



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE
WYDZIAŁ ZARZĄDZANIA

Pracownia Zastosowań Matematyki w Ekonomii

Praca dyplomowa magisterska

*Zastosowania sieci neuronowych do prognozowania ceny
komputerów przenośnych w sklepach internetowych*

Neural networks for predicting laptop prices in online shops

Autor:
Kierunek studiów:
Opiekun pracy:

Piotr Morawiecki
Informatyka i Ekonometria
dr Marcin Suder

Kraków, 2020

„Uprzedzony o odpowiedzialności karnej na podstawie art. 115 ust. 1 i 2 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.): „Kto przywłaszcza sobie autorstwo albo wprowadza w błąd co do autorstwa całości lub części cudzego utworu albo artystycznego wykonania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3. Tej samej karze podlega, kto rozpowszechnia bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystyczne wykonanie albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, wideogram lub nadanie.”, a także uprzedzony o odpowiedzialności dyscyplinarnej na podstawie art. 211 ust. 1 ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (t.j. Dz. U. z 2012 r. poz. 572, z późn. zm.) „Za naruszenie przepisów obowiązujących w uczelni oraz za czyny uchybiające godności studenta student ponosi odpowiedzialność dyscyplinarną przed komisją dyscyplinarną albo przed sądem koleżeńskim samorządu studenckiego, zwanym dalej "sądem koleżeńskim", oświadczam, że niniejszą pracę dyplomową wykonałem(-am) osobiście i samodzielnie i że nie korzystałem(-am) ze źródeł innych niż wymienione w pracy.”

Data i podpis dyplomanta

Spis treści

Wstęp	4
1. Rola zakupów online	6
1.1. Charakterystyka zakupów online i offline	6
1.2. Charakterystyka kupujących online.....	8
1.3. Najczęściej kupowane produkty online.....	9
1.4. Najczęściej używane strony online.....	10
2. Uzyskanie danych	12
2.1. Ruch na stronie.....	12
2.2. Charakterystyka strony.....	13
2.3. Web Scraping.....	15
2.3.1. Wysyłanie zapytania protokołu HTTP.....	16
2.3.2. Otrzymanie odpowiedzi protokołu HTTP	16
2.3.3. Wyciąganie informacji	17
3. Sieć neuronowa	21
3.1. Budowa neuronu w sieci neuronowej.....	22
3.2. Budowa sieci neuronowej	23
3.3. Rodzaje sieci neuronowych	24
3.4. Funkcje aktywacji	25
3.5. Nauka sieci neuronowej	26
4. Model sieci neuronowej.....	28
4.1. Opis danych	28
4.2. Przygotowanie danych	31
4.3. Przekształcenie zmiennej kategorycznej.....	32
4.4. Przekształcenie zmiennej numerycznej.....	34
4.5. Prezentacja gotowego zbioru danych	36
4.6. Budowa sieci neuronowej	37
4.7. Wyniki algorytmu	39
Zakończenie	41
Bibliografia.....	43
Spis ilustracji	44
Spis tabel	44

Wstęp

Rola zakupów online odgrywa kluczową rolę w dzisiejszych czasach. Chyba nikt nie może sobie wyobrazić życia bez sklepów internetowych. Różnorodność produktów jakie oferują sklepy internetowe jest niewyobrażalna. Większość rzeczy można zakupić online, bez wychodzenia z domu. Jest to bardzo komfortowy sposób zakupów. Wiąże się to z ogromnym zyskiem dla sklepów. W sieci występuje również ogromna konkurencja na rynku z powodu istnienia wielu sklepów internetowych. Możliwość wyceny produktów bez ingerencji człowieka byłaby znacznym ułatwieniem dla funkcjonowania sklepów online.

Celem pracy jest stworzenie modelu do wyceny laptopów za pomocą zebranych danych. Wiele wcześniejszych badań przedstawia zastosowania sieci neuronowych do wyceny nieruchomości czy ceny akcji. Wycena nieruchomości jest podjęta przez Núñez Tabales, Julia M., Caridad y Ocerin, José María, Rey Carmona, Francisco J.¹, którzy zajmowali się hiszpańskim rynkiem nieruchomości czy I. S. H. Bahia², która analizowała mieszkania w Bostonie. Ciekawe badania przeprowadzili Changshou Luo, Qingfeng Wei, Liying Zhou, Junfeng Zhang i Sufen Sun³, którzy porównali trzy modele sieci neuronowych prognozując ceny warzyw w poszczególnych miesiącach. W ich badaniu, sztuczna sieć neuronowa wykazała się największą dokładnością w prognozie cen. Kolejnym interesującą analizą cen było badanie autorów: S. Chen, E. Chou, R. Yang⁴, którzy prognozowali cenę produktów na podstawie obrazków. Oczywiście podobnych prac jest więcej. W wielu z nich autorzy twierdzą, że sieci neuronowe potrafią bardzo dobrze dopasować się do danych i świetnie nadają się do wykorzystania w prognozie cen.

Obecnie wiele sklepów internetowych oferuje różnego rodzaju sprzęty elektroniczne. Odpowiednia wycena danego sprzętu nie jest rzeczą prostą. Sklepy internetowe nie chcą z pewnością tracić na swoich produktach, czyli wystawiać przedmiotów w niższych cenach niż sprzęt jest warty. Z drugiej strony zbyt wysoka cena, może odstraszyć klientów. W sieci jest

¹ Núñez Tabales, Julia M., Caridad y Ocerin, José María, Rey Carmona i Francisco J. (2013). Artificial neural networks for predicting real estate. Universidad Pablo de Olavide, Sevilla.

² Bahia, I. S. (2013). A Data Mining Model by Using ANN for Predicting Real. Ministry of Higher Education and Scientific Research, Baghdad, Iraq.

³ Luo, C., Wei, Q., Zhou, L., Zhang, J. i Sun, S. (2011). Prediction of Vegetable Price Based on Neural Network and Genetic Algorithm. Beijing Academy of Agriculture and Forestry Sciences.

⁴ Chen, S., Chou, E. i Yang, R. (2017). *The Price is Right: Predicting Prices with Product Images*. Stanford University

też ogromna konkurencja, a klient ma łatwy dostęp do wielu rodzaju sklepów w krótkim czasie. Dostęp do danych w wybranych sklepach internetowych pozwoli wykorzystać sieć neuronową do wyceny sprzętu w danej kategorii. W badaniu dane na temat komputerów przenośnych zbierane są automatycznie, co oznacza, że nie potrzebują zbyt wysokiego nakładu ze strony człowieka. Jest to przykład zastosowania, ponieważ podobny schemat można wykorzystać na innego rodzaju produktach.

W rozdziale 1 zostanie przedstawiona rola zakupów online w dzisiejszych czasach. Choć sprzedaż internetowa jest bardzo popularna, wciąż wiele osób dokonuje zakupy jedynie w sklepach stacjonarnych. Dotyczy się to głównie osób starszych. Co więcej, niektóre kategorie produktów w ponad 90% są kupowane jedynie w sklepie stacjonarnym, co zostanie później przedstawione.

Rozdział 2 zostanie poświęcony pozyskaniu danych. Dane pochodzą ze strony internetowej sklepu RTV euro AGD i dotyczą komputerów przenośnych. Badanie wykorzystuje metody Web Scrapingu, czyli narzędzia do pobierania danych. Jest to dość popularna metoda, która jest często wykorzystywana do pozyskania interesujących zasobów danych. Pozyskane dane będą musiały zostać przekształcone odpowiednio przekształcone, aby mogły być użyte w dalszej części badania.

W rozdziale 3 zostanie przedstawiona sieć neuronowa. Historia sieci neuronowych sięga pierwszej połowy XX wieku. Obecnie do dyspozycji jest kilka rodzajów sieci neuronowych, funkcji aktywacji czy sposobów nauki sieci neuronowych. Wybranie odpowiedniego rodzaju zależy od potrzeby badania, wykorzystanych danych oraz wyniku jaki powinien zostać uzyskany w badaniu.

W rozdziale 4 oprócz przekształcania danych, zostanie zbudowany model do wyceny produktów. Jak się później okaże, przy wycenie komputerów przenośnych, jest wiele zmiennych, które powinny zostać brane pod uwagę. Aby mogły zostać one wykorzystane w badaniu powinny zostać odpowiednio przekształcone. Przygotowane dane zostaną wykorzystane w celu budowy sieci neuronowych. Wynik modelu będzie można sprawdzić i porównać z rzeczywistymi wynikami zbioru testowego.

1. Rola zakupów online

Coraz więcej osób dokonuje zakupów online. Łatwa dostępność do sieci sprawiła, że Polacy coraz chętniej wykorzystują sklepy internetowe do robienia zakupów. Jak podaje raport EY⁵, około 57% Polaków robi zakupu w sieci, a ponad 2/3 wyszukuje informacji w sieci o produktach i cenach.

1.1. Charakterystyka zakupów online i offline

Zgodnie ze wspomnianym wcześniej raportem Polacy najczęściej wykorzystują dostęp do sieci w celu zakupu elektroniki. Aż 71% badanych dokonuje zakupu sprzętu elektronicznego przy użyciu obu kanałów. Przeciwnieństwem tego są zakupy spożywcze. 90% badanych nadal preferuje zakupy stacjonarne, co przedstawia raport poniżej.

Rysunek 1 Charakterystyka kupujących online oraz offline



Źródło: <https://spidersweb.pl/bizblog/zakupy-w-internecie-badanie-ey-polska/>

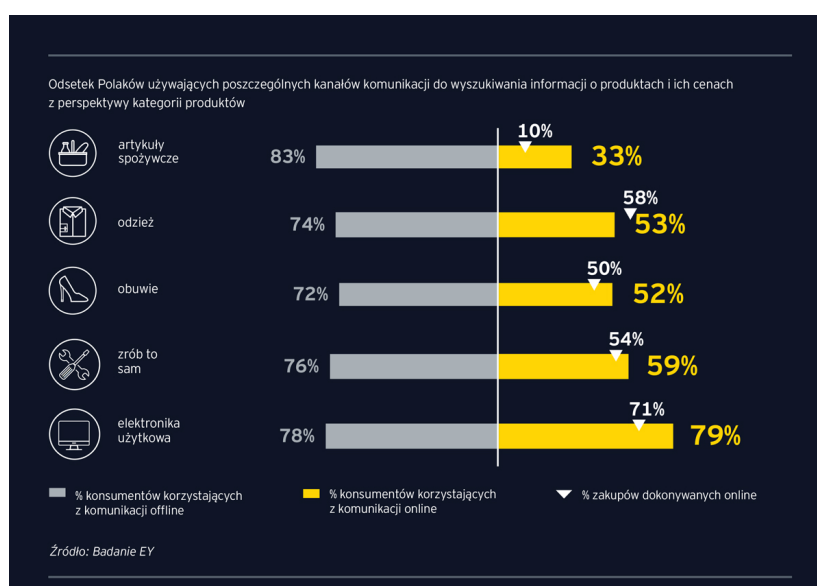
Warto również wspomnieć, że dostęp do sieci lepiej jest wykorzystywany przez ludzi młodszych. Jednie 39% badanych w wieku powyżej 60 lat korzysta z zakupów online. Dla porównania 74% badanych w wieku 15-19 lat korzysta z zakupów online oraz offline. Badani wykorzystują również dostęp do sieci w celu zdobycia informacji na temat danego produktu.

⁵ Na podstawie: https://www.ey.com/pl_pl/news/2019/08/jak-polscy-konsumenci-kupuja-w-sieci-wyniki-badania-ey (dostęp: Wrzesień 2020)

Jak przedstawia raport, 66% badanych wykorzystuje dostęp do sieci w celu weryfikacji danego produktu.

Najczęściej sprawdzaną kategorią online przez badanych jest elektronika użytkowa. 79% badanych deklaruje wykorzystanie sieci do sprawdzenia produktu we wspomnianej kategorii. Najrzadziej są oczywiście artykuły spożywcze. Jedynie 39% badanych sprawdza informacje o danym produkcie w sieci. Porównanie poszczególnych kanałów komunikacji do wyszukiwania informacji o produktach i cenach przedstawione jest poniżej.

