

## Wybrane zagadnienia analizy szeregów czasowych

Krzysztof Adamski\*

Warszawska Wyższa Szkoła Informatyki, Polska

---

### Streszczenie

Artykuł obejmuje przegląd wybranych zagadnień związanych z analizą i predykcją szeregów czasowych zawierających dane z notowaniami giełdowymi. Przedstawiona została taksonomia szeregów czasowych oraz scharakteryzowane główne kierunki spotykane w analizie danych giełdowych. Szerzej opisane zostały wybrane narzędzia analizy technicznej. W kolejnych rozdziałach dokonany został przegląd wybranych metod uczenia maszynowego, zarówno metod algorytmicznych jak i metod wykorzystujących sieci neuronowe, przydatnych w analizie i predykcji szeregów czasowych.

**Słowa kluczowe** – Szeregi czasowe, Analiza techniczna, Predykcja, Uczenie maszynowe, Sieci neuronowe

---

\* E-mail: k.s.adamski21@gmail.com

## 1. Wstęp

Nauka o danych to dziedzina, która w ostatnich latach rozwija się w szybkim tempie. Dostępność coraz większej liczby danych z różnych obszarów ludzkiej działalności stworzyła warunki do opracowywania coraz bardziej zaawansowanych technik analitycznych. Rozwojowi temu towarzyszą rosnące oczekiwania i wymagania stawiane przez różne środowiska, zarówno naukowo-badawcze jak i typowo utylitarne z biznesem, finansami, medycyną, przemysłem i obronnością na czele. Normą stało się ciągle monitorowanie wzrostu gospodarczego i jego efektywności oraz podejmowanie na tej podstawie prób optymalizacji działań w najróżniejszych obszarach i skalach – do skali globalnej włącznie. W wymienionych przypadkach, wszystkie operacje na danych odnoszą się w ogromnej większości do struktur będących szeregami czasowymi.

Dane jakie każdego dnia powstają na globalnych rynkach, charakteryzuje duży udział składnika stochastycznego. Tworzy się on jako spontaniczna reakcja zbiorowości na docierające informacje, zarówno te w postaci budującego się wykresu ceny, jak i te odnoszące się do gospodarki, polityki i różnych wydarzeń mogących mieć wpływ na indywidualne decyzje poszczególnych inwestorów. W obecnych czasach, szybkiego transferu informacji, nie do pominięcia są również, mające duży wpływ, elementy celowej dezinformacji z jakimi można się spotkać w sieci Internet. Mamy zatem do czynienia z globalnym procesem społecznym (socjologicznym), który trudno jest ująć w sztywne ramy jakiegoś szablonu czy nawet bardzo elastycznego modelu.

Przegląd zagadnień ujętych w temacie artykułu wymaga odniesienia się do tego co kryje się za sformułowaniem „wyniki predykcji”. Słowo „predykcja” jest wyjaśniane jako „przewidywanie”. Wzmianka o tym terminie pojawia się w kontekście wielu zachowań ludzkich jakie towarzyszą podejmowaniu decyzji w ograniczonym polu informacyjnym. Za osiągnięcia związane z tą tematyką zostały przyznane nagrody Nobla z ekonomii w roku 1996 za „ekonomiczną teorię motywacji w warunkach asymetrii informacji” oraz w roku 2000 za „opracowanie teorii i metod stosowanych w empirycznych analizach zachowań indywidualnych” jak również w roku 1997 za „opracowanie metody wyceny instrumentów pochodnych, zwłaszcza opcji” pozwalającej na przewidywanie kursów tych instrumentów z użyciem zaproponowanego modelu [1].

Według *Słownika Języka Polskiego*, predykcja jest terminem zarezerwowanym dla: „określenia w analizie statystycznej przewidywania przyszłych realizacji albo cech statystycznych zjawisk losowych” [2]. Wynikiem predykcji może być prognozowana wartość liczbowa lub dana nominalna (kategoryczna). Predykcją ludzkość zajmowała się już w czasach starożytnych. Wynikiem obliczeń predykcyjnych było np. przewidywanie terminów zaćmień słońca już w starożytnym Egipcie, choć nie dotyczyło to zjawiska losowego w znaczeniu przyjmowanym we współczesnych modelach opisujących różne procesy. Współcześnie formułowane modele wykorzystywane w predykcji zawierają człon resztkowy związany z losowym charakterem zjawiska. Predykcja pomimo jasnej definicji słownikowej świadczącej o jej statystycznym rodowodzie postrzegana może być również jako wynik stosowania podejścia tylko pośrednio związanego ze statystyką. Dotyczy to poszukiwania w szeregach czasowych wystąpień wzorców zapowiadających najbardziej prawdopodobny dalszy przebieg opisywanego (badanego) procesu. Metody te pomimo swojego niższego charakteru znajdują zastosowanie w odniesieniu do niektórych grup danych a w szczególności tych pochodzących z rynków i giełd.

Predykcja w odniesieniu do tak powstającego szeregu czasowego jest zatem dużym wyzwaniem, właśnie ze względu na losowość procesu z jakim mamy do czynienia. Informacje podawane przez duże światowe instytucje finansowe pokazują, że większość transakcji na rynkach finansowych zawierana jest z wykorzystaniem zawansowanego oprogramowania wykorzystującego szeroko rozumianą, sztuczną inteligencję (AI). Daje to kolejny powód do tego, by bliżej przyjrzeć się wybranym aspektom tego, nie do końca jawnego, obszaru działania świata finansów.

Celem niniejszego artykułu jest przegląd wybranych zagadnień związanych z szeregami czasowymi, ich opracowaniem i badaniem oraz predykcją odnoszącą się w szczególności do danych ekonomicznych zawierających notowania. Przegląd ten podzielony został na trzy dopełniające się części:

- Część pierwsza obejmuje typowe zagadnienia związane z szeregami czasowymi (nauka o danych). Ponieważ dane rynkowe charakteryzując się najwyższą jakością i kompletnością, są reprezentatywne dla szeregów czasowych, dlatego one zostaną omówione w pierwszej kolejności. Ta część rozpocznie się od rzutu okiem na ich historię. W dalszej kolejności zostaną omówione metody ich analizy, w tym analiza fundamentalna, analiza tech-

niczna, a także analiza ilościowa. Zostanie także poruszony temat psychologii inwestowania, która ma ogromny wpływ na tworzenie kształtowanie nastrojów wpływających w konsekwencji na formowanie danych.

- Część drugą poświęcono tematyce obejmującej wybrane algorytmiczne metody uczenia maszynowego. Na wstępie omówiono zagadnienia związane z przygotowaniem danych do analizy. W dalszej kolejności przedstawiono znane klasyczne algorytmy uczenia maszynowego.
- W ramach części trzeciej omówione zostały wybrane zagadnienia związane z sieciami neuronowymi stosowanymi do przetwarzania danych obejmującego predykcję.

## 2. Szeregi czasowe

Patrząc na historię rozwoju statystyki, temat szeregów czasowych stał się przedmiotem badań naukowych stosunkowo późno. Za początki analizy szeregów czasowych przyjmuje się lata dwudzieste XX w. Stało się to za sprawą modelu autoregresyjnego zastosowanego do analizy danych rzeczywistych dotyczących zmienności plam słonecznych [3]. Jednym z pionierów statystyki, któremu przypisuje się współautorstwo metody analizy szeregów czasowych, był George Edward Pelham Box. Opisał on i wprowadził w latach 70. XX w. powszechnie obowiązujący model szeregu czasowego [3,4].

Definicje szeregów czasowych można znaleźć zarówno w literaturze, jak i na stronach internetowych instytucji zajmujących się ich opracowywaniem i analizą [5]. Mogą być one formułowane w różny sposób. Podstawowa różnica dotyczy obszarów i zakresów zagadnień, których szeregi czasowe dotyczą. Definicje formułowane przez teoretyków nieco różnią się od tych podawanych przez praktyków. Przykładowo definicja dostępna na stronie Głównego Urzędu Statystycznego, który zajmuje się gromadzeniem danych w postaci szeregów czasowych oraz ich przetwarzaniem i analizą, sformułowana jest w sposób dostosowany do obszaru działania tej instytucji [6].

Definicja o charakterze ogólnym określa szereg czasowy jako zbiór danych (punktów) uporządkowanych chronologicznie, gdzie czas jest zmienną niezależną lub jako realizację procesu stochastycznego, którego dziedziną jest czas [5,7]. Należy przy tym dodać, że:

- Dane mogą być uporządkowane według stałego interwału czasowego, jednak często spotykane są dane niezachowujące tej reguły;
- Dane są zazwyczaj wartościami liczbowymi i/lub kategorycznymi. Szeregi czasowe mogą być tworzone również przez inne typy bądź struktury danych;
- Szeregi czasowe wyróżnia sposób zbierania danych i związany z tym błąd pomiarowy lub niepewność danych;
- Szereg czasowy charakteryzuje również cel, w jakim dane są gromadzone, oraz ich przeznaczenie.

Ze względu na swój ogólny charakter przytoczona definicja dotyczy wszystkich szeregów czasowych spotykanych w praktyce analitycznej. Przeprowadzenie ich analizy prowadzi do osiągnięcia dwóch pożądaných celów:

- Opisu badanego zjawiska oraz poznanie wszystkich charakteryzujących go parametrów. Dzięki temu możliwe będzie zrozumienie sekwencji zdarzeń, a także odtworzenie przebiegu zjawiska podczas symulacji;
- Prognozowania przyszłych wartości przy użyciu opracowanych modeli.

Szeregi czasowe spotkać można w praktycznie każdej dziedzinie ludzkiej działalności. Dotyczy to nauk ścisłych: astronomii, astrofizyki, fizyki, meteorologii, elektroniki, elektrotechniki, telekomunikacji, kryptografii, akustyki, chemii, mechaniki, techniki jak również biologii, medycyny, ekonomii, a także całego obszaru nauk nazywanych umownie humanistycznymi, wśród których dla przykładu można wymienić: historię, prawo, socjologię i nauki pokrewne. Wymienione obszary zastosowania analizy szeregów czasowych dają obraz tego, z jaką mnogością i różnorodnością danych mamy do czynienia. Każdy z wymienionych obszarów wymaga specyficznego podejścia do danych a otrzymywane wyniki pokazują jak różnorodne problemy mogą być rozwiązywane.

W przypadku ekonomii opracowany został cały zestaw metod pozwalających na: opis i analizę zjawisk gospodarczych, obliczanie wskaźników statystycznych charakteryzujących dane makroekonomiczne oraz predykcję analizowanych szeregów czasowych. Ze względu na duży udział czynnika losowego w danych, predykcja szeregów czasowych jest aktualnie jednym z najbardziej eksplorowanych badawczo kierunków analizy szeregów czasowych. Osiągane wyniki (nie zawsze szeroko publikowane, a czasem stanowiące tajemnicę firm) wykorzystywane są do osiągnięcia przewagi na rynkach lub przewidywania zagrożeń dla zbiorowości ludzkiej. Z tego

powodu zjawiska ekonomiczne opisywane szeregami czasowymi, stały się źródłem postępu w badaniach i narzędziem do testowania powstających ciągle nowych podejść, koncepcji i teorii matematycznych związanych z predykcją.

Z racji znaczenia jakie reprezentuje predykcja szeregów czasowych w ekonomii, w dalszej części zostaną przedstawione zagadnienia jakie związane są z tym obszarem badań. Na uwagę zasługuje fakt, że próby badania i składania nowych propozycji w podejściu do tego tematu, podejmują nie tylko naukowcy, ale zaangażowana jest również ogromna społeczność traderów i inwestorów działających na różnych rynkach kapitałowych. Zauważyć należy również, że wybitne osiągnięcia w zakresie opracowania matematycznych metod i modeli związanych z rynkami doczekały się nagrody Nobla, np. za opracowanie przez Fischera Blacka, Myrona Scholesa i Roberta Mertona modelu wyceny opcji Blacka-Scholesa (Nobel z ekonomii 1997) [8].

Pokazane wyżej zastosowania prowadzą do określenia analizy szeregów czasowych jako próby podsumowania i opisu zjawiska poprzez uzyskanie parametrów statystycznych ze zbioru punktów ułożonych w porządku chronologicznym. Szereg czasowy poddawany analizie rozpatruje się jako strukturę, której dekompozycja może pokazać różne składniki oraz ich wzajemne relacje. Poglądowo można to zilustrować za pomocą diagramu zamieszczonego na rys. 1.

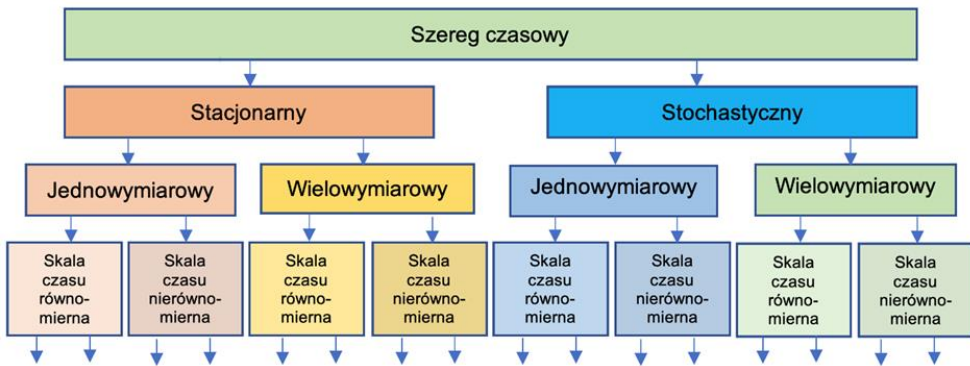


**Rysunek 1.** Model struktury szeregu czasowego wykorzystywany w dekompozycji danych [5]

Przedstawiona wcześniej różnorodność szeregów czasowych jest powodem do tworzenia ich taksonomii. Budowanie taksonomii stanowi ułatwienie w klasyfikacji danych jakie będą przedmiotem badań oraz jest pomocne w doborze metod analizy.

Diagram z rys. 2 jest przykładem ilustrującym jak rosnąca liczba kryteriów klasyfikacyjnych prowadzi do szybkiego rozrostu grafu reprezentującego taką taksonomię. Diagram ten stanowi tylko jeden z możliwych wariantów klasyfikacji jaki można utworzyć, przyjmując listę kryteriów i określając kolejność ich zastosowania. Pokazany na rys. 2 wariant taksonomii powstał z wykorzystaniem listy następujących kryteriów:

- Szereg jest stacjonarny czy stochastyczny?
- Szereg zawiera dane jednowymiarowe czy wielowymiarowe?
- Czy skala czasu jest równomierna czy nierównomierna?



Rysunek 2. Wariant taksonomii szeregów czasowych

Przyjęcie kolejnych kryteriów, symbolicznie pokazane strzałkami prowadzącymi od najniższej warstwy w dół, istotnie rozbudowuje strukturę, ukazując z jak złożonym zagadnieniem możemy mieć do czynienia. Kolejność stosowania przyjętych kryteriów również może mieć istotne znaczenie dla osiągnięcia łatwości stosowania konkretnej taksonomii.

Przykładowo w pokazanej wyżej liście kryteriów jako pierwsze występuje rozróżnienie pomiędzy szeregiem stacjonarnym i stochastycznym. W wielu przypadkach trudno jest a priori określić prawdziwy charakter danych podlegających klasyfikacji. Zaklasyfikowanie szeregu czasowego jako stochastycznego bez szczegółowej analizy, w wielu przypadkach może być nieuzasadnione. Za istotne z praktycznego punktu widzenia uważa się wyodrębnienie w danych cech określanych mianem systematycznych, takich jak: trend, cykliczność i sezonowość oraz części resztowej reprezentującej składniki przypadkowe (rys. 1) o charakterze losowym [6]:

- Trend jest tutaj rozumiany jako tendencja do jednokierunkowych zmian wartości zmiennej (wzrost lub spadek) obserwowanej w przyjętej skali czasowej.
- Cykliczność jest cechą wskazującą na istnienie stałej długookresowej wartości czasu, w którym dane liczbowe oscylują wokół pewnej wielkości stałej lub trendu.
- Sezonowość związana jest z wahaniami wartości szeregu wokół jego tendencji rozwojowej rozpatrywanej w horyzoncie czasowym nieprzekraczającym jednego roku [6]. Za losowość występującą w danych odpowiada wspomniana wyżej część resztowa i jej amplituda porównywana zwykle z wartościami pozostałych składowych.

Następnym zagadnieniem jest określenie związków jakimi mogą być połączone poszczególne składowe szeregu czasowego. Ogólna postać modelu szeregu czasowego [9] będąca podstawą do prognozowania jest opisana zależnością:

$$y_t^* = f(t, y_{t-1}, \dots, y_{t-p}, y_{t-1}^*, \dots, y_{t-p}^*, \xi_t). \quad (1)$$

W powyższym wzorze przyjęto następujące oznaczenia:

$t$  – czas,

$p$  – wartość opóźnienia,

$y_t^*, y_{t-1}^*, y_{t-p}^*$  – wartości prognoz szeregu czasowego w chwilach:  $t$ ,

$t - 1, t - p$ ,

$y_{t-1}, y_{t-p}$  – wartości szeregu czasowego w chwilach:  $t - 1, t - p$ ,

$\xi_t$  – składnik losowy.

W analizie na potrzeby predykcji brane są pod uwagę wartości danych z określonego przedziału czasu oraz wartości prognoz jakie były obliczone dla poszczególnych momentów czasu z tego przedziału. Dobór okna czasowego jest dodatkowym parametrem, który należy uwzględnić w trakcie oceny wyników wstępnych. Uwzględniany jest też składnik losowy zawarty w danych. Wyróżniane są trzy modele związków pomiędzy składowymi szeregu czasowego: addytywny, multiplikatywny oraz model mieszany. Model addytywny ma postać:

$$y_t = f(t) + g(t) + h(t) + \xi_t \quad (2)$$

Z kolei, model multiplikatywny można przedstawić jako:

$$y_t = f(t) g(t) h(t) \xi_t \quad (3)$$



Model mieszany może zostać wyrażony na kilka sposobów:

$$y_t = f(t) + g(t) + h(t) \xi_t \quad (4)$$

$$y_t = f(t) g(t) \xi_t + h(t) \quad (5)$$

$$y_t = f(t) h(t) + g(t) \xi_t \quad (6)$$

$$y_t = f(t) h(t) \xi_t + g(t) \quad (7)$$

We wzorach (2)-(7) przyjęto następujące oznaczenia:

$y_t$  – wartość danej z szeregu czasowego,

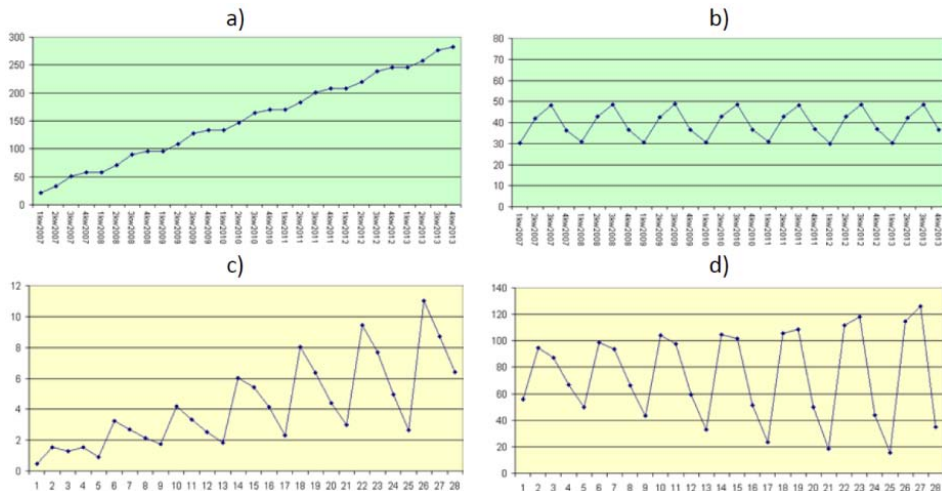
$f(t)$  – trend,

$g(t)$  – sezonowość,

$h(t)$  – cykliczność,

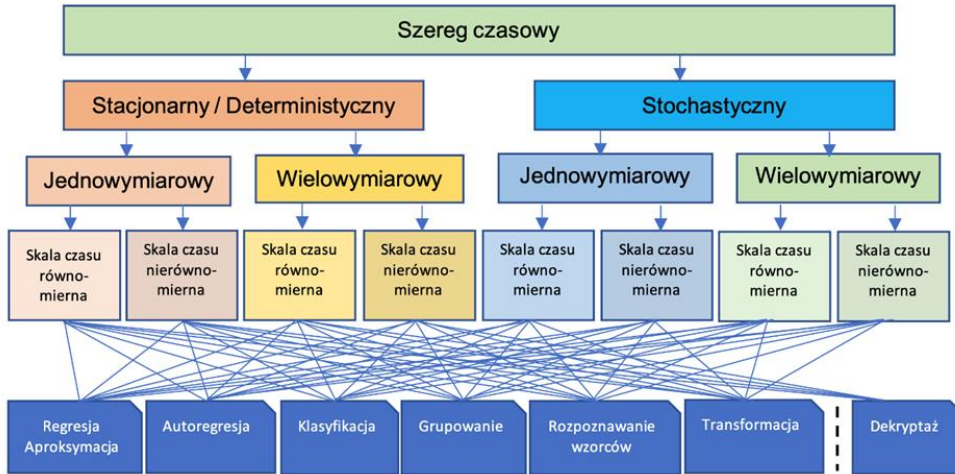
$\xi_t$  – losowość.

Na rysunku 3 pokazane zostały przykładowe wykresy dla modelu addytywnego oraz multiplikatywnego w wariantach z trendem niezerowym oraz bez trendu.



**Rysunek 3.** Wykresy szeregów czasowych z różnymi modelami zależności pomiędzy składowymi [9]: **a)** Trend i sezonowość kwartalna addytywna –  $y_t = f(t) + g(t) + h(t) + \xi_t$ ; **b)** Sezonowość kwartalna addytywna bez trendu –  $y_t = g(t) + h(t) + \xi_t$ ; **c)** Trend i sezonowość multiplikatywna –  $y_t = f(t) g(t) h(t) + \xi_t$ ; **d)** Brak trendu i sezonowość multiplikatywna –  $y_t = g(t) h(t) + \xi_t$

Wyodrębnione na rys. 3b i rys. 3d przebiegi sezonowości dla modeli addytywnego oraz multiplikatywnego pokazują jak rezultaty przetwarzania szeregu czasowego mogą pomóc w analizie i klasyfikacji surowych danych.

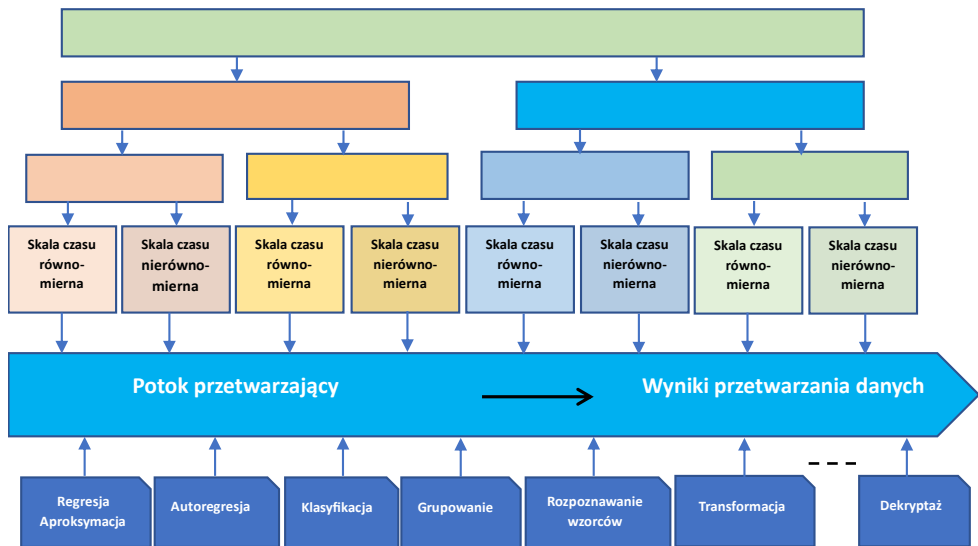


**Rysunek 4.** Ilustracja możliwości wyboru najpopularniejszych wariantów przetwarzania szeregów czasowych

Systematycznie poddawany jest również proces samej analizy szeregów czasowych. Należy wziąć pod uwagę cel jaki postawiono przed wykonaniem zadania oraz jakich technik można użyć do jego osiągnięcia. W przypadku ogólnym wykorzystując zbudowaną wcześniej taksonomię szeregów czasowych można pokazać możliwe warianty dokonywania analiz z wykorzystaniem dostępnych narzędzi i metod. Na rys. 4 schematycznie zilustrowany został stopień złożoności wyboru wariantów przetwarzania i analizy danych zawartych w szeregu czasowym. Pokazane na tym rysunku powiązania za pomocą wielu linii, obrazują możliwość jednoczesnego zastosowania w konkretnym przypadku wielu opcji zarówno przetwarzania jak i analizy.

Wielość dostępnych opcji oraz możliwość ich stosowania w różnej kolejności w potokowym modelu przetwarzania danych pokazana została na rys. 5. Łatwo zauważyć, że mamy tutaj do czynienia z wielowymiarowym problemem badawczym, którego usystematyzowanie wymaga przeprowadzenia wielu obliczeń symulacyjnych oraz dokonania analiz porównawczych. Spotykane w literaturze doniesienia dotyczące czasem bardzo zawężonych tematycznie zagadnień, są zwykle wynikiem podjętych wcześniej wielu prób i eksperymentów, mających na celu optymalizację

ścieżki opisującej sposób postępowania z danymi w dobrze zdefiniowanym i opisanym przypadku. Dotyczy to w szczególności danych otrzymywanych w naukach eksperymentalnych takich jak np. biologia czy medycyna, w których standaryzacja procedur przetwarzania danych ma kluczowe znaczenie ze względu na wiarygodność, porównywalność i wykorzystanie otrzymanych wyników.



**Rysunek 5.** Ilustracja możliwości wyboru najpopularniejszych wariantów przetwarzania szeregów czasowych w potokowym modelu przetwarzania danych

Szeregi czasowe spotykane w poszczególnych obszarach badań, ze względu na swój charakter i specyfikę, mogą być przetwarzane według indywidualnych ścieżek w potoku przetwarzającym dając pożądane rezultaty z punktu widzenia zastosowań jak również dalszych badań i analiz.

### 2.1. Rys historyczny dotyczący danych giełdowych

Historia rynków i pieniądza sięga bardzo zamierzchłych czasów. Jednak odległe epoki nie dostarczają obfitej ilości danych. Zachowały się gliniane tabliczki z Mezopotamii z drugiego tysiąclecia p.n.e., zapisane pismem klinowym, jak również chińskie papierowe banknoty z VII w. Istnieją zapisy matematyczne, jakie powstały

za czasów Hammurabiego 1792-1750 r. p.n.e. odnoszące się do pożyczek długoterminowych, sugerujące stosowanie procentu składanego. Na terenie Europy podwaliny matematyczne pod współczesne obliczenia finansowe położył Fibonacci z Pizy, publikując swoje dzieło *Liber Abaci (Księga rachunków)* w roku 1202 [1].

Później włoskie ośrodki handlowe przyczyniły się do rozprzestrzenienia bankowości oraz powstania różnych instrumentów finansowych np. pożyczek. Powstały również banki zajmujące się wymianą walut. W basenie Morza Śródziemnego centrami handlowymi były: Wenecja, Florencja, Pizya. Tam też skupione było przeprowadzanie transakcji, będące wzorem dla innych regionów i krajów. Na uwagę zasługuje przełom jaki nastąpił za sprawą odkrycia Ameryki przez Krzysztofa Kolumba. Powstały imperia kolonialne. W Europie swoją pozycję jako pieniądź ugruntowały sobie kruszce, masowo przywożone z kolonii. Moment powstania kompanii indyjskich w XVII wieku związany jest z wprowadzeniem akcji. Powstawały giełdy (Amsterdam, Londyn) pozwalające na obrót takimi papierami wartościowymi. Liczba i rodzaje instrumentów finansowych zaczęła rosnąć. Na arenę wkroczyły Stany Zjednoczone Ameryki. Liczba banków, giełd jak również kryzysów, znacząco wzrosła za sprawą tzw. baniek spekulacyjnych. W poszczególnych krajach cywilizują handel tymi nowymi instrumentami liczne regulacje prawne, jednakże nie są one w stanie zapobiec kolejnym kryzysom, cyklicznie pojawiającym się na globalnych rynkach finansowych [11,12].

