacji topologii, co też zostanie przedstawione w tej pracy.

### 3.3 Funkcje aktywacji

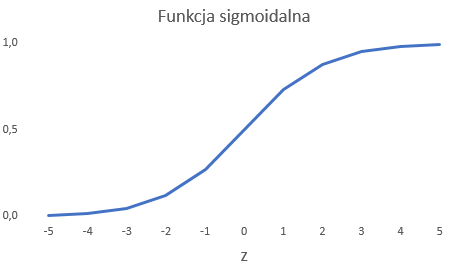
Funkcje aktywacji to funkcje służące do przekształcenia sygnału wejściowego neuronu do postaci nieliniowej. Dla każdej sieci można przypisać inne funkcje aktywacji dla poszczególnych warstw a nawet neuronów, natomiast w większości wypadków stosuje się jedną funkcję aktywacji dla całej sieci. Jako informację wejściową traktuje się sumę sygnałów wejściowych oraz przypisanych im wag[[31]](#footnote-31):

Funkcja taka musi spełniać kilka założeń[[32]](#footnote-32):

* nie może być stała
* musi być ograniczona
* musi być monotonicznie rosnąca
* oraz różniczkowalna

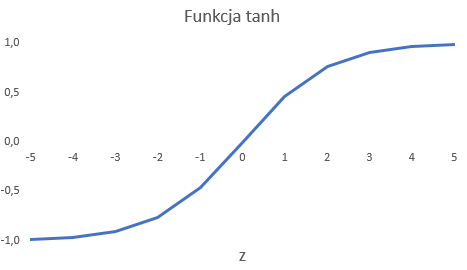
Powinna ona również odzwierciedlać rodzaj odpowiedzi czy też danych wyjściowych, których model ma dostarczać – jeśli zjawisko modelowane jest binarne takiej też odpowiedzi powinna dostarczyć funkcja. W przypadku klasyfikacji z kilkoma kategoriami również musi ona zwracać każdą z możliwych odpowiedzi. Biorąc pod uwagę powyższe własności współcześnie najczęściej stosowanymi funkcjami aktywacji są[[33]](#footnote-33):

* funkcja sigmoidalna:



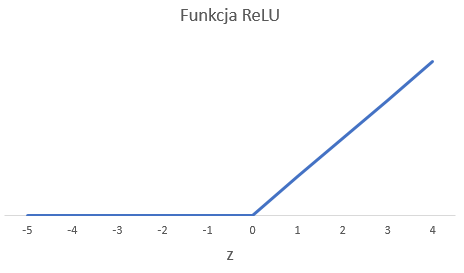
Rysunek 7 Funkcja sigmoidalna

* tangens hiperboliczny:



Rysunek 8 Funkcja tangens hiperboliczny

* ReLU (ang. *restricted linear unit*):



Rysunek 9 Funkcja ReLU

Poza faktem spełniania wyżej wymienionych założeń powyższe funkcje mają też inne zalety. Funkcja sigmoidalna dla *z* blisko zera wykazuje się liniowością, natomiast im *z* bardziej oddala się od zera tym wartości są bliższe odpowiednio 0 lub 1, stając się prawie funkcją stałą. Funkcja tangens hiperboliczny zachowuje się w bardzo podobny sposób, z tym że wartości skrajne to -1 i 1. Funkcja ReLU znajduje zastosowanie m.in. w przetwarzaniach obrazu, głównie ze względu na wydajność obliczeń. Jest ona też standardową funkcją w części zastosowań (np. w pakiecie scikit-learn[[34]](#footnote-34) języka Python).

### 3.4 Uczenie sieci

W zadaniach klasyfikacji przez model dobrze dopasowany do danych rozumie się model, który minimalizuje funkcję szacującą błąd klasyfikacji, zatem jest to zagadnienie optymalizacyjne. W przypadku sieci neuronowej błąd ten definiuje się jako[[35]](#footnote-35):

,

gdzie :

* – wartość rzeczywista *j* dla rekordu *d*
* – wartość *j* dla rekordu *d,* oszacowana przez neuron z wagami W.

Metodą, która służy do optymalizacji wag w sieci neuronowej do uzyskania minimalnej wartości błędu zdefiniowanego powyżej służy algorytm propagacji wstecznej (ang. *backpropagation algorithm*), przedstawiony po raz pierwszy w 1986 na łamach magazynu *Nature[[36]](#footnote-36).* Podstawą algorytmu jest informacja o tym, w jaki sposób zmienia się błąd E(W) przy zmianie wartości jednej wagi o jednostkę – na tej podstawie możliwe jest minimalizowanie błędu w danych wyjściowych[[37]](#footnote-37). Metoda propagacji wstecznej przeszukuje przestrzeń rozwiązań i ewaluuje zbiór wag w danym kroku na podstawie wartości funkcji E metodą spadku gradientowego (ang. *gradient descent*), który można przedstawić jako:

gdzie:

* - współczynnik szybkości uczenia, ,
* – sygnał wejściowy *i* do neuronu *j*

Jeśli neuron *j* należy do warstwy wejściowej to , w przeciwnym wypadku jest to wartość przekazana z poprzedzającej warstwy. Jeśli neuron *j* należy do warstwy wyjściowej to:

,

gdzie:

* – pochodna cząstkowa funkcji aktywacji *f* względem sumy wag .

Jeśli neuron *j* należy do warstwy ukrytej to:

,

gdzie to gradient wag warstwy wyjściowej.

Ze względu, że do obliczeń jako pierwsza wymagana jest wartość zmiana wag sieci zaczynana jest od ostatniej warstwy, przekazując tę wartość do warstw poprzednich w celu wyliczenia – czyli dokonywana jest propagacja wsteczna tej wartości, stąd nazwa algorytmu. Na podstawie powyższych równań aktualizowane są wagi dla sieci w iteracji *k*:

Kryterium stopu algorytmu jest osiągnięcie ustalonej z góry liczby iteracji lub ustalonego poziomu błędu bliskiemu zeru. Oczywistą wadą algorytmu jest złożoność obliczeniowa związana z koniecznością obliczania pochodnych cząstkowych oraz wymagania, przedstawione we wcześniejszym podrozdziale, co do funkcji aktywacji.

## 4. Zbiór danych

### 4.1 Źródło danych

Zbiór danych został udostępniony przez firmę Loyalty Partner Polska Sp. z o.o., której głównym produktem jest program lojalnościowy PAYBACK. Dane pochodzą z akcji marketingowej obejmującej wysyłkę e-mail informującej o dostępnych ofertach wyróżnionych sklepów oraz związanych z nimi promocjach. Głównym celem akcji było skłonienie do zakupu w jednym ze sklepów promowanych przez program.

Zmienne występujące w zbiorze opisują uczestników programu w dniu poprzedzającym rozpoczęcie akcji marketingowej jaką było wysyłanie wiadomości  
e-mail. Każdy wiersz odpowiada jednemu potencjalnemu odbiorcy.

### 4.2 Charakterystyki zmiennych

Zbiór składa się z 226,272 obserwacji, z czego 1772 obserwacje zostały oznaczone jako zdarzenia “pozytywne” czyli zaledwie 0.7%. Definicja zdarzenia pozytywnego to otrzymanie wiadomości e-mail, otworzenie jej oraz dokonanie zakupu w ciągu 14 dni w jednym ze sklepów komunikowanych w tej wysyłce. Poniżej przedstawiona zostanie analiza eksploracyjna wybranych zmiennych.

W zbiorze występuje łącznie 96 zmiennych, z czego 5 ma charakter tekstowy natomiast 91 numeryczny. Zmienna *event* to zmienna celu - oznacza wystąpienie badanego zdarzenia czyli dokonanie zakupu w jednym ze sklepów w 14 dni po otrzymaniu wiadomości e-mail. Załącznik nr 1 przedstawia spis wszystkich zmiennych, wraz z ich typem oraz opisem.