Rysunek 2 Charakterystyka zakupów online oraz offline



Źródło: <https://spidersweb.pl/bizblog/zakupy-w-internecie-badanie-ey-polska/>

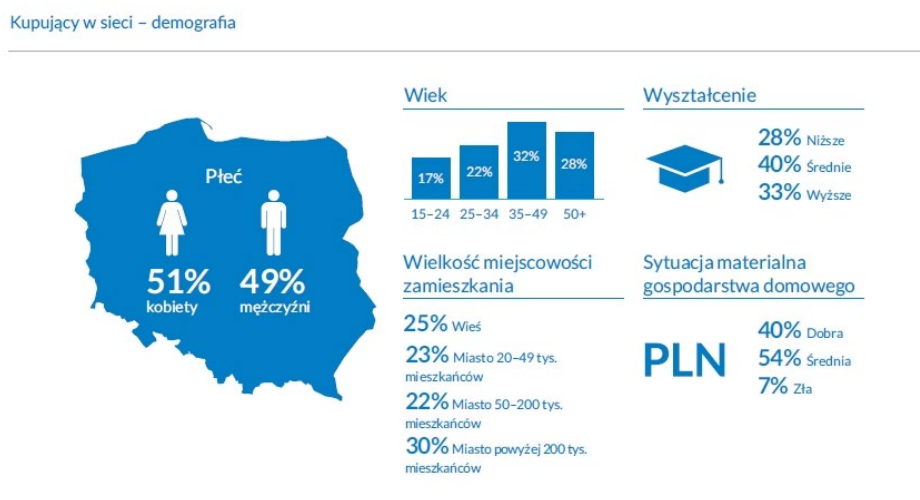
Jedynie elektronika użytkowa jest częściej sprawdzana online niż offline przez badanych. Pozostałe kategorie, choć są często wyszukiwane w sieci, nadal przejawiają ogromne zainteresowanie w sklepach stacjonarnych. Warto wspomnieć, że 61% badanych wykorzystuje oba kanały w celu zdobycia informacji o interesującym produkcie. Wyłącznie przy wykorzystaniu kanału offline poszukuje 32% badanych. Natomiast, jedynie 5% badanych ogranicza się do korzystania z Internetu jako jedyne źródła informacji o danym produkcie. Pozostali badani deklarują, że nie docierają do informacji o danym produkcie⁶.

⁶ Na podstawie: https://www.ey.com/pl_pl/news/2019/08/jak-polscy-konsumenci-kupuja-w-sieci-wyniki-badania-ey (dostęp: Wrzesień 2020)

1.2. Charakterystyka kupujących online

Jak zostało wcześniej wspomniane, 71% badanych dokonuje zakupów online. W kraju liczącym 38 mln ludzi, liczba ta przedkłada się na 27 mln obywateli korzystających z sieci w celach zakupowych. Jak podaje raport Gemius⁷, zakupy online są bardziej popularne u ludzi młodszych, co można zauważyć poniżej.

Rysunek 3 Charakterystyka kupujących online



Źródło: <https://www.gemius.pl/wszystkie-artykuly-aktualnosci/e-commerce-w-polsce-2020.html>

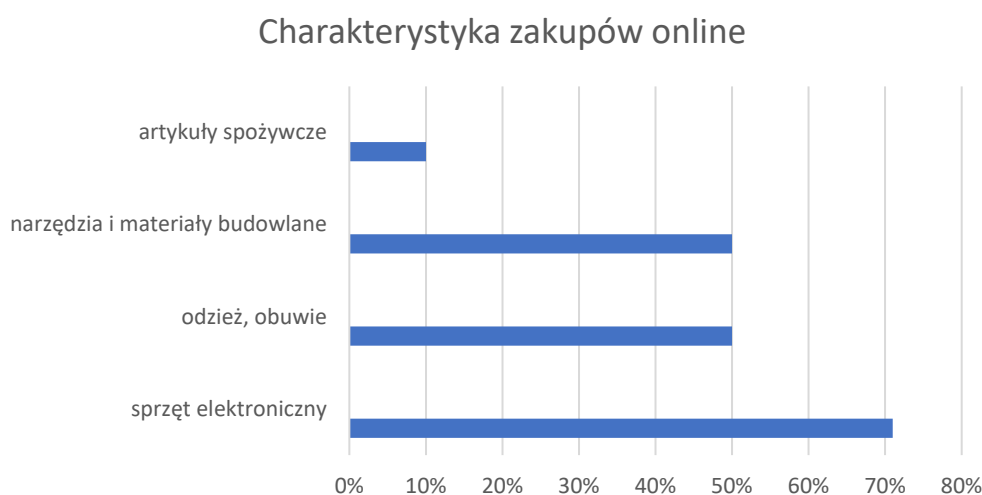
Jak podaje raport powyżej płeć nie ma wpływu na charakterystykę zakupów online. Największy wpływ ma sytuacja materialna gospodarstwa domowego. 94% badanych z dobra i średnią sytuacją materialną dokonuje zakupów online. Kolejne czynniki, u których widać różnicę, to wykształcenie oraz wiek. Ludzie w średnim wieku oraz ze średnim wykształceniem chętniej dokonują zakupów online.

⁷ Na podstawie: <https://www.gemius.pl/wszystkie-artykuly-aktualnosci/e-commerce-w-polsce-2020.html> (dostęp: Wrzesień 2020)

1.3. Najczęściej kupowane produkty online

Jak zostało wcześniej wspomniane, ludzie młodszy częściej korzystają z dostępu do Internetu w celu dokonania zakupów. Raport EY podaje⁸, że najczęściej ludzie kupują w sieci elektronikę. Mniej chętnie zakupywane są artykuły spożywcze. Można to zauważyć na poniższym wykresie.

Rysunek 4 Charakterystyka zakupów online



Źródło: Opracowanie własne na podstawie https://www.ey.com/pl_pl/news/2019/08/jak-polscy-konsumenci-kupuja-w-sieci-wyniki-badania-ey

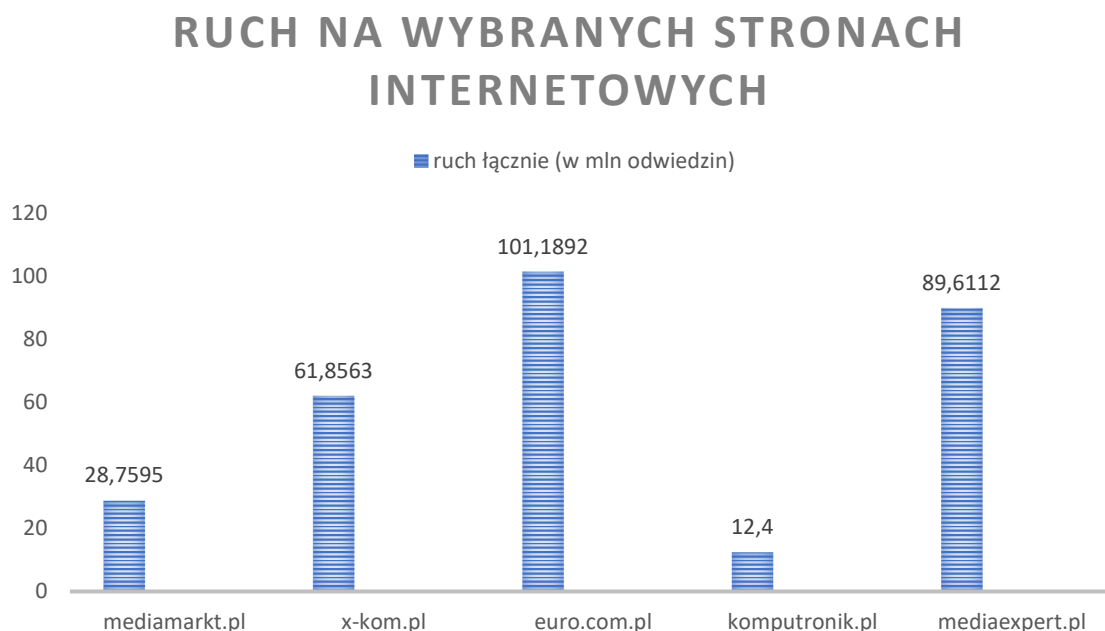
Tylko 10% ludzi dokonuje zakupu artykułów spożywczych online. Zupełnym przeciwieństwem jest zakup sprzętu elektronicznego. Aż 71% osób robi zakupy we wspomnianej kategorii online. Kategoria ta jest zdecydowanie najchętniej wybierana przez Polaków. W Polsce jest kilka stron, które dominują w sprzedaży produktów we wspomnianej kategorii. Aby sprawdzić, które są najczęściej wybierane zostanie sprawdzony ruch na każdej stronie.

⁸ Na podstawie: https://www.ey.com/pl_pl/news/2019/08/jak-polscy-konsumenci-kupuja-w-sieci-wyniki-badania-ey (dostęp: Wrzesień 2020)

1.4. Najczęściej używane strony online

Ruch na stronie pozwoli sprawdzić najczęściej używane strony online. W celu tym zostanie wykorzystane narzędzie dostępne na stronie similarweb.com. Strona ta pokazuje miesięczny ruch generowany na stronie. Jeśli ruch będzie większy, oznacza to, że użytkownicy częściej odwiedzają dana stronę, przeglądając ją. Jeśli ruch na stronie będzie mały, oznacza to, że użytkownicy rzadko odwiedzają dana strona. Wiązać się to będzie z mniejszą sprzedażą produktów na danej stronie. Raport najpopularniejszych polskich witryn ze sprzętem elektronicznym można sprawdzić poniżej. Dotyczy on informacji z września 2020 roku.

Rysunek 5 Ruch na wybranych stronach internetowych



Źródło: Opracowanie własne na podstawie www.similarweb.com

Jak można zauważyć na powyższym wykresie ruch stworzony przez użytkowników jest najwyższy dla strony euro.com.pl. Strona ta należy do polskiego sklepu RTV euro AGD. Wysoki ruch na stronie z pewnością jest spowodowany dużym asortymentem, który oferuje sklep. W sklepie tym można zakupić sprzęt elektroniczny oraz różnego rodzaju AGD do domu. Dla porównania sklep x-kom, koncentruje się głównie na komputerach, telefonach oraz telewizorach. Nie zmienia to faktu, że sklep RTV euro AGD jest jednym z największych

sklepów online w Polsce, który oferuje duży asortyment sprzętu elektronicznego oraz cieszy się ogromną popularnością w Polsce.

W badaniu zostanie wykorzystana strona sklepu RTV euro AGD w celu pobrania danych. Dane będą dotyczyć oferowanych przez sklep komputerów przenośnych. Dane zostaną wykorzystane w celu budowy sieci neuronowej, która będzie prognozować cenę wspomnianego wcześniej sprzętu na podstawie ich parametrów.

2. Uzyskanie danych

W poprzednim rozdziale zostało wspomniane, że dane będą pochodzić ze strony euro.com.pl. W celu tym zostaną wykorzystane metody Web Scrapingu, aby pobrać dane bezpośrednio ze strony. W badaniu zostanie wykorzystany język python, który oferuje możliwość wykorzystania wspomnianego wcześniej narzędzia.

2.1. Ruch na stronie

Przed przystąpieniem do budowy aplikacji będącej odpowiedzialnej za pobranie danych należy przeanalizować ruch na danej witrynie internetowej. Ma to ogromne znaczenie w celu utrzymania płynności danej strony. Aplikacje, które pobierają dane ze stron internetowych działają znacznie szybciej niż zwyczajny użytkownik. Generowanie zbyt dużego ruchu na stronie może przyczynić się do kilku możliwości:

- spowolnienie funkcjonowania strony internetowej
- przeciążenie strony
- zablokowanie użytkownika/ip

Zablokowanie użytkownika jest właściwie konsekwencją obronną witryny internetowej. Jeśli ruch z danego ip jest podejrzanie wysoki, może zostać on zablokowany w celu niegenerowania zbyt dużego ruchu na stronie, który mógłby spowodować zawieszenie się strony internetowej. Oczywiście nie jest to pożądany efekt dla witryny internetowej, ponieważ powoduje przestój w sprzedaży online.