Na osobną wzmiankę zasługuje historia notowań na giełdach ryżu w Japonii. Równolegle rozwijający się w izolacji rynek japoński, również wypracował instrumenty finansowe pozwalające na handel jak i na spekulację tymi walorami.

I tu dochodzimy do czasów współczesnych, w których handel zarówno akcjami przedsiębiorstw, obligacjami, kontraktami terminowymi, funduszami ETF, opcjami, surowcami, kontraktami na różnice kursowe (CFD), walutami, a ostatnio kryptowalutami, prowadzony jest w wielu przypadkach całodobowo w dni robocze. Wyjątkiem są rozproszone rynki krypto-walutowe, wykorzystujące technologię Blockchain, które pracują siedem dni w tygodniu przez cały rok. Handel odbywa się drogą elektroniczną z wykorzystaniem globalnej sieci teleinformatycznej oraz platform transakcyjnych, które są specjalistycznym oprogramowaniem narzędziowym do zawierania transakcji. Wyjątkowa łatwość uczestnictwa w globalnych rynkach i szybki dostęp do całego szeregu, trudnych dotąd do uzyskania, informacji, stały się przyczyną powstania przepływów finansowych na niespotykaną dotąd skalę. Przykładowo

dzienny obrót na globalnym rynku walutowym Forex wielokrotnie przekracza roczny PKB wielu państw, w tym również Polski.

Zaistniała sytuacja pozwala na gromadzenie i bezproblemowe udostępnianie danych giełdowych wszystkim zainteresowanym. Dane te należą do grupy o najwyższej jakości i kompletności, gdyż do ich gromadzenia i przetwarzania wykorzystywane są redundantne systemy informatyczne czasu rzeczywistego. Jako reprezentatywne, dane te pozwalają na pokazanie różnych typowych problemów występujących w analizie szeregów czasowych.

## **2.2. Przegląd metod analizy danych giełdowych**

Poszukiwania literaturowe dotyczące analizy danych giełdowych, a właściwie szeregów czasowych je zawierających, należy poprzedzić uwagą wskazującą, że w świecie ekonomii stosuje się różne podejścia mające na celu dokonanie takiej analizy. Wymienić tu należy następujące analizy: fundamentalną; techniczną; ilościową i analizę sentymentu.

Każde z tych podejść służy temu, by z danych uzyskać informacje pozwalające na zdobycie przewagi na rynku. Jednak śledząc literaturę i fora internetowe, łatwo można się zorientować, że każde z nich ma swoich zwolenników i zagorzałych przeciwników [13]. Poszczególne podejścia cechuje pewna specyfika, która powoduje powstawanie ograniczeń co do kręgu użytkowników lub stwarza wysoki próg wejścia do skorzystania z konkretnej metodologii. Należy odnotować, że analizie fundamentalnej zarzucana jest zbyt duża powolność. Analiza ilościowa uchodzi za zbyt skomplikowaną, co może przynieść katastrofalne skutki przy próbie jej wykorzystania przez nieprzygotowanych merytorycznie użytkowników. Z kolei analiza techniczna postrzegana jest przez wiele środowisk jako nie mająca racjonalnych podstaw samosprawdzająca się przepowiednia. Jeśli chodzi o analizę sentymentu, to uchodzi ona za zbyt subiektywną i dającą opóźnioną ocenę [14].

Wartym podkreślenia jest fakt, że dyskusje na temat, która z nich jest lepsza i dlaczego, zajmują wielu specjalistów i badaczy już od momentu ich powstania. Szczególnie dużo kontrowersji powstało na temat analizy technicznej [13]. W stosunku do niej podnoszone są argumenty o subiektywizmie stosowanych metod graficznego zobrazowania wyników analiz. Jednak coraz łatwiejszy dostęp do narzędzi analitycznych wykorzystujących analizę techniczną jak i łatwy dostęp do danych,

spowodowały mimo wszystko jej ogromne rozpowszechnienie wśród szerokich kręgów inwestorów i traderów. Analiza ta cechuje się stosunkowo niskim tzw. progiem wejścia, co oznacza dostępność dla szerokiego grona osób bez wykształcenia kierunkowego.

### **2.2.1. Analiza fundamentalna**

Analiza fundamentalna prowadzona jest przez wąskie grono specjalistów analityków gospodarczych i rynkowych. Dane, do których mają dostęp, obejmują bardzo szeroki wachlarz parametrów, wskaźników i indeksów używanych w ekonomii do klasyfikacji, ocen i wyceny poszczególnych walorów takich jak: surowce, dobra konsumpcyjne, waluty i usługi. Analizy prowadzone z wykorzystaniem tej metodologii dają wiele wskazówek dla makroekonomii, ale również dla inwestorów długoterminowych, których horyzont czasowy inwestycji wykracza często poza ramy jednego roku kalendarzowego. W przypadku tych analiz, wnioski dla rynków kapitałowych i rynków akcji nie zawsze są już tak przydatne, gdyż perspektywa czasowa podejmowanych działań często nie przekracza jednego miesiąca. Dodatkowo wynik analizy fundamentalnej zawiera dużą liczbę parametrów nie zawsze zrozumiałych dla osób bez stosownego wykształcenia i umiejętności ich wykorzystania.

### **2.2.2. Analiza techniczna**

Drugie z wymienionych na wstępie tego rozdziału podejść do analizy danych jest udziałem tzw. analizy technicznej. Jest ona w centrum zainteresowania szerokiej społeczności analityków giełdowych, traderów oraz inwestorów indywidualnych. U jej podstaw leży przekonanie, że zjawiska rynkowe przebiegają według pewnych schematów. Zakłada się przy tym, że notowania obserwowane w historii są w jakimś stopniu lub fragmencie podobne do sytuacji aktualnej. Ponad to, stosowana jest przy tym statystyczna obróbka danych, co skutkuje używaniem licznych wskaźników i sygnałów „technicznych” wspomagających podejmowanie decyzji inwestycyjnych. Należy przy tym wspomnieć o szeroko stosowanych graficznych sposobach reprezentacji danych. W tzw. zachodniej analizie technicznej szeroko stosowane na wykresach są słupki (ang. *bars*) [13]. Jednak ich czytelność i łatwość interpretacji danych zawartych na wykresie okazała się trudniejsza w porównaniu do innych technik. Wśród nich poczesne miejsce zajmują świece japońskie opracowane

najprawdopodobniej na początku XX w. a także inne techniki japońskie (wykresy Renko, Kagi, Haikin Ashi itp. [15,16].

W Japonii, w erze przed powszechnym zastosowaniem komputerów, opracowana została technika analizy wykresu Ichimoku Kinko-Hyo. Przez okres ok. trzech lat pracowało nad nią ponad dwa tysiące studentów, wykonujących obliczenia pod kierunkiem Goichi Hosody. Obecnie technika ta jest zaimplementowana na powszechnie stosowanych platformach transakcyjnych. Jej skuteczne użycie uwarunkowane jest jednak dogłębnym przestudiowaniem, siedmiu trudno dostępnych, ksiąg autorstwa Hosody. Na uwagę zasługuje tu użycie specyficznym zdefiniowanego sposobu liczenia średnich, nie spotykanego w podejściu zachodnim [17].

Japońskie techniki zobrazowania ceny obejmują również wspomniane wcześniej tzw. świece Renko (jap. – cegiełki) [15]. Wykres Renko skonstruowany jest tak, by odzwierciedlał wyłącznie ceny a eliminował skalowanie poziomej osi czasu. Niemniej jednak wykres taki pokazuje następstwo zmian ceny, umożliwiając tym samym budowanie prostych systemów transakcyjnych.

W przypadku japońskiej analizy technicznej jako ciekawostkę można przytoczyć informację, że najstarsze dane zapisywane od XI w. jako ceny giełdowe, pochodzą z japońskiej giełdy handlującej ryżem. Tam też w izolacji od świata zachodu powstały instrumenty finansowe oraz specyficzne techniki analizy wykresów. Te ostatnie zostały ujawnione dla świata zachodu w latach dziewięćdziesiątych XX w. za sprawą publikacji Steve'a Nisona [15]. Zmieniło to oblicze analizy technicznej jaka wcześniej była powszechnie stosowana.

Pod koniec XIX i na początku XX wieku powstały w USA używane do dziś teorie wspomagające inwestorów. Do najbardziej znanych należy teoria Dow'a dotycząca trendów [18,19,20], a także teoria fal Eliota, dotycząca struktury przebiegów zmiany cen na wykresie [21] czy, również opracowana w latach 30. XX wieku, teoria Wyckoffa dotycząca akumulacji i dystrybucji na rynkach [22,23].

Wiele zmian dotyczących analizy ceny i przewidywania jej zmian wniosło zastosowanie komputerów. Możliwe stało się wspomaganie obliczeniami statystycznymi w czasie rzeczywistym. Powstały liczne wskaźniki statystyczne, bardzo często reprezentowane w formie graficznej bezpośrednio na wykresie ceny. Wymienić tu należy m.in. proste średnie kroczące, wykładnicze średnie kroczące, średnie bez opóźnienia, średnie zakresu zmienności, wstęgi zmienności, widma, wskaźniki trendu, wskaźniki impetu [13,24]. Pokazane zostały różne zależności i „ukryte” relacje na

wykresach notowań. Wśród nich warto wymienić proporcje geometryczne bazujące na tzw. współczynnikach Fibonacciego [25,26], wzorce kształtów geometrycznych opisywane w literaturze jako „pattern”, np. Gartley 222, Butterfly, Bat itp. [25] oraz tzw. Price Action operujące na następujących po sobie sekwencjach akcja – konsolidacja przedzielanych od czasu do czasu sekwencją akcja – reakcja [27].

Powszechna dostępność platform transakcyjnych z możliwością użycia setek wskaźników i narzędzi analitycznych wspomagających proces „ręcznej” analizy i podejmowania decyzji inwestycyjnych nie zmniejszyła jednak stopnia trudności jakimi obarczona jest predykcja szeregów czasowych.

Środowisko analityków i inwestorów, bez względu na wewnętrzne podziały (traderzy, inwestorzy instytucjonalni, inwestorzy krótkoterminowi oraz długoterminowi, spekulanci), poszukuje coraz lepszych narzędzi wspomagających proces podejmowania decyzji. Dbałość o szeroką edukację w obszarze analizy technicznej przejawia się powstaniem krajowych stowarzyszeń. W Stanach Zjednoczonych już w roku 1972 zostało utworzone Market Technician Association – MTA [16].

W relacjach pochodzących od wszystkich tych grup przewija się mocno akcentowany wpływ czynnika związanego z ludzką psychiką. Większość inwestorów zdaje sobie sprawę z ogromnego udziału tego czynnika w podejmowanych decyzjach. Statystyki podawane przez instytucje finansowe i biura maklerskie pokazują, że w zależności od rynku od 70% do 98% inwestorów zamiast oczekiwanych zysków ponosi straty [28]. Duży wysiłek skierowany jest zatem na wyeliminowanie czynnika ludzkiego i związanych z tym emocji. Rozwiązania upatruje się w stosowaniu do zawierania transakcji, algorytmów oraz metod sztucznej inteligencji (AI).

Z dostępnych danych wiadomo, że transakcje dokonywane przez algorytmy osiągnęły już udział ok. 92% ogólnej ich liczby w roku 2019 [29]. Zaznaczyć jednak należy, że tendencja w tym zakresie jest rosnąca. Przedstawione informacje jednoznacznie pokazują w jakim kierunku zmierza światowa zbiorowość w swoich wysiłkach na rynkach inwestycyjnych i finansowych.

#### ***2.2.2.1. Przegląd wybranych narzędzi analizy technicznej***

Przedstawione wyżej ogólne spojrzenie na analizę techniczną wymaga jednak uszczegółowienia wybranych zagadnień, ze względu na ich kluczowe znaczenie. Pomocne w dalszych rozważaniach będzie omówienie terminu „świeca japońska” oraz niektórych pojęć jakie są z nią związane.



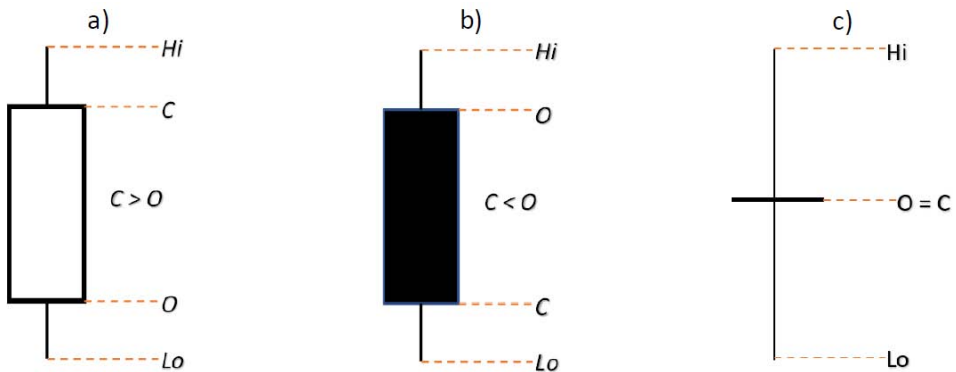
Świeca japońska jest formą sparametryzowanego piktogramu, którego wygląd oraz zasadę tworzenia przedstawia rys. 6. Piktogram świecy japońskiej utworzony jest przez prostokąt zwany korpusem oraz pionowe kreski nazywane cieniami lub knotami. Wysokość korpusu obliczana jest jako różnica pomiędzy ceną otwarcia i ceną zamknięcia:

$$K = |O - C|, \quad (8)$$

gdzie:

$K$  – wysokość korpusu świecy w jednostkach wyrażających cenę instrumentu,  
 $O$  – kurs otwarcia notowań w chwili rozpoczęcia okresu czasu objętego świecą,  
 $C$  – kurs zamknięcia notowań dla okresu czasu objętego świecą.

Przyjmuje się, że świeca wzrostowa (rys. 6a) ma korpus biały, natomiast świeca spadkowa (rys. 6b) ma korpus czarny. Wyjątkiem jest świeca doji, pokazana na rys. 6c, w której wysokość korpusu wynosi zero, gdyż kurs otwarcia równy jest kursowi zamknięcia.



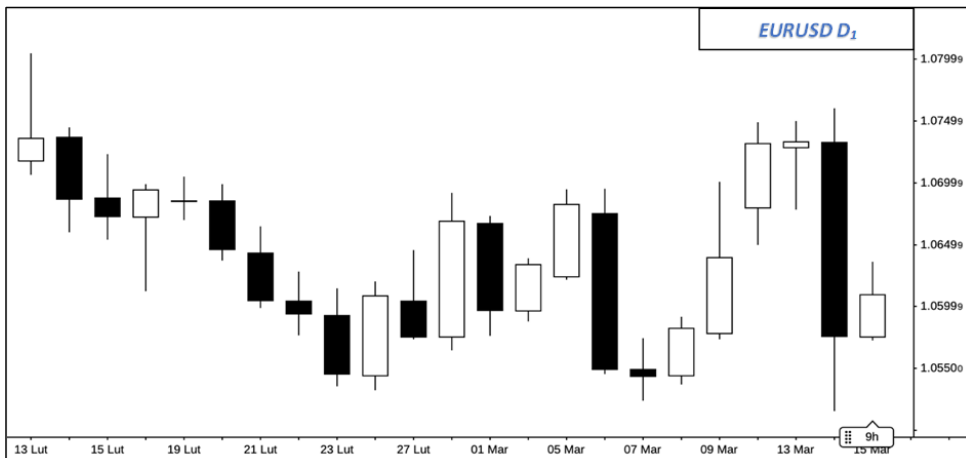
**Rysunek 6.** Świeca japońska przedstawiająca wahania kursu w zadanym okresie czasu [15,30]: **a)** świeca wzrostowa; **b)** świeca spadkowa; **c)** świeca doji. Na rysunku przyjęto następujące oznaczenia:  $O$  – kurs otwarcia,  $C$  – kurs zamknięcia,  $Hi$  – kurs maksymalny,  $Lo$  – kurs minimalny.

Opisana wyżej, w ślad za japońskim oryginałem, konwencja nadawania kolorów białego i czarnego poszczególnym świecom jest umowna. Na platformach analitycznych i transakcyjnych spotykane są często inne przypisania kolorów poszczególnym świecom, np.: zielony lub żółty dla świec wzrostowych a czerwony dla świec spad-

kowych. Długość górnego cienia świecy jest różnicą wartości Hi oraz C lub O zależnie od tego czy świeca jest wzrostowa czy spadkowa. Analogicznie dla cienia dolnego jest to różnica O lub C i wartości Lo. Wysokość całej świecy (amplituda) liczona jest natomiast jako różnica pomiędzy wartościami Hi i Lo. Przykładowy wykres świecowy pokazany został na rys. 7.

Piktogram świecy japońskiej umieszczany jest na wyskalowanym wykresie tak, aby jego charakterystyczne punkty odpowiadały wartościom kursów na osi rzędnych i jednostkom czasu dla osi odciętych. Świece japońskie niemal zupełnie wyparły używaną wcześniej na Zachodzie notację słupkową (ang. *bars*).

Pokazany na rys. 7 tzw. „czysty wykres” jest punktem wyjścia do przeprowadzania różnych analiz oraz wyliczania wskaźników pomocnych w ocenie sytuacji na rynku. Każda świeca na wykresie jest zatem graficzną reprezentacją jednego punktu w szeregu czasowym nazywanym OHLC od pierwszych liter odpowiadających wartościom: Open, Hi, Lo, Close. Każdą czwórkę takich liczb poprzedza znacznik czasu, określający precyzyjnie moment otwarcia świecy. Opcjonalnie mogą też występować w szeregu czasowym inne wartości liczbowe jak np. wolumen, który również może być pokazywany na platformach transakcyjnych.

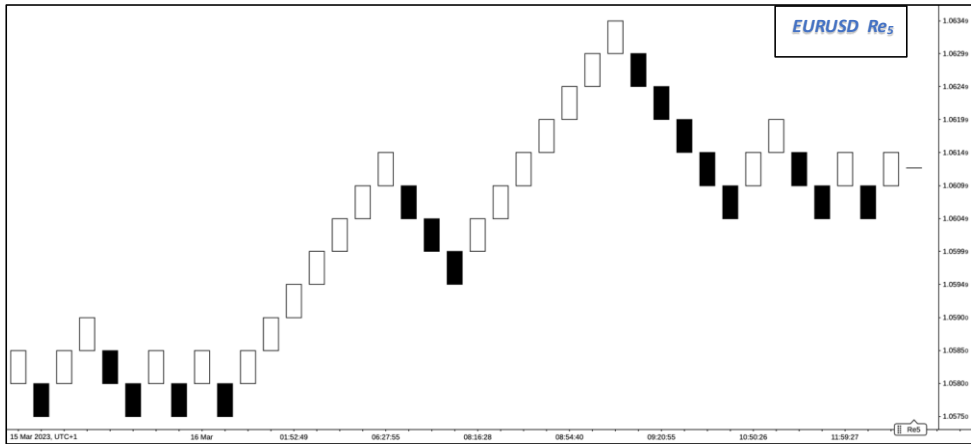


**Rysunek 7.** Wykres notowań z wykorzystaniem świec japońskich [15,16,31], sporządzony na platformie cTrader [32]

Każdemu szeregowi czasowemu przypisany jest jeden interwał czasowy (rama czasowa) dotyczący wszystkich zawartych w nim świec. Dane dla poszczególnych interwałów czasowych zapisywane są w plikach jako oddzielne szeregi czasowe,

np. notowań ze świecami dziennymi (rama czasowa D1), jednogodzinnymi (rama czasowa h1) czy jednominutowymi (rama czasowa m1). Opisany wyżej sposób zapisu danych liczbowych zawierających znaczniki czasu oraz notowania stał się już standardem.

Koncepcja świecy japońskiej nie wyczerpuje zagadnienia graficznej reprezentacji notowań. Na uwagę zasługuje również wykres ceny tworzony z użyciem tzw. klocków Renko (w języku japońskim „renko” oznacza cegłę).

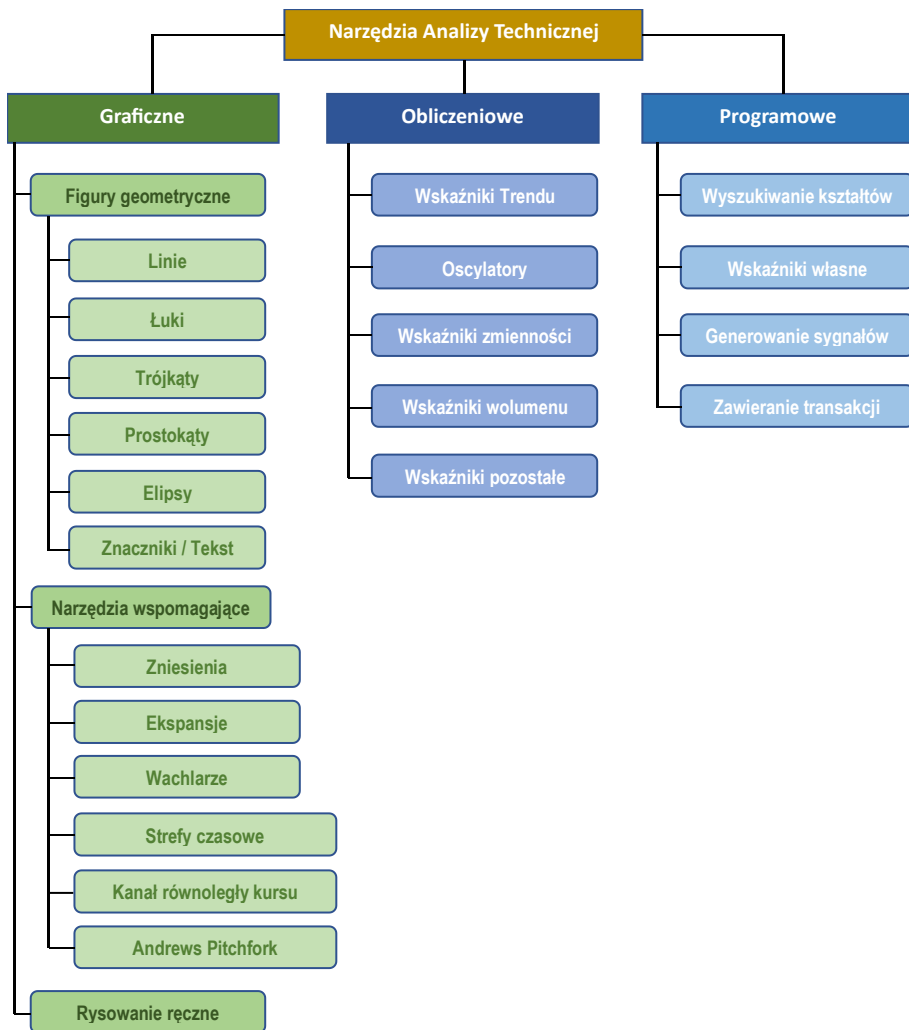


**Rysunek 8.** Przykładowy wykres notowań z wykorzystaniem klocków Renko [15], sporządzony na platformie cTrader [32]

Wykres Renko pokazany na rys. 8, tworzony jest według prostej reguły rysowania symbolicznego prostokąta w sytuacji, gdy cena „idąc z trendem” pokonała zdefiniowany a priori przedział wartości, np.: dla wykresu renko EURUSD  $Re_5$  z rys. 8, wysokość prostokąta „cegiełki” wynosi 5 pipsów. Charakterystyczna dla tego typu wykresów jest nieliniowa skala osi czasu. Oś ta prezentuje następstwo ruchów ceny, zaś jej skala jest nieregularna i ma charakter orientacyjny. W przypadku wzrostów ceny, prostokąt na wykresie Renko jest biały a dla spadków jest wypełniany kolorem czarnym, analogicznie do konwencji opisaney wcześniej dla klasycznych świec japońskich.

Analiza techniczna wykorzystuje przedstawione reprezentacje danych liczbowych, w procesie ich opracowywania z wykorzystaniem technologii informatycznych. W erze przed powszechnym zastosowaniem komputerów, wyniki ręcznych

obliczeń rysowane były na arkuszach papieru. Obecnie wszystkie narzędzia analityczne zostały przeniesione na komputerowe platformy transakcyjne.



**Rysunek 9.** Systematyka narzędzi stosowanych w analizie technicznej [16]

Wśród typowych narzędzi, które tworzą swoisty niezbędnik wymienić należy:

- bogaty zestaw wskaźników wyliczanych i wyświetlanych na ekranie w czasie rzeczywistym (średnie kroczące, oscylatory, indeksy);

- narzędzia graficzne do nanoszenia różnych figur geometrycznych na wykres (proste, kanały czyli proste równoległe, trójkąty, okręgi, różne oznaczenia i symbole);
- narzędzia graficzne wspomagające poszukiwania proporcji i związków geometrycznych pomiędzy ruchami ceny (zniesienia wewnętrzne, zniesienia zewnętrzne, łuki, wachlarze i strefy czasowe Fibonacciego).

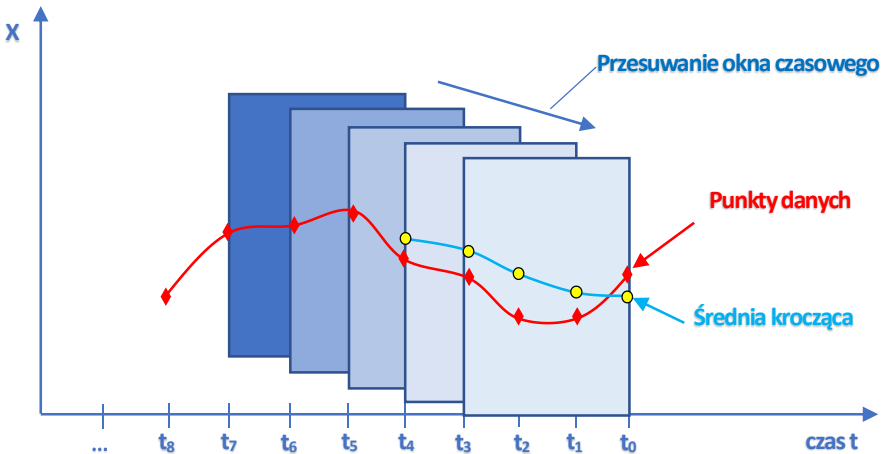
Taksonomię wymienionych narzędzi można przedstawić w postaci graficznej. Pozwala ona na łatwiejszą orientację w tym bardzo obszernym arsenale. Pokazany diagram na rys. 9 jest przykładem takiej systematyki sporządzonej w oparciu o informacje zaczerpnięte z literatury oraz platform transakcyjnych: [15,16,31,33,34,35]. Jest on wynikiem ogólnego spojrzenia na warsztat analityka i nie wyczerpuje wszystkich aspektów tematu. Wachlarz pokazanych narzędzi uzupełnia możliwość zaprogramowania dowolnych procedur wspomagających wyszukiwanie i analizę wzorców oraz robotów transakcyjnych automatyzujących handel.

Bardzo istotnym a trudnym do ujęcia w tej systematyce jest sposób użycia poszczególnych narzędzi. Obszerna wiedza kryjąca się za wieloma z tych terminów decyduje o skutecznym ich wykorzystaniu. Przykładowo zagadnienia poziomów i zniesień Fibonacciego są tematem wielu opracowań i obszernych monografii [25,26,36,37], pozwalających na głębsze poznanie i ich zastosowanie. W wielu pracach można również znaleźć informacje dotyczące tzw. wzorców (ang. *patterns*) [25], które bezpośrednio wiążą się ze zniesieniami i ekspansjami Fibonacciego.