Ogół zmiennych można podzielić na kilka kategorii. Te oznaczone numerami 1 do 21 charakteryzują transakcyjność danego uczestnika – liczbę punktów zebranych w programie (*TRX\_all\_L12M\_PUR\_PTS*) czy kwotę wydaną na zakupy w sklepach Online w ciągu ostatnich 12 miesięcy (*TRX\_aff\_L12M\_PUR\_AMT*). Wszystkie one mają charakter numeryczny. Nie występują braki danych, natomiast kilka kolumn przyjmuje tylko wartości 0. Poniższa tabela prezentuje te zmienne oraz liczbę obserwacji, które przyjmują wartość inną niż zero. Można dostrzec, że zmienne związane ze spalaniem punktów w sklepach Online (*TRX\_aff\_L12M\_RED\_CNT, TRX\_aff\_L12M\_RED\_PTS, TRX\_aff\_L3M\_RED\_CNT, TRX\_aff\_L3M\_RED\_PTS)* dla wszystkich obserwacji są wypełnione zerami, więc zostaną one usunięte ze zbioru danych, jako nie wnoszące żadnej informacji.

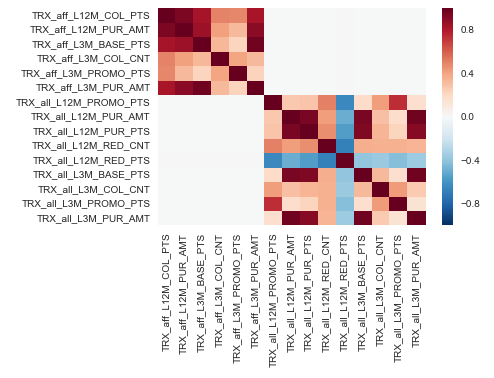
Tabela Zmienne 1-21 występujące w zbiorze danych wraz z liczebnością obserwacji róznych od zera

|  |  |
| --- | --- |
| **Nazwa zmiennej** | **Liczebność** |
| TRX\_aff\_L12M\_COL\_PTS | 4340 |
| TRX\_aff\_L12M\_PUR\_AMT | 3843 |
| TRX\_aff\_L12M\_RED\_CNT | 0 |
| TRX\_aff\_L12M\_RED\_PTS | 0 |
| TRX\_aff\_L3M\_BASE\_PTS | 3469 |
| TRX\_aff\_L3M\_COL\_CNT | 3997 |
| TRX\_aff\_L3M\_PROMO\_PTS | 3010 |
| TRX\_aff\_L3M\_PUR\_AMT | 3490 |
| TRX\_aff\_L3M\_RED\_CNT | 0 |
| TRX\_aff\_L3M\_RED\_PTS | 0 |
| TRX\_all\_L12M\_PROMO\_PTS | 12900 |
| TRX\_all\_L12M\_PUR\_AMT | 18789 |
| TRX\_all\_L12M\_PUR\_PTS | 18818 |
| TRX\_all\_L12M\_RED\_CNT | 7451 |
| TRX\_all\_L12M\_RED\_PTS | 7441 |
| TRX\_all\_L3M\_BASE\_PTS | 14454 |
| TRX\_all\_L3M\_COL\_CNT | 14764 |
| TRX\_all\_L3M\_PROMO\_PTS | 8463 |
| TRX\_all\_L3M\_PUR\_AMT | 14494 |
| TRX\_all\_L3M\_RED\_CNT | 3441 |
| TRX\_all\_L3M\_RED\_PTS | 3435 |

Powyższe zmienne charakteryzują się również wysoką skośnością, co prezentuje poniższa tabela zawierająca podstawowe statystyki opisowe. Zawarto również 95-ty percentyl dla odpowiedniej ilustracji asymetryczności.

Tabela Statystyki opisowe zmiennych transakcyjnych

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Nazwa Zmiennej** | **Średnia** | **Odchylenie standardowe** | **Minimum** | **50%** | **95%** | **Maksimum** |
| **TRX\_aff\_L12M\_COL\_PTS** | 62 | 1568 | 0 | 0 | 0 | 381560 |
| **TRX\_aff\_L12M\_PUR\_AMT** | 51 | 2491 | 0 | 0 | 0 | 762614 |
| **TRX\_aff\_L3M\_BASE\_PTS** | 14 | 707 | 0 | 0 | 0 | 213603 |
| **TRX\_aff\_L3M\_COL\_CNT** | 0 | 2 | 0 | 0 | 0 | 542 |
| **TRX\_aff\_L3M\_PROMO\_PTS** | 11 | 183 | 0 | 0 | 0 | 23400 |
| **TRX\_aff\_L3M\_PUR\_AMT** | 19 | 1265 | 0 | 0 | 0 | 394954 |
| **TRX\_all\_L12M\_PROMO\_PTS** | 150 | 1315 | 0 | 0 | 144 | 219079 |
| **TRX\_all\_L12M\_PUR\_AMT** | 541 | 9924 | 0 | 0 | 1208 | 4196505 |
| **TRX\_all\_L12M\_PUR\_PTS** | 188 | 3056 | -83319 | 0 | 486 | 1201205 |
| **TRX\_all\_L12M\_RED\_CNT** | 0 | 1 | 0 | 0 | 0 | 55 |
| **TRX\_all\_L12M\_RED\_PTS** | -342 | 3169 | -310235 | 0 | 0 | 3499 |
| **TRX\_all\_L3M\_BASE\_PTS** | 48 | 1258 | -157 | 0 | 83 | 555835 |
| **TRX\_all\_L3M\_COL\_CNT** | 1 | 7 | 0 | 0 | 3 | 1947 |
| **TRX\_all\_L3M\_PROMO\_PTS** | 43 | 433 | 0 | 0 | 0 | 45890 |
| **TRX\_all\_L3M\_PUR\_AMT** | 142 | 4203 | 0 | 0 | 204 | 1907204 |
| **TRX\_all\_L3M\_RED\_CNT** | 0 | 0 | 0 | 0 | 0 | 51 |
| **TRX\_all\_L3M\_RED\_PTS** | -128 | 1630 | -283403 | 0 | 0 | 0 |



Rysunek Wykres korelacji pomiędzy zmiennymi opisującymi transakcyjność uczestników

Powyższy wykres przedstawia korelacje pomiędzy zmiennymi opisującymi transakcyjność uczestników programu. Można zauważyć, że zmiennie opisujące aktywność w sklepach online (z przedrostkiem *TRX\_aff)* posiadają zerową korelacje ze zmiennymi opisującymi transakcyjność w całym programie. Oznaczałoby to, że z zmienne opisujące całość transakcji danej osoby nie zawierają informacji o aktywności w sklepach online. Istnieje więc silne podejrzenie, że zmienne te nie są poprawne więc nie zostaną one użyte w modelowaniu.

Druga grupa zmiennych, od 22 do 30, opisuje stan wirtualnego konta danego uczestnika programu – stan salda punktowego (*ACCT\_TOT\_COL\_AMT*) czy liczba kart płatniczych przypisanych do konta (*CARDA\_PAYM\_CNT*). Wszystkie zmienne z tej kategorii są typu numerycznego, nie występują również braki danych. Poniższa tabela przedstawia dla nich podstawowe statystyki opisowe.

Tabela Statystyki opisowe zmiennych charakteryzujących konto uczestnika

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Nazwa zmiennej** | **Średnia** | **Odchylenie** | **Min** | **50%** | **75%** | **Max** |
| **ACCT\_BAL\_MOD\_CNT** | 457 | 572 | 0 | 298 | 645 | 56182 |
| **ACCT\_TOT\_COL\_AMT** | 31317 | 55256 | 0 | 14968 | 39132 | 4455410 |
| **ACCT\_UNBLOCKED\_POINTS** | 4619 | 8757 | 0 | 1759 | 5266 | 831637 |
| **TIME\_FROM\_ENRL** | 1885 | 831 | 31 | 2105 | 2696 | 2755 |
| **TIME\_FROM\_FIRST\_DT** | 1999 | 799 | 32 | 2304 | 2718 | 2758 |
| **TIME\_FROM\_LAST\_DT** | 174 | 375 | 1 | 22 | 137 | 2737 |
| **CARDA\_PAYM\_CNT** | 0 | 1 | 0 | 0 | 0 | 33 |
| **ACCT\_PTS\_EXPIRED** | -330 | 1605 | -99039 | 0 | 0 | 0 |
| **ACCT\_PTS\_TO\_EXPIRE** | 124 | 681 | 0 | 0 | 0 | 88779 |

Kolejną grupę stanowią zmienne demograficzne i geo-lokalizacyjne – są to na przykład wiek uczestnika (*CUST\_AGE)*, jego płeć (*CUST\_GENDER)* oraz liczba ludności miejscowości zamieszkania (*CUST\_GFK\_POPULATION)*. W tej kategorii znajdują się zmienne deklaratywne – są to wspomniane już wiek i płeć oraz wykonywany zawód (*CUST\_PROFESSION\_CD)*. Dekodowanie wartości poszczególnych zmiennych przedstawia poniższa tabela.