Strona sklepu RTV euro AGD jest bardzo popularna. Jest bardzo często odwiedzana przez użytkowników, co przekłada się na bardzo wysoki ruch miesięczny w granicach 100 mln odwiedzin miesięcznie. Jest to duża liczba, a strona jest z pewnością przygotowana na wysoki ruch. Aplikacja, która zostanie wykorzystana w badaniu może bez problemu wykorzystać wspomnianą wcześniej stronę w celu pobrania danych.

2.2. Charakterystyka strony

Przed przystąpieniem budowy aplikacji należy zapoznać się z charakterystyką strony. Ważnym aspektem przy budowy aplikacji do Web Scrapingu jest wizualizacja tego, jak będzie pracowała aplikacja. Należy sobie odpowiedzieć na pytania związane z rodzajem pobieranych produktów, źródłem danych, źródłem szczegółowych informacji o produktach oraz o stworzonym ruchu przez aplikację.

Źródłem danych, tzn. miejscem gdzie zostanie wysłane pierwsze zapytanie jest strona <https://www.euro.com.pl/laptopy-i-netbooki.bhtml>. Na głównej stronie jest podana liczba rekordów w tej kategorii produktów. Oczywiście nie wszystkie produkty znajdują się na jednej stronie. Aplikacja więc będzie musiała poruszać się po kolejnych stronach w podanej kategorii. Przykładowa liczba stron znajduje się poniżej.

Rysunek 6 Liczba stron w kategorii Laptopy na stronie euro.com.pl (stan na 29.10.2020)

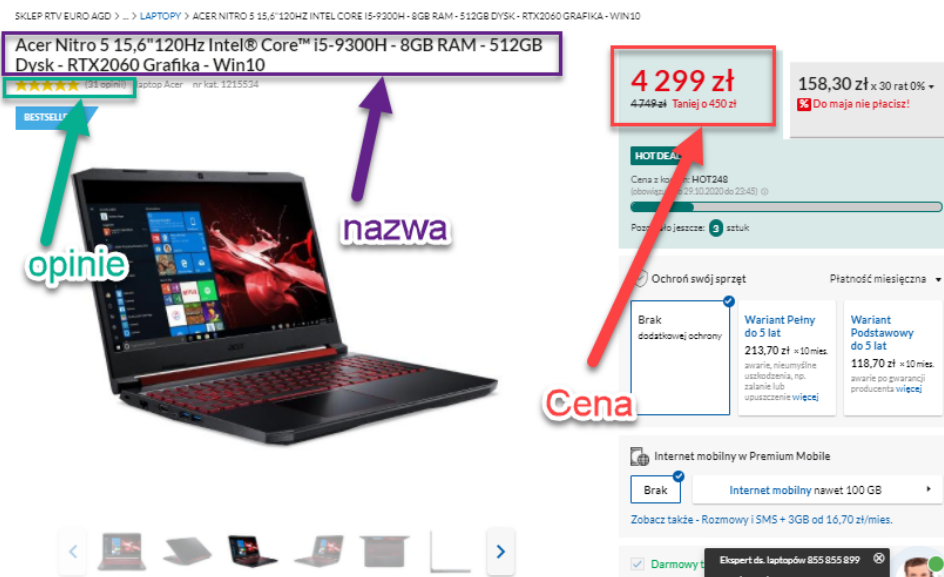


Źródło: Opracowanie własne na podstawie euro.com.pl

Aplikacja otwierając każdą ze stron musi pobrać informację na temat witryn url, która będzie prowadzić do informacji szczegółowych o danym laptopie. Na każdej stronie jest około 25-30 laptopów. Aby otrzymać całkowitą liczbę produktów, należy więc pomnożyć liczbę stron przez liczbę produktów na stronie.

Po uzyskaniu listy witryn url na temat laptopów, aplikacja powinna wysłać zapytanie do każdej z nich w celu pobrania szczegółowych informacji. Przykładowe miejsca oraz charakterystyka pobierania informacji znajdują się na obrazkach poniżej.

Rysunek 7 Strona internetowa laptopa Acer sklepu RTV euro AGD



Źródło: Opracowanie własne na podstawie euro.com.pl

Rysunek 8 Dane techniczne laptopa Acer sklepu RTV euro AGD

Dane techniczne laptop Acer Nitro 5 15,6"120Hz Intel® Core™ i5-9300H - 8GB RAM - 512GB Dysk - RTX2060 Grafika - Win10	
PARAMETRY EKRANU	
Ekran	15,6 cala, 1920 x 1080 pikseli
Typ matrycy	matowa LED, IPS
Częstotliwość odświeżania obrazu	120 Hz
PROCESOR	
Model procesora	Intel® Core™ i5 9gen 9300H
Częstotliwość taktowania	2,4 - 4,1 GHz
Liczba rdzeni procesora	4
Pamięć podręczna CACHE	8 MB
Zintegrowany układ graficzny	Intel® UHD Graphics 630
PAMIĘĆ RAM	
Pamięć RAM	8 GB
Typ pamięci RAM	DDR4
Możliwość rozszerzenia pamięci RAM do	32 GB
KARTA GRAFICZNA	
Model karty graficznej	nVidia® GeForce RTX2060
Pamięć karty graficznej	6 GB

Źródło: Opracowanie własne na podstawie euro.com.pl

Dane z rysunku 7 oraz 8 pochodzą z tej samej strony. Rysunek 7 przedstawia początkowy widok jaki zastaje użytkownik po otwarciu witryny url danego laptopa. Zostały na nim zaznaczone informacje jakie powinna zbierać aplikacja. Rysunek 8 przedstawia szczegółowe dane techniczne laptopa. Wszystkie te dane powinny zostać pobrane przez aplikację. Choć na

rysunku przedstawia się to w jasny sposób, to stworzenie uniwersalnej aplikacji nie będzie łatwe. Należy przewidzieć, że informacje zawarte są w dokumencie HTML, a ścieżki oraz miejsce może się różnić w zależności od strony produktu. Warto również zaznaczyć, że liczba informacji może się różnić w zależności od laptopa. Niektóre laptopy będą miały przedstawioną pełną specyfikację, a niektóre tylko szczątkowe informacje, lub niektóre informacje będą niekompletne. Aplikacja powinna poprawnie zapisać dane niezależnie od różnicy w wyglądzie stron.

2.3. Web Scraping

Początku Web Scrapingu sięgają lat 90. XX wieku. Matthew Gray, student fizyki z Massachusetts, w 1993 roku postanowił gromadzić dane na temat witryn internetowych. W owym czasie było ich niewiele. W czerwcu 1993 roku na całym świecie było około stu stron internetowych. Matthew stworzył program o nazwie „World Wide Web Wanderer”, który został napisany w języku Perl. Program był dość kontrowersyjny, ponieważ jego wczesne wersje, gwałtownie zbierając dane, obniżały wydajność stron internetowych. Było to spowodowane tym, że program odwiedzał te same strony setki razy dziennie. W późniejszym czasie autor zmienił sposób działania programu. Matthew, który jako pierwszy stworzył robota do przeszukiwania sieci, miał potencjał, aby stworzyć pierwszą wyszukiwarkę internetową, kilka lat wcześniej niż Yahoo czy Google. Jednak nie to było celem autora, ale takim sposobem zainspirował wiele programistów do kontynuowania idei robotów internetowych.⁹

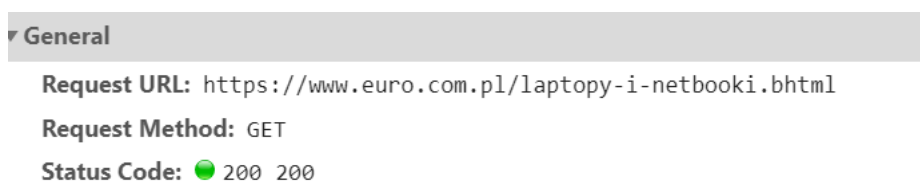
W dzisiejszych czasach Web Scraping znany jest jako metoda do pobierania danych ze stron internetowych. Obecnie na całym świecie jest niezliczona liczba stron internetowych, a dane są niezwykle cenne. Mogą one zawierać dane na temat produktów w sklepach, ceny mieszkań, ofert hoteli itp. Metoda to polega na wysłaniu zapytania do strony internetowej, otrzymaniu odpowiedzi zwrotnej oraz wyciągnięcie interesujących informacji ze strony.

⁹ Na podstawie: <https://history-computer.com/Internet/Conquering/Wanderer.html> (Listopad 2020)

2.3.1. Wysyłanie zapytania protokołu HTTP

Pierwszym krokiem w budowanie aplikacji do Web Scraping jest zweryfikowanie URL oraz metody jakiej wykonuje dany URL. Strony o laptopach na stronie sklepu RTV euro AGD znajdują się pod adresem: <https://www.euro.com.pl/laptopy-i-netbooki.bhtml>. Wykorzystuje ona metody GET, co można zauważyć na poniższym obrazku.

Rysunek 9 Zapytanie do strony euro.com.pl



Źródło: Opracowanie własne

Metoda GET charakteryzuje się tym, że parametry podane są w adresie. W tym przypadku parametrem jest ścieżka w domenie strony euro.com.pl, do której uda się aplikacja. Druga popularna metoda wykorzystywanych na wielu stronach internetowych jest metoda POST. Odróżnia się od metody GET tym, że parametry wysyłane są w wiadomości i nie są widoczne w adresie URL. Metoda POST często wykorzystana jest przy logowaniu danej witryny internetowej. Pozostałe metody http, które są zdecydowanie rzadziej używane to: HEAD, PUT, DELETE, OPTIONS, TRACE, CONNECT, PATCH.¹⁰

Po wysłaniu zapytania użytkownik bądź aplikacja otrzymuje odpowiedź, a więc zawartość strony internetowej. Rodzaj odpowiedzi jest uzależniona od wielu czynników, a typ odpowiedzi jest przedstawiany w postaci kodu HTTP.

2.3.2. Otrzymanie odpowiedzi protokołu HTTP

Kody odpowiedzi można podzielić na: informacyjne, powodzenia, przekierowania, błędy aplikacji klienta oraz błędy serwera HTTP.¹¹ Każdy rodzaj odpowiedzi ma charakterystyczny dla siebie kod. Kod składa się z trzech cyfr. Odpowiedni rodzaj kodu ma wydzielony przedział

¹⁰ Na podstawie: https://pl.wikipedia.org/wiki/Hypertext_Transfer_Protocol (dostęp: Listopad 2020)

¹¹ Na podstawie: https://pl.wikipedia.org/wiki/Kod_odpowiedzi_HTTP (dostęp: Listopad 2020)

cyfr. Kody informacyjne przyjmują wartość z numerem 1 z przodu, tzn. 100-199. Kody powodzenia z numerem 2 z przodu. Kody przekierowania z numerem 3, kody błędu aplikacji klienta z 4, a kody błędu serwera z 5. Zdecydowanie najbardziej interesujące dla użytkownika są trzy rodzaje kodu: kod powodzenia, kod błędu aplikacji klienta oraz kod błędu serwera http. Kod powodzenia informuje o prawidłowym zapytaniu użytkownika oraz poprawna przesłana odpowiedzi. Kod błędu aplikacji klienta informuje o wysłaniu złego zapytania klienta bądź o zablokowaniu klienta. Kod błędu serwera oznacza zerwania zapytania bądź odpowiedzi z przyczyn technicznych serwera.