Ważnym zagadnieniem w analizie technicznej jest również umiejętność posługiwania się elementami geometrycznymi nanoszonymi ręcznie na wykres. Ręczne poprowadzenie prostej związanej z linią trendu lub umieszczenie prostokąta obrazującego strefę wsparcia lub oporu jest nieodzowną umiejętnością, od której zależeć może właściwa ocena sytuacji rynkowej. Aktualnie duże znaczenie ma również umiejętność sformułowania własnych algorytmów i zaprogramowania procedur pozwalających na uzyskanie dodatkowych narzędzi indywidualizujących przeprowadzane analizy lub automatyzujących podejmowane działania. Jednymi z najczęściej wykorzystywanych elementów pomocnych w analizie technicznej są tzw. wskaźniki. Kategoria ta obejmuje wiele grup różniących się między sobą sposobem obliczania, zastosowaniem jak i graficzną reprezentacją na ekranie platform transakcyjnych [13].

### 2.2.2.2. Średnia krocząca (ruchoma)

Jest to jedna z najbardziej znanych idei jakie są wykorzystywane do wspomaganie analiz. Używana jest ona do tworzenia różnych wskaźników i oscylatorów. Obliczanie średniej kroczącej dla szeregu czasowego korzysta z danych, które z oczywistych względów należą już do historii. W konsekwencji wynik obliczeń odzwierciedla w jakimś stopniu dynamikę zmian zawartą w danych. Sposób tworzenia średniej kroczącej bazuje na uwzględnianiu w obliczeniach wartości, które są odległe od punktu bieżącego nie więcej niż przyjęta a priori pewna liczba obserwacji liczona wstecz wzdłuż osi czasu. Na rys. 10 pokazana została zasada przesuwania się tzw. okna czasowego ograniczającego zakres punktów branych do obliczeń.



**Rysunek 10.** Obliczanie średniej kroczącej. Ilustracja zasady przesuwania okna czasowego obejmującego w każdym kroku stałą liczbę punktów danych wziętych do obliczenia średniej

Oznaczone kolorem czerwonym poszczególne punkty danych objęte aktualnym oknem, są uśredniane z wykorzystaniem jednej z przyjętych formuł. Obliczone wartości umieszczane są na wykresie razem z danymi z analizowanego szeregu czasowego. Formuły stosowane do obliczeń w zależności od typu średniej definiowane są w odmienny sposób. Najpopularniejszą jest prosta krocząca średnia arytmetyczna (SMA) wyrażona równaniem:

$$SMA = \frac{x(t_0) + x(t_1) + \dots + x(t_k)}{k+1} \quad (9)$$

W wyrażeniu (9), definiującym SMA przyjęto następujące oznaczenia:

*SMA* – prosta średnia krocząca (arytmetyczna),

$x(t_0)$  – wartość zmiennej  $x$  w chwili  $t_0$ ,

$x(t_k)$  – wartość zmiennej  $x$  w chwili  $t_k$ ,

$k$  – liczba okresów objętych aktualnym oknem czasowym,

$k + 1$  – liczba punktów danych objętych oknem czasowym.

Średnia ta charakteryzuje się istotnym opóźnieniem, które na wykresie jest wyraźnie widoczne jako przesunięcie linii oraz tworzonych przez nie ekstremów, pomiędzy oboma wykresami. Dane z historii w istotny sposób mogą zmieniać wartość takiej średniej. Jest to szczególnie widoczne w sytuacjach występowania dużych wahań i nagłych zmian wartości aktualnej. Jest to powodem wielu ograniczeń w jej stosowaniu.

Zjawisko wpływu wartości historycznych na wartość wyliczanej średniej zostało zredukowane w pewnym stopniu w przypadku wykładniczej średniej kroczącej (EMA), która jest odmianą średniej ważonej [38,39].

Średnia ta obliczana jest z wykorzystaniem równania:

$$EMA = \frac{x(t_0) + x(t_1)(1-\alpha) + x(t_2)(1-\alpha)^2 + \dots + x(t_k)(1-\alpha)^k}{1 + (1-\alpha) + (1-\alpha)^2 + \dots + (1-\alpha)^k} \quad (10)$$

Oznaczenia we wzorze (10) są następujące:

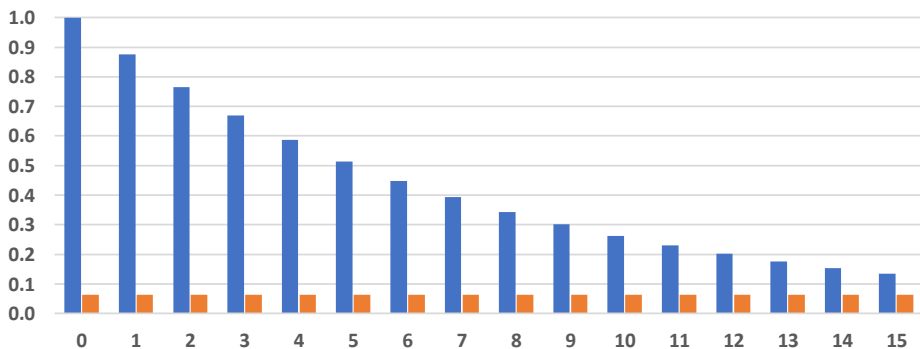
*EMA* – wykładnicza średnia krocząca,

$x(t_0)$  – wartość zmiennej  $x$  w chwili  $t_0$ ,

$x(t_k)$  – wartość zmiennej  $x$  w chwili  $t_k$ ,

$k$  – liczba okresów objętych aktualnym oknem czasowym,

$\alpha = \frac{2}{k+1}$  – współczynnik wagowy.



**Rysunek 11.** Porównanie wartości wag  $w = (1 - \alpha)^k$  dla kolejnych próbek  $n$  przy  $k = 15$ , w trakcie obliczania średniej EMA(15) – kolor niebieski oraz wag  $w$  dla średniej SMA(15) – kolor pomarańczowy

Współczynniki wagowe  $(1 - \alpha)^k$  wyliczane w równaniu (10) zmieniają się wykładniczo. Powoduje to szybszy zanik wpływu wartości historycznych na wartość obliczanej średniej. Porównanie zmian wartości wag w funkcji numeru próbki dla średnich kroczących SMA oraz EMA ilustruje wykres z rys. 11.

Zastosowanie średniej EMA zmniejsza wielkość opóźnienia średniej w stosunku do wartości rzeczywistych. Porównanie przebiegów średnich SMA i EMA dla okna czasowego o identycznej wielkości równej dziesięciu interwałom, pokazane zostało na rys. 12.

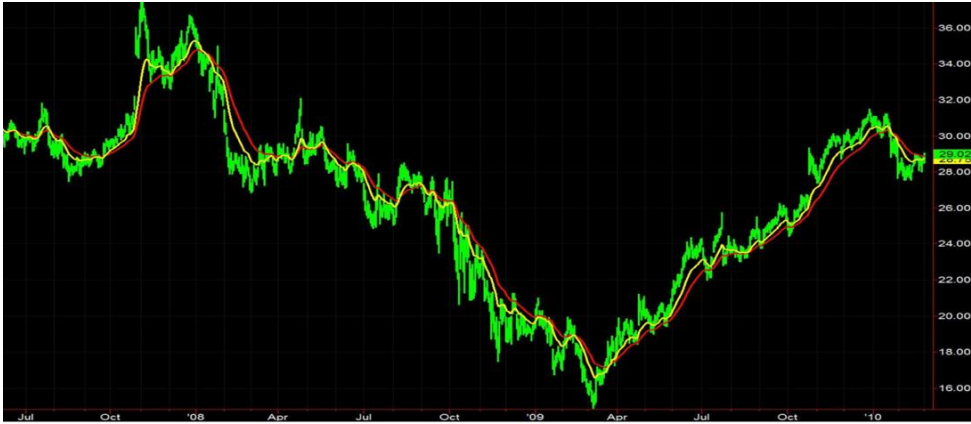
Wśród możliwości obliczania średniej kroczącej godne uwagi jest zastosowanie metody najmniejszych kwadratów (MNK). Metoda ta została szerzej przedstawiona w podrozdziale 3.3. Ten sposób obliczeń tworzy średnią kroczącą szeregu czasowego (TSMA) [40]. Średnia ta wykazuje jeszcze mniejsze opóźnienie w stosunku do wartości rzeczywistych, co zostało pokazane na rys. 12.



**Rysunek 12.** Porównanie przebiegów średnich kroczących SMA(10), EMA(10) i TSMA(10), na przykładzie notowań pary walutowej EURUSD dla ramy czasowej h4. Wykresy sporządzone na platformie cTrader [32]

Wieloletnie poszukiwania nowych rozwiązań zaowocowały opracowaniem koncepcji średniej bez opóźnienia (ZLMA), którą w latach 80. XX w. zaproponowali John Ehlers i Ric Way [16,41,42,43]. Idea stojąca za konstrukcją tej średniej sprowadza się do szacowania błędów regresji i na tej podstawie obliczania poprawek do predykcji. Porównanie przebiegu średniej ZLMA w stosunku do średniej liczonej klasycznie pokazuje rys. 13. Wykorzystanie przecięć tej średniej ze średnią EMA pozwala na ułatwienie w konstruowaniu systemów do handlu automatycznego.





**Rysunek 13.** Porównanie przebiegu średniej bez opóźnienia ZLMA i średniej SMA dla notowań spółki MSFT zaznaczonych kolorem czerwonym. Kolorem żółtym zaznaczona jest średnia ZLMA a kolorem czerwonym średnia EMA [42]

Przegląd średnich kroczących wymaga jeszcze omówienia średnich opracowanych na użytek analiz giełdowych w Japonii. Średnie te również wykorzystują okno czasowe obejmujące określoną stałą liczbę świec. Różnica polega na odmiennym obliczaniu średniej. Jest ona wyznaczana jako średnia arytmetyczna z minimalnej i maksymalnej ceny zamknięcia świec objętych oknem, zgodnie z równaniem (11). Odpowiada to wyśrodkowaniu wartości pomiędzy ekstremami w zadanym okresie czasu:

$$avr_k = \frac{MIN(C_k) + MAX(C_k)}{2} \quad (11)$$

gdzie:

$avr_k$  – średnia krocząca ekstremów,

$C_k$  – wartości zamknięcia świec o indeksach od 1 do  $k$ ,

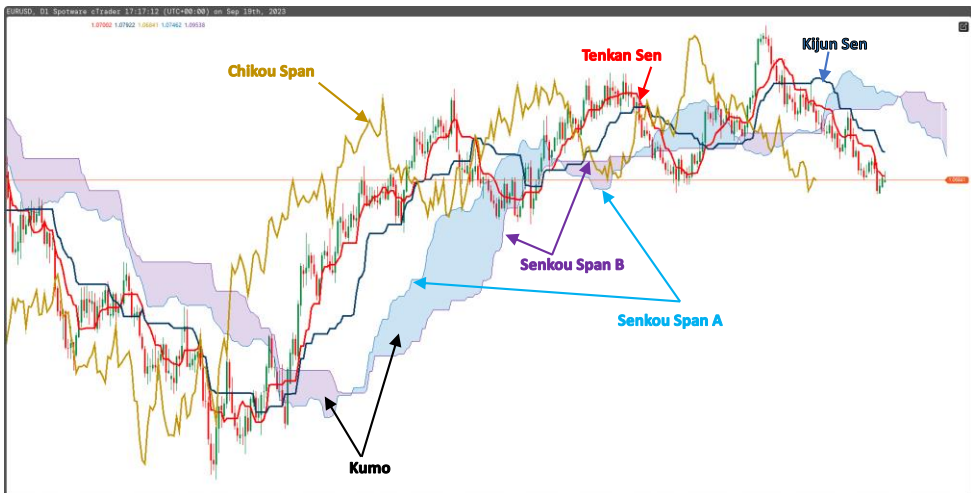
$MIN(C_k)$  – minimalna wartość zamknięcia świec o indeksach od 1 do  $k$ ,

$MAX(C_k)$  – maksymalna wartość zamknięcia świec o indeksach od 1 do  $k$ ,

$k$  – liczba świec objętych aktualnym oknem czasowym.

Z tym specyficznym sposobem obliczania średniej wiąże się też skokowy charakter zmian jej wartości. O zmianie wartości średniej może zdecydować nowa świeca, która została objęta przesuniętym właśnie oknem lub wypadnięcie ostatniej świecy z takiego okna, nazywane „wymieraniem świecy”. Średnie liczone w ten sposób, dla

ściśle określonych okien czasowych, zestawione zostały na jeden wykres przez japońskiego inwestora Goichi Hosodę. Zestawienie to znane na świecie pod nazwą *Ichimoku Kinko Hyo* pokazane zostało na rys. 14 w formie wskaźnika na platformie transakcyjnej. Wskaźnik *Ichimoku* obejmuje pięć linii, z których cztery są średnimi wyznaczanymi w opisany wcześniej sposób. Linie *Tenkan Sen* i *Kijun Sen* wyznaczane są według zależności (11) obejmującej odpowiednio 9 i 26 świec. Z kolei linia *Senkou Span A* powstaje jako średnia *Tenkan Sen* i *Kijun Sen* za 26 okresów i jest przesunięta do przodu o 26 okresów. Linia *Senkou Span B* jest średnią linii *Tenkan Sen* i *Kijun Sen*, lecz obliczaną z 52 okresów i również jest przesunięta do przodu o 26 okresów. Linie te tworzą tzw. chmurę *Kumo*, która jest cieniowana zależnie od położenia obu linii względem siebie. Ostatnia z linii *Chikou Span* reprezentuje ceny zamknięcia świec i jest przesunięta o 26 okresów do tyłu. Opisane wyżej linie w zależności od położenia względem siebie oraz przecięć jakie tworzą na wykresie, mogą być pomocą w ocenie sytuacji rynkowej oraz źródłem sygnałów transakcyjnych. Sposób interpretacji oraz definicje poszczególnych sygnałów określa strategia *Ichimoku* [17].



**Rysunek 14.** Przykładowy wygląd wskaźnika Ichimoku Kinko Hyo dla notowań pary walutowej EURUSD (D1). Opis poszczególnych linii, za japońskim oryginałem, odpowiada określeniom używanym przez analityków giełdowych. Wykresy sporządzone na platformie cTrader [32]

Przy okazji omawiania średnich ruchomych warto nadmienić, że można je stosować do konstrukcji innych wskaźników jak również oscylatorów. Przykładem mogą być

umieszczane na wykresie razem z danymi tzw. wstęgi, które obrazują zakres zmienności np. wstęga Bollingera [16] czy kanał Keltnera [44]. Składają się one z trzech linii. Linia środkowa liczona jest jako średnia krocząca np. EMA. Linie górna i dolna są oddalone od linii środkowej o wyliczoną odległość na podstawie innego wskaźnika np. odchylenia standardowego (dla wstęgi Bollingera), ATR (dla kanału Keltnera) lub tzw. koperty (przyjętej stałej wartości procentowej). Przykładowy wykres zawierający wskaźnik ze wstęgami Bollingera jest pokazany na rys. 15.



**Rysunek 15.** Przykład wskaźnika ze wstęgą Bollingera dla pary EURUSD (D1). Wykresy sporządzone na platformie cTrader [32]

Kolejną omawianą kategorią narzędzi wykorzystywanych w analizie technicznej, są oscylatory. W odróżnieniu od wskaźników rysowanych na tle wykresu danych, wykresy oscylatorów umieszczane są pod wykresem. Tworzą one ramki z wykresami wstęgowymi. Oscylatory wskazują sytuacje na wykresie, dając przewagę przy podejmowaniu decyzji, szczególnie w przypadkach braku wyraźnego trendu na rynku. Oscylatory spotykane na platformach transakcyjnych można podzielić na kilka grup zadaniowych. Dobiera się je indywidualnie do potrzeb dyktowanych przeprowadzaną analizą. Wymieniane są trzy sytuacje dla których są one konstruowane:

- Osiągają wartości skrajne, zbliżone do ekstremów wahań co może być świadectwem wykupienia (wartości bliskie maksimum) lub wyprzedania (wartości bliskie minimum) na rynku;

- Wykazują dywergencję zmian wartości w stosunku do zmian ceny w okolicy ekstremum, co może sygnalizować zbliżającą się zmianę kierunku trendu;
- Przecinają linię poziomu zerowego dając ważny sygnał do zajęcia pozycji zgodnej z trendem.

Wśród najbardziej popularnych zadań realizowanych za pomocą oscylatorów wymienić można: mierzenie impetu wyrażającego tempo zmiany cen, mierzenie względnej zmiany (ROC), mierzenie siły względnej (RSI), wykrywanie anomalii MACD [13,16].

MACD (*Moving Average Convergence / Divergence*) jest przykładem zastosowania średnich ruchomych do tworzenia oscylatorów. Wykorzystywany jest on do analizy zarówno ceny jak i innych danych np. dotyczących wolumenu czy liczby kontraktów. Oscylator MACD tworzony jest z wykorzystaniem dwóch średnich kroczących o różnych oknach czasowych. Średnia o węższym oknie czasowym (szybsza) tworzy punkty przecięcia z linią średniej o większym oknie czasowym (wolniejsza). Dodatkowo obliczana jest różnica pomiędzy wartościami obu średnich, co umożliwia zobrazowanie zarówno za pomocą linii jak i słupków histogramu. W praktyce wybór stosowanego wariantu MACD podyktowany jest indywidualnymi preferencjami analityka. Oba warianty tego oscylatora pokazane zostały na rys. 16.



**Rysunek 16.** Przykład zobrazowań oscylatora MACD dla pary walutowej EURUSD (D1): **a)** linie dające sygnały przy przecięciu; **b)** histogramem pokazujący różnicę pomiędzy średnimi. Wykresy sporządzone na platformie cTrader [32]

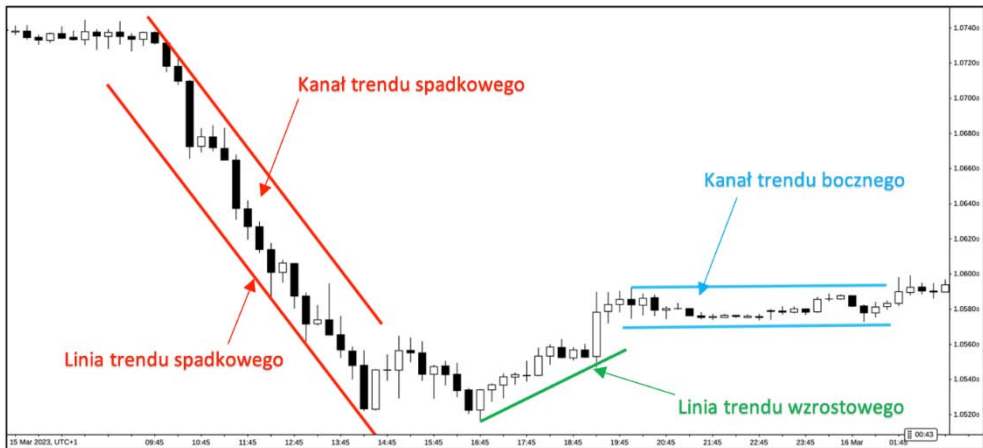
Specjalnej uwagi w przeglądzie wskaźników wymaga grupa przeznaczona do określania trendu. Zagadnienie to jest w centrum uwagi już od ponad stu lat. Ukute na przełomie wieków XIX i XX w Stanach Zjednoczonych powiedzenie: „Trend is your friend” odnosi się do inwestowania zgodnie z trendem. Już wtedy zaobserwowano, że decyzje „kontrariańskie” (przeciw trendowi) kończą się w przeważającej części dotkliwymi porażkami. Powstały fundamentalne teorie Dowa i Eliota [21]. Inwestorzy mają do wyboru wiele narzędzi wspomagających określenie panującego na rynku trendu. Ich wykorzystanie pozwala zmniejszyć wpływ subiektywnej oceny na podejmowane decyzje w tzw. handlu uznaniowym.

W wielu przypadkach prawidłowe rozpoznanie trendu nie jest zadaniem łatwym, ze względu na duży poziom szumu transakcyjnego, jaki widać na wykresach. Dostępne na platformach transakcyjnych wskaźniki trendu można podzielić na takie, które stosują narzędzia statystyczne oraz na geometryczne, wynikające z tradycji ręcznego rysowania trendów na papierze. Jednymi z najstarszych wskaźników trendu są opisane wcześniej średnie kroczące, wynikające z obróbki statystycznej szeregu czasowego i pozwalające w kontrolowany sposób dopasowywać wskaźnik do charakteru danych i rodzaju prowadzonej analizy. Do tej grupy należą wszystkie rodzaje średnich kroczących (rys. 12), wstęgi (rys. 15), oscylator MACD (rys. 16), a także oscylatory, używane jako narzędzie pomocnicze, gdyż uprawdopodobniają możliwość zmiany kierunku trendu.



**Rysunek 17.** Przykład rozrysowania trendów elementarnych dla pary EURUSD (D1). Strzałka oznacza tworzenie się struktury trendu spadkowego wyższego rzędu. Wykresy sporządzone na platformie cTrader [32]

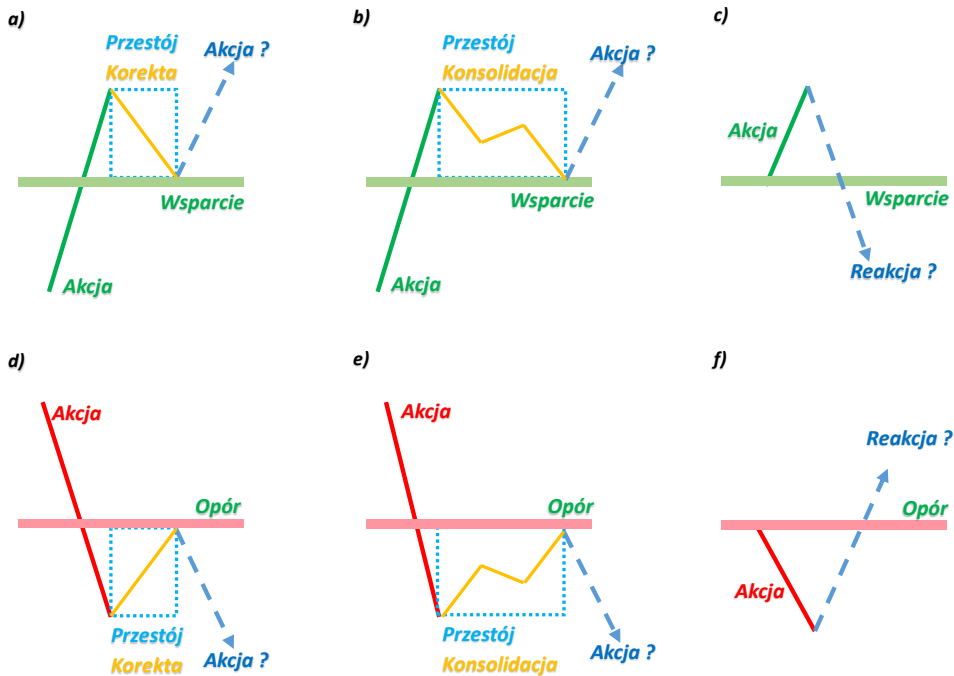
Druga grupa wskaźników bazująca na wykreślnym (geometrycznym) wyznaczeniu trendu, wykorzystuje zwykle prostą definicję stosowaną jako kryterium kwalifikacyjne. Jednym z bardziej znanych jest określenie trendu wzrostowego jako takich zmian ceny, w których zarówno maksima jak i minima są coraz wyższe. Symetrycznie dla trendu spadkowego mowa o coraz niższych ekstremach. Zmiany ceny są tu rozumiane jako przebieg ceny w pewnym okresie czasu, w którym występują zarówno wzrosty jak i spadki notowań [21]. Przykład zamieszczony na rys. 17 pokazuje kreślowanie trendów elementarnych wykorzystujące taką definicję. Można tu zauważyć tworzenie się struktury o tendencji spadkowej, utworzonej ze składowych niższego rzędu.



**Rysunek 18.** Stosowanie pomocniczych elementów geometrycznych do wyznaczania trendów dla pary EUR/USD m15. Wykresy sporządzone na platformie cTrader [32]

Często wykorzystywanym sposobem w analizach jest kreślenie pojedynczej linii pomocniczej łączącej dołki lub szczyty notowań na wykresie, odpowiednio dla wzrostów lub spadków notowań. Przedstawione sytuacje ilustruje rys. 18

Utrzymywanie się ceny powyżej linii łączącej minima tzw. dołki jest interpretowane jako oznaka trendu wzrostowego. Analogicznie dla trendu spadkowego linia łącząca maksima tzw. szczyty leży ponad wartościami ceny. Stosowane są również obie linie jednocześnie tworząc tzw. kanał, który może reprezentować wzrosty lub spadki. Przypadek, gdy linie kanału przebiegają poziomo jest określany jako trend boczny [16].



**Rysunek 19.** Wybrane schematy cenowe wykorzystywane w analizie Price Action (PA) [27]:  
**a) i d)** prawdopodobna akcja po korekcie; **b) i e)** prawdopodobna akcja po konsolidacji;  
**c) i f)** możliwa reakcja po akcji

Największym ograniczeniem w określaniu trendu jest fakt, że moment jego zakończenia i zmiany na przeciwny można zauważyć i potwierdzić tylko *post factum*. Uważa się, że mając do dyspozycji wyłącznie dane historyczne nie można wykonać skutecznej predykcji ruchów ceny. Charakter wyprzedzający mają tylko nieliczne oscylatory. Jednak ich skuteczność w predykcji jest bardzo ograniczona. Pomimo dostrzeganych trudności dotyczących predykcji, inwestorzy stosują jednak wymienione rozwiązania, godząc się z ich ułomnościami.

Uważna obserwacja wykresów oraz przebiegu notowań na poszczególnych sesjach giełdowych legły u podstaw kolejnego sposobu analizy, a mianowicie analizy ruchu ceny PA (ang. *Price Action*) [16], zakładającej występowanie ściśle określonych następstw wynikających z ruchu ceny. W praktyce oznacza to, że po dużym ruchu ceny, zwanym akcją, występuje jedna z możliwości:

- faza przestoju, w postaci korekty lub konsolidacji;

- reakcja, będąca akcją w przeciwnym kierunku.

Istotną rolę w prowadzeniu analizy PA mają tzw. wsparcia i opory. Ich umieszczenie na wykresie dopełnia analizowany układ cenowy. Wynikające z tych obserwacji schematy pokazane zostały na rys. 19.

Strategie inwestycyjne wykorzystujące PA posiłkują się jeszcze dodatkowymi elementami potwierdzającymi wartość powstałego sygnału. Jednym z nich jest analiza wielointerwałowa. To bardzo często stosowana technika pozwalająca na obserwację danych z różnej perspektywy czasowej. Zakłada się w tym przypadku samopodobieństwo wykresów dla różnych ram czasowych (fraktalność). Ramy te różnią się między sobą o współczynnik krotności, najczęściej przyjmujący wartość z przedziału od czterech do sześciu, np. h4, h1, m15. Pojawienie się koincydencji schematów cenowych (traktowanych jako sygnały) z różnych ram czasowych, jest traktowane jako uprawdopodobnienie zapowiedzi wystąpienia ruchu ceny w określonym kierunku, a w konsekwencji kierunku spodziewanego trendu.

### 2.2.3. Analiza ilościowa

Analiza ilościowa reprezentuje podejście, które okrzepło w erze stosowania komputerów do obliczeń statystycznych. Analizowane dane obrabiane są statystycznie, co pozwala na uzyskanie numerycznych ocen badanego procesu, rynku, zbioru danych. Tak otrzymane wyspecjalizowane analizy ze zbiorem obliczonych estymatorów w czasie rzeczywistym, mogą dać podstawę do oceny sytuacji i podejmowania decyzji inwestycyjnych. Wykorzystanie wyników tej analizy możliwe jest zatem tylko w kręgu instytucji finansowych, które dysponują odpowiednio dużymi zasobami sprzętowymi, oprogramowaniem oraz dostępem do danych.