Tabela Zmienne deklaratywne nominalne i ich dekodowanie

|  |  |
| --- | --- |
| **Nazwa zmiennej** | **Dekodowane wartości** |
| CUST\_PROFESSION\_CD | 1 - Manager,  2 - Samozatrudniony,  3 - Pracownik biurowy,  4 - Pracownik fizyczny, 5 - Pracuje w domu,  6 - Nie pracuje,  7 - Student, 99 - brak danych |
| CUST\_GENDER | 1 - mężczyzna,  2 - kobieta,  0 – nieokreślona |

Niestety zmienna określająca zawód uczestnika, jako deklaratywna, posiada znaczną ilość braków danych (oznaczonych liczbą 99), co prezentuje poniższa tabela. W związku z tym faktem zmienna ta również zostanie usunięta ze zbioru.

Tabela Kategorie zmiennej CUST\_PROFESSION\_CD

|  |  |  |
| --- | --- | --- |
| Kategoria | Liczebność | Procent zbioru |
| 99 | 154576 | 68% |
| 3 | 33829 | 15% |
| 2 | 11529 | 5% |
| 4 | 9463 | 4% |
| 1 | 7728 | 3% |
| 7 | 4417 | 2% |
| 6 | 3077 | 1% |
| 5 | 1653 | 1% |

Zmienne geo-lokalizacyjne obejmują zmienne binarne informujące o położeniu miejsca aktywności uczestnika w obszarze aktywności danego sklepu oraz odległość od centrum takiego obszaru, co prezentuje poniższa tabela.

Tabela Zmienne geolokalizacyjne

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Nazwa zmiennej** | **Średnia** | **Odchylenie standardowe** | **Min** | **50%** | **75%** | **Max** |
| **CUST\_CA\_AAA\_FLG** | 0,603451 | 0,489182 | 0 | 1 | 1 | 1 |
| **CUST\_CA\_BBB\_FLG** | 0,967822 | 0,176473 | 0 | 1 | 1 | 1 |
| **CUST\_CA\_CCC\_FLG** | 0,630719 | 0,482611 | 0 | 1 | 1 | 1 |
| **CUST\_CA\_BBB\_KM\_CNT** | 5,735212 | 32,123715 | 0 | 1,99 | 5,14 | 999 |
| **CUST\_CA\_AAA\_KM\_CNT** | 17,17591 | 37,780421 | 0 | 6,34 | 19,81 | 999 |
| **CUST\_CA\_CCC\_KM\_CNT** | 23,45034 | 43,725027 | 0 | 6,61 | 34,41 | 999 |

Ostatnią kategorię stanowią zmienne charakteryzujące aktywność uczestnika na stronie internetowej programu PAYBACK, w kilku jej sekcjach, w różnych przedziałach czasowych. Stanowią one najliczniejszą grupę – aż 50 zmiennych. Poniższa tabela przedstawia zmienne reprezentujące wszystkie sekcje w przedziale czasowym ostatnich 6 miesięcy.

Tabela Zmienne opisujące aktywność uczestnika na stronie internetowej programu w ciągu ostatnich 6 miesięcy

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Nazwa zmiennej** | **Średnia** | **Odchylenie standardowe** | **Min** | **25%** | **50%** | **75%** | **Max** |
| **WEB\_ALL\_TIME\_SPEND\_L6M** | 3286 | 15182 | 0 | 0 | 298 | 3111 | 3149094 |
| **WEB\_AFF\_TIME\_SPEND\_L6M** | 711 | 4070 | 0 | 0 | 0 | 151 | 731304 |
| **WEB\_REW\_TIME\_SPEND\_L6M** | 1079 | 4156 | 0 | 0 | 0 | 799 | 1302762 |
| **WEB\_OCC\_TIME\_SPEND\_L6M** | 463 | 2756 | 0 | 0 | 2 | 297 | 867372 |
| **WEB\_HIS\_TIME\_SPEND\_L6M** | 227 | 2453 | 0 | 0 | 0 | 60 | 811922 |

W tej kategorii nie występują braki danych. Można zauważyć że i w tej kategorii występuje wysoka skośność a drugi kwartyl stanowią uczestnicy mało aktywni – nie będący ani razu na żadnej z sekcji strony internetowej Programu. Ostatnie 2 zmienne w zbiorze to numer oferty na którą odpowiedziała dana osoba oraz kategoria tej oferty. Ze względu na to, iż są one przypisane tylko do osób, które odpowiedziały na mailing nie będą one użyte w modelu.

### 4.3 Przekształcenia zmiennych

Sieci neuronowe to zbiór modeli, wymagający zbioru danych w postaci numerycznej – takie też są wszystkie zmienne w zbiorze. Jak wynika z porównania statystyk opisowych zakres wartości dla wszystkich zmiennych jest bardzo duży – pomiędzy 0 i 1 dla jednych, aż po kilkadziesiąt tysięcy poniżej lub powyżej zera dla innych. Aby ograniczyć wpływ tak wysokich wartości na modele za pomocą modułu *preprocessing* z pakietu *sklearn* przeprowadzona została standaryzacja do procesu Gaussowskiego o średniej 0 i odchyleniu standardowym 1, według poniższego wzoru[[38]](#footnote-38):

Takie przeskalowanie pomaga zapewnić równe traktowanie danych wejściowych względem wag w procesie uczenia oraz bardziej jednorodny wybór ich początkowego zakresu[[39]](#footnote-39). W zbiorze nie występują zmienne kategoryczne, więc nie ma potrzeby przeprowadzenia kodowania takich zmiennych, co jest zwyczajową procedurą w przygotowywaniu zbioru dla sieci neuronowych.

Analizowany zbiór składa się z zaledwie 0.7% pozytywnych zdarzeń. Taki rozkład zmiennej celu oznacza, że model klasyfikujący wszystkie zdarzenia jako 0 osiągnie ponad 99% skuteczność. W celu zniwelowania tej niepożądanej właściwości zastosowane zostało zbilansowanie zbioru za pomocą próbkowania losowego[[40]](#footnote-40). W tym celu wykorzystany został algorytm *RandomUnderSampler* z pakietu *imblearn[[41]](#footnote-41)*. Po tym przekształceniu liczebność zbioru to 5746, gdzie liczba obserwacji o kategorii 1 wynosi 1326 czyli około 30% liczebności nowego zbioru. Natomiast liczebność obserwacji o kategorii 0 to 4420 czyli około 70% zbioru.