Najczęściej zwracany kod informacyjny a zarazem najbardziej pożądanym przez użytkownika jest kod odpowiedzi powodzenia z numerem 200. Oznacza on, że odpowiedź została pozytywnie przesłana do użytkownika. Taki wynik jest również pożądanym przez aplikację w celu pobrania danych. Aby dane zostały poprawnie pobrane ze strony, aplikacja musi wysłać zapytanie do serwera na temat każdego laptopa. Jeśli zapytanie będzie niepoprawne bądź odpowiedź będzie inna niż pozytywna, dane nie zostaną pobrane lub pobrane zostaną w sposób niepełny.

W wyniku pozytywnej odpowiedzi z numerem 200, z serwera jest przesyłana zawartość strony, czyli dokument zawierający informacje znajdujące się na stronie. Dokument przyjmuje postać HTML. W celu poruszania się po dokumencie należy dobrze zrozumieć funkcjonowanie ścieżek dokumentu HTML, aby dostać się do konkretnych informacji.

2.3.3. Wyciąganie informacji

Jak zostało wcześniej wspomniane odpowiedź ze strony jest dokumentem HTML. Aby wyciągnąć informacje należy sprawnie poruszać się po wspomnianym dokumencie. Informacje w dokumencie HTML znajdują się w znacznikach. Przykład takiego dokumentu można zauważyć poniżej.

Rysunek 10 Przykład dokumentu HTML ze strony w3schools.com

```
<!DOCTYPE html>
<html lang="en">
<head>
  <title>Title of the document</title>
</head>
<body>

<h1>This is a heading</h1>
<p>This is a paragraph.</p>

</body>
</html>
```

This is a heading

This is a paragraph.

Źródło: Opracowanie własne na podstawie w3schools.com

Fragment rysunku przedstawiony po prawej to faktyczny wygląd strony. Część po lewej przedstawia dokument HTML, który tworzy stronę internetową. Napis „This is a heading” znajduje się w znaczniku „title”, zaś „this is a paragraph” w znaczniku „p”. Jest to oczywiście bardzo prosty przykład, a strony internetowe są zdecydowanie bardziej skomplikowane. Dla przykładu sprawdźmy, gdzie znajduje się cena wspomnianego wcześniej laptopa Acer.

Rysunek 11 Strona internetowa laptopa Acer sklepu RTV euro AGD

SKLEP RTV EURO AGD > ... > LAPTOPY > ACER NITRO 5 15,6" 120Hz INTEL CORE i5-9300H - 8GB RAM - 512GB DYSK - RTX2060 GRAFIKA - WIN10

**Acer Nitro 5 15,6" 120Hz Intel® Core™ i5-9300H - 8GB RAM - 512GB
Dysk - RTX2060 Grafika - Win10**

★★★★★ (31 opinii) laptop Acer nr kat. 1215534

BESTSELLER



4 299 zł

~~4 549 zł~~ - Taniej o 400 zł

158,30 zł x 30 rat 0% +

Do maja nie płacisz!

HOT DEAL

Cena z kodem: HOT248

(obowiązuje do 29.10.2020 do 23:45) @

Pozostało jeszcze: 9 sztuk

Ochroni swój sprzęt

Płatność miesięczna

Brak
dodatkowej ochrony

**Wariant Pełny
do 5 lat**

213,70 zł x 10 mies.
awaria, nieumyślne
uszkodzenia, np.
zalewanie lub
upuszczenie [więcej](#)

**Wariant
Podstawowy
do 5 lat**

118,70 zł x 10 mies.
awaria po gwarancji
producenta [więcej](#)

Źródło: Opracowanie własne na podstawie euro.com.pl

Rysunek 12 Fragment kodu źródłowego strony laptopa Acer sklepu RTV euro AGD

```
<div id="product-upsetting" class="product-upsetting standard">
  <div class="product-detail-prices">
    <ul class="price-tabs">
      <li data-context="standard" class="price-tab price-promotion-tab is-active">
        <a href="#" class="tab-button">
          <div class="product-price">
            4 299 zł
          </div>
        </a>
      </li>
      <li data-context="instalment" class="price-tab price-instalment-tab promo-installment">
      </li>
    </ul>
    <div class="price-tab-content">

```

Źródło: Opracowanie własne na podstawie kodu źródłowego strony euro.com.pl

Zaznaczona cena na rysunku 11 znajduje się w znaczniku „div” w klasie „produkt-price”. Rysunek 12 jest to tylko fragment dokumentu HTML. Całkowita ścieżka do ceny przedstawia się następująco:

`/html/body/main/section[2]/div[2]/div[3]/div[1]/ul/li[1]/a/div[1]`

Jest ona dość skomplikowana. Przy niewielkiej zmianie strony dana ścieżka może przekierować do zupełnie innego miejsca. Aby aplikacja działała uniwersalnie powinna być odporna na tego typu zmiany. Powinna ona szukać znacznika „div” z klasą „produkt-price” gdziekolwiek na stronie internetowej, a następnie pobrać informacje z danego znaczniku.

Niektóre jednak informacje nie są tak łatwo dostępne. Dotyczy się to głównie szczegółowych danych technicznych laptopów. Dane te są zapisane w dość prosty sposób i ciężko jest odróżnić, gdyż znaczniki nie posiadają charakterystycznych dla nich klas. Można to zauważyć na przykładzie informacji na temat ekranu laptopa Acer, która znajduje się na obrazku poniżej.

Rysunek 13 Fragment kodu źródłowego laptopa Acer sklepu RTV euro AGD

```
<tbody>
  <tr class="group-name"></tr>
  <tr>
    <td>
      <a title="Przekątna matrycy - laptopy" class="js-dictionary" href="/sownik.bhtml?definitionId=357249498">
        Ekran
      </a> == $0
    </td>
    <td>
      15,6 cala, 1920 x 1080 pikseli
    </td>
  </tr>
  <tr>
    <td></td>
    <td>
      matowa LED, IPS
    </td>
  </tr>
</tbody>
```

Źródło: Opracowanie własne na podstawie kodu źródłowego strony euro.com.pl

W tym przypadku aplikacja powinna wyszukać znacznik „a” z dokładnym tekstem „Ekran”, a następnie pobrać informację z kolejnego znacznika i przypisać do odpowiedniej kolumny. Podobny sposób powinien zostać zaaplikowany na pozostałe informacje o laptopie.

Dane, które zostaną pobrane przez aplikację będą surowe. Bardzo możliwe, że będą one nie pełne, ponieważ z pewnością niektóre parametry nie zostały w pełni podane dla wszystkich laptopów. Może być to związane ze specyfikacją laptopa, bądź po prostu brakiem danych informacji. Aby dane mogły zostać wykorzystane muszą zostać odpowiednio przerobione.

Przetworzone dane mogą zostać wykorzystane do budowy sieci neuronowej. Sieć ta jest wykorzystywana w wielu podobnych zagadnieniach w przeszłości. Wielu autorów przyczyniło się do rozwoju sieci neuronowych, takich jak: McCulloch oraz Pitts czy McClelland i Rumelhart. Kolejny rozdział zostanie poświęcony wyjaśnieniu budowy sieci neuronowej.

3. Sieć neuronowa

Model sieci neuronowej został zaproponowany przez McCullocha i Pillsa¹² w 1943 r. jako alternatywna metoda programowania algorytmów badań Karla Lasheya w latach 20. XX wieku. Ich idea stanowi podstawowe ogniwo większości używanych modeli sieci neuronowych. Istotnym elementem modelu zaproponowanego przez autorów jest sumowanie sygnałów wejściowych odpowiednią wagą oraz poddanie otrzymanej sumy działaniu odpowiedniej funkcji aktywacji.

Kolejnym ciekawym autorem był Hebb¹³, który przedstawił swoje badanie w 1949 r. Według autora informacje mogą być przechowywane w strukturze połączeń między neuronami. Zaproponował także metodą uczenia wraz ze zmianą wag w połączeniach między neuronami. Metoda ta jest znana to dzisiaj pod postacią „reguła Hebba”.

W latach 50. XX wieku powstały pierwsze sieci neuronowe. Przykładem takiej sieci jest perceptron, który zostały zaprojektowany przez Franka Rosenblatta¹⁴ w 1958 roku. Sieć miała za zadanie klasyfikację obrazów. Sieć, zaprojektowana przez autora, była zdolna do nauki przez modyfikację połączeń prowadzących do układów progowych.

Znaczące badania w budowie sieci neuronowej przeprowadzili David Hubel i Torsten Wiesel¹⁵ w 1959 roku. Naukowcy nagrywali zachowanie neuronów wzrokowych kota manewrując światłem. W swoich badaniach wykazali, że w korze wzrokowej znajduje się mapa topograficzna, która reprezentuje pole widzenia, gdzie pobliskie komórki przetwarzają informacje z sąsiednich pól. Co więcej, w ich pracy ustalono, że neurony w korze wzrokowej są ułożone według precyzyjnej architektury. Komórki o podobnych funkcjach są zorganizowane w kolumny, małe maszyny obliczeniowe, które przekazują informacje do wyższego obszaru mózgu, gdzie powstaje obraz wizualny. Podsumowując, ich praca ujawniła, w jaki sposób neurony kory wzrokowej zakodowały cechy obrazu, podstawowe właściwości obiektów, które pomagają człowiekowi budować percepcję otaczającego świata. Badanie to

¹² McCulloch i Pitts. (1943). A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN. University of Illinois

¹³ HEBB, D. O. (1949). The Organization of Behavior.

¹⁴ Rosenblatt, F. (1958). THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN. Cornell Aeronautical Laboratory.

¹⁵ HUBEL, D. H. i WIESEL, T. N. (1959). RECEPTIVE FIELDS OF SINGLE NEURONES IN THE CAT'S STRIATE CORTEX. Wilmer Institute, The Johns Hopkins Hospital and University, Baltimore.

przyczyniło się do rozwoju sieci konwolucyjnych, które wykorzystywane jest do rozpoznawania obrazów, a autorzy zostali uhonorowani Nagrodą Nobla w 1981 roku.

W 1965 roku Ivakhnenko i Lapa¹⁶ opublikowali pierwszy algorytm przypominający głębokie uczenie. Charakteryzował się kilkoma warstwami ukrytymi. W każdej warstwie wybrali najlepsze cechy metodami statystycznymi i przekazali je do następnej warstwy. Nie używali propagacji wstecznej do trenowania sieci, ale stosowali dopasowywanie metodą najmniejszych kwadratów warstwa po warstwie, gdzie poprzednie warstwy były dopasowywane niezależnie od późniejszych warstw. Ivakhnenko¹⁷ w 1971 roku opisał głęboką sieć z ośmioma warstwami. Warstwy były przyrostowo rozwijane i trenowane za pomocą analizy regresji, a następnie zbędne neurony zostały ucinane.

W 1986 r. Rumelhart i McClelland¹⁸ przedstawili sieć neuronową jako zbiór neuronów z niewielką pojemnością danych połączonych w jednym kierunku złączami zwanymi aksonami. Węzły działają z lokalnymi danymi dostarczonymi przez aksony oraz posiadają odpowiednie wagi parametrów. Reguła propagacji określa, w jaki sposób dane wejściowe do neuronu są wartościowane i przetwarzane.

Sieć neuronowa została zaprojektowana tak, aby odzwierciedlać działanie ludzkiego mózgu. Wydaje się, że wspomniani wcześniej autorzy, zapoczątkowali temat sieci neuronowych, który jest dziś powszechnie znany.