Efektom analizy ilościowej jest powstanie specyficznych technik transakcyjnych, wśród których najbardziej znane i podawane w źródłach [5], to:

- handel spreadami i arbitraż (ang. *Spread trading & arbitrage*);
- złożona regresja i uczenie maszynowe (ang. *Complex regression & machine learning*);
- metody głębokiego uczenia się do przewidywania (ang. *Deep learning methods for prediction*);
- technika wysokiej częstotliwości i animacja rynku (ang. *High-frequency technic and market making*).



Na uwagę szczególną zasługuje fakt opracowania i stosowania w ostatnim czasie wysoce specjalistycznych rozwiązań do przetwarzania w czasie rzeczywistym, określanym jako HFT (ang. *High Frequency Trading*). Rozwiązania te operują na rynkach finansowych i giełdach, z czasami reakcji na poziomie pojedynczych mikrosekund. To nieosiągalne dla istot żywych tempo pracy przynosi jednak firmom wielomilionowe zyski. Stosowanie tego typu technik wymaga jednak spełnienia szeregu warunków, które decydują o sukcesie. Jednym z nich jest zapewnienie minimalnej dozwolonej przepisami odległości (długości łączy) mierzonej pomiędzy produkcyjnymi serwerami transakcyjnymi giełdy a terminalami i serwerami firmy wykorzystującej HFT. Zabieg ten stosuje się po to, by nie dopuścić do uprzywilejowania jednego z konkurentów przez powstanie zauważalnych opóźnień i różnic w czasach propagacji transmitowanych sygnałów i zleceń. Pozwala to wyrównać szanse wśród konkurentów zainteresowanych stosowaniem rozwiązań wykorzystujących tego typu technologię.

#### **2.2.4. Analiza sentymentu i aspekt psychologiczny**

Przegląd metod analitycznych zamyka analiza sentymentu. Termin ten dotyczy przetwarzania danych nieustrukturyzowanych wraz z tradycyjnymi danymi liczbowymi. Analizy tego typu zaczęły być wyodrębniane jako oddzielna grupa, po pomyślnym zastosowaniu technik uczenia maszynowego. Głównym źródłem danych jest raport COT (ang. *Commitments of Traders*) traktowany jako wskaźnik nastrojów. Dodatkowo mogą być również brane pod uwagę inne zbiory danych i informacji, np.:

- stosunek liczby opcji PUT i CALL na rynkach opcji;
- wyniki uzyskane za pomocą TEXT-MINING i algorytmów wyszukiwania słów o pozytywnym wydźwięku;
- indeks VIX (ang. *Volatility Index*) zwany też indeksem strachu;
- indeks UMCSI (ang. *United States Michigan Consumer Sentiment Index*);
- notatki i anonse prasowe;
- media społecznościowe [45].

W trakcie obróbki takich danych tworzony jest ranking poszczególnych informacji, który po konsolidacji daje obraz sentymentu całej zbiorowości. Pokaz tego typu podejścia do przygotowywania decyzji inwestycyjnych odbył się w trakcie konkursu ogłoszonego przez Kaggle (zajmuje się współzawodnictwem w zakresie AI). Wyniki

współzawodnictwa pokazały, że zastosowanie takiej analizy może dać pozytywne rezultaty inwestycyjne [46].

Analiza sentymentu wiąże się więc ze społecznym oddziaływaniem na tworzenie się notowań rynkowych. Zauważyć jednak należy, że zbiorowość tworzą jednostki podejmujące konkretne decyzje inwestycyjne. Dochodzimy tu do pojedynczego człowieka jako w pewnym stopniu niezależnego podmiotu decydującego o swoich środkach, które zaangażuje w proces handlu lub inwestycji. Wspomniana niezależność, w dobie szybkiego przepływu informacji, jest jednak pojęciem bardzo dyskusyjnym. Na uwagę zasługują tu liczne prace z dziedziny psychologii i finansów behawioralnych. Opisują one mechanizmy oddziałujące na psychikę człowieka stojącego przed problemem podjęcia ważnej decyzji inwestycyjnej lub przed zadaniem oceny i wyborem najlepszej (z punktu widzenia decydującego) strategii inwestycyjnej.

Powstała w latach 70. XX wieku teoria finansów behawioralnych zakłada, że decyzje inwestycyjne w środowisku ryzyka nie muszą być racjonalne. Daniel Kahneman i Amos Tversky, dwaj psychologowie amerykańscy, w roku 1979 opublikowali pracę, która dała podwaliny do obalenia panującego wcześniej przekonania o racjonalności ludzkich wyborów [47]. Przyczyną takiego stanu rzeczy są występujące liczne mechanizmy oddziałujące na psychikę. Wśród czynników, jakie mają wpływ na podejmowane decyzje, wymieniane są: strach, nadzieja, euforia, rozpacz, panika, depresja, optymizm czy pesymizm. Tkwią one głęboko w ludzkiej psychice, przekładając się na powstawanie błędów i uprzedzeń, wśród których najczęściej rozważane są:

- nadmierna pewność siebie – dotyczy przeceniania własnej wiedzy i umiejętności, powodując podejmowanie zbyt dużego ryzyka i dokonanie niefortunnej inwestycji;
- awersja do straty – to tendencja do niesprzedawania walorów, które przynoszą straty a jednocześnie sprzedawania walorów, które zyskują;
- iluzja kontroli – polega na subiektywnym przeświadczeniu, że szansa odniesienia sukcesu jest dużo większa niż jest prawdopodobna w rzeczywistości. Występuje tu przesadna pewność możliwości kontrolowania skutków decyzji, a przecież skutek podjętej decyzji zależy nie tylko od kompetencji i umiejętności, ale również od zewnętrznych czynników losowych;

- pułapka potwierdzania – jest tendencją do wyszukiwania informacji potwierdzających, a jednocześnie ignorowanie informacji zaprzeczających;
- błąd post factum – to niewłaściwe przekonanie, że zaobserwowany po fakcie wynik mógł być przewidziany wcześniej. U podstaw wystąpienia tej skłonności leży obarczanie za niepowodzenie losu lub innych czynników zewnętrznych;
- efekt utopionych kosztów – powoduje, że inwestor jest skłonny trwać przy podjętej złej decyzji, gdy wiązała się ona ze sporym zaangażowaniem kapitału [47].

Łatwy dostęp do informacji spowodował, że inwestor powinien również oceniać swoją łatwowierność i podatność na manipulację. Zapewnienie dopływu jak najświeższych informacji wiąże się również z opracowaniem procedur sprawdzających i weryfikujących otrzymane informacje.

Obserwacja zachowań inwestycyjnych ludzi jako jednostek pokazała, że podejmowane przez nich decyzje nie zawsze są racjonalne. Wpływają na to różne stany emocjonalne poszczególnych osób, ale też fakt, że rynkiem finansowym zainteresowana jest specyficzna zbiorowość. Jego uczestnik poddaje się zazwyczaj presji zbiorowości giełdowej. Ulega emocjom panującym w tej grupie, co wpływa na racjonalność podejmowanych przez niego decyzji. Zdolność do logicznego myślenia jest tłumiona przez zbiorowość uczestników gier giełdowych.

Taką tezę zdają się potwierdzać dane z biur maklerskich i giełd, mówiące o tym, że tylko nieliczni odnoszą sukcesy w handlu instrumentami finansowymi. Odsetek osób ponoszących straty w zależności od instrumentu oraz rynku waha się w granicach od ok. 80% do nawet 98%. Zwarzywszy na fakt, że przy zajmowaniu pozycji na rynku inwestor ma do rozważenia tylko dwie możliwości zmiany ceny (wzrost lub spadek) oraz że obie mogą być równo prawdopodobne, sporym zaskoczeniem jest tak duży odsetek pozycji stratnych. Zdaje się to tylko potwierdzać słuszność tezy, jaką przyjęto w teorii finansów behawioralnych.

Pewne światło na to zagadnienie dają wywiady z traderami, którzy na co dzień są uczestnikami rynków kapitałowych i nauczyli się trudnej sztuki podejmowania racjonalnych decyzji w warunkach rynkowej presji oraz występowania czynnika losowego, który tylko zwiększa odczuwany stres i podejmowane ryzyko. Tajemnice warsztatów ich pracy sprowadzają się w większości do stwierdzenia, że najważniejszym czynnikiem w osiągnięciu sukcesu rynkowego jest przestrzeganie procedur oraz

branie pod uwagę faktów, a nie przypuszczeń lub oczekiwań budowanych na podstawie iluzji i złudzeń (często optycznych), jakie powstają w trakcie obserwacji tworzącego się wykresu ceny oraz informacji rynkowych napływających za pośrednictwem różnych kanałów. W ramach przestrzegania procedur podkreślane jest również duże znaczenie braku odstępstw od strategii, która została przyjęta do realizacji. Rzutuje to bezpośrednio na jakość wykonywanych analiz oraz na zyskowość zajmowanych na rynkach pozycji.

Chęć wyeliminowania z operacji rynkowych czynnika związanego z emocjami i psychiką spowodowała, że już od wielu lat w świecie finansów dąży się do zautomatyzowania większości działań, wykorzystując w tym celu algorytmy oraz szeroko rozumianą i intensywnie rozwijaną sztuczną inteligencję (AI). Odnotować należy, że wszystkie rodzaje analiz wymienione w punktach 2.4.1 do 2.4.4 tego rozdziału, używają zaawansowanego oprogramowania, celem zdobycia przewagi rynkowej w czasie rzeczywistym i uzyskania wyników jakościowych nieosiągalnych wcześniej.

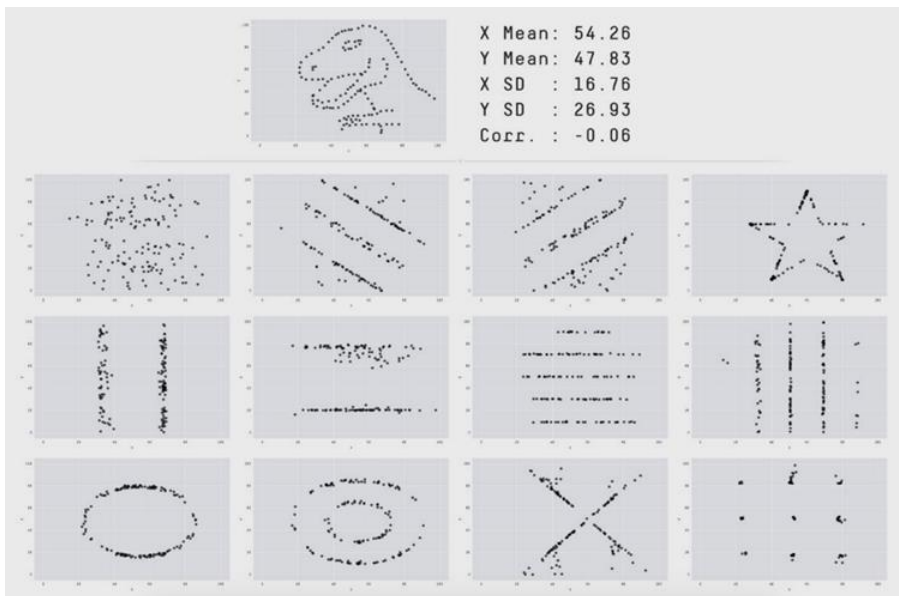
### **3. Przegląd wybranych algorytmicznych metod uczenia maszynowego**

Na użytek tego artykułu przegląd metod uczenia maszynowego został podzielony na dwie części. Pierwsza z nich dotyczy metod, które wykorzystują klasycznie rozumiane algorytmy wykonujące zadania w określony i zdefiniowany z góry sposób. Przedstawione tu zostały algorytmy używane do wstępnej obróbki i oceny danych, a także bardziej złożone algorytmy wykorzystywane w analizie szeregów czasowych.

Druga część odnosi się do metod określanых ogólnie jako sieci neuronowe, w których następuje iteracyjny proces uczenia, specyficzny dla każdego z typów sieci. W rezultacie, dla struktury tworzącej sieć neuronową, obliczany jest zestaw charakterystycznych parametrów, które decydują o wyniku jaki po przetworzeniu informacji wejściowej otrzymany zostanie na wyjściu. Zakres omawianych w przeglądzie metod został ograniczony, co było podyktowane ich praktycznym zastosowaniem w przetwarzaniu szeregów czasowych. Nie zostały szerzej omówione zagadnienia dotyczące modeli deskrypcyjnych, które dotyczą wyjaśniania zależności znajdujących się w danych. Głównym obszarem zainteresowania stały się modele predykcyjne uwzględniające kolejność analizowanych danych oraz tematy powiązane.

### 3.1. Wstępna ocena danych

Przygotowanie danych do przetwarzania stanowi pierwszy etap w procesie ich analizy. Wymaga ono wzięcia pod uwagę tego, jaki typ eksperymentu na danych jest aktualnie planowany. W przypadku modeli predykcyjnych wymaga się zwracania bardzo wiarygodnych i dokładnych wyników, pokazujących przewidywania w określonym horyzoncie czasowym. W ramach przygotowania przeprowadzane jest sprawdzenie kompletności danych oraz poszukiwanie wartości odstających. Pewnym wyzwaniem jest natomiast identyfikacja i poprawianie wartości błędnych. Ma to szczególne znaczenie w przypadku uczenia maszynowego na potrzeby modeli predykcyjnych. Użycie zbioru danych uczących zawierającego wartości błędne, wpływa na użyteczność otrzymanego modelu, ocenianą przez jakość uzyskiwanych wyników predykcji [49]. Jest to zgodne z zasadą, która mówi, że błędne dane dają błędne wyniki [50]. Dodatkowo należy mieć na uwadze to, że w bazie danych mogą być przechowywane informacje zbędne, nieodpowiednio sformatowane oraz wartości niezgodne z zasadami lub zdrowym rozsądkiem [50]. Surowe dane z bazy danych wymagają zatem oczyszczenia i / lub przekształcenia.

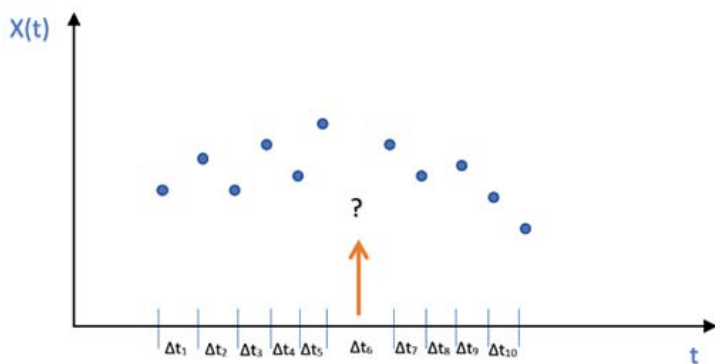


**Rysunek 20.** Graficzna prezentacja zbiorów danych o takich samych statystykach zbiorczych (dokładność dwa miejsca po przecinku) obejmujących wartości średnie X Mean oraz Y Mean, odchylenia standardowe X SD i Y SD, a także współczynnik korelacji Pearsona: Corr [48]

Na początkowym etapie badania danych często dokonuje się wstępnej oceny zależności na drodze sporządzenia diagramów korelacyjnych. W latach 70. XX wieku wykazano, że różne zbiory danych mogą mieć identyczne parametry statystyczne [51]. Współcześnie, pojawiają się kolejne przykłady takich zbiorów danych. Przykładowy zestaw opublikowany w roku 2017 pokazany został na rysunku 20 [48]. Stworzyło to nowe podejście do graficznej oceny danych, decydujące często o doborze narzędzi analitycznych oraz procedur badawczych.

Sprawdzenie kompletności danych może być realizowane na różne sposoby, zależne od struktury zbioru danych. Jedną z takich możliwości jest – przykładowo – wykrywanie wartości *NULL* w kolumnach tabel [49]. W przypadku szeregów czasowych sprawdzenie kompletności danych wykorzystuje dodatkowo prosty iteracyjny algorytm obliczający interwały czasowe  $\Delta t$  pomiędzy kolejnymi sąsiadującymi punktami. Wychwycenie odstępstwa od schematu określającego kolejne terminy akwizycji danych (harmonogramu) jest sygnałem traktowanym jako wskaźnik braku kompletności danych w badanym szeregu czasowym.

Na rysunku 21 pokazana została ogólna zasada obowiązująca w trakcie obliczeń. Badane są wartości pewnej zmiennej  $X$ . Oś odciętych na wykresie jest osią czasu. Pokazane różnice czasowe  $\Delta t$  pomiędzy punktami danych są wyliczane dla całego szeregu.



**Rysunek 21.** Zasada badania kompletności danych szeregu czasowego o równomiernych odstępach

Warunek jaki spełniają kompletne dane, w których zachowana jest zasada równomierności odstępów czasowych, przedstawia poniższa zależność:

$$\text{Max}(\Delta t_1; \Delta t_n) = \text{Min}(\Delta t_1; \Delta t_n) \quad (12)$$

W powyższym równaniu  $(\Delta t_1; \Delta t_n)$  oznacza przedział zawierający wszystkie wartości różnic czasowych. Z kolei, dla szeregu czasowego zawierającego dane o nierównomiernych odstępach czasowych pomiędzy poszczególnymi punktami, wspomniany warunek ma postać następującą:

$$\text{Max}(\Delta t_1; \Delta t_n) \geq k_{min} \text{Min}(\Delta t_1; \Delta t_n) \quad (13)$$

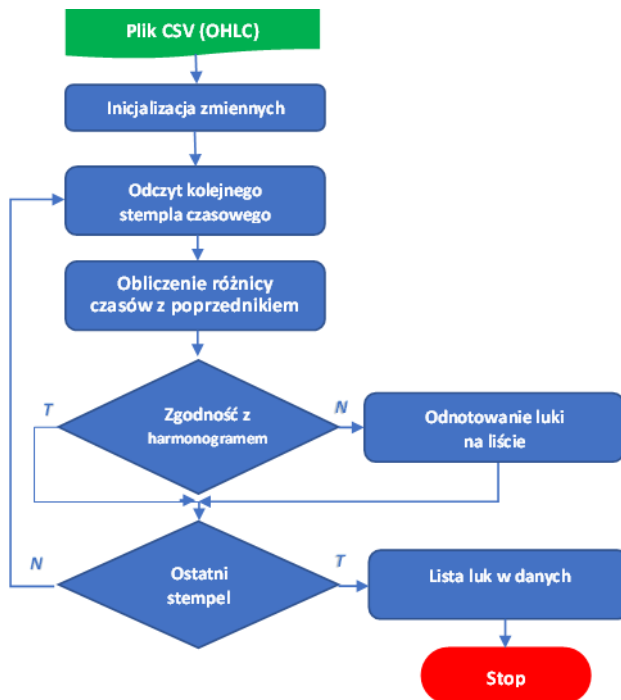
Tutaj  $(\Delta t_1; \Delta t_n)$  oznacza przedział zawierający wszystkie wartości różnic czasowych, zaś  $k_{min}$  jest współczynnikiem wielokrotności wynikający z harmonogramu gromadzenia danych.

Ważne dla szeregów czasowych jest również to, by w danych znajdowały się kolejne uporządkowane chronologicznie punkty należące do jednej ramy czasowej. Oznacza to, że interwały czasowe  $\Delta t$  nie mogą przyjmować wartości pośrednich, które nie są wielokrotnością interwału bazowego obowiązującego dla przyjętej ramy czasowej [49]. W przypadku, gdy zbiór danych zawiera wartości z różnych ram czasowych, w celu identyfikacji dominującej ramy czasowej konieczne jest określenie liczności poszczególnych interwałów czasowych w tym zbiorze. Na tej podstawie można ocenić, które dane są nadmiarowe i kwalifikują się do usunięcia.

Zadaniem algorytmu sprawdzania kompletności danych przedstawionego na rysunku 22 jest sprawdzenie zgodności obliczonych różnic czasowych z zadaniem harmonogramem. Wykorzystywany w tym celu harmonogram może mieć postać zdefiniowaną w tabeli 1 (przykład dla rynku Forex).

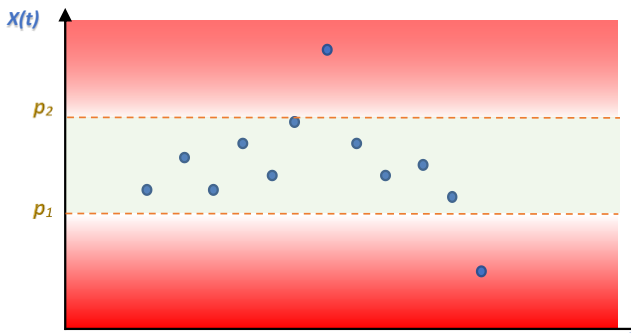
**Tabela 1.** Zestawienie harmonogramów notowań na rynku Forex dla typowych ram czasowych [53]

| L.p. | Interwał | Początek notowań |       | Początek ostatniej świecy |       |
|------|----------|------------------|-------|---------------------------|-------|
|      |          | Dzień            | Czas  | Dzień                     | Czas  |
| 1    | $D_1$    | Niedziela        | 23:00 | Czwartek                  | 23:00 |
| 2    | $h_{12}$ | Niedziela        | 23:00 | Piątek                    | 11:00 |
| 3    | $h_4$    | Niedziela        | 23:00 | Piątek                    | 19:00 |
| 4    | $h_1$    | Niedziela        | 23:00 | Piątek                    | 22:00 |
| 5    | $m_{15}$ | Niedziela        | 23:00 | Piątek                    | 22:45 |
| 6    | $m_5$    | Niedziela        | 23:00 | Piątek                    | 22:55 |
| 7    | $m_1$    | Niedziela        | 23:00 | Piątek                    | 22:59 |



**Rysunek 22.** Schemat uproszczonej sieci działań algorytmu do badania kompletności danych w szeregu czasowym

Kolejne sprawdzenie dotyczące tzw. wartości odstających polega na ocenie, czy wartość w konkretnym punkcie danych rażąco odbiega od zdefiniowanej wcześniej i przyjętej normy. W najprostszym przypadku sprowadza się to do arbitralnego wyznaczenia wartości progowych, co ilustruje rysunek 23.



**Rysunek 23.** Zasada wyznaczania wartości odstających poprzez arbitralne ustalenie wartości progowych  $p_1$  oraz  $p_2$



W tym przypadku wartości progowe tworzą przedział poza którym traktowane są jako odstające.

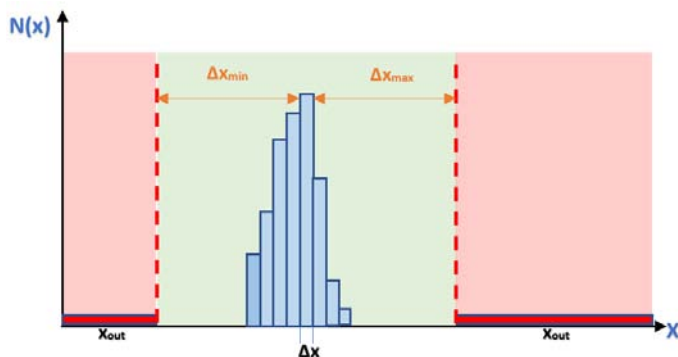
Innym stosowanym sposobem wyznaczenia granic jest podejście statystyczne. Polega ono na obliczeniu kwartyli  $Q_1$  oraz  $Q_3$  dla wartości z szeregu a następnie na ich podstawie wyznaczeniu rozstępu kwartylowego  $R_Q = Q_3 - Q_1$  i ostatecznie określeniu szukanych granic przedziału  $[Q_1 - 3R_Q; Q_3 + 3R_Q]$  zawierającego wyłącznie wartości nieodstające [52].

Spotykane jest również podejście wykorzystujące średnie kroczące. W tym przypadku dąży się do wyliczenia wstęgi zmienności, której wartości tworzą pewne krzywe. Wyznaczają one nieprzekraczalny margines wartości minimalnej i maksymalnej dla danych. Przekroczenie tego marginesu interpretowane jest jako wystąpienie wartości odstającej.

Dla danych wielowymiarowych obliczenia przeprowadzane są indywidualnie dla każdego z wymiarów. Poszukiwanie sprowadza się do sprawdzenia czy konkretny punkt mieści się w przyjętej wstędze odchyień, użytej do klasyfikacji danych. Przekroczenie odchylenia o zadany współczynnik jest sygnalizowane jako wartość odstająca.

Kolejną metodą pozwalającą wyodrębnić dane odstające, jest analiza rozkładu wartości, zilustrowana na rysunku 24. Rozkład taki otrzymuje się na drodze zliczania wystąpień poszczególnych wartości występujących w danych. Wartości zliczane są na podstawie ich przynależności do jednego z przedziałów  $\Delta x$ . W tym przypadku wartości odstające  $x_{out}$  zlokalizowane są w odległości większej niż graniczne wyznaczone przez  $\Delta x_{min}$  i  $\Delta x_{max}$ . Odległości te, jako kryterium kwalifikacyjne dla wartości odstających  $x_{out}$ , mogą być przyjęte arbitralnie na podstawie oceny rozkładu lub też wyznaczone jako wielokrotność odchyień standardowych w rozkładzie. Zakresy wartości sklasyfikowanych jako odstające zaznaczono na wykresie z kolorem czerwonym.

W statystyce opracowano różne metody wykrywania wartości odstających, które mogą być stosowane dla wielu rodzajów danych. Wyróżnia się w szczególności metodę dźwigni, badanie odległości Cooka [54] oraz wykresy residuów i pierwiastków residuów [55].



**Rysunek 24.** Zasada wyznaczania wartości odstających  $x_{out}$  poprzez wyznaczenie parametrów statystycznych rozkładu danych

### 3.2. Standaryzacja i normalizacja danych

Oprócz pokazanych wyżej metod służących wstępnej ocenie danych, wymienić należy również metody przetwarzania wykorzystywane na etapie przygotowawczym. W razie zaistnienia takiej konieczności zwykle stosowana jest standaryzacja lub normalizacja, jako sposób unifikacji danych. Ma to szczególne znaczenie, gdy wartości liczbowe dla poszczególnych wymiarów znacząco różnią się pomiędzy sobą. W niektórych sytuacjach dane mogą się różnić o całe rzędy wielkości.

Operacja standaryzacji to rodzaj ujednolicenia zmiennej losowej  $x$ , w wyniku działania której zmienna ta przyjmuje średnią wartość oczekiwaną równą zero i odchylenie standardowe wynoszące jeden [56,57]. Przeprowadzenie tej operacji na danych pozwala uzyskać dopasowanie wartości do wymagań różnych algorytmów [50]. W tym przypadku obliczenia standaryzujące wykonywane są według zależności:

$$z = \frac{x - \mu}{\sigma} \quad (14)$$

gdzie:

- $z$  – wartość standaryzowana,
- $x$  – zmienna niestandaryzowana,
- $\mu$  – średnia populacji,
- $\sigma$  – odchylenie standardowe.

Odchylenie standardowe  $\sigma$  obliczane jest najczęściej jako odchylenie standardowe z próby  $SD$ :

$$SD = \sqrt{\frac{\sum_{i=1}^N (x - \bar{x})^2}{N - 1}} \quad (15)$$

W powyższym wzorze przyjęto następujące oznaczenia:

$SD$  – odchylenie standardowe z próby,

$x$  – zmienna niestandardyzowana,

$\bar{x}$  – wartość średnia z próby,

$N$  – licznosc próby.

Wartość średnią  $\mu$  rozumianą jako średnią arytmetyczną  $\bar{x}$  definiuje równanie:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} \quad (16)$$

gdzie:

$\bar{x}$  – wartość średniej arytmetycznej z próby;

$x$  – zmienna niestandardyzowana;

$N$  – licznosc próby.