## 5. Modelowanie

### 5.1 Optymalizowana topologia sieci neuronowej

W poniższej pracy algorytmy genetyczne zostały użyte do optymalizacji topologii sieci neuronowej. Implementacja rozwiązania została stworzona w języku programowania Python. Do stworzenia sieci został wykorzystany moduł *scikit-learn* z jego implementacją perceptronu wielowarstwowego (*MLPClassifier*)*[[42]](#footnote-42)*. Implementacja ta pozwala na określenie parametrów sieci neuronowej, z których te najważniejsze stały się przedmiotem optymalizacji. Są to:

* Funkcja aktywacji
* Rodzaj współczynnika uczenia sieci
* Wartość startowego współczynnika uczenia
* Liczba warstw ukrytych
* Liczba neuronów w warstwie ukrytej

Wśród optymalizowanych funkcji aktywacji znalazły się funkcja logistyczna, tangens hiperboliczny oraz ReLU, które zostały omówione w rozdziale 3.3 niniejszej pracy. Rodzaj współczynnika uczenia określony został jako stały (ang. *constant)* lub zmniejszający się wraz z przebiegiem procesu uczenia (ang. *invscaling*). Trzecim parametrem jest wartość startowego współczynnika uczenia, który może przyjmować wartości od 5e-4 do 5e-1. Liczba warstw ukrytych przyjmuje wartości od 1 do 5, natomiast liczba neuronów w warstwie od 1 do 128. Wartości te zostały przyjęte po części na podstawie ogólnej praktyki[[43]](#footnote-43) mówiącej o całkowitej liczbie neuronów w warstwie ukrytej sieci pomiędzy 5 a 100, a po części w celu sprawdzenia czy dużo wyższa liczba neuronów może prowadzić do wyższej mocy predykcyjnej. Powyższe wartości parametrów pozwalają na osiągnięcie od 1 do 640 neuronów w warstwie ukrytej.

### 5.2 Implementacja algorytmu genetycznego

Podstawową kwestią związaną z implementacją algorytmu genetycznego było kodowanie parametrów sieci – czyli takie ich przedstawienie, jako pojedynczego chromosomu, aby możliwe było zastosowanie wszystkich operatorów algorytmu. Ze względu na liczbę parametrów sieci oraz ich postać (liczby od 0 do 1001) przyjęte zostało kodowanie za pomocą liczb naturalnych. Za gen została uznana liczba reprezentująca dany parametr sieci, przyjmująca odpowiedni zakres. Przykładowy chromosom przedstawiony został poniżej:

[ 1, 0, 714, 3, 80]

Chromosom reprezentowany jest przez listę zawierającą liczby w zakresie zależnym od danego parametru. Pierwszy element to rodzaj funkcji aktywacji – gdzie 1 oznacza funkcję logistyczną, 2 – tangens hiperboliczny, 3 – ReLU. Drugi element listy to rodzaj współczynnika uczenia – 0 reprezentuje stały współczynnik, natomiast 1 – zmniejszający się w trakcie uczenia. Trzeci element to początkowa wartość współczynnika uczenia – tutaj reprezentowana przez liczby od 1 do 1000. Ta wartość zostaje przekształcona przez poniższe równanie w celu otrzymania pożądanych wartości z przedziału 5e-4 do 5e-1:

,

gdzie A to trzeci element chromosomu.

Czwarty element listy to liczba warstw ukrytych, natomiast piąty liczba neuronów w każdej warstwie.

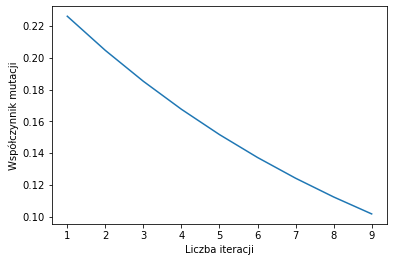
Mutacja została zaimplementowana na poziomie chromosomu, nie na poziomie pojedynczego genu tzn. dla konkretnego chromosomu sprawdzano czy należy dokonać mutacji. Jeśli tak to został zastosowany operator mutacji na jednym z genów danego chromosomu. Współczynnik mutacji został określony następująco:

,

gdzie *i* to numer obecnej generacji.

Taka implementacja została zapożyczona z algorytmu wyżarzania symulowanego i stosowanego tam współczynnika temperatury, który maleje wraz z liczbą iteracji. Współczynnik mutacji pełni rolę sposobu na wychodzenie z minimum lokalnego, ale może też powodować uzyskiwanie nieoptymalnych rozwiązań, jeśli obecne optymalne rozwiązanie będzie zmieniać się zbyt często.

Dzięki takiej definicji prawdopodobieństwo mutacji maleje wraz z liczbą iteracji algorytmu co przedstawia poniższy wykres:



Rysunek Wykres prawdopodobieństwa mutacji wraz ze wzrostem liczby iteracji algorytmu

Kolejną istotną częścią algorytmu genetycznego jest funkcja przystosowania. W przypadku modelowania zagadnienia związanego z marketingiem bezpośrednim, czy innym opartym o zbiór z niską liczbą odpowiedzi pozytywnych, określenie sukcesu jako klasyfikacja największej ilości obserwacji poprawnie nie miałaby sensu, gdyż model przypisałby wszystkie obserwacje jako zdarzenie negatywne i osiągnął 99.2% dokładność. Dlatego też funkcją przystosowania stała się funkcja zysku (straty) zdefiniowana jako iloczyn wartości danej kategorii zaklasyfikowania i liczby obserwacji zaklasyfikowanych do tej kategorii a reprezentowana przez poniższą macierz kosztu:

Tabela Macierz kosztu dla klasyfikacji odpowiedzi na kampanię marketingową

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Wynik | Klasyfikacja | Rzeczywista odpowiedź | Zysk | Uzasadnienie |
| True Positive | Odpowiedź | Odpowiedź | 19.9 | Zysk minus koszt wysyłki |
| False Positive | Odpowiedź | Brak odpowiedzi | 0.1 | Koszt wysyłki |
| True Negative | Brak odpowiedzi | Brak odpowiedzi | 0 | Brak kontaktu, brak straconego zysku |
| False Negative | Brak odpowiedzi | Odpowiedź | -20 | Stracony zysk |

Wartości przypisane do poszczególnych kategorii nie odzwierciedlają prawdziwych kosztów wysyłki w firmie PAYBACK. Są one umowne i mają na celu reprezentowanie faktu, że zysk z odpowiedzi na wysyłkę jest znacznie większy niż koszt wysłania mailingu. Tak więc funkcja przystosowania będzie mieć następującą postać:

Taka postać pozwala na promowanie topologii sieci, które poprawnie klasyfikują zdarzenia pozytywne oraz optymalizuje zysk z kampanii marketingowej.

W zastosowanym podejściu przeprowadzono 10 iteracji algorytmu genetycznego (czyli 10 generacji), gdzie w każdej iteracji populacja składała się z 10 chromosomów – sieci neuronowych – co daje łącznie trening 100 modeli.

### 5.3 Wyniki

Poniżej przedstawione zostaną wyniki modelowania odpowiedzi na kampanię marketingową a zatem również weryfikacja hipotezy o przydatności ewolucyjnych sieci neuronowych w badanym zjawisku.

#### 5.3.1 Model odniesienia

Standardowym modelem używanym do optymalizacji kampanii marketingowych jest model regresji logistycznej[[44]](#footnote-44), głównie ze względu na swoją prostotę, niewielkie zapotrzebowanie na moc obliczeniową oraz łatwą możliwość interpretacji uzyskanych wyników. Do trenowania i testowania modelu użyto tych samych danych co w przypadku sieci neuronowej, również poddanych standaryzacji oraz próbkowaniu losowemu. Podobnie jak w przypadku sieci zbiór treningowy składał się z 5756 obserwacji, natomiast testowy z 56,568. Podstawową miarą jakości modelu jest wartość funkcji przystosowania, bezpośrednio określająca przychód generowany przez model, uzyskana na podstawie macierzy trafności.

Tabela Macierz trafności dla modelu regresji logistycznej, Źródło: opracowanie własne

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Klasyfikacja | |
|  |  | 0 | 1 |
| Odpowiedź | 0 | 49,501 | 6,621 |
| 1 | 109 | 337 |

Powyższa tabela przedstawia wyniki klasyfikacji uzyskane przez model regresji logistycznej. W oparciu o macierz trafności obliczono również czułość klasyfikatora (ang. *sensitivity)*, która jest miarą informującą o stosunku liczby obserwacji poprawnie klasyfikowanych jako pozytywne do liczby naprawdę pozytywnych obserwacji (określana również jako *ang. true positive rate)*:

Model regresji zatem uzyskał całkiem zadowalający wynik - niemalże 76% obserwacji pozytywnych jest przez niego klasyfikowanych jako pozytywne. Wartość funkcji przystosowania dla tego modelu wyniosła 3853 z maksymalnej możliwej 8875 dla modelu idealnie klasyfikującego co stanowi 43% tej wartości.