3.1. Budowa neuronu w sieci neuronowej

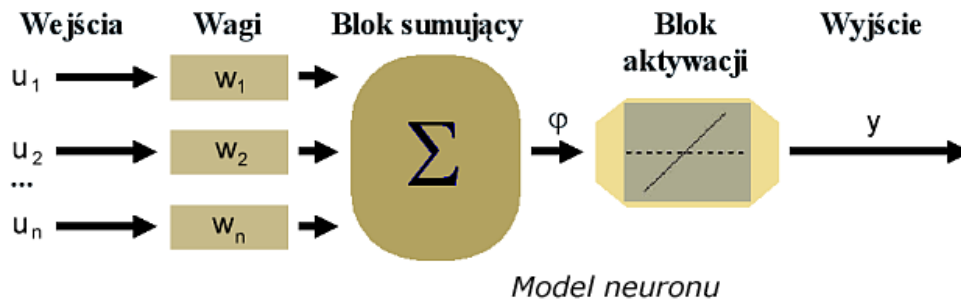
Na podstawie działania neurony w mózgu człowieka można zbudować schemat działania sztucznego neuronu w sieci neuronowej. Każdy neuron charakteryzuje się tym, że na początku dostaje dane, zwane danymi wejściowymi. Dane te są modyfikowane przez odpowiednie wagi. Następnie są sumowane oraz poddane funkcji aktywacji w bloku aktywacji. Dane na końcu przesyłane są dalej do kolejnej warstwy w sieci neuronowej bądź są wynikiem sieci neuronowej, jeśli dany neuron był neuronem wyjściowym. Schemat działania neurony można zauważyć na poniższym rysunku.

¹⁶ Ivakhnenko, A. G. i Lapa, V. G. (1965). *Cybernetic Predicting Devices*. CCM Information Corp.

¹⁷ Ivakhnenko, A. G. (1971). *Polynomial Theory of Complex Systems*. IEEE Transactions on Systems, Man and Cybernetics.

¹⁸ Rumelhart, D. i McClelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge: MIT.

Rysunek 14 Model neuronu

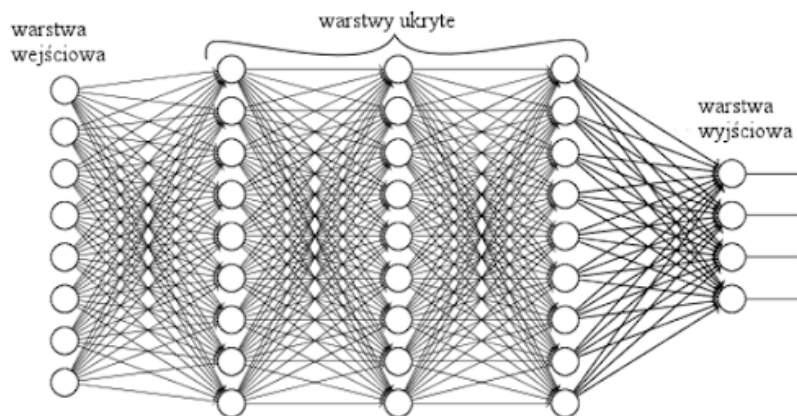


Źródło: <http://galaxy.uci.agh.edu.pl/>

3.2. Budowa sieci neuronowej

Podstawowy model sieci neuronowej posiada warstwę wejściową, jedną lub kilka warstw ukrytych oraz warstwę wyjściową. W warstwach tych znajdują się neurony. Połączenie pomiędzy warstwami jest możliwe dzięki synapsom. Budowa sieci neuronowej została przedstawiona poniżej.

Rysunek 15 Budowa sieci neuronowej



Źródło: <http://www.deltami.edu.pl/>

Warstwa wejściowa charakteryzuje się tym, że dostają dane z zewnątrz, przetwarza je oraz wysyła dane dalej do neuronów w warstwie ukrytej. Przeliczają one wartości z poprzedniej warstwy i przekazują je do kolejnej warstwy ukrytej, bądź warstwy końcowej, czyli wyjścia.

Jest to uzależnione od ilości warstw w modelu sieci neuronowej. Warstwa wyjścia dokonuje ostatecznych obliczeń, a następnie wysyła wynik uzyskany przez model.

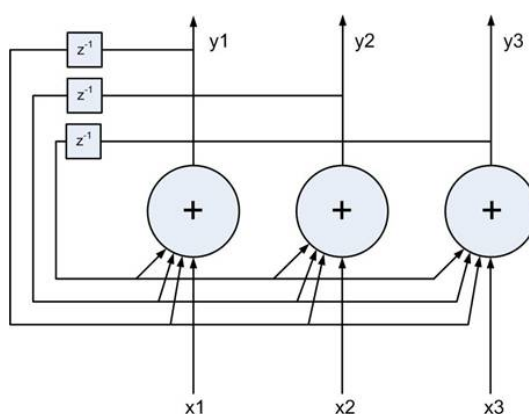
Przedstawiony schemat dotyczy podstawowej sieci jednokierunkowej. Jest to bardzo często wykorzystywana metoda w budowie sieci neuronowej, lecz nie jedyna. Pozostałe sieci to: rekurencyjna oraz komórkowa. Obie sieci są bardziej skomplikowane, a wykorzystanie ich jest uzależnione od potrzeby budowy modelu.

3.3. Rodzaje sieci neuronowych

Sieć jednokierunkowa jest bardzo intuicyjna i prosta w działaniu. Ma ona szereg zastosowań i jest chętnie wybierana w różnego rodzaju badaniach. Na początku przyjmuje dane wejściowe, przetwarza je i przesyła dalej do warstw ukrytych i tak aż do warstwy wyjściowej, która przesyła wynik końcowy. Działania odbywają się w jednym kierunku.

Sieć rekurencyjna, zwana również dynamiczną, jest znacznie bardziej skomplikowana od sieci jednokierunkowej. Przepływ sygnałów jest dwukierunkowy, a w sieci występują sprzężenia zwrotne. Oznacza to, że wyjścia neuronów mogą być połączone z wejściami neuronów w tej samej lub warstwie pośredniej. Przykładem sieci rekurencyjnej jest sieć Hopfielda, którą przedstawia poniższy rysunek.

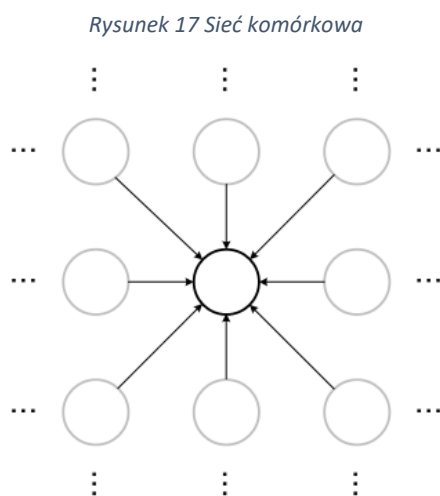
Rysunek 16 Sieć rekurencyjna



Źródło: <http://galaxy.agh.edu.pl/>

Sieć komórkowa charakteryzuje się wzajemnym sprzężeniem między elementami najbliższego sąsiada. Połączenia są w ogólności nieliniowe oraz opisane są poprzez układ równań nieliniowych. W metodzie tej trudno jest znaleźć idealną równowagę pomiędzy

skutecznością, efektywnością oraz uniwersalnością. Przykładem tego typu sieci jest sieć Kohonena, która jest przedstawiona na poniższym rysunku.



Źródło: <http://galaxy.agh.edu.pl/>

Znając rodzaje sieci neuronowych można wybrać tą, która wydaje się najbardziej optymalna do wybranego zagadnienia. Każda jednak sieć potrzebuje dane, które będzie przetwarzała. Za przekształcanie danych odpowiedzialny jest blok aktywacji, który może zostać opisany za pomocą kilku funkcji aktywacji. Wybór powinien zależeć od rodzaju danych oraz zagadnienia, które dotyczą.

3.4. Funkcje aktywacji

Wybór funkcji aktywacji jest istotną kwestią w budowie sieci neuronowej. Funkcja ta oblicza wartość wyjścia neuronów w sieci neuronowej. Jest kilka rodzaju funkcji aktywacji. Jednym z kryteriów wyboru są dane jakich dotyczą oraz wartość wyjścia, jaki powinien uzyskać dany neuron.

Jedną z podstawowych funkcji aktywacji jest funkcja liniowa. Wartość funkcji wyjścia nie jest ograniczona, a funkcja przedstawia się następującym wzorem:

Równanie 1 Wzór funkcji liniowej

$$f(x) = ax + b$$

Funkcja skoku jednostkowego unipolarna, przyjmuje wartość wyjścia 0 lub 1, najczęściej używana jest w modelach podejmujących decyzję, a wzór przedstawia się następująco:

Równanie 2 Wzór funkcji skoku jednostkowego unipolarnej

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Funkcja skoku jednostkowego bipolarna różni się od unipolarnej o wartość wyjścia jakie przyjmuje. Funkcja prezentuje się następująco:

Równanie 3 Wzór funkcji skoku jednostkowego bipolarnej

$$f(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

Funkcja sigmoidalna unipolarna, która znana jest również jako funkcja logistyczna. Przedstawia się ona następująco

Równanie 4 Wzór funkcji sigmoidalnej unipolarnej

$$f(x) = \frac{1}{1 + e^{-\beta x}}$$

Tanges hiperboliczny prezentujący się następująco:

Równanie 5 Wzór tangesa hiperbolicznego

$$f(x) = \frac{2}{1 + e^{-\beta x}} - 1 = \frac{1 - e^{-\beta x}}{1 + e^{\beta x}}$$

3.5. Nauka sieci neuronowej

Zaletą sieci neuronowej jest możliwość nauki. Sieć potrafi dostosować wagi połączeń pomiędzy neuronami tak, aby uzyskać jak najlepszy wynik. Metody nauki można podzielić na trzy grupy, które nieco się od siebie różnią.

Pierwszą metoda nauki jest uczenie z nadzorem. W metodzie tej celem jest minimalizacja pewnej funkcji kosztu, która jest obliczana przez sieć neuronową. Funkcja kosztu powinna brać

pod uwagę wyniki obliczane przez sieć neuronową oraz prawidłowe wyniki dla danych wejściowych, które są potem porównywane. Sieć jest trenowana na zbiorze danym zwanym zbiorem uczącym, a testowana jest na innym zbiorze danym, na których nie była wcześniej trenowana, zwanym testowym. Ideą tego typu uczenia jest minimalizacja wspomnianej wcześniej funkcji kosztu.

Kolejną metodą jest nauka bez nadzoru. Główną różnicą od nauki z nadzorem jest brak prawidłowego wyniku, do którego mógłby się porównać algorytm. Sieć stara się nauczyć pewnych wzorców, dzięki którym może odróżnić od siebie dane. Sieć nie będzie w stanie przedstawić konkretnego wyniku, ale będzie w stanie dane przypisać do danej grupy. Przykładem takiej typu metody jest analiza skupień.

Ostatnim rodzajem nauki jest nauka ze wzmocnieniem. Dla tego typu nauki, zbiór treningowy nie jest od razu dostępny w całości. W algorytmie natomiast występują akcje oraz przyznawane są nagrody. Sieć znajduje się w interakcji ze środowiskiem i wykonuje pewne akcje na podstawie obserwacji. Każda akcja powoduje przejście do innego stanu oraz otrzymaniu nagrody ze związaną wcześniej akcją. Algorytm skupia się na wykonywaniu akcji, które spowodują otrzymanie najwyższej nagrody. Sieć może być uczona na temat akcji jakie ma wykonywać, aby zmaksymalizować nagrodę.

Nauka ze wzmocnieniem jest zdecydowanie trudniejszą metodą w porównaniu do metod nauki z nadzorem bądź bez nadzoru. Nauka bez nadzoru jest wykorzystywana w zagadnieniach grupowania danych, w których nie jest dostępny wynik końcowy, zaś nauka z nadzorem jest wykorzystywana ze zbiorami danych, w której znany jest wynik, a algorytm powinien dopasować dane w jak najlepszy sposób.