Drugim sposobem stosowanym do ujednoczenia danych we wstępnej obróbce jest normalizacja. W ogólnym przypadku, ma ona na celu sprowadzenie wartości danych liczbowych zmiennej  $x$  do przedziału  $[a, b]$ , który umożliwi dalsze jednolite obliczenia [58]. Zastosowanie do obliczeń formuły określonej wzorem (17), pozwala na znormalizowanie danych zgodnie z wymogami przyjętej procedury przetwarzania:

$$x' = a + \frac{(x - x_{min})(b - a)}{(x_{max} - x_{min})} \quad (17)$$

We wzorze (17) przyjęto następujące oznaczenia:

$x'$  – zmienna po normalizacji,

$x$  – zmienna przed normalizacją,

$a, b$  – granice przedziału obejmującego wartości  $x'$  (po znormalizowaniu),

$x_{min}$  – wartość minimalna zmiennej przed normalizacją,

$x_{max}$  – wartość maksymalna zmiennej przed normalizacją.

Wybór sposobu ujednoczenia danych każdorazowo podyktowany jest wymogami jakie narzucają kolejne etapy przetwarzania. Ważne jest tu dokonanie spójnego i logicznego wyboru dotyczącego wszystkich przetwarzanych zmiennych.

### 3.3. Regresja

Ogólna definicja podaje, że regresja polega na rozwiązaniu statystycznego problemu estymacji warunkowej wartości oczekiwanej [59,60]. Oznacza to dopasowanie zbioru parametrów zadanej formuły matematycznej na podstawie danych wejściowych, tzw. modelu regresyjnego, tak aby w wyniku uzyskać formułę (model) jak najdokładniej opisującą dane wejściowe. Metoda ta pozwala na zbadanie związku istniejącego pomiędzy zmiennymi objaśniającymi, a zmienną zależną (objaśnianą) oraz na wykorzystanie zdobytej w ten sposób wiedzy do przewidywania nieznanymi wartości jednych wielkości na podstawie innych wielkości zadanych.

Regresja jest narzędziem statystycznym szeroko wykorzystywanym do rozwiązywania zadań predykcji, jak również stanowi jeden z elementów w uczeniu maszynowym. Historia regresji sięga początków XIX wieku, kiedy to matematycy Adrien-Marie Legendre i Carl Friedrich Gauss opublikowali swoje prace dotyczące metody najmniejszych kwadratów (MNK) [61]. Historia samego terminu „regresja”, ocenianego przez niektórych badaczy jako niefortunne, wywodzi się z badań biologicznych z XIX w., w których badane zjawisko wykazywało tendencję do powrotu do średniej (trendu). Tendencja ta określana jako „regresja w kierunku średniej” przeniosła się na inne dziedziny badań w tym statystykę [61].

Współcześnie regresja jest obiektem wielu badań, pomimo swojej ponad dwustuletniej historii. Świadczy o tym choćby nagroda Nobla w dziedzinie ekonomii za badania nad analizą regresji w obecności heteroskedastyczności przyznana w roku 2003 [62]. Regresja w zależności od wariantu posługuje się właściwym dla siebie modelem. Jego ogólna postać zapisywana jest zwykle równaniem:

$$Y = f(X, \beta) + \varepsilon \quad (18)$$

gdzie:

- $Y$  – zmienna objaśniana,
- $X$  – wektor zmiennych objaśniających (predyktorów),
- $\beta$  – wektor współczynników regresji,
- $\varepsilon$  – błąd losowy, o rozkładzie być może zależnym od  $X$ .

Pojęcie regresji obejmuje wiele form [60], wśród których te najbardziej znane, to:

- regresja liniowa [63];
- regresja procentowa;

- regresja liniowa z przekształceniami;
- regresja kwantylowa;
- regresja nieliniowa;
- regresja logistyczna;
- regresja krokowa;
- regresja odporna [64,65];
- regresja nieparametryczna.

Regresja liniowa utożsamiana jest często z metodą najmniejszych kwadratów. Założeniem w tej metodzie jest istnienie liniowej kombinacji zmiennych objaśniających, która w wyniku tworzy wartości zmiennej objaśnianej. Dla  $k$  zmiennych objaśniających równanie (18) staje się modelem tzw. wielokrotnej regresji liniowej [66]:

$$\hat{Y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + \varepsilon \quad (19)$$

W powyższym modelu przyjęto następujące oznaczenia:

- $\hat{Y}$  – zmienna objaśniana (zależna),
- $x_1 \dots x_k$  – wektory zmiennych objaśniających (zmiennie niezależne),
- $b_0 \dots b_k$  – współczynniki modelu,
- $\varepsilon$  – błąd losowy.

Modelowi (19) odpowiada układ równań, w którym poszczególne punkty pomiarowe  $x$  tworzą wektory danych wartości wejściowych:

$$\begin{aligned} \hat{y}_1 &= b_0 + b_1x_{11} + \dots + b_kx_{1k} \\ \hat{y}_2 &= b_0 + b_1x_{21} + \dots + b_kx_{2k} \\ &\vdots \\ \hat{y}_n &= b_0 + b_1x_{n1} + \dots + b_kx_{nk} \end{aligned} \quad (20)$$

W powyższym układzie równań przyjęto oznaczenia:

- $y_1 \dots y_n$  – wartości wektora zmiennej objaśnianej (zmienna zależna),
- $x_{i1} \dots x_{ik}$  – wektory wartości objaśniających (zmiennie niezależne od  $x_1$  do  $x_k$  dla  $i$ -tej próbki),
- $b_0 \dots b_k$  – współczynniki modelu.

Układ równań (20) można przedstawić w postaci macierzowej:

$$\hat{y} = Xb \quad (21)$$

gdzie:

$b = [b_0, b_1, \dots, b_k]^T$  – wektor parametrów modelu,

$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}$  – macierz wartości objaśniających,

$\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]^T$  – wektor wartości objaśnianych (model).

Za miarę błędu modelu przyjmuje się wielkość  $Q$ , która jest kwadratem normy euklidesowej wektora błędów:

$$Q = \sum_i (y_i - \hat{y}_i)^2 \quad (22)$$

Minimalizując tę sumę kwadratów można uzyskać układ równań normalnych:

$$X^T X b = X^T y \quad (23)$$

Rozwiązanie tego układu wyznacza poszukiwane parametry modelu (19). Do wyznaczenia wektora parametrów  $b$  modelu prowadzi formalne przekształcenie równania macierzowego (23) do następującej postaci:

$$b = (X^T X)^{-1} X^T y \quad (24)$$

W praktyce współczynników modelu (19) nie znajduje się przez znajdowanie macierzy odwrotnej  $(X^T X)^{-1}$ , a raczej przez rozwiązanie układu równań normalnych (23) przy pomocy algorytmów numerycznych.

Różne formy regresji w większości są pochodną metody najmniejszych kwadratów (MNK), która jest modyfikowana w kierunku pokonania jakiejś zdefiniowanej trudności aplikacyjnej. Jedną z nich jest regresja procentowa wykorzystywana w ekonomii. Metoda ta minimalizuje sumę kwadratów błędów względnych modelu regresyjnego, odniesionych do wartości zmiennej wyjaśnianej:

$$Q = \sum_i \left( \frac{\hat{y}_i - y_i}{y_i} \right)^2 \quad (25)$$

Tak postawiony problem również jest rozwiązywany metodą najmniejszych kwadratów. Układ równań normalnych dla tego przypadku ma następującą postać [67]:

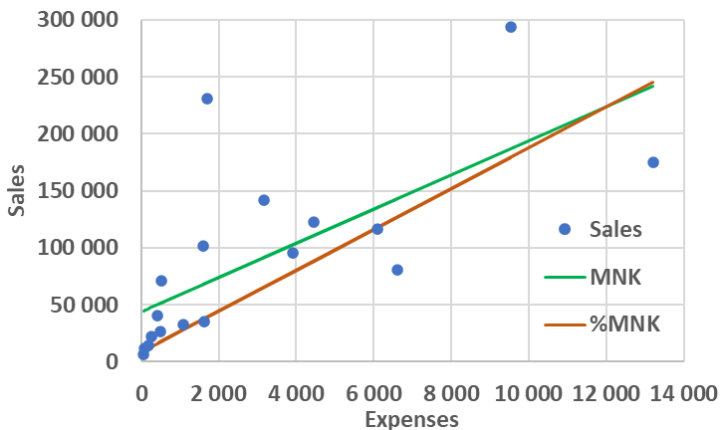
$$X^T D^2 X b = (DX)^T D y \quad (26)$$

Różnica pomiędzy równaniem (23), a równaniem (26) przejawia się w tym, że w układzie równań normalnych (26) pojawiła się kwadratowa diagonalna macierz  $D$ , zawierającą na przekątnej kolejne odwrotności wartości zmiennej objaśnianej ( $1/y_i$ ). Analogicznie do zależności (24), w regresji procentowej wektor  $b$  parametrów modelu (19), formalnie można przedstawić przy pomocy równania.

$$b = (X^T D^2 X)^{-1} X^T D^2 y \quad (27)$$

Tu również należy pamiętać, że w praktyce, parametry  $b$  znajduje się nie przez wyliczanie macierzy  $(X^T D^2 X)^{-1}$ , a przez numeryczne rozwiązanie układu równań normalnych (26).

Na rys. 25 porównano model otrzymany klasyczną metodą najmniejszych kwadratów z modelem uzyskanym przy pomocy procentowej metody najmniejszych kwadratów. Obydwa modele znajdowano dla danych zaczerpniętych z pracy [67]. Na płaszczyźnie wykresu dane te są reprezentowane w postaci niebieskich punktów. Z kolei, klasyczny model uzyskany metodą najmniejszych kwadratów reprezentuje linia zielona, zaś model otrzymany procentową metodą najmniejszych kwadratów jest reprezentowany linią czerwoną.



**Rysunek 25.** Porównanie liniowych modeli  $\hat{y} = b_1x + b_0$  dla danych z pracy [67] (niebieskie punkty), otrzymanych klasyczną metodą najmniejszych kwadratów (zielona linia MNK:  $\hat{y} = 14,994x + 439542$ ,  $R^2 = 0,4783$ ) oraz procentową metodą najmniejszych kwadratów (czerwona linia %MNK:  $\hat{y} = 17,88x + 8817$ ,  $R^2 = 0,4783$ ) [68]

Warto jeszcze wspomnieć o sposobie oceny dokładności regresji. Możliwą miarą jest kwadrat współczynnika korelacji Pearsona  $R$  definiowany zależnością:

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (28)$$

gdzie:

$y_i$  – wartość szeregu czasowego, zmiennej zależnej,

$\bar{y}$  – średnia wartości szeregu czasowego,

$\hat{y}_i$  – wartość estymowana zmiennej objaśnianej.

W praktyce stosuje się kwadrat współczynnika Pearsona  $R^2$ , nazywany współczynnikiem dopasowania. Określa on jak duża część wariancji zmiennej objaśnianej została wyjaśniona przez model regresyjny [69]. Współczynnik ten może przyjmować wartości z przedziału  $[0, 1]$ , przy czym wartość bliska jedności świadczy o wysokim dopasowaniu modelu do danych.

Regresja liniowa z przekształceniami dopuszcza istnienie w modelu dowolnych przekształceń zmiennych pod warunkiem, że całość pozostaje kombinacją liniową. Poszczególne przekształcenia zmiennych mogą odzwierciedlać zależności nieliniowe. Przykład takiego modelu przedstawia równanie:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 \sqrt{X} + \beta_4 \log(X) + \varepsilon \quad (29)$$

gdzie:

$Y$  – zmienna objaśniana (zależna),

$X$  – wektory wartości objaśniających (zmienna niezależna),

$\beta_0 \dots \beta_4$  – współczynniki modelu,

$\varepsilon$  – błąd losowy.

Regresja kwantylowa zaproponowana w roku 1978, jest metodą statystyczną pokrewną regresji obliczanej metodą najmniejszych kwadratów [55]. Jest stosowana, gdy w wyniku obliczeń oczekuje się wartości mediany lub określonego kwantylu zmiennej objaśnianej. Regresja ta pozwala na uzyskanie modelu, który na podstawie wartości zmiennej objaśniającej określa centyle lub kwantyle zmiennej zależnej. Do zalet tej regresji zalicza się to, że nie są stawiane wymogi i założenia dotyczące rozkładu zmiennej wyjaśnianej oraz to, że jest odporna na wpływ wartości odstających. Obliczenia regresji kwantylowej warunkowane są jedynie ilościowym charakterem zmiennej wyjaśnianej oraz zmiennych objaśniających.



W przypadku nominalnych zmiennych objaśniających, zwanych też jakościowymi lub kategoriowymi, możliwe jest sztuczne utworzenie ich odpowiedników w wartościach liczbowych. Stosowane są wówczas różne metody kodowania danych nominalnych, które skutkują nadaniem rang liczbowych poszczególnym kategoriom. Wśród rodzajów rang wymieniane są:

- rangi proste;
- rangi wiązane;
- rangi ułamkowe;
- rangi procentowe [70];
- oraz rangi zespolone [71].

Rangi zespolone są przedmiotem aktualnych propozycji i badań. Oryginalne podejście wykorzystujące liczby zespolone do kodowania danych nominalnych opisane jest w pracy [71]. Pozwala ono na kodowanie wartości nominalnych za pomocą zespolonych rang. Moduły tych rang uwzględniają statystyczną siłę kodowanych wartości nominalnych. Rangi przypisane w ten sposób stwarzają możliwość dalszego przetwarzania, gdyż zakodowane dane można traktować jak kolejną liczbową zmienną niezależną.

Regresja nieliniowa reprezentuje modele, które nie sprowadzają się do kombinacji liniowej wyrazów. Modele nieliniowe buduje się na potrzeby indywidualnego przypadku i często są przybliżane na drodze obliczeń numerycznych. W niektórych przypadkach model nieliniowy można przekształcić do modelu liniowego wykorzystując logarytmowanie i/lub podstawienia zastępcze. Dotyczy to np. modeli wielomianowych, potęgowych (np. Coba-Douglasa) i wykładniczych. Postać przykładowego modelu nieliniowego przedstawia poniższa zależność [60].

$$\hat{y} = b_0 + \frac{b_1 X}{b_2 + X} + \varepsilon \quad (30)$$

gdzie:

- $\hat{y}$  – zmienna objaśniana (zależna),
- $X$  – wartości objaśniające (zmienna niezależna),
- $b_0 \cdots b_2$  – współczynniki modelu,
- $\varepsilon$  – błąd losowy.

Regresja logistyczna wykorzystywana jest w przypadkach, gdy spodziewane wartości  $\hat{y}$  określają prawdopodobieństwo zajścia zjawiska lub przynależności do klasy.

Pożądane wartości przyjmowane przez zmienną  $\hat{y}$  ograniczają się do wartości z przedziału  $[0, 1]$ . Regresja logistyczna reprezentowana jest modelem opisanym następującą zależnością:

$$\hat{y} = \frac{a}{1 + be^{-cx}} \quad (31)$$

W powyższym równaniu przyjęto oznaczenia:

$\hat{y}$  – zmienna objaśniana (zależna),

$x$  – wartości objaśniające (zmienna niezależna),

$a, b, c$  – współczynniki modelu (wartości większe od zera).

Wyznaczanie współczynników modelu (31) wymaga użycia metod numerycznych, gdyż występują trudności z analitycznym obliczaniem współczynnika  $b$  tego modelu.

Regresja krokowa jest procedurą wykorzystującą regresję, w której zmienne wyjaśniające dobierane są do zadanego modelu. Dla regresji krokowej postępującej, w kolejnych krokach wybiera się zmienne najsilniej skorelowane ze zmienną objaśnianą aż do wyczerpania listy zmiennych objaśniających lub do uzyskania braku istotności kolejnych zmiennych. W przypadku regresji krokowej wstecznej procedura rozpoczyna się od modelu obejmującego wszystkie możliwe zmienne objaśniające. Następnie na drodze eliminacji kolejnych zmiennych dąży się do uzyskania najlepszego modelu. Do oceny modelu w trakcie procedury dobierania zmiennych wyjaśniających zwykle stosuje się współczynnik informacyjny Akaike (AIC) lub bayesowski współczynnik informacyjny (BIC) [55].

Regresja odporna stosowana jest dla danych o znacznej liczbie wartości odstających. Jednym z wariantów takiej regresji jest tzw. metoda najmniejszych przyciętych kwadratów (LTS). Po wykonaniu estymacji MNK zliczane są residua a następnie odrzucane największe z nich i wykonywana jest kolejna estymacja z tak przyciętymi danymi. Wykonanie kolejnych kroków przebiega wg takiego samego schematu. Prowadzi się je aż do uzyskania zadawalającego modelu odpornego na wartości odstające [55].

Regresja nieparametryczna jest koncepcją alternatywną do regresji ze ściśle określonym modelem parametrycznym. Jest bardziej elastyczna w poszukiwaniu rozwiązań dzięki osłabieniu założeń wstępnych dotyczących parametryzacji funkcji estymowanej. Jest to jednak okupione bardziej złożonym matematycznym opisem modelu oraz trudniejszym do wyznaczenia przedziałem ufności dla wartości prognozowanej [55, 72].

### 3.4. Autoregresja

Autoregresję można traktować jako szczególny przypadek modelu regresji wielokrotnej, opisanej w poprzednim podpunkcie. Autoregresja (AR) odznacza się tym, że regresja obliczana jest z użyciem ważonej kombinacji starszych wartości (obliczonych wcześniej prognoz) tej samej zmiennej  $y$ . W procedurze obliczającej autoregresję parametrem  $p$  jest głębokość w historii, na jaką sięga procedura obliczeniowa, by przewidzieć kolejną wartość. W wielu szeregach czasowych zauważalna jest autokorelacja z kolejnymi wartościami opóźnionymi (autokorelacja seryjna), co świadczy o liniowym związku pomiędzy obserwacjami opóźnionymi. W procesie AR modelowana jest warunkowa wartość  $y_t$  jako funkcja przewidywanych wartości szeregu czasowego  $\hat{y}_{t-1}, \hat{y}_{t-2}, \dots, \hat{y}_{t-p}$ , wyznaczonych dla kolejnych chwil w przeszłości. Oś czasu jest w tym przypadku skalowana wartościami dyskretnymi o stałej różnicy  $\Delta t$  pomiędzy kolejnymi próbkami (danymi).

Analogicznie jak w przypadku modelu regresji wielokrotnej, przewidującej na podstawie liniowej kombinacji wartości cech i przypisanych im poszukiwanych wag, autoregresja przewiduje na podstawie ważonej kombinacji wartości wyznaczonych modelem autoregresji w kolejnych krokach z przeszłości determinowanej parametrem  $p$  [73].

Autoregresja stanowi podstawę do konstrukcji całej grupy modeli parametrycznych wykorzystywanych w analizie szeregów czasowych. Wśród nich wymienić można takie jak:

- ARIMA (ang. *Autoregressive Integrated Moving Average*) – model ten został zaproponowany przez George'a Boxa i Gwilyma Jenkinsa w roku 1976. Jest metodą statystyczną służącą do analizowania szeregów czasowych o liniowych zależnościach między danymi. Model łączy autoregresję (AR) ze średnią krocząca (MA), wykorzystując dodatkowo człon integracyjny (I). model jest tworzony przy założeniu, że składnik losowy mający średnią równą zero i stałą wariancję jest niezależny (nieskorelowany) [52];
- SARIMA (ang. *Seasonal Autoregressive Integrated Moving Average*) – model sezonowych autoregresyjnych zintegrowanych średnich kroczących uwzględnia sezonowość zawartą w szeregu czasowym [74];

- ARIMAX (ang. *Autoregressive Integrated Moving Average*) – model uzupełnia ARIMA o dane heterogeniczne, które są uwzględniane w predykcji [75];
- ARMA (ang. *Autoregressive Moving Average*) – model łączy autoregresję (AR) ze średnią krocząca (MA) [76];
- ARCH (ang. *Autoregressive Conditional Heteroscedastic*) – model zaproponowany przez Roberta Engle’a w roku 1982 (nagroda Nobla z ekonomii w roku 2003). Pozwala na opisywanie niejednorodności składnika losowego w czasie. Jest wykorzystywany w analizie finansowych szeregów czasowych, w których zmienność podlega okresowym fluktuacjom i grupowaniu [77];
- GARCH (ang. *Generalized Autoregressive Conditional Heteroscedastic*) – zaproponowany przez Tima Bollersleva w roku 1986 uogólniony model ARCH [78];
- EGARCH (ang. *Exponential GARCH*) – model zaproponował Nelson w roku 1991. Jest wykładniczym uogólnionym modelem GARCH [78];
- SWARCH (ang. *Switching ARCH*) – model zakłada współistnienie wielu modeli, pomiędzy którymi można dokonywać warunkowego przełączania, dobierając najlepszy model dla danego przypadku, odpowiadający aktualnej fazie, w jakiej znajduje się rynek [79];
- HARCH (ang. *Heterogenous Interval Autoregressive Conditional Heteroscedasticity*) – model uwzględnia różne interwały czasowe wykorzystywane do pomiaru wariancji warunkowej, co odpowiada zróżnicowaniu horyzontów czasowych poszczególnych grup inwestorów wpływających na heterogeniczność (niejednorodność) rynku. Model ten w założeniu odpowiada założeniu o tzw. fraktalności rynku [79];
- FIGARCH (ang. *Fractionally Integrated ARCH*) – model należy do odrębnej, młodszej gałęzi modeli ARCH. Zakłada się tu zróżnicowanie pamięci dla wyznaczonej wariancji krótko i długoterminowej, co różni go od typowego modelu GARCH [79];
- HYGARCH (ang. *Hypebolic GARCH*) – model będący uogólnieniem modelu FIGARCH [79].

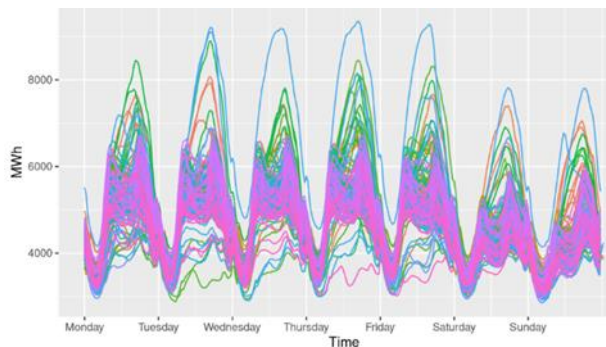
Modele te stosowane są zależnie od spodziewanych cech badanego szeregu czasowego i wymagają przeprowadzenia wstępnej oceny danych celem właściwego ich doboru.

### 3.5. *Usuwanie sezonowości i okresowości z danych*

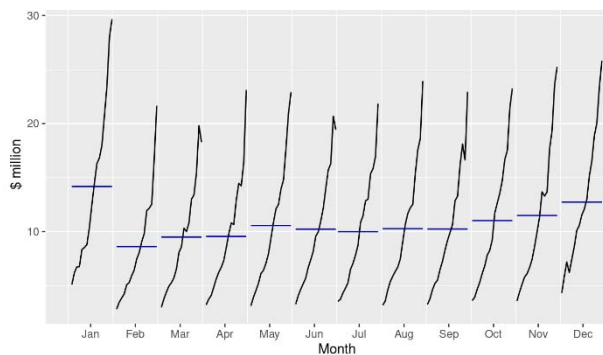
Kolejnymi zagadnieniami, z którymi zmagają się analitycy jest wyodrębnienie a często również usunięcie sezonowości i okresowości z danych reprezentowanych przez szereg czasowy. Pozwala to na lepszą ocenę trendu jaki reprezentują dane a także na określenie stopnia regularności w wyodrębnionych składowych mających znamiona cykliczności. W przypadku szeregów czasowych zawierających notowania towarów i surowców, można się spodziewać występowania tych składowych w znaczącym stopniu. Dokonanie ekstrakcji składowej sezonowej i/lub okresowej ułatwia predykcję notowań.

Wśród metod pozwalających na analizę okresowości czy sezonowości wyróżnić można te polegające na ocenie sporządzonego wykresu oraz te wykorzystujące aparat obliczeniowy umożliwiający dekompozycję szeregu czasowego na poszczególne składowe opisane zależnościami (2), (3). Metodami najczęściej wymienianymi są:

- Sporządzenie wykresu danych w funkcji czasu, umożliwiające wizualną ocenę ich charakteru [81];
- Sporządzenie wykresów sezonowych, w których nakłada się na jeden wykres dane z analogicznych sezonów wziętych z analizowanego okresu czasu (jak na rysunkach 26 i 27), np.: okres jednej dekady utworzy dziesięć rocznych wykresów naniesionych na ten sam układ współrzędnych, w którym oś czasu obejmuje dwanaście miesięcy [82];
- Sporządzanie złożenia częściowych wykresów sezonowych (alternatywa dla wykresów sezonowych), w których dla każdego sezonu tworzony jest wykres cząstkowy (jak na rys. 27) z analogicznych danych i umieszczany na osi czasu, np.: dla dekady utworzonych jest dwanaście wykresów cząstkowych odpowiadających kolejnym miesiącom, a w obrębie każdego miesiąca zebrane są dane z każdego roku analizowanej dekady [83];
- Estymacja gęstości spektralnej pozwalająca na uzyskanie składowych częstotliwościowych (analiza harmoniczna), zawartych w szeregu czasowym, wyznaczonych za pomocą dyskretnej transformaty Fouriera [84];



**Rysunek 26.** Wykres przedstawiający tygodniowe wzorce sezonowe dla zapotrzebowania na energię elektryczną [80]



**Rysunek 27.** Sezonowy wykres podserii miesięcznej sprzedaży leków przeciwcukrzycowych w Australii [83]

- Autokorelacja postrzegana jako matematyczne narzędzie do znajdowania powtarzających się wzorców w tym sygnałów okresowych jest korelacją  $R_k$  sygnału  $y_T$  z jego opóźnioną kopią  $y_{T-k}$  w funkcji opóźnienia  $k$  [85], obliczaną wg zależności:

$$R_k = \frac{\sum_{T=k+1}^T (y_T - \bar{y})(y_{T-k} - \bar{y})}{\sum_{T=1}^T (y_T - \bar{y})^2} \quad (32)$$

W powyższym równaniu oznaczenia są następujące:

$R_k$  – współczynnik autokorelacji szeregu czasowego z jego wersją opóźnioną o  $k$  próbek,

$y_T$  – wartości szeregu czasowego,

$\bar{y}$  – wartość średnia szeregu czasowego,

$T$  – długość szeregu czasowego,

$k$  – parametr reprezentujący opóźnienie liczone od wartości końcowej szeregu czasowego dla poszczególnych współczynników korelacji  $R_1$  do  $R_k$ , gdzie  $k < T$ ;

- Metody średnich kroczących i ważonych średnich kroczących pozwalają na eliminowanie składnika sezonowego z danych, gdy liczba uśrednianych próbek odpowiada długości rozpatrywanego sezonu; ważne średnie kroczące dają jednak płynniejsze oszacowanie cyklu trendu [86];
- Metoda X11 opracowana przez *US Census Bureau* i *Statistics Canada* w celu dekompozycji i wyrównań statystycznych obejmujących dane miesięczne oraz kwartalne, dokonywanych przez urzędy statystyczne [87];
- Procedura seats (ang. *Seasonal extraction in ARIMA time series*) została opracowana w banku Hiszpanii i stosowana jest przez agencje rządowe na całym świecie, lecz tylko w odniesieniu do sezonowości miesięcznej i kwartalnej [87];
- Metoda STL (ang. *Seasonal and trend decomposition using LOESS*) opracowana przez zespół badaczy w roku 1990, uważana jest za wszechstronną i niezawodną metodę dekompozycji szeregów czasowych, gdyż radzi sobie z sezonowościami różnych typów (np. kwartalne, miesięczne, tygodniowe, dzienne, godzinowe); jest odporna na wartości odstające; analityk może kontrolować tempo zmian składnika sezonowego; możliwe jest przeprowadzenie dekompozycji z modelem multiplikatywnym (3) stosując pomocniczo transformację Boxa-Coxa [88,89].