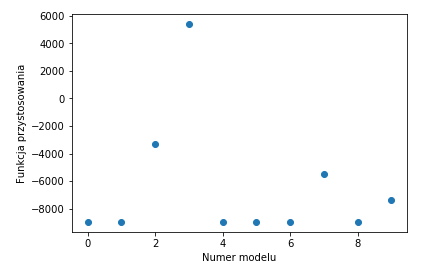
#### 5.3.2 Modele sieci neuronowych

W trakcie 10 iteracji algorytmu genetycznego treningowi oraz ewaluacji poddane zostało łącznie 100 sieci. Pierwszą populację stanowiły modele o zupełnie losowych topologiach, co przedstawia poniższa tabela:

Tabela Parametry populacji sieci dla pierwszej iteracji algorytmu

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Wartości parametrów | | | | |
| Numer sieci | 1 | 2 | 3 | 4 | 5 |
| 1 | 1 | 0 | 985 | 4 | 21 |
| 2 | 1 | 0 | 970 | 3 | 54 |
| 3 | 3 | 1 | 753 | 3 | 81 |
| 4 | 3 | 0 | 392 | 2 | 22 |
| 5 | 1 | 1 | 895 | 3 | 11 |
| 6 | 1 | 0 | 469 | 5 | 90 |
| 7 | 1 | 0 | 512 | 3 | 98 |
| 8 | 3 | 1 | 711 | 3 | 44 |
| 9 | 1 | 1 | 126 | 4 | 103 |
| 10 | 3 | 1 | 335 | 4 | 114 |

Niestety w pierwszej iteracji do populacji nie został wybrany gen reprezentujący funkcję aktywacji tangens hiperboliczny (parametr 1), jak również liczba warstw ukrytych jest w zakresie od 2 do 5 a nie jak zakładano od 1 do 5, co nieco ogranicza przestrzeń dostępnych rozwiązań. Poniższy wykres przedstawia wartość funkcji przystosowania dla poszczególnych modeli w tej iteracji. Możemy zaobserwować, że dla 9 spośród 10 modeli wartość funkcji przystosowania jest poniżej 0 a aż 6 z nich przyjmuje wartość poniżej -8000. Jeśli przyjrzymy się wartościom macierzy trafności okazuje się, że modele z ujemną wartością tejże funkcji klasyfikują wszystkie obserwacje jako jedną z kategorii, zupełnie ignorując drugą czyli są równoważne modelom naiwnym.



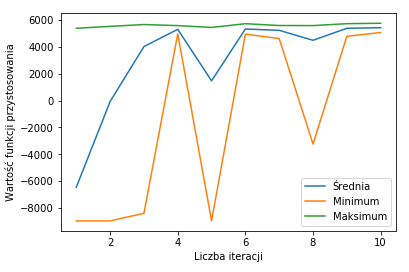
Rysunek Wykres wartości funkcji przystosowania dla modeli z pierwszej iteracji

Tabela Macierz trafności dla 10 modeli pierwszej iteracji algorytmu

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Klasyfikacja | | | |
| Numer sieci | TN | FN | FP | TP |
| 1 | 56122 | 0 | 446 | 0 |
| 2 | 56122 | 0 | 446 | 0 |
| 3 | 47175 | 8947 | 282 | 164 |
| 4 | 45290 | 10832 | 60 | 386 |
| 5 | 56122 | 0 | 446 | 0 |
| 6 | 56122 | 0 | 446 | 0 |
| 7 | 56122 | 0 | 446 | 0 |
| 8 | 55166 | 956 | 357 | 89 |
| 9 | 56122 | 0 | 446 | 0 |
| 10 | 55752 | 370 | 405 | 41 |

Jeden z modeli, osiągnął wartość funkcji przystosowania ponad 5 tysięcy – co jest już całkiem dobrym wynikiem. Można się spodziewać, że w kolejnych iteracjach jego topologia będzie najbardziej dominującą w całej populacji.

Poniższy wykres przedstawia wartości funkcji przystosowania dla wszystkich iteracji algorytmu, w podziale na średnią, minimum oraz maksimum. Wartość średnia, podobnie jak minimalna, dla pierwszej populacji jest najniższa zgodnie z tym czego oczekiwalibyśmy po losowym doborze parametrów. Już w drugiej iteracji średnia osiąga wartość bliską zeru, a w trzeciej i kolejnych ponad 4 tysiące, czyli lepiej niż model odniesienia. W badanym przypadku już od czwartej iteracji średnia nie ulega znacznemu polepszeniu. Jeśli chodzi o minimalną wartość funkcji przystosowania widzimy że mutacja w pokoleniach 5 i 8 sprawiła, że jest ona niezwykle niska, natomiast w pozostałych nie różni się wiele od średniej.



Rysunek Wykres wartości funkcji przystosowania w kolejnych iteracjach

Wartość maksymalna funkcji przystosowania wydaje się być stałą wartością, co świadczy o tym, że losowy sposób doboru parametrów może czasem zaowocować znalezieniem rozwiązania bliskiego optymalnemu lub przynajmniej bardzo dobrego w pierwszej iteracji lub w jednej z początkowych.

Ostatecznie najlepszym modelem okazała się sieć neuronowa z iteracji numer 10, której wartość funkcji przystosowania wynosi 5768 a jej czułość 0.886 czyli poprawnie klasyfikuje 88.6% obserwacji pozytywnych. W porównaniu do modelu regresji logistycznej jest to o 12.6% lepsza dokładność klasyfikacji oraz o 1915 wyższa wartość funkcji przystosowania. Poniższa tabela przedstawia macierz trafności dla tego modelu:

Tabela Macierz trafności dla najlepszego modelu sieci neuronowej.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Klasyfikacja | |
|  |  | 0 | 1 |
| Odpowiedź | 0 | 45445 | 10677 |
| 1 | 41 | 395 |

Topologia tej sieci składa się z funkcji ReLU, stałego współczynnika uczenia z wartością początkową 392, 3 warstw ukrytych w których znajdują się po 22 neurony. Jest ona niemalże identyczna jak najlepsza sieć w pierwszej generacji – różni się od niej jedynie większą liczbą warstw ukrytych.

6. Wnioski

## Zakończenie

Bibliografia

*\*wersja robocza*

1. “Big Data-Driven Marketing: How Machine Learning Outperforms Marketers’

Gut-Feeling,” http://web.media.mit.edu/~yva/papers/sundsoy2014big.pdf

2. Artificial Intelligence for Marketing. Practical Applications. Jim Sterne, 2017, Wiley

3. Evolutionary Optimization Algorithms. Dan Simon, 2013

4. Adaptation in Natural and artificial systems, John H. Holland, MIT Press, 1992

5. Algorytmy ewolucyjne i ich zastosowania, Ewa Figielska, na: [*http://zeszyty-naukowe.wwsi.edu.pl/zeszyty/zeszyt1/Algorytmy\_Ewolucyjne\_I\_Ich\_Zastosowania.pdf*](http://zeszyty-naukowe.wwsi.edu.pl/zeszyty/zeszyt1/Algorytmy_Ewolucyjne_I_Ich_Zastosowania.pdf)

6. Variants of Evolutionary Algorithms for Real World Applications, Raymond Chiong, Thomas Weise, Zbigniew Michalewicz, Springer, 2012