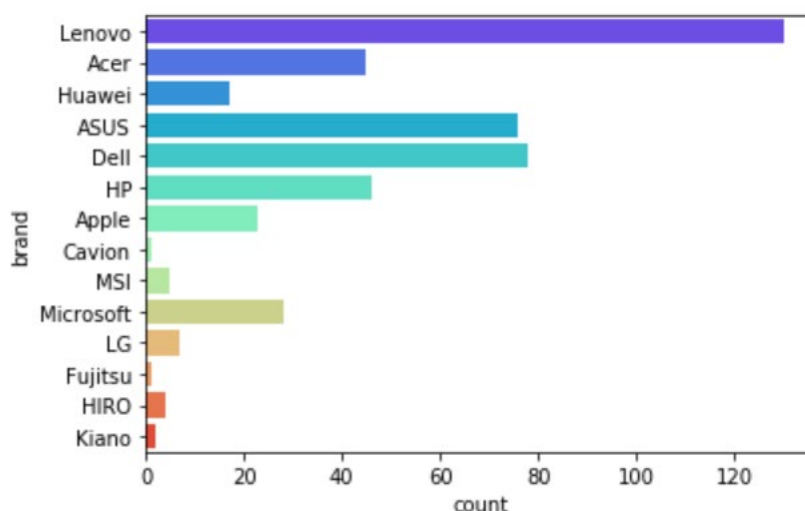
4. Model sieci neuronowej

Sieć neuronowa zostanie wykorzystana w celu prognozy cen laptopów w sklepie internetowym. Sieć wykorzystuje dane wejściowe w celach nauki oraz przystosowania neuronów, aby jak najdokładniej prognozować cenę danego laptopa.

4.1. Opis danych

Dane użyte w badaniu pochodzą ze strony sklepu internetowego RTV EURO AGD. Dotyczą one laptopów, które znajdują się w ofercie danego sklepu. Zostały pobrane w dniu 19 czerwca 2020 roku. Na dany moment w sklepie internetowym było dostępne 459 laptopów. Sklep zdecydowanie najwięcej oferuje laptopów marki Lenovo, co można zauważyć na poniższym wykresie.

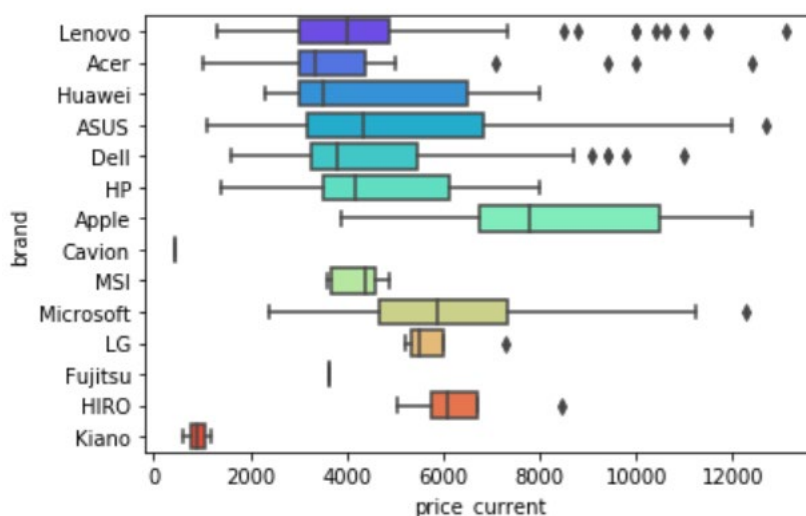
Rysunek 18 Zestawienie marek laptopów w ofercie sklepu RTV euro AGD



Źródło: Opracowanie własne

W swojej ofercie sklep posiada ponad 120 laptopów marki Lenovo. Do najpopularniejszych marek należą także: Dell, ASUS, Acer, HP oraz Microsoft. Oczywiście ilość nie wiąże się z ceną. Najdroższy laptop w ofercie sklepu RTV euro AGD należy do Lenovo, jednak to Apple oferuje produkty z wyższej półki, czyli sprzęt dla którego średnia cena jest najwyższa. Można to zauważyć na poniższym wykresie.

Rysunek 19 Średnie ceny marek laptopów sklepu RTV euro AGD



Źródło: Opracowanie własne

Rysunek 19 pokazuje, że najdroższy laptop jest dostępny za ponad 12 tys. zł. Natomiast średnia cena produktu Apple znajduje się w granicach 7-8 tys. zł. W danym porównaniu jest zdecydowanie z przodu przed Microsoft oraz HIRO. Marka Lenovo, która okazała się najpopularniejszą w ofercie sklepu, oferuje laptopy z średnią ceną w granicach 4 tys. zł.

Oczywiście jest wiele wskaźników, które mogą wpływać na daną cenę laptopa. Badanie polega na zebraniu wszystkich cech laptopów, znalezieniu najistotniejszych zmiennych oraz wykorzystaniu ich w celach prognozy cen. Zbiór danych zawiera 28 zmiennych, które mogą zostać wykorzystane w analizie ceny laptopów. Zmienne te zostaną odpowiednio przekształcone, aby mogły być wykorzystane w budowaniu sieci neuronowej. Współczynniki wraz z opisem zostały przedstawione poniżej.

Tabela 1 Zmienne wykorzystane w badaniu

Nazwa zmiennej	Opis
brand	Marka laptopa
screen	Rozmiar ekranu
screen_type	Typ ekranu
screen_freq	Częstotliwość odświeżania ekranu
CPU	Rodzaj procesora
CPU_clock_rate	Taktowanie procesora
cores	Liczba rdzeni procesora
cache	Pamięć cache procesora
ram	Pamięć ram
ram_type	Rodzaj pamięci ram
graphic_card	Rodzaj karty graficznej
graphic_card_memory	Pojemność karty graficznej
disk_SSD	Pojemność dysku SSD
disk_connection_type	Rodzaj połączenia dysku z płytą główną
disk_HDD	Pojemność dysku HDD
rotation_speed	Prędkość dysku
software	Dostępne oprogramowanie
camera	Kamera wbudowana
speakers	Typ głośników
speakers_number	Liczba głośników
speakers_power	Moc głośników
microphone	Dostępność mikrofonu
battery	Rodzaj baterii
dimension	Rozmiar laptopa
weight	Waga laptopa
warranty	Gwarancja laptopa
battery_warranty	Gwarancja baterii
additional_information	Dodatkowe informacje

Źródło: Opracowanie własne

Tabela 1 przedstawia listę zmiennych, które zostaną wykorzystane w celu analizy ceny sprzętu. Oczywiście, dane na stronach internetowych nie są kompletne. Wiele rekordów ma brakujące informacje, z którymi należy sobie poradzić. Kolejnym krokiem będzie przekształcenie zmiennych tak, aby mogły być wykorzystane w dalszej analizie.

4.2. Przygotowanie danych

Pierwszym krokiem będzie sprawdzenie ilości brakujących rekordów. Aby maksymalnie wykorzystać dostępne dane należy pominąć jak najmniej rekordów, a ewentualne braki odpowiednio przekształcić. Tabela poniżej przedstawia brakujące rekordy.

Tabela 2 Zestawienie brakujących rekordów

Nazwa zmiennej	Liczba brakujących rekordów
brand	0
screen	0
screen_type	39
screen_freq	399
CPU	0
CPU_clock_rate	0
cores	3
cache	9
ram	0
ram_type	13
graphic_card	0
graphic_card_memory	97
disk_SSD	0
disk_connection_type	7
disk_HDD	423
rotation_speed	429
software	0
camera	5
speakers	17
speakers_number	3
speakers_power	359
microphone	2
battery	101
dimension	13
weight	8
warranty	0
battery_warranty	109
additional_information	1

Źródło: Opracowanie własne

Tabela 2 pokazuje, że w danych jest sporo kolumn z brakującymi danymi. Niektóre kolumny mają pojedyncze brakujące wartości. Dane w takich kolumnach można odpowiednio wypełnić. Niektóre jednak kolumny, wręcz przeciwnie, mają sporo brakujących wartości. Uzupełnianie danych może spowodować więcej szkód niż dobrego. Takie kolumny należy

wykluczyć z analizy. Kolumny jak: screen_freq, disk_HDD, rotation_speed oraz speakers_power zostają wykluczone na pierwszym etapie, ponieważ posiadają zbyt dużą liczbę brakujących wartości.

Kolejnym krokiem w analizie będzie poradzenie sobie z brakującymi wartościami. Kolumny można podzielić na katégoryczne i numeryczne. Jak wcześniej było wspomniane, jest kilka możliwości poradzenia sobie z brakującymi danymi. Jednym z nich była średnia, dla kolumn numerycznych, oraz wykorzystanie najczęściej występującej wartości, dla kolumn katégorycznych. W kolejnych etapach zostanie przedstawione po jednym przykładzie odpowiedniego przekształcenia zmiennej katégorycznej i numerycznej.

4.3. Przekształcenie zmiennej katégorycznej

Przykładem zmiennej katégorycznej w danych jest zmienna screen_type. Na początek warto spojrzeć jakie dane znajdują się w tej kolumnie. Zostanie to pokazane na rysunku poniżej.

Rysunek 20 Zestawienie wartości w zmiennej screen_type

```
array(['matowa LED, IPS', 'blyszczaca LED', 'matowa LED',  
      'matowa LED, SVA', 'matowa LED, TN', 'blyszczaca LED, IPS',  
      'blyszczaca LED, TN', nan, 'blyszczaca OLED', 'blyszczaca AMOLED',  
      'matowa LED, VA'], dtype=object)
```

Źródło: Opracowanie własne

Rysunek 20 pokazuje, że w kolumnie screen_type znajduje się aż 10 unikalnych wartości oraz są także rekordy z brakującymi danymi. Typ matrycy jest dość istotny w wyborze laptopa, więc brakujące dane zastąpimy wartością „inny”. Pozostałe wartości należy odpowiednio przekształcić. Warto sprawdzić ile razy występują dane wartości, co przedstawi tabela poniżej.

Tabela 3 Zestawienie występowania wartości w zmiennej screen_type

Wartość	Liczba wystąpień
matowa LED, IPS	195
matowa LED	77
błyszczaca LED	63
błyszczaca LED, IPS	54
matowa LED, TN	28
błyszczaca LED, TN	3
matowa LED, SVA	1
matowa LED, VA	1
błyszczaca OLED	1
błyszczaca AMOLED	1

Źródło: Opracowanie własne

W tabeli 3 można zauważyć, że najczęściej występuje wartość „matowa LED” oraz „błyszczaca LED”. W sumie tych wartości jest 420, pozostałe to „błyszczaca OLED” oraz „błyszczaca AMOLED”. Oczywiście są też brakujące wartości, co zostało wcześniej wspomniane. Wartości te można odpowiednio przekształcić, co będzie zaproponowane w tabeli poniżej.

Tabela 4 Przekształcenie wartości w zmiennej screen_type

Aktualna wartość	Nowa wartość
matowa LED, IPS	matowa LED
matowa LED	matowa LED
błyszczaca LED	błyszczaca LED
błyszczaca LED, IPS	błyszczaca LED
matowa LED, TN	matowa LED
błyszczaca LED, TN	błyszczaca LED
matowa LED, SVA	matowa LED
matowa LED, VA	matowa LED
błyszczaca OLED	inny
błyszczaca AMOLED	inny

Źródło: Opracowanie własne

Tabela 4 prezentuje nowe przekształcenia wartości. Można wyróżnić dwa główne typy „matowa LED” oraz „błyszczaca LED”. Wartości „błyszczaca OLED” oraz „błyszczaca AMOLED” zostały przypisane do wartości „inny”, ponieważ mają jedynie po jednej wartości. Dzięki nowym przekształceniom w kolumnie screen_type można wyróżnić trzy nowe wartości, które prezentuje tabela 5.

Tabela 5 Zestawienie wartości w zmiennej screen_type po przekształceniach

Wartość	Liczba wystąpień
matowa LED	302
błyszczaca LED	120
inny	37

Źródło: Opracowanie własne

Kolumna screen_type może zostać teraz wykorzystana w dalszej analizie. Aby zmienna ta była poprawnie wykorzystana, musi zostać zamieniona w zmienną zero-jedynkową. Dzięki takiej transformacji w danych pojawią się dwie nowe kolumny matowa LED oraz błyszczaca LED. Kolumna screen_type natomiast zniknie. Przykładowe dane dla odpowiednich wartości są przedstawione w tabeli poniżej.