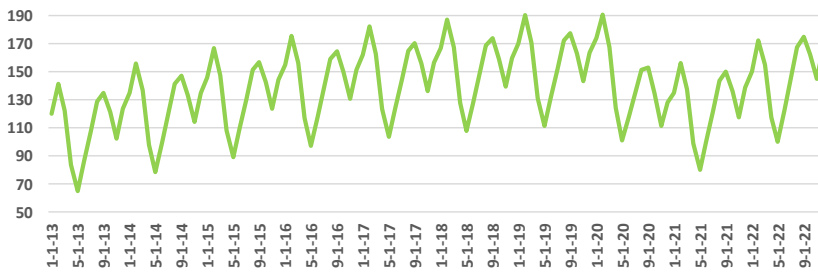
### 3.6. Określanie trendu reprezentowanego przez dane

Bardzo ważnym zagadnieniem w analizie szeregów czasowych jest określenie trendu, jaki występuje w danych. W ogólnym przypadku trend rozumiany jako tendencja wzrostowa lub spadkowa, można go uzyskać po usunięciu z danych składowych związanych z sezonowością i okresowością. Prawidłowe określenie poszukiwanego trendu umożliwia ocena składowej obejmującej zmiany losowe. Zakłada się, że wartość średnia składowej losowej wynosi zero, jeśli spełnia ona założenia dotyczące normalności swojego rozkładu. W ekonomii rozważa się również tzw. trend

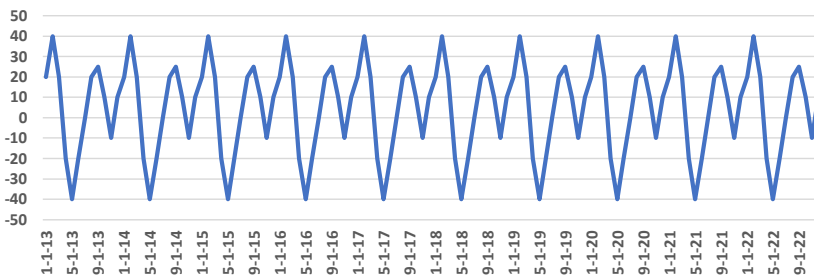
boczny, odpowiadający przypadkowi, kiedy nie obserwuje się wyraźnego trendu wzrostowego lub spadkowego.

Wyznaczanie trendu zawartego w szeregu czasowym metodami liczbowymi jest mocno związane z usuwaniem okresowości i sezonowości, opisanymi w podrozdziale 3.5. Zakłada się, bazując na modelach definiowanych równaniami (2) do (7), że dekompozycja szeregu czasowego na poszczególne składowe oraz usunięcie tak wyznaczonych wartości ze zmiennej objaśnianej pozwala na wyznaczenie składnika reprezentującego trend. Sposób przetwarzania danych zależy od przyjętego modelu.

Przykładowe wykresy ilustrujące proces wyznaczania trendu dla danych reprezentujących obroty przedsiębiorstwa handlowego pokazane zostały na rysunkach 28-31. Obraz szeregu czasowego z danymi, które stały się podstawą do przeprowadzenia obliczeń zawiera rys. 28. Wyodrębnioną składową związaną z sezonowością pokazuje rys. 29. Składowa losowa zawarta w danych przedstawiona została na rys. 30, a uzyskany obraz trendu ilustruje rys. 31.

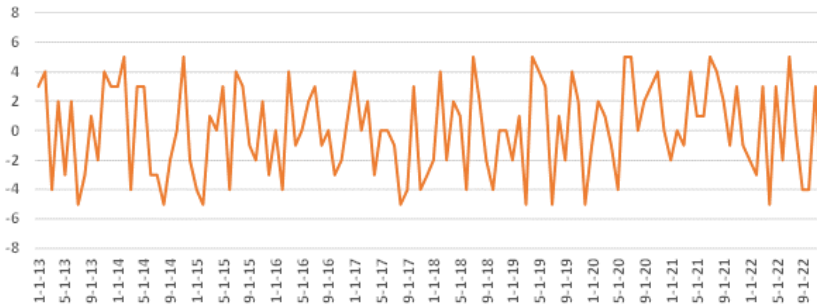


**Rysunek 28.** Wykres obrotów miesięcznych przedsiębiorstwa w funkcji czasu za okres 10 lat

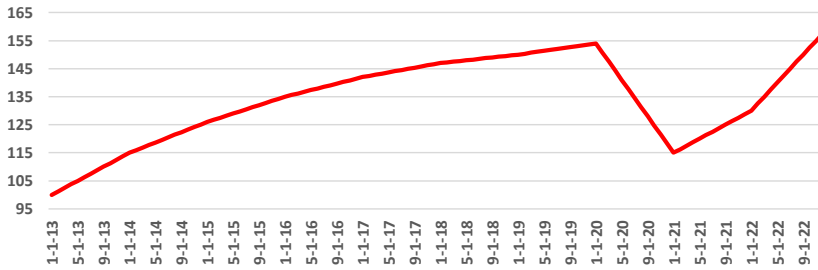


**Rysunek 29.** Wykres składowej sezonowej obrotów miesięcznych przedsiębiorstwa w funkcji czasu za okres 10 lat





**Rysunek 30.** Wykres składowej losowej obrotów miesięcznych przedsiębiorstwa w funkcji czasu za okres 10 lat



**Rysunek 31.** Wykres składowej trendowej obrotów miesięcznych przedsiębiorstwa w funkcji czasu za okres 10 lat

Procedura obliczeniowa konstruowana jest w zależności od charakteru danych jakie zawiera analizowany szereg czasowy oraz parametrów składowej losowej (szumu). Do wyeliminowania składowej sezonowej, w prostych przypadkach, stosowana jest często metoda średnich kroczących lub jej modyfikacja z wagami. Średnia ruchoma liczona jest jednak inaczej niż średnia krocząca opisana równaniem (9) w podrozdziale dotyczącym analizy technicznej. Różnica polega na symetrycznym indeksowaniu punktów danych względem punktu aktualnie obliczanego. Zatem uśredniana jest próbka środkowa i symetrycznie po  $k$  próbek z obu stron [86]:

$$\hat{Y}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j} \quad (33)$$

W równaniu (33) przyjęto następujące oznaczenia:

$\hat{Y}_t$  – wartość średniej ruchomej szeregu czasowego dla próbki  $t$ ,

$y_{t+j}$  – wartości z szeregu czasowego o indeksie  $t+j$ ,

$m$  – rząd średniej ruchomej  $m=2k+1$ ,

$j$  – indeks bieżący dla sumowania,

$k$  – parametr reprezentujący granice przesunięcia indeksu od wartości bieżącej  $t$ .

Analogicznie prowadzone są obliczenia w przypadku średniej ruchomej ważonej. Tu również liczba sumowanych punktów szeregu czasowego jest nieparzysta. Istotne w tym przypadku jest to, by suma wag wynosiła jeden a ich wartości o symetrycznych indeksach względem próbki środkowej były jednakowe.

Jednak nie wszystkie szeregi czasowe mają parametry statystyczne pozwalające na użycie tak prostych środków. Stąd w wielu sytuacjach istnieje konieczność zastosowania modeli regresyjnych bądź autoregresyjnych wymienionych w punktach 3.3 oraz 3.4. Wiele z tych metod jest jeszcze przedmiotem aktualnych badań i udoskonalień, o których donosi literatura przedmiotu. W szczególności dotyczy to niestacjonarnych szeregów czasowych oraz procesów, w których czynnik losowy odgrywa dużą rolę a jego parametry statystyczne odbiegają znacząco od założeń przyjmowanych w dotychczas stosowanych modelach.

Dodatkowo należy tu wspomnieć o metodach wyznaczania trendu stosowanych w analizie technicznej, a nie mające znamion obróbki statystycznej. Wynikają one z różnych podejść i teorii jakie powstały w czasach, gdy jeszcze nie było do dyspozycji komputerów (przełom XIX i XX wieku). Możliwość zastosowania takich procedur i algorytmów we wstępnej obróbce danych, może również dać zadowalające rezultaty, bez angażowania dużych zasobów obliczeniowych. Metody te opisane zostały w punkcie 2.3, dotyczącym analizy technicznej.

### **3.7. Klasyfikacja**

Metody klasyfikacji są kolejną grupą narzędzi analityka danych. Mogą być one wykorzystywane jako narzędzie wspomagające w analizie danych jakimi są szeregi czasowe. W metodach tych określana jest tzw. bliskość rozumiana jako stopień podobieństwa lub niepodobieństwa pomiędzy poszczególnymi punktami zmiennej

objaśnianej. Przy założeniu pewnego stopnia współmierności zmiennych objaśniających stosuje się dobór odpowiedniej metryki, która jest miarą niepodobieństwa. Wybór metryki jest kluczowym zagadnieniem rozstrzygającym o rezultacie stosowanej metody. W wielu przypadkach stosowana jest odległość Euklidesowa lub ważona odległość Euklidesowa, a w niektórych tzw. odległość Manhattan, odległość kątowna czy odległość (metryka) Minkowskiego [50]. Przypadki danych niewspółmiernych wymagają zastosowania normalizacji. W grupie metod klasyfikacji wymieniane są:

- Algorytm k-najbliższych sąsiadów (k-nn), pozwalający na określenie prawdopodobieństwa, z jakim konkretny punkt można zakwalifikować do pewnej grupy (klasy) [90];
- Maszyna wektorów nośnych (ang. *Support vector machines*, SVM), która należy do metod uczenia nadzorowanego; pozwala na przeprowadzenie klasyfikacji zarówno liniowej jak i nieliniowej; SVM wyznacza hiperpłaszczyznę rozdzielającą grupy danych, co wykorzystuje się w klasyfikacji [91];
- Naiwny klasyfikator Bayesa, reprezentujący grupę klasyfikatorów probabilistycznych, bazujących na twierdzeniu Bayesa, z wykorzystaniem dodatkowo silnego (naiwnego) założenia, dotyczącego niezależności pomiędzy zmiennymi niezależnymi [92];
- Drzewa decyzyjne, które dokonują cyklicznego podziału klasyfikacyjnego wektorów danych na podstawie szeregu warunków logicznych, sprawdzanych w kolejnych krokach algorytmu [90];
- Lasy decyzyjne, które w trakcie uczenia konstruują wiele drzew decyzyjnych, a w fazie roboczej jako wynik obliczeń określają klasę dominującą, wybieraną „przez głosowanie” – należą one do tzw. metod zespołowych i są jedną ze skuteczniejszych metod klasyfikacji [49].

Do klasyfikacji wykorzystywane są również sieci neuronowe, będące strukturami złożonymi ze sztucznych neuronów, przetwarzające dane wejściowe w kolejnych warstwach, dające na wyjściu prawdopodobieństwo przynależności punktu danych (punktu pomiarowego) do predefiniowanej klasy. Sieci neuronowe zostaną omówione w dalszej części niniejszego artykułu.

### 3.8. Grupowanie

Grupowanie (ang. *Clustering*) jest zadaniem polegającym na podzieleniu danych wejściowych (objaśniających) według kryterium podobieństwa. Istotnym zagadnieniem jest również wybór metryki określającej podobieństwo. W trakcie grupowania maksymalizowane jest podobieństwo rekordów wewnątrz grupy a minimalizowane podobieństwo pomiędzy poszczególnymi grupami. Do metod grupowania zaliczane są m.in.:

- Algorytm k-średnich, będący prostym i efektywnym sposobem znajdowania grup w zbiorze danych wymaga podania a priori liczby grup w danych; obliczenia wykonywane są iteracyjnie aż do momentu osiągnięcia takiego podziału, który minimalizuje odległości poszczególnych punktów od centroidów w obrębie poszczególnych grup i dalsze zmiany w doborze centroidów dla tych grup nie prowadzą do poprawy podziałów [49];
- Grupowanie hierarchiczne, które tworzy strukturę drzewiastą na drodze rekurencyjnego dzielenia (wariant rozdzielający) lub w drugim wariancie – rekurencyjnego łączenia elementów zbioru. Jest on wtedy nazywany aglomeracyjnym i najczęściej implementowany bywa w pakietach oprogramowania do obliczeń statystycznych [49].

Grupowanie wykorzystuje się często jako krok wstępny w procesie eksploracji danych. Grupy utworzone w wyniku takich obliczeń użyte zostają jako dane wejściowe do kolejnego kroku przetwarzania, realizowanego inną techniką, np. siecią neuronową czy algorytmem klasyfikacji. W analizie szeregów czasowych grupowanie ma znaczenie pomocnicze ze względu na cele jakie stawiane są przed przetwarzaniem takich danych.

### 3.9. Rozpoznawanie wzorców

Rozpoznawanie wzorców stanowi dużą pomoc w sytuacjach analizy wielkich zbiorów danych, kiedy poszukiwane są z góry określone informacje odpowiadające zadanym wzorcom. W przypadku szeregów czasowych ma to kluczowe znaczenie. W rozpoznawaniu wzorców dąży się do sklasyfikowania obserwacji i określenia jej pozycji w wielowymiarowej przestrzeni cech tworzonej przez kategorie danych wejściowych. Dziedziny takie jak rozpoznawanie mowy, rozpoznawanie obrazów (w tym

identyfikacja twarzy czy pisma ręcznego), biometria oraz identyfikacja określonych ciągów w danych np. giełdowych czy pochodzących z serwisów społecznościowych, należą do szeroko znanych obszarów zastosowań technik rozpoznawania wzorców.

Dane wejściowe trafiające do przetwarzania wymagają jednak wstępnej obróbki ze względu na ich „objętość”. Najczęstszym wymogiem jest kompresja danych, mająca na celu nie tylko zwykłą redukcję rozmiaru plików, ale też redukcję liczby wymiarów (zmiennych niezależnych), których udział w wyjaśnianiu zmiennej zależnej (objaśnianej) jest mały i może być pominięty.

W takich sytuacjach stosowana jest analiza głównych składowych (ang. *Principal Component Analysis*, PCA). Algorytm PCA może być stosowany w wariancie wykorzystującym kowariancję, w którym na wynik najbardziej wpływa zmienna o największej wariancji. Cecha ta może być pożądana przy analizie porównywalnych danych, np. takich jak procentowe zmiany kursów giełdowych. Z kolei, PCA w wersji wykorzystującej macierz korelacji odnosi się do danych wstępnie wystandaryzowanych, by uzyskać na wejściu podobną wariancję dla wszystkich zmiennych, jak np. w przypadku analiz cen na rynkach towarowych, gdzie wartości liczbowe w poszczególnych kategoriach mogą mieć różne rzędy wielkości [93].

Jednym z często wymienianych narzędzi wspomagających rozpoznawanie wzorców, jest maszyna wektorów nośnych (SVM), zaproponowana w latach 60. XX wieku, a zmodyfikowana o możliwość tworzenia klasyfikacji nieliniowych w latach 90. [91]. Algorytm SVM pomocny jest w identyfikacji, klasyfikacji i klasteryzacji obrazów oraz kategoryzacji tekstów.

Odrębną grupą narzędzi pomocną w rozpoznawaniu wzorców są sieci neuronowe, np. sieci CNN rozpoznające obrazy. Zostaną one opisane szerzej w dalszej części niniejszego artykułu.

### **3.10. Odkrywanie reguł**

Odkrywanie reguł związane jest z wyszukiwaniem asocjacji występujących w zgromadzonych danych. Asocjacje są zależnościami lub korelacjami występującymi pomiędzy danymi zawartymi w dużych zbiorach informacji. Wynikiem procesu odkrywania asocjacji jest zbiór reguł asocjacyjnych, które opisują ukryte lub trudne do identyfikacji związki pomiędzy danymi. Prowadzi to do pokazania wzorców zachowań ludzi jako zbiorowości lub wskazania zdarzeń, które zazwyczaj są obserwowane

w czasie trwania jakiegoś procesu. Odnosić się to może do przewidywania zmian pogody, diagnostyki medycznej, bankowości, rynków kapitałowych jak również szeroko rozumianego marketingu z tworzeniem wskazówek do stymulowania wyników sprzedaży.

Modelowanie wykorzystywane w odkrywaniu reguł posługuje się tablicami obserwacji stworzonymi z atrybutów i obserwacji. Tablice takie pozwalają na modelowanie wielu sytuacji ze świata rzeczywistego. Powstałe reguły asocjacyjne cechuje ważność nazywana wsparciem (ang. *support*) i siła związana z ufnością (ang. *confidence*), które są miarami wyznaczanymi statystycznie. Ważność reguły jest utożsamiana z prawdopodobieństwem zajścia jakiegoś zdarzenia natomiast siła reguły związana jest z prawdopodobieństwem warunkowym jego zajścia.

Reguły asocjacyjne klasyfikuje się zwykle ze względu na: typ przetwarzanych danych, wymiarowość przetwarzania oraz stopień abstrakcji reprezentowanej przez dane. Typ danych implikuje reguły asocjacyjne binarne związane ze współwystępowaniem danych oraz reguły asocjacyjne ilościowe reprezentujące współwystępowanie wartości danych. Wymiarowość przetwarzania określa reguły jednowymiarowe, gdy dane należą do tej samej dziedziny oraz wielowymiarowe, gdy dziedziny są różne. W przypadku stopnia abstrakcji rozróżnia się jednopoziomowe i wielopoziomowe reguły asocjacyjne [94].

Algorytmy używane do odkrywania reguł są intensywnie analizowane, by poprawić ich wydajność. Algorytm naiwny operujący na zbiorze elementów  $A$  i bazie danych, polega na wyznaczeniu wszystkich podzbiorów zbioru  $A$  oraz obliczeniu wsparcia w bazie danych dla każdego z nich. W następnym kroku wyznaczane są binarne reguły asocjacyjne dla wygenerowanych wcześniej podzbiorów a następnie dla reguł o odpowiednio wysokiej wartości wsparcia obliczana jest ich ufność. Stosowanie tego algorytmu wiąże się z aprobatą wykładniczego wzrostu wielkości zadania obliczeniowego w odpowiedzi na liniowy wzrost liczby elementów zbioru  $A$ . Z punktu widzenia efektywności tego algorytmu, czynnikiem o największym znaczeniu, jest tworzenie zbiorów częstych ze zbioru  $A$  o zadanej minimalnej wartości wsparcia. Prace nad tym zagadnieniem trwają od roku 1993, gdy algorytm zaproponowano po raz pierwszy [94]. Istnieje już wiele opracowanych algorytmów służących do odkrywania reguł asocjacyjnych [95], przykładem mogą być: Apriori, Eclat, FP-Growth czy Partition. Algorytmy te zmniejszają efekt wykładniczego wzrostu

zadania obliczeniowego bazując na antymonotoniczności, która dotyczy obliczania wsparcia zbiorów częstych [95].

### **3.11. Predykcja i jej ocena**

We wstępie artykułu został zdefiniowany termin „predykcja”. Predykcja (inaczej przewidywanie) jest bardzo ważna w obszarze badania szeregów czasowych. Do predykcji wykorzystywane są bardzo różne algorytmy. Trudno tu mówić o jednym wyspecjalizowanym algorytmie czy nawet ich grupie. Metody predykcji mogą wykorzystywać różnorodne koncepty i teorie, a w szczególności podejście statystyczne wykorzystujące algorytmy o różnym stopniu złożoności. Jest to związane z występowaniem składnika losowego w modelach danych, sprawiającym, że dokonanie predykcji jest zadaniem złożonym i trudnym.

Jednak wraz z rozwojem metod przetwarzania informacji jakość predykcji, w sensie dokładności przewidywania przyszłych wartości szeregu czasowego, ulega stopniowej poprawie. Zarówno prognozowanie pogody, diagnostyka medyczna, rozpoznawanie mowy, rozpoznawanie obrazów i innych danych z niezerową losowością przeprowadzane jest z coraz większą precyzją. Opisane w podrozdziale 2.2 cztery różne podejścia do analizy szeregów czasowych (analiza fundamentalna, techniczna, ilościowa oraz sentymentu) podporządkowane są głównie wykonywaniu predykcji.

Najbardziej rozległym tematycznie wydaje się być wachlarz narzędzi uformowany przez analizę techniczną, która swoimi początkami sięga wieku XIX. Współcześnie analiza techniczna, jak i podejścia alternatywne, adaptowała wiele metod i narzędzi opracowanych w ostatnich latach. Odnotować również należy eksplozję zastosowań sieci neuronowych w tym zakresie. Zagadnienia związane z sieciami neuronowymi omówione zostaną w rozdziale 4 niniejszego artykułu, gdyż reprezentują odmienne podejście w stosunku do metod „klasycznych” związanych z przetwarzaniem informacji.

Opisane wcześniej metody i procedury postępowania reprezentują różny stopień właściwości predykcyjnych. Wśród nich tylko część, którą stanowią metody sprzed ery komputerów, jest niealgorytmiczna lub nie wymaga przeprowadzania żmudnych obliczeń arytmetycznych. Jednak metody wykorzystujące algorytmy i komputerowe

przetwarzanie danych stanowią przytłaczającą większość. Wymienić tu można te najbardziej popularne:

- średnie kroczące;
- metoda najmniejszych kwadratów;
- regresja;
- autoregresja;
- rozpoznawanie wzorców.

Predykcja szeregów czasowych wykorzystująca wymienione metody wymaga jednak przejścia czasem żmudnego procesu doboru parametrów definiujących dopasowywane modele. Pomimo zaawansowanego aparatu pojęciowego i obliczeniowego, uzyskiwane wyniki reprezentują tylko wartości z pewnego przedziału wyznaczonego jako przedział ufności prognozy. Dodatkowym elementem utrudniającym prognozowanie jest występująca w różnych przypadkach heteroskedastyczność badanego procesu, charakteryzująca się trudną do zamodelowania zmiennością wariancji obserwowaną w danych. W przypadku takich danych podejmowane są próby redukcji heteroskedastyczności, np. poprzez różnicowanie w szeregu czasowym. W założeniach wielu metod prognostycznych występuje wymaganie braku heteroskedastyczności. Stąd właśnie wynika konieczność jej redukcji.

Możliwość użycia większej grupy algorytmów otwiera dodatkowo możliwość skorzystania z zalet przetwarzania (uczenia) zespołowego (ang. *ensemble learning*), co w praktyce skutkuje poprawą jakości predykcji. Wynik końcowy jest uzyskiwany za pomocą agregacji wyników pojedynczych modeli, zazwyczaj metodą prostego „głosowania”. W niektórych przypadkach używane są również warianty głosowania uwzględniające wagi dla poszczególnych wyników cząstkowych z użyciem algorytmu większości ważonej (WMA) [96].

Uzyskiwane w ten sposób wyniki wykorzystywane są w różnych dziedzinach ludzkiej aktywności, np. w teledetekcji satelitarnej, wykrywaniu zmian w obrazach, rozpoznawaniu twarzy, bezpieczeństwie komputerowym, rozpoznawaniu emocji, wykrywaniu oszustw czy podejmowaniu decyzji finansowych [97].

Ostatni z wymienionych obszarów działań jest szczególnie związany z analizą szeregów czasowych, gdyż dotyczy przewidywania zarówno niepowodzeń biznesowych, jak i zbliżających się kryzysów finansowych lub gospodarczych. Spektakularne wyniki osiągnęte są w zakresie manipulacji cenami akcji na giełdach, gdzie



analiza zespołowa odkrywa próby wpływania nieuczciwych spekulantów na zmiany kursu za pomocą sztucznie kreowanego popytu lub podaży wybranego waloru [97].

Ważnym zagadnieniem związanym z predykcją jest ocena jakości otrzymanych przewidywań. Popularnym parametrem stosowanym do takiej oceny jest współczynnik determinacji (dopasowania)  $R^2$ . Obliczany jest jako kwadrat współczynnika korelacji  $R$  Pearsona podanej zależnością (28) i omówioną w punkcie 3.3.

Współczynnik determinacji nie jest jedynym sposobem oceny predykcji. Wśród estymatorów pomocnych w ocenie wymieniane są również:

- Analiza odchylenia standardowego składnika resztowego  $S$ , rozumianego jako przeciętne odchylenie będące różnicą pomiędzy wartością rzeczywistą  $y(t)$ , a odpowiadającą jej wartością teoretyczną  $\hat{y}(t)$  wyznaczoną za pomocą ocenianego modelu [98];
- Współczynnik zbieżności  $\varphi^2 = 1 - R^2$  pokazujący stopień niedopasowania modelu do wartości zmiennej zależnej [98];
- Współczynnik zmienności losowej [98];
- Accuracy – parametr określający dokładność, będący różnicą jedynki i straty obliczanej jako MSE, stosowany często w przypadku oceny wyników pracy sieci neuronowych [99];
- Odchylenie od charakterystyki monotonicznej (liniowej) [100];
- Weryfikacja hipotezy o istotności współczynnika korelacji wielorakiej  $r$  związanej z wartością krytyczną statystyki  $f$ , która wynika z rozkładu Fishera-Snedecora [98];
- Zestaw metryk do oceny strategii inwestycyjnej [101].

Tradycyjnym sposobem oceny wyników jest również wzrokowa analiza i porównywanie wykresów z naniesionymi wynikami predykcji. Stosowanie poszczególnych metod oceny związane jest ze specyfiką szeregu czasowego reprezentującego badany proces oraz celów jakim ma służyć modelowanie [101].

#### 4. Przegląd wybranych sieci neuronowych

Wstępny przegląd literatury przedmiotu pokazał, że temat sieci neuronowych a wraz z nim temat uczenia maszynowego pomimo swojej stosunkowo krótkiej historii jest bardzo obszerny [90,102]. Jest to powodem dokonania tylko ograniczonego przeglądu struktur sieci neuronowych na tle ogólnej ich taksonomii. Szerzej omówione

zostaną te z sieci, które mogą być wykorzystane do wykonywania predykcji w szeregach czasowych.

#### 4.1. *Rys historyczny*

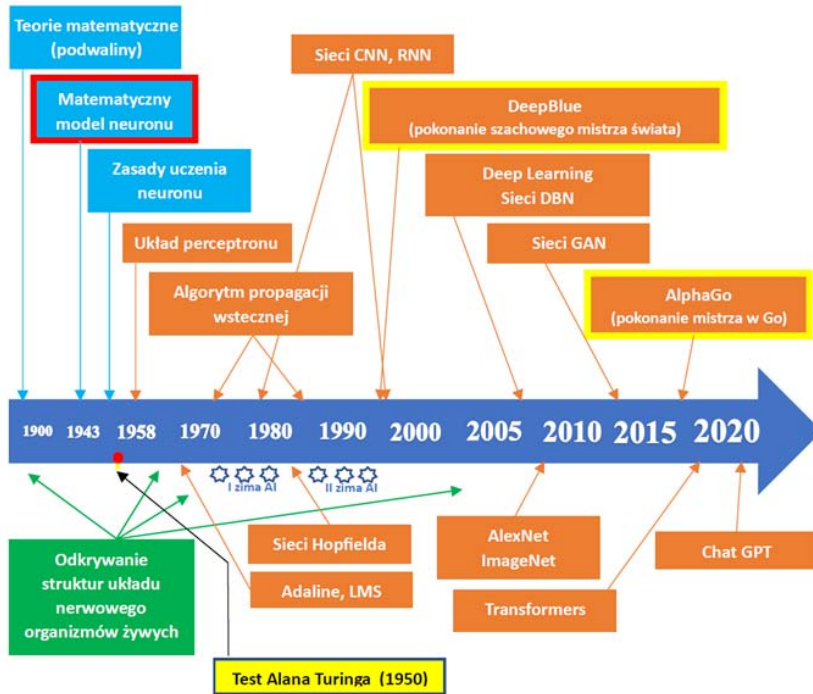
Historia sieci neuronowych rozpoczęła się od badań neurofizjologicznych związanych z działaniem układu nerwowego a w szczególności ludzkiego mózgu, prowadzonych na przełomie XIX i XX w. Świadczą o tym nagrody Nobla przyznane za kolejne odkrycia w wieku XX [102]. Badania te trwają również obecnie, przynosząc czasem zaskakujące informacje związane z budową bądź działaniem biologicznego układu nerwowego.