7. Z. Michalewicz. ***Algorytmy genetyczne + struktury danych = programy ewolucyjne.*** WNT, Warszawa 2003.

8. D.E. Goldberg. ***Algorytmy genetyczne i ich zastosowania.*** WNT, Warszawa 1995.

9. Arabas J., 2004, Wykłady z algorytmów ewolucyjnych, Wydawnictwo Naukowo-Techniczne, Warszawa.

10. Computation Intelligence Techniques for Trading and Investment, CH. Dunis, New York, Routledge, 2014

|  |  |  |  |
| --- | --- | --- | --- |
| **Lp.** | **Zmienna** | **Typ** | **Opis** |
| **1** | TRX\_aff\_L12M\_COL\_PTS | Num | Liczba wszystkich punktów zebrana w ciągu ostatnich 12 miesięcy w Sklepach Online |
| **2** | TRX\_aff\_L12M\_PUR\_AMT | Num | Kwota wydana na transakcje zakupowe w ciągu ostatnich 12 miesięcy w Sklepach Online |
| **3** | TRX\_aff\_L12M\_RED\_CNT | Num | Liczba transakcji z płatnością punktami dokonana w ciągu ostatnich 12 miesięcy w Sklepach Online |
| **4** | TRX\_aff\_L12M\_RED\_PTS | Num | Liczba punktów wykorzystanych w zakupach w ciągu ostatnich 12 miesiecy w Sklepach Online |
| **5** | TRX\_aff\_L3M\_BASE\_PTS | Num | Liczba punktów podstawowych zebrana w ciągu ostatnich 3 miesiecy w Sklepach Online |
| **6** | TRX\_aff\_L3M\_COL\_CNT | Num | Liczba transakcji kolekcyjnych dokonanych w ciagu ostatnich 3 miesiecy w Sklepach Online |
| **7** | TRX\_aff\_L3M\_PROMO\_PTS | Num | Liczba punktów promocyjnych zebrana w ciągu ostatnich 3 miesięcy w Sklepach Online |
| **8** | TRX\_aff\_L3M\_PUR\_AMT | Num | Kwota wydana podczas transakcji zakupowych w ciagu ostatnich 3 miesiecy w Sklepach Online |
| **9** | TRX\_aff\_L3M\_RED\_CNT | Num | Liczba transakcji redempcyjnych dokonana w ciągu ostatnich 3 miesięcy w Sklepach Online |
| **10** | TRX\_aff\_L3M\_RED\_PTS | Num | Liczba punktów redempcyjnych wydanych w ciągu ostatnich 3 miesiecy w Sklepach Online |
| **11** | TRX\_all\_L12M\_PROMO\_PTS | Num | Liczba punktów promocyjnych zebrana w ciągu ostatnich 12 miesięcy w PAYBACK |
| **12** | TRX\_all\_L12M\_PUR\_AMT | Num | Kwota wydana podczas transakcji zakupowych w ciagu ostatnich 12 miesiecy w PAYBACK |
| **13** | TRX\_all\_L12M\_PUR\_PTS | Num | Liczba punktów z transakcji zakupowych zebrana w ciągu ostatnich 12 miesiecy w PAYBACK |
| **14** | TRX\_all\_L12M\_RED\_CNT | Num | Liczba transakcji redempcyjnych dokonana w ciągu ostatnich 12 miesięcy w PAYBACK |
| **15** | TRX\_all\_L12M\_RED\_PTS | Num | Liczba punktów redempcyjnych wydanych w ciągu ostatnich 12 miesiecy w PAYBACK |
| **16** | TRX\_all\_L3M\_BASE\_PTS | Num | Liczba punktów podstawowych zebrana w ciągu ostatnich 12 miesiecy w PAYBACK |
| **17** | TRX\_all\_L3M\_COL\_CNT | Num | Liczba transakcji kolekcyjnych dokonanych w ciagu ostatnich 12 miesiecy w PAYBACK |
| **18** | TRX\_all\_L3M\_PROMO\_PTS | Num | Liczba punktów promocyjnych zebrana w ciągu ostatnich 3 miesięcy w PAYBACK |
| **19** | TRX\_all\_L3M\_PUR\_AMT | Num | Kwota wydana podczas transakcji zakupowych w ciagu ostatnich 3 miesiecy w PAYBACK |
| **20** | TRX\_all\_L3M\_RED\_CNT | Num | Liczba transakcji redempcyjnych dokonana w ciągu ostatnich 3 miesięcy w PAYBACK |
| **21** | TRX\_all\_L3M\_RED\_PTS | Num | Liczba punktów redempcyjnych wydanych w ciągu ostatnich 3 miesiecy w PAYBACK |
| **22** | ACCT\_BAL\_MOD\_CNT | Num | Liczba modyfikacji salda konta |
| **23** | ACCT\_TOT\_COL\_AMT | Num | Liczba zebranych punktów (ever) w programie |
| **24** | ACCT\_UNBLOCKED\_POINTS | Num | Liczba niezablokowanych punktów na koncie |
| **25** | TIME\_FROM\_ENRL | Num | Czas od zarejestrowania konta |
| **26** | TIME\_FROM\_FIRST\_DT | Num | Czas od pierwszej akcji w programie |
| **27** | TIME\_FROM\_LAST\_DT | Num | Czas od ostatniej akcji w programie |
| **28** | CARDA\_PAYM\_CNT | Num | Liczba kart płatniczych przypisanych do konta |
| **29** | ACCT\_PTS\_EXPIRED | Num | Liczba wygaszonych punktów |
| **30** | ACCT\_PTS\_TO\_EXPIRE | Num | Liczba punktów do wygaśnięcia |
| **31** | CUST\_GFK\_POPULATION | Num | Liczba mieszkańców w miejscu zamieszkania uczestnika |
| **32** | CUST\_GFK\_HOUSEHOLDS | Num | Liczba gospodarstw domowych w miejscu zamieszkania uczestnika |
| **33** | CUST\_HH\_SIZE | Num | Wielkość gospodarstwa domowego uczestnika |
| **34** | CUST\_PROFESSION\_CD | Num | Zawód uczestnika |
| **35** | CUST\_AGE | Num | Wiek uczestnika |
| **36** | CUST\_GENDER | Num | Płeć uczestnika |
| **37** | CUST\_CA\_AAA\_FLG | Char | Czy uczestnik mieszka w pobliżu sieci partnera AAA |
| **38** | CUST\_CA\_BBB\_FLG | Char | Czy uczestnik mieszka w pobliżu sieci partnera BBB |
| **39** | CUST\_CA\_CCC\_FLG | Char | Czy uczestnik mieszka w pobliżu sieci partnera CCC |
| **40** | CUST\_CA\_AAA\_KM\_CNT | Num | Odległość (km) od najbliższego punktu sprzedaży partnera AAA |
| **41** | CUST\_CA\_BBB\_KM\_CNT | Num | Odległość (km) od najbliższego punktu sprzedaży partnera BBB |
| **42** | CUST\_CA\_CCC\_KM\_CNT | Num | Odległość (km) od najbliższego punktu sprzedaży partnera CCC |
| **43** | WEB\_ALL\_UNQ\_SES\_CNT\_L2W | Num | Liczba unikalnych sesji w ciagu ostatnich 2 tygodni na stronie PAYBACK |
| **44** | WEB\_ALL\_TIME\_SPEND\_L2W | Num | Liczba minut spędzona w ostatnich 2 tygodniach na stronie PAYBACK |
| **45** | WEB\_ALL\_UNQ\_SES\_CNT\_L4W | Num | Liczba unikalnych sesji w ciagu ostatnich 4 tygodni na stronie PAYBACK |
| **46** | WEB\_ALL\_TIME\_SPEND\_L4W | Num | Liczba minut spędzona w ostatnich 4 tygodniach na stronie PAYBACK |
| **47** | WEB\_AFF\_UNQ\_SES\_CNT\_L2W | Num | Liczba unikalnych sesji w ciagu ostatnich 2 tygodni na stronie Sklepów Online |
| **48** | WEB\_AFF\_TIME\_SPEND\_L2W | Num | Liczba minut spędzona w ostatnich 2 tygodniach na stronie Sklepów Online |