Tabela 6 Przykładowe wartości w przekształconej zmiennej screen_type

screen type	matowa LED	błyszczaca LED
matowa LED	1	0
błyszczaca LED	0	1
inny	0	0

Źródło: Opracowanie własne

4.4. Przekształcenie zmiennej numerycznej.

Przykładem zmiennej numerycznej jest zmienna screen. Zawiera ona wartość rozmiaru ekranu. Na początek warto przyrzeć się dokładniej wartościom w tej kolumnie. Przedstawione są one na rysunku poniżej.

Rysunek 21 Zestawienie wartości w zmiennej screen

```
array(['15.6', '14', '17.3', '13.3', '10.1', '11.6', '12.1', '10', '16',
      '12.3', '15', '15.4', '13.5', '13', '13.9', '13.4', '12.5', '10.5',
      '17', '14.1'], dtype=object)
```

Źródło: Opracowanie własne

Jak widać na rysunku powyżej w danych mamy do czynienia z różnymi rozmiarami ekranu. Rozmiar ekran zależy od producenta i jest to bardzo elastyczna wartość. Można oczywiście sprawdzić, które wartości są najpopularniejsze, co przedstawia tabela poniżej.

Tabela 7 Zestawienie występowania wartości w zmiennej screen

Rozmiar ekranu	Liczba wystąpień
15,6	213
14	96
13,3	58
17,3	31
12,3	12
13,9	10
13,5	8
13	7
15,4	5
16	4
11,6	3
13,4	2
12,1	2
10,1	2
10	2
17	2
15	2
14,1	2
10,5	1
12,5	1

Źródło: Opracowanie własne

Tabela powyżej pokazuje, że można wyróżnić kilka najpopularniejszych rozmiarów ekranu, tak jak: 15,6, 14, 13,3 oraz 17,3. Tworząc ze zmiennej screen zmienną kategorię będzie bardzo ciężko przypisać wartości, które występują rzadziej. Oczywiście rozwiązaniem byłoby stworzenie przedziałów i przypisanie wartości do odpowiedniego przedziału, jednak wydaje się, że dobrym rozwiązaniem będzie pozostawienie tej zmiennej jako zmienna numeryczna.

4.5. Prezentacja gotowego zbioru danych

Przekształcając wszystkie dostępne zmienne w sposób przedstawiony powyżej można uzyskać aż 71 zmiennych, co przedstawia tabela poniżej.

Tabela 8 Zestawienie zmiennych użytych w modelu

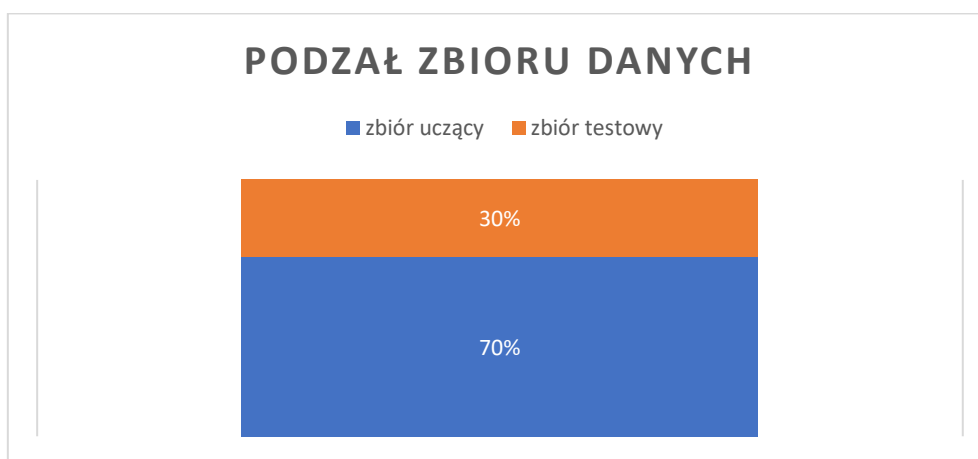
Index	Column	Typ zmiennej	Index	Column	Typ zmiennej
1	price_current	numeryczna	37	Metalowa obudowa	zero-jedynkowa
2	screen	numeryczna	38	DTS X Ultra	zero-jedynkowa
3	cores	numeryczna	39	NFC	zero-jedynkowa
4	cache	numeryczna	40	ScreenPad Plus	zero-jedynkowa
5	ram	numeryczna	41	Acer	zero-jedynkowa
6	software	zero-jedynkowa	42	Apple	zero-jedynkowa
7	camera	zero-jedynkowa	43	Dell	zero-jedynkowa
8	speakers	zero-jedynkowa	44	HP	zero-jedynkowa
9	speakers_number	numeryczna	45	Huawei	zero-jedynkowa
10	microphone	zero-jedynkowa	46	Inny	zero-jedynkowa
11	battery	numeryczna	47	LG	zero-jedynkowa
12	weight	numeryczna	48	Lenovo	zero-jedynkowa
13	warranty	numeryczna	49	MSI	zero-jedynkowa
14	battery_warrant	zero-jedynkowa	50	Microsoft	zero-jedynkowa
15	CPU_clock_rate_min	numeryczna	51	blyszczaca	zero-jedynkowa
16	CPU_clock_rate_max	numeryczna	52	matowa	zero-jedynkowa
17	dimension_total	numeryczna	53	3 GB	zero-jedynkowa
18	Podswietlana klawiatura	zero-jedynkowa	54	4 GB	zero-jedynkowa
19	TouchPad	zero-jedynkowa	55	6 GB	zero-jedynkowa
20	Klawiatura numeryczna	zero-jedynkowa	56	8 GB	zero-jedynkowa
21	DTS Audio	zero-jedynkowa	57	wspoldzielona z pamiecia	zero-jedynkowa
22	Aluminiowa obudowa	zero-jedynkowa	58	AMD Ryzen 5	zero-jedynkowa
23	Czytnik linii papilarnych	zero-jedynkowa	59	AMD Ryzen 7	zero-jedynkowa
24	Dolby Vision	zero-jedynkowa	60	Intel Core i3	zero-jedynkowa
25	glosniki z Dolby Atmos	zero-jedynkowa	61	Intel Core i5	zero-jedynkowa
26	HDR	zero-jedynkowa	62	Intel Core i7	zero-jedynkowa
27	Touch ID	zero-jedynkowa	63	Intel Core i9	zero-jedynkowa
28	Ekran dotykowy	zero-jedynkowa	64	inny	zero-jedynkowa
29	ScreenPad	zero-jedynkowa	65	DDR4	zero-jedynkowa
30	glosniki Bang & Olufsen	zero-jedynkowa	66	Intel	zero-jedynkowa
31	glosniki Harman Kardon	zero-jedynkowa	67	Other	zero-jedynkowa
32	Touch Bar	zero-jedynkowa	68	nVidia	zero-jedynkowa
33	Zawias obrotowy 360 stopn	zero-jedynkowa	69	512GB>=	zero-jedynkowa
34	HP Sure View	zero-jedynkowa	70	=<128	zero-jedynkowa
35	Podswietlana klawiatura R	zero-jedynkowa	71	Empty	zero-jedynkowa
36	TrackPoint	zero-jedynkowa			

Z tabeli powyżej można zauważyć, że dominują zmienne zero-jedynkowe. Przy dużej ilości istotnych informacji w zmiennej kategorycznej tworzy się kilka zmiennych zero-jedynkowych. Skutkuje to powstaniem wielu zmiennych, które powinny zostać wykorzystane w modelu. Na szczęście sieć neuronowa doskonale sobie z tym radzi. W kolejnym etapie zostanie przedstawiona budowa sieci neuronowej.

4.6. Budowa sieci neuronowej

Pierwszym krokiem w budowie sieci neuronowej będzie podzielenie zbioru danych na zbiór uczący oraz zbiór testowy. Jest to znana praktyka. Zbiór uczący zostanie wykorzystany w celach nauki sieci neuronowej, a zbiór testowy zostanie wykorzystany w sprawdzeniu wyników. Podział ten w badaniu przedstawia rysunek poniżej.

Rysunek 22 Podział zbioru danych



Źródło: Opracowanie własne

Kolejnym etapem będzie transformacja zmiennych. Jak wiadomo, niektóre wartości zawierają bardzo mały przedział jak na przykład pamięć ram (2 GB, 4 GB, 6 GB itp.). Niektóre jednak zawierają bardzo duże wartości, na przykład całkowity rozmiar laptopa. W celu upodobnienia danych zostanie zastosowanie estymator min max. Estymator ten skaluje oraz przekształca każdą wartość, tak aby znajdowała się w danym zakresie.

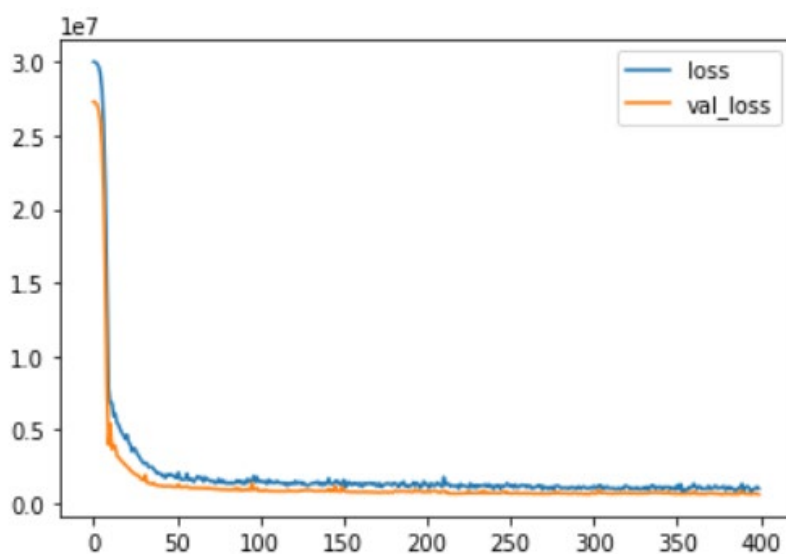
Przekształcone dane można wykorzystać w procesie budowania sieci neuronowej. Sieć neuronowa będzie się składać z neuronów aktywacji, neuronów ukrytych i neuronów wyjścia.

Model będzie zawierał 70 neuronów aktywacji, ponieważ w danych jest dostępne 70 zmiennych. Neuron wyjścia będzie 1, ponieważ model będzie szacował jedną wartość. Kolejną kwestią jest ilość neuronów ukrytych. To one w głównej mierze odpowiadają za dopasowanie modelu. Ustalenie liczby rzędów oraz liczby neuronów w każdym rzędzie zależy w dużej mierze od zbioru danych oraz osoby projektującej sieć. W tym przypadku będą to trzy rzędy ukrytych sieci neuronowej po 140 neuronów w każdym rzędzie. Dzięki takiemu rozwiązaniu wartości przy neuronach będą dobrze dopasowane.

Kolejną kwestią, którą należy rozważyć jest zbytne dopasowanie modelu do istniejących danych. W etapie uczenia jest to zbytne dopasowanie do zbioru uczącego. Aby tego uniknąć można zastosować pominięcie wybranych neuronów w etapie uczenia. W modelu będzie zastosowana funkcja, która co każdy krok losowo w każdym rzędzie neuronów ukrytych wyłączy 20% neuronów. Dzięki temu zastosowaniu neurony nie będą zbyt dopasowane do zbioru uczącego, a wyniki na zbiorze testowym powinny okazać się lepsze.

Ostatnią kwestią, którą należy są dwa parametry używane w procesie tworzenia modelu. Są to „batch_size” oraz „epochs”. Pierwszy parametr definiuje przedział jaki bierze algorytm w celach nauki. Jeśli parametr ten zostanie ustawiony na 64 to algorytm w pierwszym kroku nauki weźmie 64 rekordy a w kolejnym kolejne 64. Drugi parametr oznacza liczbę pełnych przejść przez zbiór uczący. W badaniu parametr ten został ustawiony na 400, a parametr „batch_size”, jak w przykładzie, na 64. Wykres poniżej przedstawia wyniki reszt dla algorytmu.