Wyniki badań neurofizjologów i biofizyków przeniknęły z pewnym opóźnieniem do środowisk inżynierskich, co zaowocowało zaproponowaniem matematycznego modelu neuronu w roku 1943 przez Warrena S. McCullocha i Waltera H. Pittsa. Niedługo później w roku 1949 Donald O. Hebb opublikował zasadę uczenia sztucznych sieci neuronowych. W ślad za tym Frank Rosenblatt w roku 1957 wysunął koncepcję i skonstruował sztuczną sieć neuronową nazwaną perceptronem. Był to pierwszy opis sieci neuronowej korzystającej z algorytmu [103]. Z biegiem czasu świat nauki donosił o kolejnych odkryciach związanych ze sztuczną inteligencją (AI) i sieciami neuronowymi w szczególności. Kolejne kroki milowe w tej dziedzinie umieszczone na osi czasu pokazane zostały na rys. 32, natomiast bardziej szczegółowe kalendarium zestawione zostało w tabeli 2.

Odnotować należy fakt, że jeszcze w XIX wieku podwaliny dla wielu koncepcji stosowanych w AI powstawały w głowach matematyków, którzy nie byli świadomi perspektyw i możliwych zastosowań swoich odkryć. Jak pokazuje historia, rozwój koncepcji sztucznych sieci neuronowych ANN (ang. *Artificial Neural Networks*) oraz sposobu ich uczenia (uczenie maszynowe – ang. *machine learning*, ML) przebiegał falami, gdyż okresowo zmieniało się nastawienie do tego fenomenu, zarówno w środowiskach naukowych, jak i w opinii publicznej.

Na osi czasu zaznaczone zostały gwiazdkami dwa okresy tzw. „zimy AI”, które zostały odnotowane na przestrzeni lat 1974-1980 oraz 1987-1993. Obserwowany był w tym czasie regres nakładów finansowych na badania nad sztuczną inteligencją. Pierwszy z nich wywołany został ukazaniem się w roku 1969 publikacji *Perceptrons: An Introduction to Computational Geometry* autorstwa Marviniego Minsky’ego

i Seymoura Paperta, którzy poddali krytyce perspektywy rozwoju deklarowane przez środowiska zajmujące się AI oraz ANN.



**Rysunek 32.** Historia rozwoju sztucznej inteligencji (AI) oraz sztucznych sieci neuronowych (ANN) z zaznaczonymi „kamieniami milowymi” wyznaczającymi rozwój dyscypliny [104,105]

Okresy te przyniosły pewne spowolnienie prac, ale nie doprowadziły do ich całkowitego wygaszenia. W rezultacie tej krytyki powstały nowe koncepcje rozwijające dotychczasowe trendy i stosowane architektury sieci. Już wkrótce po tym przestoju zaczęły się pojawiać spektakularne sukcesy. Jednym z nich było zwycięstwo AI w rozgrywce z szachowym mistrzem świata Garrim Kasparovem.

W kolejnych latach rosła liczba doniesień o znaczących postępach w rozwiązywaniu wielu problemów, do tej pory uważanych za nierozwiązywalne. Do najbardziej znaczących zaliczane są: rozpoznawanie twarzy, rozpoznawanie mowy, tłumaczenie maszynowe z języków takich jak chiński, japoński czy arabski, zwycięstwo w grze w Go oraz powstanie modelu Chat GPT prowadzącego swobodny dialog

z człowiekiem na niespotykanym dotąd poziomie, zarówno językowym jak i merytorycznym. W opinii wielu obserwatorów, model ten jest bardzo bliski pozytywnego przejścia testu Alana Turinga ogłoszonego w roku 1950.

**Tabela 2.** Chronologia rozwoju sieci neuronowych [103,105,106]

| Rok  | Opis osiągnięcia  |
|------|---|
| 1906 | Nagroda Nobla za badania nad strukturą układu nerwowego, odkrycie neuronów w mózgu (koniec XIX w.) – <i>Santiago Ramón y Cajal, Camillo Golgi</i> |
| 1943 | Model sztucznego neuronu zainicjował dyscyplinę i rozwój sieci neuronowych – <i>Warren Sturgis McCulloch, Walter Harry Pitts</i>                  |
| 1949 | Zasada uczenia sztucznych sieci neuronowych na poziomie synaptycznym – <i>Donald Olding Hebb</i>  |
| 1950 | Ogłoszenie testu dla sztucznej inteligencji – <i>Alan Turing</i>  |
| 1952 | Model dynamiczny neuronu (Nagroda Nobla w Fizjologii i Medycynie 1963) – <i>Alan Lloyd Hodgkin, Andrew Fielding Huxley</i>                        |
| 1957 | Powstanie perceptronu i algorytmu trenowania jednowarstwowej sieci neuronowej – <i>Frank Rosenblatt</i>   |
| 1960 | Adaline: adaptacyjna liniowa sieć neuronowa – <i>Bernard Widrow, Marcial Edward "Ted" Hoff</i>  |
| 1969 | Krytyka perceptronu z podaniem ograniczenia dla funkcji XOR – <i>Marvin Lee Minsky, Seymour Papert</i>  |
| 1972 | Liniowa pamięć asocjacyjna – <i>Teuvo Kohonen, James A. Anderson</i>  |
| 1974 | Model propagacji wstecznej do trenowania perceptronów wielowarstwowych – <i>Paul Werbos</i>   |
|      | Algorytm propagacji wstecznej z regułą łańcuchową – <i>Paul Werbos</i>  |
| 1980 | Koncepcja wielowarstwowej sieci neuronowej pozwalającej modelować bardziej złożone funkcje  |
|      | Powstanie architektury sieci spłotowych CNN (Neocognitron) – <i>Kunihiko Fukushima</i>  |
|      | Upadek perceptronu spowodowany jego ograniczeniami oraz powstaniem sieci wielowarstwowych   |
| 1982 | Algorytm propagacji wstecznej z regułą łańcuchową zaimplementowany na komputer – <i>Paul Werbos</i>   |
|      | Powstanie sieci Hopfielda jako zaczątku sieci rekurencyjnych RNN – <i>John Hopfield</i>   |
| 1984 | Wprowadzenie sieci Boltzmana jako stochastycznych sieci neuronowych – <i>Geoffrey Hinton, Terrence Sejnowski, David Ackley</i>                    |
| 1985 | Koncepcja głębokich sieci neuronowych pozwalająca na uczenie hierarchiczne  |

Wybrane zagadnienia analizy szeregów czasowych

| Rok  | Opis osiągnięcia  |
|------|---|
| 1986 | Formalna propozycja algorytmu propagacji wstecznej dla sieci ANN – <i>David Rumelhart, Geoffrey Hinton, Ronald J. Williams</i>                |
|      | Implementacja algorytmu propagacji wstecznej dla wielowarstwowych sieci neuronowych jako metody standardowej                                  |
| 1987 | Sieci impulsowe – <i>Wolfgang Maass</i>   |
|      | NetTalk – sieć neuronowa uczy się mowy na zbiorze danych zamiany tekstu na mowę   |
| 1989 | CNN używa do nauki algorytmu propagacji wstecznej osiągając lepszą zdolność rozpoznawania obrazów   |
|      | Udowodniono twierdzenie o aproksymatorach uniwersalnych pokazujące zdolność ANN do aproksymacji dowolnej funkcji ciągłej                      |
|      | Zaproponowany został model Q-learning dla uczenia ze wzmocnieniem (RL) – <i>Chris Watkins</i>   |
| 1991 | Zidentyfikowanie zanikającego gradientu utrudniającego proces uczenia sieci neuronowych   |
| 1997 | Opracowanie LSTM dedykowanego problemowi zanikającego gradientu w rekurencyjnych sieciach neuronowych RNN                                     |
|      | Pokonanie szachowego mistrza świata Garriego Kasparova przez IBM DeepBlue   |
| 1998 | Maszyna wektorów nośnych SVM<br><i>Vladimir Naumovich Vapnik, Alexey Yakovlevich Chervonenkis</i>   |
|      | Początki głębokich sieci spłotowych – <i>Yann LeCun</i>   |
| 2002 | Opracowanie ograniczonej sieci Boltzmanna jako sieci generującej – <i>Geoffrey Hinton</i>   |
| 2006 | Opracowanie wielowarstwowych głębokich sieci przekonań (DBN) jako sieci generujących – <i>Geoffrey Hinton</i>                                 |
| 2007 | Zastosowanie procesorów graficznych GPU do przyspieszenia nauki sieci neuronowych   |
| 2008 | Rozpoznawanie mowy przez rekurencyjne sieci RNN – <i>Alex Graves</i>  |
| 2009 | Wprowadzenie IMAGENET dającego wielkoskalowe zbiory do klasyfikacji obrazów   |
| 2010 | Walka z zanikającym gradientem w głębokich sieciach neuronowych   |
| 2012 | Głęboka sieć CNN ALEXNET do klasyfikacji obrazów, eksplozja w uczeniu głębokim (DL) – <i>Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton</i> |
| 2013 | Zaproponowany został Deep Q-Networks (DQN) dla sieci RL – <i>Volodymyr Mnih i inni</i>  |
| 2014 | Powstanie sieci generujących GAN wprowadza nową klasę modeli generujących – <i>Ian Goodfellow</i>   |
| 2016 | ALPHAGO pokonuje mistrza świata w grze Go   |

| Rok  | Opis osiągnięcia   |
|------|--|
| 2018 | Nagroda Turinga za wkład w powstanie uczenia głębokiego DL – <i>Yoshua Bengio, Geoffrey Hinton, Yann LeCun</i> |
| 2022 | Sukces modelu ChatGPT służącego do zaawansowanej konwersacji z człowiekiem                                     |

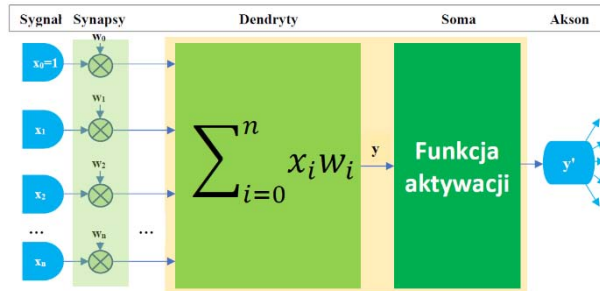
Sieci neuronowe wykazały wysoką emergencję rozumianą jako uzyskiwanie złożonych zachowań systemu zbudowanego z dużej liczby prostych elementów, które takich cech mogą nie posiadać. Aktualnie obserwowany jest przyspieszający ciągle rozwój całej dyscypliny określanej jako AI. Świadczą o tym ogłaszane wyniki prac naukowych, które pojawiają się z coraz większą częstotliwością. Dostępne są również praktyczne rezultaty tych prac w postaci aplikacji oraz usług świadczonych w wielu dziedzinach życia, zarówno konsumenckiego, jak i biznesowego [107,108].

Największe korporacje prowadzą swoisty wyścig w tej dziedzinie. Nieco zaskakującym w tym kontekście jest fakt ogłoszenia memorandum na temat zagrożeń, jakie może wywołać niczym niepojęty rozwój AI, które wystosowali do świata przedstawiciele firm wiodących w tej branży. Ostatnio (kwiecień 2023 r.), pojawiają się doniesienia mówiące, że sprawą wykorzystania AI bardzo zainteresowały się kręgi polityczne. W Unii Europejskiej rozważane są możliwości wprowadzenia ograniczeń w niektórych obszarach zastosowań AI dotyczących szeroko pojętego bezpieczeństwa ludzi.

## 4.2. Sieci neuronowe i ich podział

Przystępując do opisu taksonomii sieci neuronowych należy przyjrzeć się bliżej budowie tych struktur. Opublikowany w roku 1943 model matematyczny neuronu, który zapoczątkował dalsze badania w tej dziedzinie, sprowadza się do wykonania kilku prostych z matematycznego punktu widzenia operacji. Pierwotnie, neuron biologiczny odkryty pod koniec XIX w., składa się z dendrytów, somy (ciała komórki) i aksonu. Ciało komórki pośredniczy w przekazywaniu sygnałów z wielu dendrytów do jednego aksonu. Rolę złączy akson-dendryt przekazujących impulsy pomiędzy neuronami pełnią synapsy, które przesyłając sygnał wprowadzają do niego określoną wagę. Synapsy pełnią zatem podwójną rolę w systemie nerwowym.

Model przetwarzania sygnałów przedstawiono na rysunku 33. Stał się on pierwowzorem do tworzenia innych konstrukcji, których autorzy usiłowali zredukować niedostatki protoplasty. Powstające kolejne warianty struktur modelujących neuron biologiczny przyczyniły się do gwałtownego rozwoju wielu gałęzi nauki, poprzez fakt tworzenia, badania i zastosowania sieci wykorzystujących takie modele.



**Rysunek 33.** Ilustracja struktury modelującej pojedynczy neuron biologiczny [102]

Pokazana wyżej struktura modelująca neuron odpowiada jednemu równaniu algebraicznemu z  $n$  „zmiennymi”, gdyż wartość dla linii  $x_0$  jest stała i równa jeden:

$$y = \sum_{i=0}^n x_i w_i. \quad (34)$$

W powyższym wyrażeniu przyjęto następujące oznaczenia:

- $n$  – liczba zmiennych objaśniających,
- $i$  – indeks zmiennych,
- $x_i$  –  $i$ -ta zmienna niezależna (objaśniająca),
- $w_i$  – współczynnik wagi wnoszony przez  $i$ -tą synapsę,
- $y$  – pośrednia wartość zmiennej zależnej (suma ważona),
- $y'$  – zmienna zależna (objaśniana).

Sygnały  $x_1$  do  $x_n$ , pokazane na rys. 33, pochodzące z czujników lub z aksonów innych neuronów, w synapsach poddawane są mnożeniu przez odpowiedni współczynnik wagowy  $w_i$ . Sygnał na wejściu  $x_0$  razem ze współczynnikiem  $w_0$  utożsamiany jest z wartością wagi obciążającej neuron (ang. *bias*). Dendryty pełnią rolę przewodów doprowadzających tak obrobiony sygnał do ciała komórkowego (somy), gdzie

następuje sumowanie wartości (iloczynów cząstkowych  $x_i w_i$ ) ze wszystkich dendrytów. W kolejnym kroku, za pomocą funkcji aktywacji, dokonywana jest walidacja sumy ważonej  $y$ . Wartość  $y'$ , na wyjściu neuronu, otrzymywana jest po przejściu przez blok z funkcją aktywacji (zazwyczaj nieliniową). Wybór funkcji aktywacji ma kluczowe znaczenie, gdyż rzutuje na właściwości użytkowe całej sieci jak również odpowiada za podatność na wystąpienie niekorzystnego zjawiska jakim jest zanik gradientu.

Przyglądając się obliczeniowej stronie zagadnienia, łatwo zauważyć, że zmienne  $x_i$  w równaniu (34) są w rzeczywistości wartościami danymi w konkretnej chwili fazy uczenia sieci. Znalezienie rozwiązania polega tu na wyznaczeniu wartości współczynników wagowych dla poszczególnych synaps. Dla równania opisującego jeden neuron z  $n+1$  wejściami  $x_i$ , liczba współczynników  $w_i$  jest równa liczbie wejść  $x_i$ . W przypadku, gdy mamy do czynienia z siecią neuronową model staje się układem równań, w którym każde równanie odpowiada jednemu neuronowi. Teraz mamy jednak liczbę współczynników wagowych zwielfokrotnioną o liczbę równań. Zatem liczba wag, które należy znaleźć wynosi  $(n + 1)k$ , gdzie  $k$  jest liczbą neuronów tworzących sieć. Jest to w istocie zadanie odwrotne do rozwiązywania układu równań algebraicznych, w którym poszukiwanym rozwiązaniem byłoby tylko  $n$  wartości zmiennych  $x_i$ . Przedstawiona sytuacja jest powodem, dla którego obliczenia związane z ustaleniem współczynników wagowych  $w_i$  dla sieci podczas etapu jej „uczenia”, są zazwyczaj bardzo złożone i wymagają znacznych mocy obliczeniowych.

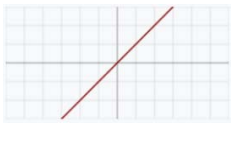



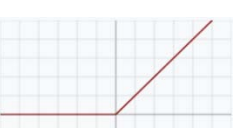
Pokazany na rysunku 33 modelowy neuron jest budulcem dla sieci typu MLP (ang. Multi Layer Perceptron). Neuron o takiej strukturze jest powszechnie stosowany w większości konfiguracji w odróżnieniu od tzw. neuronów radialnych odmiennie operujących na danych wejściowych i używających niemonotonicznej funkcji Gaussa jako funkcji aktywacji [102,105].

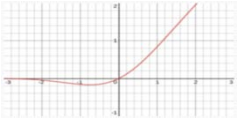





Funkcje aktywacji decydujące o funkcjonalności sieci neuronowej mogą być dobierane spośród wielu zależności matematycznych możliwych do zastosowania. W tabeli 3 zestawiono wykresy wybranych funkcji aktywacji stosowanych w syntezie sieci ANN oraz ich definicje z zakresami wartości zmiennej zależnej, jakie mogą pojawić się na wyjściu neuronu. Większość z tych funkcji jest nieliniowa, a ich dobór w konstruowanych sieciach neuronowych podyktowany jest doświadczeniem użytym na drodze badań kolejnych konfiguracji i typów takich sieci.



W tym miejscu warto zwrócić uwagę na zagadnienie różniczkowalności funkcji aktywacji. Ma to kluczowe znaczenie w procesie „uczenia” struktur zawierających neurony z konkretnym rodzajem funkcji aktywacji. Badacze zauważyli, że zjawisko zaniku gradientu powstające w trakcie treningu i przysparzające wiele kłopotów aplikacyjnych, jest związane z typem zastosowanej funkcji aktywacji. Odnosi się to szczególnie do funkcji typu ReLU (poz. 5 w tabeli 3), której cechy spowodowały poszukiwania jej substytutów. Stąd w tabeli 3 występuje tak liczna reprezentacja funkcji podobnych do ReLU, lecz z innymi własnościami swoich pochodnych (pozycje 6, 7, 8, 9, 10 w tabeli 3).

**Tabela 3.** Zestawienie typowych funkcji aktywacji wykorzystywanych w sieciach ANN [109,110]

| L.p. | Nazwa  | Wykres  | Równanie funkcji   | Zakres wartości     |
|------|--|---|--|---------------------|
| 1    | Funkcja liniowa  |    | $f(x) = x$   | $(-\infty, \infty)$ |
| 2    | Skok jednostkowy (binarna)                                   |   | $f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$ | $\{0, 1\}$          |
| 3    | Logistyczna (sigmoid)  |  | $\sigma(x) = \frac{1}{1 + e^{-x}}$                             | $(0, 1)$            |
| 4    | Tangens hiperboliczny (tgh)                                  |  | $\text{tgh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$            | $(-1, 1)$           |
| 5    | Jednostronnie obcięta funkcja liniowa (ReLU) – prostownikowa |  | $f(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$ | $[0, \infty)$       |

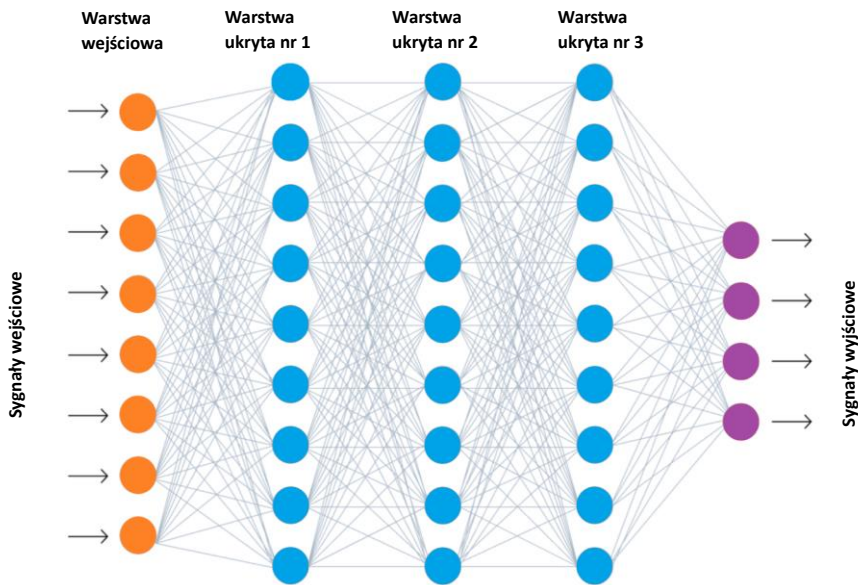
| L.p. | Nazwa  | Wykres  | Równanie funkcji  | Zakres wartości          |
|------|--|---|---|--------------------------|
| 6    | Złożenie f. liniowej i błędu Gaussa (GELU)                           |    | $f(x) = \frac{1}{2}x \left( 1 + \operatorname{erf} \left( \frac{x}{\sqrt{2}} \right) \right)$ | $(-0,17 \dots, \infty)$  |
| 7    | Softplus   |    | $f(x) = \ln(1 + e^x)$   | $(0, \infty)$            |
| 8    | Złożenie f. liniowej i wykładniczej (ELU)                            |    | $f(x) = \begin{cases} \alpha(e^x - 1), & x \leq 0 \\ x, & x > 0 \end{cases}$                  | $(-\alpha, \infty)$      |
| 9    | Sparametryzowana jednostronnie obcięta f. liniowa (PReLU, LeakyReLU) |    | $f(x) = \begin{cases} \alpha x, & x < 0 \\ x, & x \geq 0 \end{cases}$                         | $(-\infty, \infty)$      |
| 10   | Złożenie f. liniowej i Sigmoidu (SiLU)                               |   | $f(x) = \frac{x}{1 + e^{-x}}$   | $(-0,278 \dots, \infty)$ |
| 11   | Gaussowska   |  | $f(x) = e^{-x^2}$   | $(0, 1]$                 |
| 12   | Znormalizowana f. wykładnicza (Softmax)                              |   | $\sigma(x)_i = \frac{e^{x_i}}{\sum_{k=1}^K e^{x_k}}$<br>$k = 1, \dots, K$                     | $(0, 1)$                 |

Funkcje aktywacji mogą stanowić jedno z kryteriów podziału neuronów oraz tworzonych przez nie sieci. Przykładowo funkcja liniowa oraz funkcja skoku jednostkowego stosowane były w pierwszych pionierskich modelach sieci neuronowych.

Następne pozycje w tabeli 3 związane są z bardziej zaawansowanymi modelami, które badano tworząc kolejne rozwiązania AI. Wyraźnie odbiegająca swoim charakterem funkcja Gaussa jest stosowana w tzw. neuronach radialnych.

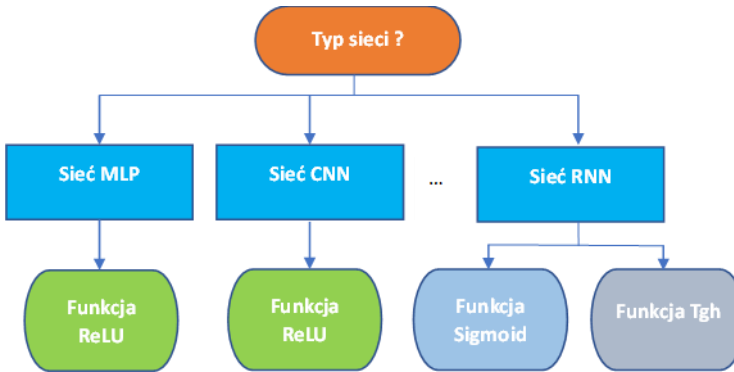
Omawiając budowę matematycznego modelu neuronu używane w tle jest pojęcie sieci neuronowej. Jest ona rozumiana jako połączenie wielu neuronów w ściśle określony sposób. Właśnie sposoby łączenia sieci wiążą się z kolejnymi odkryciami w obszarze AI.

Przykładowa struktura ilustrująca połączenia neuronów w sieć z trzema warstwami została przedstawiona na rysunku 34. Kolorami zaznaczone zostały neurony należące do różnych typów warstw tworzących taką sieć. Neurony warstw ukrytych (kółka koloru niebieskiego) mają zazwyczaj jednolitą budowę. Natomiast neurony warstw wejściowej i wyjściowej mogą być zbudowane zależnie od potrzeb aplikacji.



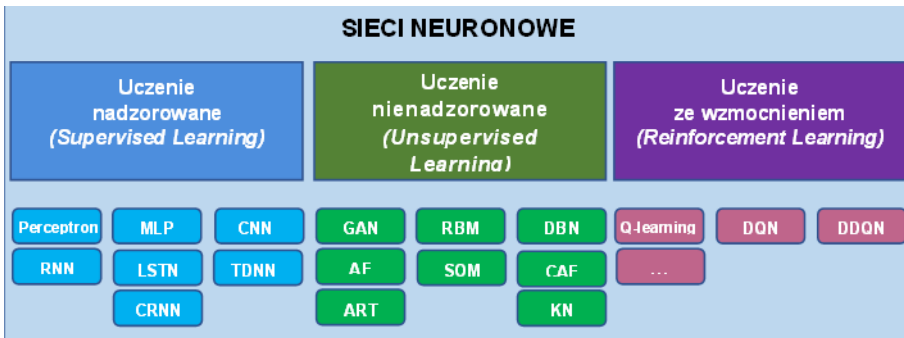
**Rysunek 34.** Przykład połączeń wielu neuronów w sieć z trzema warstwami ukrytymi [111]

Na rysunku 35 został pokazany przykładowy schemat doboru funkcji aktywacji dla warstw ukrytych wybranych sieci neuronowych. Sugestie dotyczące doboru zmieniały się z biegiem lat, gdy rosło doświadczenie zebrane podczas treningu wcześniejszych generacji tych struktur.



**Rysunek 35.** Schemat doboru funkcji aktywacji dla warstw ukrytych wybranych sieci neuronowych [112]

Opracowane do tej pory sieci neuronowe są bardzo obszerną grupą, której taksonomię można tworzyć stosując jeszcze inne kryteria podziału. Najprostsze ze stosowanych kryteriów uwzględnia sposób uczenia sieci. Zauważyć należy, że tą samą sieć można użyć do rozwiązywania zadań skrajnie różniących się między sobą. Decydującym jest tu sposób „uczenia” takiej struktury. Obraz klasyfikacji uzyskanej po zastosowaniu takiego kryterium przedstawia rysunek 36.