| **49** | WEB\_AFF\_UNQ\_SES\_CNT\_L4W | Num | Liczba unikalnych sesji w ciagu ostatnich 4 tygodni na stronie Sklepów Online |
| **50** | WEB\_AFF\_TIME\_SPEND\_L4W | Num | Liczba minut spędzona w ostatnich 4 tygodniach na stronie Sklepów Online |
| **51** | WEB\_HIS\_UNQ\_SES\_CNT\_L2W | Num | Liczba unikalnych sesji w ciagu ostatnich 2 tygodni na stronie Historii Transakcji |
| **52** | WEB\_HIS\_TIME\_SPEND\_L2W | Num | Liczba minut spędzona w ostatnich 2 tygodniach na stronie Historii Transakcji |
| **53** | WEB\_HIS\_UNQ\_SES\_CNT\_L4W | Num | Liczba unikalnych sesji w ciagu ostatnich 4 tygodni na stronie Historii Transakcji |
| **54** | WEB\_HIS\_TIME\_SPEND\_L4W | Num | Liczba minut spędzona w ostatnich 4 tygodniach na stronie Historii Transakcji |
| **55** | WEB\_OCC\_UNQ\_SES\_CNT\_L2W | Num | Liczba unikalnych sesji w ciagu ostatnich 2 tygodni na stronie eKupony |
| **56** | WEB\_OCC\_TIME\_SPEND\_L2W | Num | Liczba minut spędzona w ostatnich 2 tygodniach na stronie eKupony |
| **57** | WEB\_OCC\_UNQ\_SES\_CNT\_L4W | Num | Liczba unikalnych sesji w ciagu ostatnich 4 tygodni na stronie eKupony |
| **58** | WEB\_OCC\_TIME\_SPEND\_L4W | Num | Liczba minut spędzona w ostatnich 4 tygodniach na stronie eKupony |
| **59** | WEB\_REW\_UNQ\_SES\_CNT\_L2W | Num | Liczba unikalnych sesji w ciagu ostatnich 2 tygodni na stronie Sklepu z Nagrodami |
| **60** | WEB\_REW\_TIME\_SPEND\_L2W | Num | Liczba minut spędzona w ostatnich 2 tygodniach na stronie Sklepu z Nagrodami |
| **61** | WEB\_REW\_UNQ\_SES\_CNT\_L4W | Num | Liczba unikalnych sesji w ciagu ostatnich 4 tygodni na stronie Sklepu z Nagrodami |
| **62** | WEB\_REW\_TIME\_SPEND\_L4W | Num | Liczba minut spędzona w ostatnich 4 tygodniach na stronie Sklepu z Nagrodami |
| **63** | WEB\_ALL\_UNQ\_SES\_CNT\_L3M | Num | Liczba unikalnych sesji w ciagu ostatnich 3 miesięcy na stronie PAYBACK |
| **64** | WEB\_ALL\_TIME\_SPEND\_L3M | Num | Liczba minut spędzona w ostatnich 3 miesiącach na stronie PAYBACK |
| **65** | WEB\_ALL\_UNQ\_SES\_CNT\_L6M | Num | Liczba unikalnych sesji w ciagu ostatnich 6 miesięcy na stronie PAYBACK |
| **66** | WEB\_ALL\_TIME\_SPEND\_L6M | Num | Liczba minut spędzona w ostatnich 6 miesiącach na stronie PAYBACK |
| **67** | WEB\_ALL\_UNQ\_SES\_CNT\_L12M | Num | Liczba unikalnych sesji w ciagu ostatnich 12 miesięcy na stronie PAYBACK |
| **68** | WEB\_ALL\_TIME\_SPEND\_L12M | Num | Liczba minut spędzona w ostatnich 12 miesiącach na stronie PAYBACK |
| **69** | WEB\_AFF\_UNQ\_SES\_CNT\_L3M | Num | Liczba unikalnych sesji w ciagu ostatnich 3 miesięcy na stronie Sklepów Online |
| **70** | WEB\_AFF\_TIME\_SPEND\_L3M | Num | Liczba minut spędzona w ostatnich 3 miesiącach na stronie Sklepów Online |
| **71** | WEB\_AFF\_UNQ\_SES\_CNT\_L6M | Num | Liczba unikalnych sesji w ciagu ostatnich 6 miesięcy na stronie Sklepów Online |
| **72** | WEB\_AFF\_TIME\_SPEND\_L6M | Num | Liczba minut spędzona w ostatnich 6 miesiącach na stronie Sklepów Online |
| **73** | WEB\_AFF\_UNQ\_SES\_CNT\_L12M | Num | Liczba unikalnych sesji w ciagu ostatnich 12 miesięcy na stronie Sklepów Online |
| **74** | WEB\_AFF\_TIME\_SPEND\_L12M | Num | Liczba minut spędzona w ostatnich 12 miesiącach na stronie Sklepów Online |
| **75** | WEB\_HIS\_UNQ\_SES\_CNT\_L3M | Num | Liczba unikalnych sesji w ciagu ostatnich 3 miesięcy na stronie Historii Transakcji |
| **76** | WEB\_HIS\_TIME\_SPEND\_L3M | Num | Liczba minut spędzona w ostatnich 3 miesiącach na stronie Historii Transakcji |
| **77** | WEB\_HIS\_UNQ\_SES\_CNT\_L6M | Num | Liczba unikalnych sesji w ciagu ostatnich 6 miesięcy na stronie Historii Transakcji |
| **78** | WEB\_HIS\_TIME\_SPEND\_L6M | Num | Liczba minut spędzona w ostatnich 6 miesiącach na stronie Historii Transakcji |
| **79** | WEB\_HIS\_UNQ\_SES\_CNT\_L12M | Num | Liczba unikalnych sesji w ciagu ostatnich 12 miesięcy na stronie Historii Transakcji |
| **80** | WEB\_HIS\_TIME\_SPEND\_L12M | Num | Liczba minut spędzona w ostatnich 12 miesiącach na stronie Historii Transakcji |
| **81** | WEB\_OCC\_UNQ\_SES\_CNT\_L3M | Num | Liczba unikalnych sesji w ciagu ostatnich 3 miesięcy na stronie eKupony |
| **82** | WEB\_OCC\_TIME\_SPEND\_L3M | Num | Liczba minut spędzona w ostatnich 3 miesiącach na stronie eKupony |
| **83** | WEB\_OCC\_UNQ\_SES\_CNT\_L6M | Num | Liczba unikalnych sesji w ciagu ostatnich 6 miesięcy na stronie eKupony |
| **84** | WEB\_OCC\_TIME\_SPEND\_L6M | Num | Liczba minut spędzona w ostatnich 6 miesiącach na stronie eKupony |
| **85** | WEB\_OCC\_UNQ\_SES\_CNT\_L12M | Num | Liczba unikalnych sesji w ciagu ostatnich 12 miesięcy na stronie eKupony |
| **86** | WEB\_OCC\_TIME\_SPEND\_L12M | Num | Liczba minut spędzona w ostatnich 12 miesiącach na stronie eKupony |
| **87** | WEB\_REW\_UNQ\_SES\_CNT\_L3M | Num | Liczba unikalnych sesji w ciagu ostatnich 3 miesięcy na stronie Sklepu z Nagrodami |
| **88** | WEB\_REW\_TIME\_SPEND\_L3M | Num | Liczba minut spędzona w ostatnich 3 miesiącach na stronie Sklepu z Nagrodami |
| **89** | WEB\_REW\_UNQ\_SES\_CNT\_L6M | Num | Liczba unikalnych sesji w ciagu ostatnich 6 miesięcy na stronie Sklepu z Nagrodami |
| **90** | WEB\_REW\_TIME\_SPEND\_L6M | Num | Liczba minut spędzona w ostatnich 6 miesiącach na stronie Sklepu z Nagrodami |
| **91** | WEB\_REW\_UNQ\_SES\_CNT\_L12M | Num | Liczba unikalnych sesji w ciagu ostatnich 12 miesięcy na stronie Sklepu z Nagrodami |
| **92** | WEB\_REW\_TIME\_SPEND\_L12M | Num | Liczba minut spędzona w ostatnich 12 miesiącach na stronie Sklepu z Nagrodami |
| **93** | event | Num | Zdarzenie |
| **94** | id | Num | Identyfikator obserwacji |
| **95** | oferta | Char | Numer oferty z danej wysyłki |
| **96** | kategoria | Char | Kategoria sklepu |