Rysunek 23 Wynik reszt modelu



Źródło: Opracowanie własne

Na rysunku powyżej można zauważyć, że największy skok w nauce jest w przedziale od 0 do 50 dla parametru „epochs”. W dalszym etapie uczenia wartość reszty spada niewiele. Nie ma więc sensu ustawiać większej wartości wspomnianego wcześniej parametru. Model został stworzony. W kolejnej sekcji zostaną przedstawione wyniki dla zbudowanego algorytmu.

4.7. Wyniki algorytmu

Na początku poprzedniej sekcji zbiór danych został podzielony na uczący i testowy. Nadszedł czas aby sprawdzić wyniki. W tym celu zostanie wykorzystany zbiór testowy. Prognozy zostaną porównane z prawdziwymi wynikami. Współczynniki, które zostaną użyte w ocenie prognoz to średni błąd bezwzględny oraz średni błąd kwadratowy. Model zostanie także oceniony przez współczynnik R^2 , który opisuje w jakim stopniu model wyjaśnia zmienna objaśnianą. Wyniki znajdują się w tabeli poniżej.

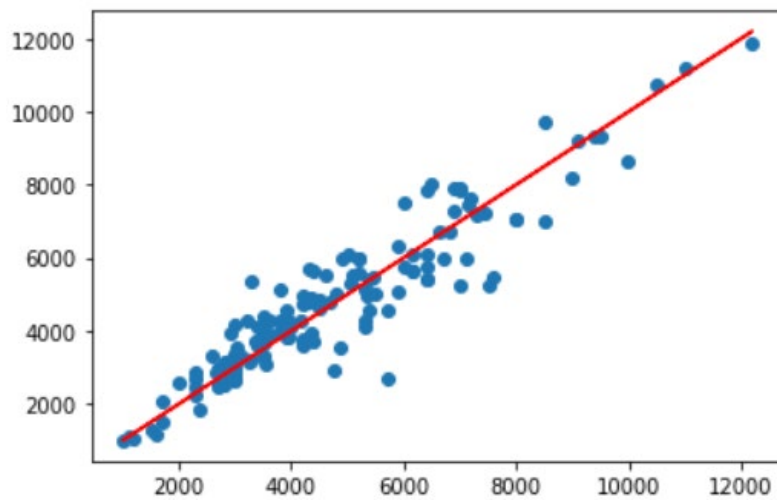
Tabela 9 Wartości wybranych współczynników modelu

Współczynnik	Wartość
MAE	571,49
RMSE	775,93
R^2	86,88%

Źródło: Opracowanie własne

Współczynnik R^2 przyjął wartość 86,88%. Jest to dość dobry wynik. Oznacza, że zmienna objaśniana została poprawnie oszacowana za pomocą zmiennych objaśniających w 86,88%. Wartość MAE wyniosła 571,49. Oznacza ona średnią różnicę między wartością prognozowaną przez model a wartością rzeczywistą. Współczynnik RMSE przyjął natomiast wartość 775,93. Warto przypomnieć, że średnia cena laptopa w zbiorze danych wynosi 4839, 86 zł. Aby lepiej zobrazować sobie wyniki wizualnie, warto spojrzeć na wykres poniżej, który prezentuje prognozy na zbiorze testowym.

Rysunek 24 Prognoza na zbiorze testowym



Źródło: Opracowanie własne

Linia czerwona na powyższym wykresie przedstawia dokładną cenę danego laptopa. Niebieskie kropki oznaczają zaś prognozy zbioru testowego. Im odległość jest większa od czerwonej linii, tym błąd w szacunku jest większy. Oczywiście można zauważyć pojedyncze przypadki, gdzie niebieskie kropki są dość daleko od czerwonej linii. Sporo jednak punktów znajduje się w sąsiedztwie czerwonej linii.

Zakończenie

Podstawowym celem pracy było pozyskanie danych ze strony internetowej oraz wycena produktów. Dany zostały pobrane z wykorzystaniem metody Web Scrapingu, które polegają na uzyskaniu danych bezpośrednio z kodu źródłowego strony. Odnosiły się one do komputerów przenośnych na stronie RTV euro AGD. Program, który został napisany umożliwia zbieranie danych okresowo, dopóki strona nie zostanie zmieniona, co daje możliwość aktualizowania nowego zbioru danych. Surowe dane zostały odpowiednio przekształcone tak, aby mogły być wykorzystane w budowie sieci neuronowej. Jest wiele czynników, które powinny zostać wykorzystane w wycenie sprzętu komputerów przenośnych. W konsekwencji oznaczało to wykorzystanie wielu zmiennych w modelu.

Dokładność modelu została oceniona za pomocą kilku wybranych współczynników. Jednym z nich był współczynnik R^2 , który przedstawia jak dobrze model został dopasowany. Uzyskał on wartość 86,88%. Jest to dość dobry wynik, jednak nie jest najlepszy. Jeśli chodzi o wycenę produktów, oczywistym faktem jest, że każdy chce dopasować cenę jak najlepiej. Oczywiście jest kilka sposobów, aby zwiększyć wartość wspomnianego wcześniej współczynnika, czyli dokładność modelu. Jednym z nich jest korzystanie z większego zbioru danych, co pozwoli lepiej dopasować model. Innym sposobem jest lepsze przekształcenie zmiennych, co oznaczałoby wykorzystanie zmiennych kategoriowych zamiast niektórych zmiennych numerycznych.

Wykorzystany schemat analizy mógłby zostać wykorzystany na innych sklepach internetowych. Skompletowanie danych ze wszystkich popularnych polskich sklepów internetowych dałoby pełny przegląd produktów. Analiza tak dużego zbioru danych umożliwiłaby dokładniejszą wycenę produktów na tle konkurencji. Oferowałaby również sprzedaż produktów z najlepszą ceną na rynku oraz wybranie tych produktów, które oferują najlepszą jakość do ceny.

Metody użyte w badaniu mogą zostać wykorzystane również na innych produktach. W ofertach sklepów internetowych jest wiele podobnych produktów, które mogą zostać poddane podobnej analizie. Wycena produktów na podstawie parametrów pomoże lepiej oszacować wartość sprzętu biorąc pod uwagę parametry produktu. Każdy właściciel sklepu chce oferować swój sprzęt w najlepszej cenie na rynku, więc może skupić się na tych produktach, które

przyniosą mu najlepszy zysk. Z drugiej strony każdy konsument chce kupić sprzęt w jak najlepszej cenie.

Przeprowadzone badanie pokazało, że obie użyte metody są bardzo pomocne w dzisiejszych czasach. Pozyskiwanie danych ze stron przy pomocy Web Scrapingu daje możliwość dostępu do szerokiego zbioru danych. Sieć neuronowa pozwala wykorzystać zebrane dane, w celu wyceny danego produktu.

Bibliografia

- (2020, Czerwiec). Pobrano z lokalizacji <https://www.edureka.co/blog/web-scraping-with-python/#lib>
- (2020, Wrzesień). Pobrano z lokalizacji https://www.ey.com/pl_pl/news/2019/08/jak-polscy-konsumenci-kupuja-w-sieci-wyniki-badania-ey
- (2020, Wrzesień). Pobrano z lokalizacji <https://www.gemius.pl/wszystkie-artykuly-aktualnosci/e-commerce-w-polsce-2020.html>
- (2020, Wrzesień). Pobrano z lokalizacji <https://spidersweb.pl/bizblog/zakupy-w-internecie-badanie-ey-polska/>
- (2020, Czerwiec). Pobrano z lokalizacji <https://realpython.com/python-web-scraping-practical-introduction/>
- (2020, Listopad). Pobrano z lokalizacji <http://galaxy.uci.agh.edu.pl/~vlsi/AI/wstep/>
- (2020, Listopad). Pobrano z lokalizacji https://pl.wikipedia.org/wiki/Funkcja_aktywacji
- (2020, Listopad). Pobrano z lokalizacji <https://history-computer.com/Internet/Conquering/Wanderer.html>
- Bahia, I. S. (2013). *A Data Mining Model by Using ANN for Predicting Real*. Ministry of Higher Education and Scientific Research, Baghdad, Iraq.
- Chen, S., Chou, E. i Yang, R. (2017). *The Price is Right: Predicting Prices with Product Images*. Stanford University.
- HEBB, D. O. (1949). *The Organization of Behavior*.
- HUBEL, D. H. i WIESEL, T. N. (1959). *RECEPTIVE FIELDS OF SINGLE NEURONES IN THE CAT'S STRIATE CORTEX*. Wilmer Institute, The Johns Hopkins Hospital and University, Baltimore.
- Ivakhnenko, A. G. (1971). *Polynomial Theory of Complex Systems*. IEEE Transactions on Systems, Man and Cybernetics.
- Ivakhnenko, A. G. i Lapa, V. G. (1965). *Cybernetic Predicting Devices*. CCM Information Corp.
- Luo, C., Wei, Q., Zhou, L., Zhang, J. i Sun, S. (2011). *Prediction of Vegetable Price Based on Neural Network and Genetic Algorithm*. Beijing Academy of Agriculture and Forestry Sciences.
- McCulloch i Pitts. (1943). *A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY*. University of Illinois.
- Mitchell, R. (2018). *Web Scraping with Python*. O'REILLY.
- Núñez Tabales, Julia M., Caridad y Ocerin, José María, Rey Carmona i Francisco J. (2013). *Artificial neural networks for predicting real estate*. Universidad Pablo de Olavide, Sevilla.
- Rosenblatt, F. (1958). *THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN*. Cornell Aeronautical Laboratory.
- Rumelhart, D. i McClelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge: MIT.

Spis ilustracji

Rysunek 1 Charakterystyka kupujących online oraz offline	6
Rysunek 2 Charakterystyka zakupów online oraz offline	7
Rysunek 3 Charakterystyka kupujących online	8
Rysunek 4 Charakterystyka zakupów online	9
Rysunek 5 Ruch na wybranych stronach internetowych	10
Rysunek 6 Liczba stron w kategorii Laptopy na stronie euro.com.pl (stan na 29.10.2020).....	13
Rysunek 7 Strona internetowa laptopa Acer sklepu RTV euro AGD	14
Rysunek 8 Dane techniczne laptopa Acer sklepu RTV euro AGD	14
Rysunek 9 Zapytanie do strony euro.com.pl.....	16
Rysunek 10 Przykład dokumentu HTML ze strony w3schools.com	18
Rysunek 11 Strona internetowa laptopa Acer sklepu RTV euro AGD	18
Rysunek 12 Fragment kodu źródłowego strony laptopa Acer sklepu RTV euro AGD	19
Rysunek 13 Fragment kodu źródłowego laptopa Acer sklepu RTV euro AGD	20
Rysunek 14 Model neuronu	23
Rysunek 15 Budowa sieci neuronowej.....	23
Rysunek 16 Sieć rekurencyjna	24
Rysunek 17 Sieć komórkowa	25
Rysunek 18 Zestawienie marek laptopów w ofercie sklepu RTV euro AGD	28
Rysunek 19 Średnie ceny marek laptopów sklepu RTV euro AGD	29
Rysunek 20 Zestawienie wartości w zmiennej screen_type	32
Rysunek 21 Zestawienie wartości w zmiennej screen	34
Rysunek 22 Podział zbioru danych	37
Rysunek 23 Wynik reszt modelu	38
Rysunek 24 Prognoza na zbiorze testowym	40

Spis tabel

Tabela 1 Zmienne wykorzystane w badaniu	30
Tabela 2 Zestawienie brakujących rekordów	31
Tabela 3 Zestawienie występowania wartości w zmiennej screen_type.....	33
Tabela 4 Przekształcenie wartości w zmiennej screen_type.....	33
Tabela 5 Zestawienie wartości w zmiennej screen_type po przekształceniach	34
Tabela 6 Przykładowe wartości w przekształconej zmiennej screen_type	34
Tabela 7 Zestawienie występowania wartości w zmiennej screen	35
Tabela 8 Zestawienie zmiennych użytych w modelu	36
Tabela 9 Wartości wybranych współczynników modelu.....	39