Załącznik Zmienne w zbiorze danych

Załącznik 2 – kod implementacji algorytmu i sieci

1. [https://sjinsights.net/2014/09/29/new-research-sheds-light-on-daily-ad-exposures/] [↑](#footnote-ref-1)
2. *Marketing bezpośredni. Koncepcja – zarządzanie – instrumenty*. M. Trojanowski, Warszawa, 2010, PWE [↑](#footnote-ref-2)
3. *Komunikacja marketingowa. Instrumenty i metody.* red. B. Szymoniuk, Warszawa, 2006, PWE [↑](#footnote-ref-3)
4. *Marketing bezpośredni*, P. Dobski, M. Szuman-Dobska, Warszawa, 1999 Wydawnictwo Prawno-Ekonomiczne [↑](#footnote-ref-4)
5. https://stat.gov.pl/obszary-tematyczne/nauka-i-technika-spoleczenstwo-informacyjne/spoleczenstwo-informacyjne/spoleczenstwo-informacyjne-w-polsce-w-2016-roku,2,6.html (data pobrania 4 listopada 2017 r.) Strona internetowa Głównego Urzędu Statystycznego – Portal Informacyjny [↑](#footnote-ref-5)
6. Błażewicz G., Rewolucja z marketing automation. Jak wykorzystać potencjał Big Data, PWN, Warszawa 2016, str. 72. [↑](#footnote-ref-6)
7. E. Figielska, *Algorytmy ewolucyjne i ich zastosowanie,* „Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki”, nr 1, 2006, s. 81-92  [↑](#footnote-ref-7)
8. J. H. Holland, *Adaptation in Natural and artificial systems*, Cambridge 1975, MIT Press [↑](#footnote-ref-8)
9. K. Deb, *Multi-Objective Optimization using Evolutionary Algorithms,*  2001, Wiley [↑](#footnote-ref-9)
10. T. D. Gwiazda, *Algorytmy ewolucyjne w rozwiązywaniu nieliniowych problemów decyzyjnych,* 2002, Wydawnictwa Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego [↑](#footnote-ref-10)
11. [http://edu.pjwstk.edu.pl/wyklady/nai/scb/wyklad10/w10.htm] [↑](#footnote-ref-11)
12. *Algorytmy genetyczne, ewolucyjne i metaheurystyki,* T. Trzaskalik (red.), Katowice, Wydawnictwo Akademii Ekonomicznej w Katowicach, 2005, s.13-14 [↑](#footnote-ref-12)
13. *Metody i modele eksploracji danych*, D.T. Larose, Hoboken, Wiley 2006 [↑](#footnote-ref-13)
14. *An analysis of the behavior of a class of genetic adaptive systems, praca doktorska,* K. De Jong, University of Michigan, Ann Arbor, 1975 [↑](#footnote-ref-14)
15. *Genetic algorithms with sharing for multi-modal function optimization,* D. Goldberg, J. Richardson w: *Genetic Algorithms and Their Applications: Proceedings of the 2nd International Conference on Genetic Algorithms,* J. Greffenstette (red.), Lawrence Erlbaum Associates, Hillsdale, 1987 [↑](#footnote-ref-15)
16. *Algorytmy genetyczne, ewolucyjne i metaheurystyki,* T. Trzaskalik (red.), Katowice, Wydawnictwo Akademii Ekonomicznej w Katowicach, 2005, s.18 [↑](#footnote-ref-16)
17. *Applied Simulated Annealing,* R. V. V. Vidal, Berlin, Springer,1993 [↑](#footnote-ref-17)
18. *Algorytmy genetyczne i ich zastosowania****,*** D.E. Goldberg, Warszawa, WNT, 1995. [↑](#footnote-ref-18)
19. J. H. Holland, *Adaptation in Natural and artificial systems*, Cambridge, MIT Press, 1975 [↑](#footnote-ref-19)
20. *An Introduction to Genetic Algorithm,* M. Mitchell, Cambridge, MIT Press, 1996 [↑](#footnote-ref-20)
21. *An experimental comparison of binary and floating point representations in genetic algorithms.* C.Z. Janikow, Z. Michalewicz, w: *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann, 1991 [↑](#footnote-ref-21)
22. *Data Mining and Predictive Analytics,* D. T.Larose, C. Larose, Hoboken, Wiley, 2015 [↑](#footnote-ref-22)
23. *Modelowanie matematyczne i symulacje komputerowe w naukach społecznych,* K. Winkowska-Nowak (red.), Warszawa, Wydawnictwo SWPS Academica, 2007 [↑](#footnote-ref-23)
24. *O celowości zastosowania sieci neuronowych w problemach związanych z elektrotechniką,* R. Tadeusiewicz, w: *Przegląd Elektrotechniczny,* R. 85NR 2/2009 [↑](#footnote-ref-24)
25. *Neural Networks for Target Selection in Direct Marketing,* R. Potharst, U. Kaymak, W.H.L.M. Pijls, w: *ERIM Report Series Research in Management*, na: <http://hdl.handle.net/1765/83> [↑](#footnote-ref-25)
26. *Applying Neural Networks and other statistical models to the classification of serious offenders and the prediction of recidivism,* M. Yang, Y. Liu, J. Coid, w: *Ministry of Justice Research Series,* nr 6/10, 2010 [↑](#footnote-ref-26)
27. *Computation Intelligence Techniques for Trading and Investment*, CH. Dunis, New York, Routledge, 2014 [↑](#footnote-ref-27)
28. *Neural Networks and Learning Machines,* S. Haykin, Pearson Prentice Hall, Upper Saddle River, 2009 [↑](#footnote-ref-28)
29. *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, T. Hastie, R. Tibshirani, J. Friedman, Springer, New York, 2009, s.393 [↑](#footnote-ref-29)
30. *Introduction to Radial Basis Function Networks,* M. J. L. Orr, na: https://www.cc.gatech.edu/~isbell/tutorials/rbf-intro.pdf [↑](#footnote-ref-30)
31. *Data Mining – Theories, Algorithms and Examples,* N. Ye, New York, CRC Press, 2014, s. 63 [↑](#footnote-ref-31)
32. *Neural Networks and Learning Machines,* S. Haykin, Pearson Prentice Hall, Upper Saddle River, 2009 [↑](#footnote-ref-32)
33. *Fundamentals of Deep Learning,* N. Buduma, Sebastopol, O’Reilly Media, 2017 [↑](#footnote-ref-33)
34. http://scikit-learn.org/stable/modules/generated/sklearn.neural\_network.MLPClassifier.html [↑](#footnote-ref-34)
35. *Data Mining – Theories, Algorithms and Examples,* N. Ye, New York, CRC Press, 2014, s. 80 [↑](#footnote-ref-35)
36. *Learning Representations by BackPropagating Errors,* D .E. Rumelhart, G. E. Hinton, R. J. Williams, w: *Nature* nr 323, 1986 [↑](#footnote-ref-36)
37. *Inteligentna Sieć, Algorytmy przyszłości,* D. McIlwrath, H. Marmanis, D. Babenko, Gliwice, Helion, 2017 [↑](#footnote-ref-37)
38. https://github.com/scikit-learn/scikit-learn/blob/a24c8b46/sklearn/preprocessing/data.py#L461 [↑](#footnote-ref-38)
39. *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, T. Hastie, R. Tibshirani, J. Friedman, Springer, New York, 2009, s.400 [↑](#footnote-ref-39)
40. *Zaawansowane metody analiz statystycznych,* red. E. Frątczak, Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie, 2013, Warszawa, s.597 [↑](#footnote-ref-40)
41. <http://contrib.scikit-learn.org/imbalanced-learn/stable/generated/> imblearn.under\_sampling.RandomUnderSampler.html [↑](#footnote-ref-41)
42. <http://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html> [↑](#footnote-ref-42)
43. *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, T. Hastie, R. Tibshirani, J. Friedman, Springer, New York, 2009, s.400 [↑](#footnote-ref-43)
44. *Modele regresji logistycznej : zastosowania w medycynie, naukach przyrodniczych i społecznych,* A. Stanisz, Kraków, 2016 [↑](#footnote-ref-44)