**Rysunek 36.** Podział sieci neuronowych ze względu na sposób ich „uczenia”

Uczenie sieci neuronowej jest procesem wymagającym przeprowadzenia wielu prezentacji zbioru informacji uczącej (wzorców). Ocenia się, że nauczenie sieci w zależności od problemu wymaga wykonania następującej liczby iteracji [113]:

$10^9$  – w przypadku prognozowania finansowego;

$10^{10}$  – dla syntezy mowy;

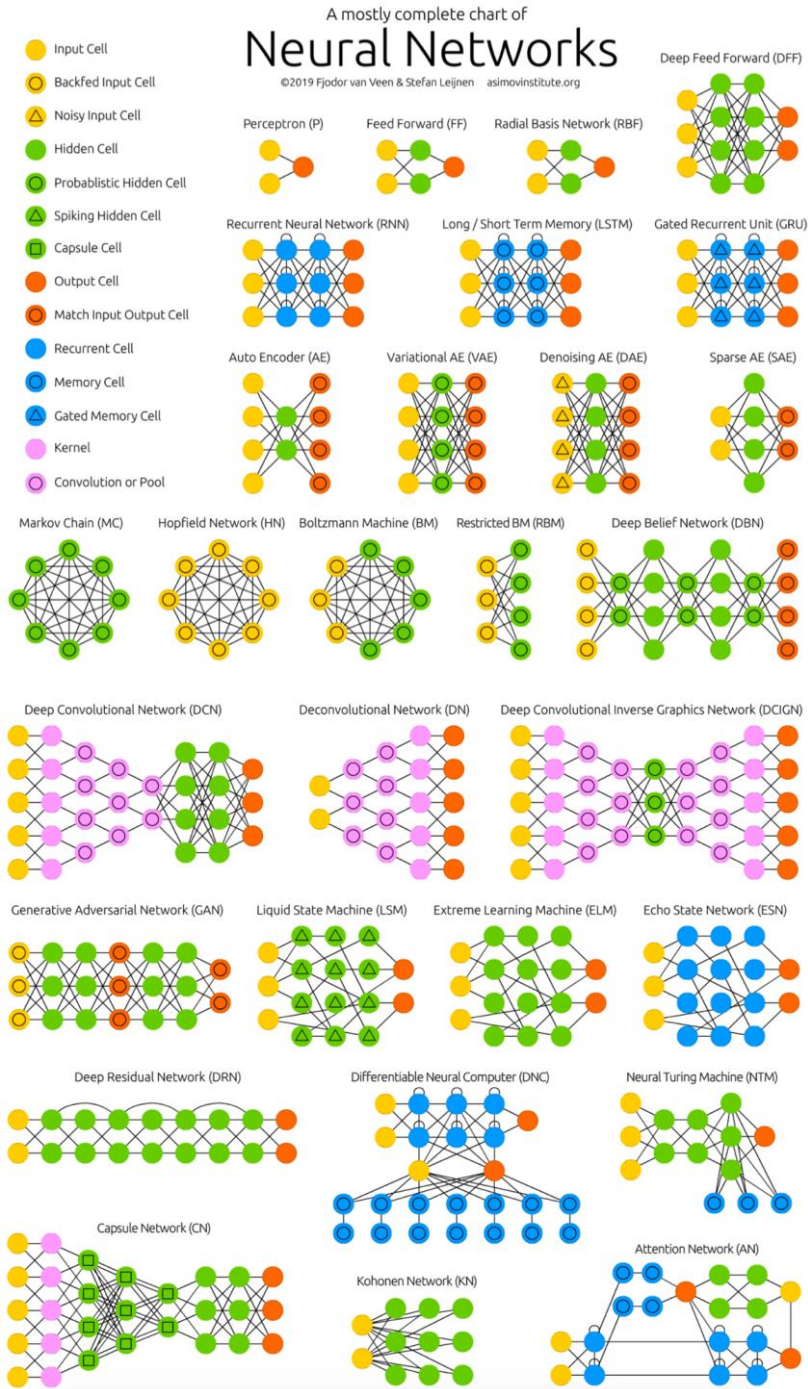
$10^{12}$  – dla rozpoznawania mowy lub pisma ręcznego;

$10^{13}$  – dla rozpoznawania japońskich znaków Kanji.

Podział pokazany na rys. 36 pomocny jest we wstępnej klasyfikacji tak, aby można łatwo ocenić podstawowe cechy sieci neuronowej. Skrócowa charakterystyka poszczególnych grup sieci jest następująca:

- Sieci z uczeniem nadzorowanym są używane do rozwiązywania zadań, gdy dane wejściowe są dobrze zaetykietowane; w trakcie uczenia sieć koryguje współczynniki wagowe tak, by zredukować popełniany błąd, który wynika z różnicy uzyskanego rezultatu i wartości reprezentowanej przez etykietę, w rezultacie otrzymywane na wyjściu sieci wartości dobrze przewidują wynik na podstawie danych wejściowych; sieci te zapewniają dobre rezultaty w zadaniach takich jak np.: rozpoznawanie obrazów, klasyfikacja tekstu, przewidywanie cen akcji giełdowych.
- Sieci z uczeniem nienadzorowanym stosuje się w przypadku danych nieetykietowanych, gdyż samodzielnie odnajdują one struktury i zależności w danych, pozwalając na automatyczne kategoryzowanie obiektów; wynikiem jest analiza skupień pokazująca podobieństwa lub niepodobieństwa wśród danych oraz grupowanie pozwalające sklasyfikować obiekt; sieci te wykrywają anomalie, oszustwa bankowe i ubezpieczeniowe, pomagają w automatycznej klasyfikacji i oznaczaniu nieobrobionych danych, mogą być również użyte do generowania danych oraz redukcji wymiarów.
- Sieci uczenia ze wzmacnianiem są stosowane do rozwiązywania zadań takich jak gry i strategię postępowania. Nauka sieci polega na uwzględnianiu w obliczeniach wartości wag, uzyskanej kary lub nagrody; proces nauki prowadzony jest metodą prób i błędów, co powoduje, że jest on z reguły relatywnie dłuższy. Sieci te skutecznie rozwiązują problemy z zakresu: robotyki, automatyki, zarządzania portfelem inwestycyjnym oraz gier strategicznych takich jak np. szachy, go, gry komputerowe.

Rozszerzona taksonomia uwzględniająca szczegóły architektury sieci prowadzi do obszernych diagramów, które można znaleźć w różnych specjalistycznych opracowaniach. Przykład takiego zestawienia zawiera rysunek 37 [114].



**Rysunek 37.** Podział sieci neuronowych uwzględniający ich architekturę [114]

Legenda odnosi się do właściwości neuronów jako elementów składowych poszczególnych typów sieci. Kolorem żółtym oznaczona została grupa elementów biorąca udział we wprowadzaniu sygnałów wejściowych do struktur sieciowych. Neurony odpowiedzialne za wyprowadzenie uzyskanych wyników przetwarzania zaznaczone są kolorem czerwonym. Typowe elementy tworzące warstwy ukryte mają kolor zielony. Kolor niebieski reprezentuje neurony z pamięcią, które są zwykle wykorzystywane w warstwach rekurencyjnych. Ostatnią z wyróżnionych grup neuronów (kolor fioletowy) stanowią elementy wykorzystywane w sieciach konwolucyjnych (CNN), do tworzenia warstw splotowych oraz warstw filtrów. Pokazane w formie graficznej sieci neuronowe identyfikowane są zarówno za pomocą nazwy opisowej, ale również skrótów powszechnie używanych w literaturze przedmiotu.

Jak podkreślają autorzy tej panoramy sieci neuronowych, przedstawiona różnorodność jest tylko migawką stanu rzeczy, aktualną na chwilę obecną. Dynamiczny rozwój tej dyscypliny, odnotowywany w ostatnich latach, pozwala przypuszczać, że przytoczony na rys. 37 diagram będzie podlegał sukcesywnemu rozszerzaniu podczas kolejnych aktualizacji.

Niektóre z pokazanych struktur sieciowych mają już znaczenie historyczne, gdyż zostały gruntownie przebadane i opisane tworząc fundament dla konstruowania nowych topologii i rozwoju całej dyscypliny AI. Spodziewać się zatem należy powstawania kolejnych rozbudowanych drzew z taksonomią dla poszczególnych typów sieci neuronowych tworzących swoiste katalogi obejmujące kolejne struktury i nowe architektury.

### **4.3. Sieci neuronowe rekurencyjne (RNN)**

Powstanie sieci Hopfielda w roku 1982 stało się zaczątkiem sieci rekurencyjnych RNN. Rekurencyjna sieć neuronowa (ang. *Recurrent Neural Network*) jest odmianą sieci opracowanej do przetwarzania sekwencji danych krok po kroku. W trakcie pracy sieć ta zachowuje swoje poprzednie stany wewnętrzne, by uwzględnić je w analizie danych, które pojawią się w następnym kroku. Neuron z połączeniem rekurencyjnym pracuje tak jak komórka pamięci [90]. Typową sieć RNN tworzy jedna lub więcej warstw z neuronami rekurencyjnymi, których zadaniem jest przetworzenie sekwencji danych podanej na wejście. Sygnał wyjściowy jest generowany przez

co najmniej jedną gęsto połączoną warstwę neuronów uwzględniających stany wewnętrzne sieci. Opisuje się to następującą ogólną zależnością [115]:

$$y_n(t) = \varphi(x_n(t-1), u_n(t-1), \vartheta), \quad (35)$$

gdzie:

$x(t-1)$  – zmienna niezależna (objaśniająca) w chwili  $t-1$ ,

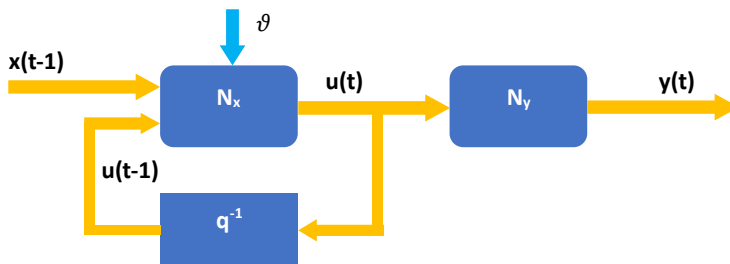
$u(t-1)$  – wektor poprzedniego stanu pośredniego,

$\vartheta$  – wagi dla poszczególnych zmiennych,

$\varphi$  – funkcja wyrażająca przetwarzanie realizowane przez całą strukturę,

$y_n(t)$  – wartość wektora zmiennej zależnej (objaśnianej) w chwili  $t$ .

Ogólny schemat struktury zawierającej neuron z rekurencją został pokazany na rysunku 38. Wektor danych  $x$  podawany jest na wejście sieci. W wyniku przejścia informacji przez człon  $N_x$  tworzy się stan  $u(t)$ . Stan ten, poprzez blok  $q^{-1}$ , z opóźnieniem o jeden cykl trafia na wejście sieci  $N_x$ . Zatem wartość sygnału  $u(t)$  zawiera również informację utworzoną w stanie poprzednim  $t-1$ . W przypadku ogólnym wartość stanu  $u(t)$  może podlegać finalnemu przetworzeniu przez opcjonalny człon  $N_y$ , który wyprowadza na wyjście rezultat końcowy  $y(t)$ .



**Rysunek 38.** Schemat ogólnej struktury z neuronem rekurencyjnym [115]

Kolejne kroki pracy sieci RNN w trybie przetwarzania danych wyglądają następująco:

- pobranie wektora danych wejściowych  $x$ ;
- przetwarzanie danych wejściowych  $x$  i poprzedniego sygnału  $u(t-1)$  w bloku  $N_x$ ;
- wygenerowanie stanów wewnętrznych przekazywanych do kolejnego kroku  $u(t)$ ;



- wygenerowanie wektora danych wyjściowych  $y(t)$ .

Stan wewnętrzny w tej strukturze tworzą informacje obejmujące to, co wcześniej „widziała” sieć. Umożliwia to uwzględnienie historii zawartej w analizowanych sekwencjach danych (kontekstu).

Proces trenowania sieci RNN jest fazą początkową i prowadzony jest jako uczenie nadzorowane. W trakcie jego trwania na wejście wprowadzane są kolejne wartości danych (wektory), a uzyskane wyniki każdorazowo są komparowane z pożądaną wartością wyjścia sieci. Algorytm uczenia modyfikuje wartości wszystkich wag trenowanej sieci na podstawie obliczonej różnicy pomiędzy wyjściem i wartością pożądaną. Proces uczenia minimalizuje funkcję straty, uzyskując tym samym najlepsze dopasowanie odpowiedzi sieci względem oczekiwań użytkownika.

Danymi wejściowymi mogą być czasowo-zależne dane sekwencyjne takie jak: zdigitalizowany sygnał audio, obraz ruchomy, sygnał z aparatury diagnostycznej EKG i EEG, ciąg tekstowy, a także dane zawarte w szeregach czasowych różnego pochodzenia, np. kursy giełdowe, notowania walut, dane pogodowe.

Zadania stawiane przed taką siecią neuronową polegają na:

- modelowaniu sekwencji, w wyniku której możliwe jest między innymi generowanie dźwięków, generowanie informacji tekstowych, sterowanie robotem, etykietowanie obrazów;
- przetwarzaniu sekwencji danych prowadzących do rozpoznawania mowy, rozpoznawania pisma, tłumaczenia maszynowego czy obróbki strumienia wideo;
- klasyfikowaniu danych zależnych od czasu, które umożliwia wykrywanie wzorców w sekwencjach czy predykcję wartości szeregu dla kolejnego kroku czasowego związaną np. z prognozowaniem cen, przewidywaniem kursów giełdowych czy przewidywaniem katastrof [105].

Pomimo wielu zalet, jakimi są obdarzone sieci RNN, należy zwrócić uwagę na fakt, że wynik działania sieci może być podatny na błędy, które propagują się w czasie. Ma to związek z sekwencyjnym przetwarzaniem danych, gdyż błędy popełnione w jednym kroku czasowym mogą mieć wpływ na wynik przetwarzania w kolejnym. Problem ten redukowany jest przy pomocy takich technik, jak LSTM (ang. *Long Short-Term Memory*).

#### 4.4. Sieci neuronowe splotowe (CNN)

Propozycja architektury splotowych (konwolucyjnych) sieci neuronowych CNN (ang. *Convolution Neural Network*) pojawiła się w roku 1980 jako inspirowany biologicznie wariant wielowarstwowej sieci MLP. Architektura ta reprezentuje sieci z głębokim uczeniem nadzorowanym, ze względu na wielość ukrytych warstw neuronowych. Sieć CNN charakteryzuje odmienny sposób przetwarzania informacji doprowadzonej do wejścia neuronu. Zamiast mnożenia wartości wejściowych przez macierz odpowiadających im wag, dokonywana jest operacja splotu. W wersji dyskretnej zapisywana zależnością:

$$s(t) = (f * g)(t) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(t - \tau). \quad (36)$$

W powyższym wyrażeniu symbol  $*$  jest operatorem splotu, zaś:

$f(\tau)$  – wartości wejściowe,

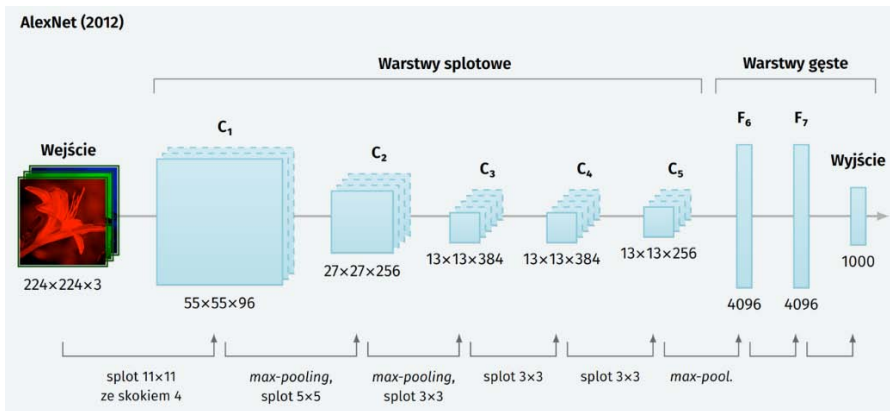
$g(t - \tau)$  – filtr (jądro),

$s(t)$  – funkcja wynikowa operacji splotu funkcji  $f$  oraz  $g$  – określana też jako mapa atrybutów lub aktywacji [116].

Sieci CNN cechuje zdolność do generalizacji sygnałów różnych klas, które mają wewnętrzne relacje przestrzenne. Dotyczy to sygnałów mogących zawierać różną liczbę wymiarów:

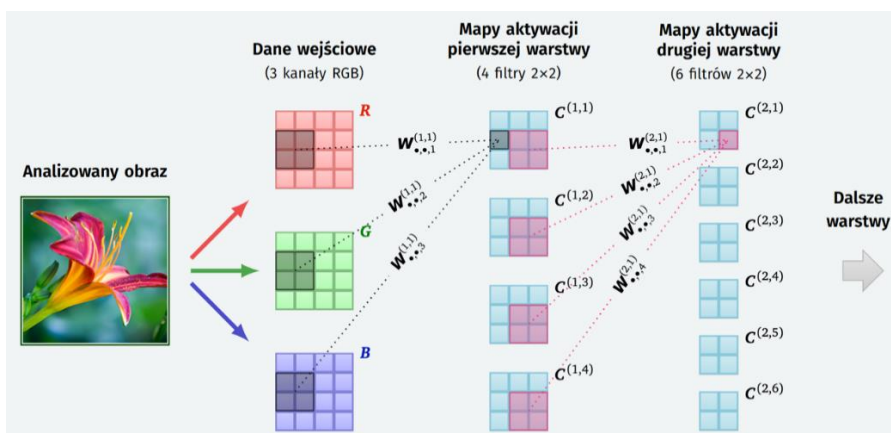
- 1D – sygnały czasowe;
- 2D – obrazy;
- 3D – video, obrazy RGB, fMRI (ang. *functional Magnetic Resonance Imaging*) – funkcjonalne obrazowanie metodą rezonansu magnetycznego;
- sygnały wielokanałowe [105].

Generalizacja rozumiana jest tu jako zdolność do uzyskania prawidłowego wyniku dla zestawu danych wejściowych, które nie były prezentowane podczas procesu „nauczania”. Dzięki zastosowaniu operacji splotu kolejne warstwy ukryte mogą mieć coraz mniejsze rozmiary (rys. 39). Przetwarzanie w takiej sieci polega na przekazywaniu wyniku uzyskanego w jednej warstwie do kolejnej warstwy. Warstwy końcowe mają już zwykle klasyczną budowę i służą do uzyskania oczekiwanej struktury wyników.



**Rysunek 39.** Schemat architektury sieci neuronowej CNN przeznaczony do obróbki danych obrazowych, zaproponowanej przez Alex Krizhevsky'ego w roku 2012 [116]

Należy zwrócić uwagę na redukcję wielkości kolejnych warstw przetwarzających. Dobór właściwych filtrów dla operacji splotowych, jak również funkcji aktywacji dla poszczególnych warstw neuronów, ma kluczowe znaczenie dla efektywnego działania tej sieci. Sieci splotowe tworzone są jako warstwy przetwarzania wstępnego. Warstwy gęste mają za zadanie przygotować wynik końcowy. Analiza działania tej architektury jest wyzwaniem dla badaczy, ze względu na liczne hiperparametry oraz wspomniane wcześniej możliwości doboru filtrów i funkcji aktywacji poszczególnych warstw.



**Rysunek 40.** Schemat pokazujący działanie warstw splotowych sieci neuronowej CNN, przeznaczony do obróbki danych obrazowych [116]

Działanie sieci z warstwami spłotowymi CNN przedstawia schemat umieszczony na rysunku 40. Jest tam widoczna redukcja wielkości map współczynników, otrzymanych po operacjach spłotu wykonanych przez kolejne warstwy sieci neuronowej. Końcowy etap przetwarzania prowadzony za pomocą warstw gęstych został na tym schemacie pominięty.

Obecny kształt sieci CNN jest wynikiem wielu badań i licznych modyfikacji wprowadzanych w kolejnych latach. Zdiagnozowane zostały różne słabości tych struktur a w ślad za tym powstały liczne techniki mające na celu ich zredukowanie. Sieci CNN wykazują między innymi takie cechy jak:

- rzadka reprezentacja – oznacza, że filtr spłotowy może być dużo mniejszy w porównaniu z rozmiarem całego sygnału wejściowego;
- współdzielenie parametrów wag – redukuje liczbę parametrów warstwy spłotowej;
- równoważność względem przesunięcia sygnału – odpowiada za aktywację tego samego filtra, pomimo że cecha, na którą reaguje może znajdować się w różnych miejscach analizowanego obrazu;
- możliwość użycia sygnału wejściowego o zmiennym rozmiarze – polega na zmianie rozmiaru map tworzonych operacjami spłotu proporcjonalnie do zmian wielkości danych na wejściu [105].

Wśród metod i technik używanych do poprawy działania i optymalizacji sieci neuronowych, a sieci CNN w szczególności, wymienić należy:

- *Pooling*, oznaczający redukcję danych, która prowadzi do poprawy „uogólniania” dokonywanego przez sieć; stosowane są często: *max pooling*, polegający na uwzględnianiu wartości maksymalnych występujących w określonych sąsiedztwach, *avg pooling*, uwzględniający wartości średnie z sąsiedztw oraz *norm pooling*, posługujący się normą z sąsiednich wyjść;
- *Dropout*, polegający na redukowaniu losowo wybranych połączeń w sieci a w etapie końcowym uśrednieniu otrzymanych w ten sposób wyników; wymusza to wykrywanie przez neurony bardziej wartościowych cech w danych, kosztem spowolnienia tempa treningu;
- Lokalną normalizację odpowiedzi, która ma za zadanie przeciwdziałać nadmiernej dominacji któregoś z neuronów i wykorzystywana jest do normalizacji jasności sąsiadujących map;

- Techniki regularyzacji, obejmujące augmentację danych stosującą przesunięcia, odbicia, ograniczoną fragmentację oraz losowe modyfikacje fragmentów obrazów;
- *Batch normalization*, polegająca na normalizacji wejść względem wartości średniej i wariancji wzdłuż wybranego wymiaru dla minipakietu danych; dzięki niej uzyskiwana jest lepsza generalizacja, wzrost szybkości i stabilności uczenia; umożliwia stosowanie większych kroków treningowych, przy czym zminimalizowane jest ryzyko zaniku lub eksplozji gradientu;
- Incepcję, która stosuje rozbcie dużych filtrów splotowych i ich zastąpienie złożeniami filtrów o mniejszych rozmiarach, co w efekcie redukuje liczbę obliczanych parametrów;
- Połączenia rezydualne, wprowadzane w celu uniknięcia załamania właściwości sieci przy jej dużej rozbudowie [116];
- *Downsampling*, który jest używany przy kosztownej obliczeniowo obróbce obrazów o wysokiej rozdzielczości;
- *Upsampling*, stosowany w fazie odtwarzania obrazu o wysokiej rozdzielczości [105].

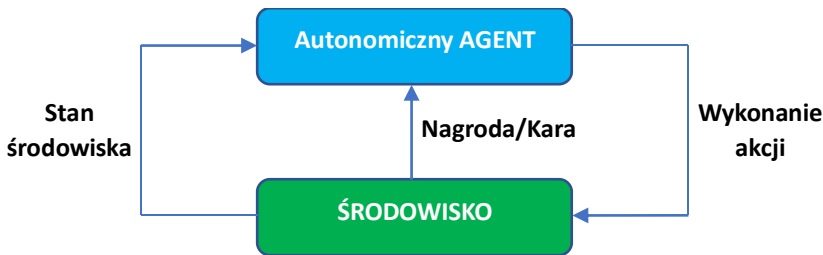
Przedstawiona lista metod poprawy parametrów sieci neuronowych nie wyczerpuje jednak wszystkich możliwości, gdyż w dalszym ciągu trwają badania i pojawiają się nowe koncepcje pozwalające uzyskiwać coraz lepsze wyniki. Obecnie sieci CNN pomocne są w analizie różnych rodzajów szeregów czasowych, gdyż z powodzeniem modelują one dane zawierające zarówno nagrania dźwięków, obrazy, treści video, badania medyczne jak i dane liczbowe z wartościami notowań czy też całą gamę wskaźników ekonomicznych.

#### 4.5. Sieci neuronowe ze wzmocnieniem (RL)

Historia idei sieci neuronowych ze wzmocnieniem (RL) (ang. *Reinforcement Learning*) sięga lat 50. XX wieku. Jednak dopiero w latach 80. i 90. pojawiły się prace opisujące pierwsze modele RL. Opublikowanie w roku 1989 przez Chrisa Watkina pracy dotyczącej modelu Q-learning rozpoczęło szersze badania w tym zakresie. Pomimo tak długiej historii, sieci tego typu są nadal intensywnie rozwijane. W roku 2013 Volodymyr Mnih z zespołem zaproponował popularne obecnie DQN (ang. *Deep Q-Networks*).

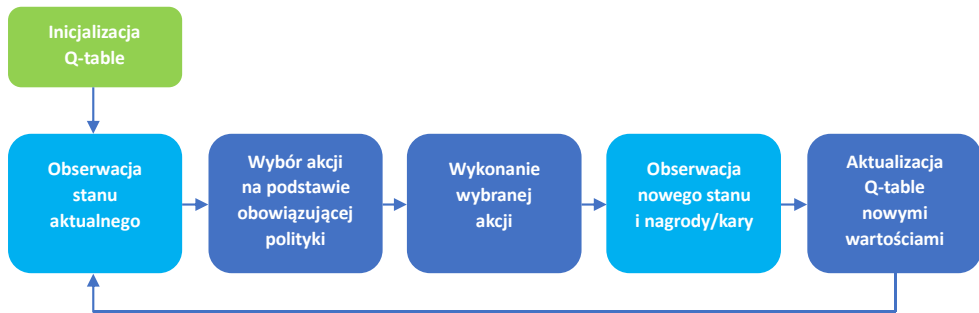
Sieci RL stanowią grupę architektur i modeli, w których obowiązuje pojęcie kary i nagrody. W sieciach tego typu „uczenie” jest procesem stymulowanym osiągniętym wcześniej wynikiem. W przypadku porażki (uzyskanie negatywnego rezultatu) do sieci wraca sygnał odpowiadający karze. Osiągnięcie sukcesu jest kwitowane nagrodą. Mechanizm przyznawania nagrody i kary, będący sprzężeniem zwrotnym w całej strukturze, powoduje, że sieć dochodzi do pożądanego wyniku metodą prób i błędów. Naturalnym środowiskiem dla funkcjonowania takich sieci są wszelkiego rodzaju gry i współzawodnictwa mające na celu uzyskanie przewagi mierzonej punktami, pozycją w rankingu czy też wynikiem finansowym.

Sieci RL w typowej architekturze utworzone są z dwóch podstawowych elementów: sieci neuronowej i algorytmu uczenia ze wzmocnieniem. Sieć neuronowa używana jest do modelowania strategii podejmowania decyzji przez agenta na podstawie stanu środowiska oraz zadanej polityki, która narzuca agentowi priorytety w działaniu. Natomiast sam algorytm RL jest wykorzystywany do uczenia agenta. Trening taki polega na określeniu jakie akcje agent powinien podjąć w danym momencie (sytuacji) w celu zmaksymalizowania oczekiwanej nagrody (rys. 41).



**Rysunek 41.** Schemat interakcji wykorzystującej uczenie ze wzmocnieniem [117]

Uczenie sieci neuronowej ze wzmocnieniem prowadzone jest według przyjętego algorytmu, który łatwo przedstawić na przykładzie klasycznego Q-learning. Schemat przedstawiający poszczególne fazy działania tego algorytmu został pokazany na rysunku 42.



**Rysunek 42.** Schemat budowy algorytmu Q-learning [117]

Sieci RL mają wiele zastosowań. Wśród nich wymienić należy takie jak: robotyka, automatyka przemysłowa, gry komputerowe jak również zarządzanie portfelem inwestycyjnym. Ich popularność wynika ze zdolności do nauczenia agenta podejmowania optymalnych decyzji w dynamicznym środowisku, nawet w przypadku, gdy zachowanie środowiska jest losowe i trudne do przewidzenia lub zależy od wyników działania innych. Sieć tego typu święciła swoje zwycięstwo w roku 2016, w czasie w pojedynku z mistrzem świata w grze Go. Uzyskany wynik otwiera szerokie perspektywy aplikacyjne dla struktur sieci neuronowych (RL) bazujących na uczeniu ze wzmocnieniem.

#### 4.6. Własności sieci neuronowych w odniesieniu do analizy szeregów czasowych

Sieci neuronowe są szeroko stosowane w wielu dziedzinach, wykonując zadania w odniesieniu do danych o różnym charakterze. Klasa danych reprezentowana przez szeregi czasowe stwarza szczególne wymagania wobec tej technologii.

W tabeli 4 zestawione zostały różne rodzaje sieci neuronowych według kryterium skutecznego ich zastosowania w odniesieniu do szeregów czasowych.

Przetwarzanie za pomocą sieci neuronowych silnie związane jest ze sposobem przygotowania danych. Na pierwszy plan wysuwają się oczywiście normalizacja i różnicowanie danych wejściowych. Jednak stosowane są również inne techniki, polegające np. na tworzeniu zakodowanych obrazów reprezentujących dane liczbowe.

**Tabela 4.** Ranking sieci neuronowych stosowanych do przetwarzania szeregów czasowych

| L.p. | Nazwa                                 | Opis  |
|------|---------------------------------------|---|
| 1    | Sieci rekurencyjne (RNN)              | Posiadają mechanizm pamięci i są w stanie analizować sekwencje danych, w tym szeregi czasowe. Najczęściej stosowanymi typami RNN są Long Short-Term Memory (LSTM) i Gated Recurrent Units (GRU). Są bardzo wszechstronne i znajdują zastosowanie w różnych dziedzinach, takich jak przetwarzanie języka naturalnego, rozpoznawanie mowy, generowanie tekstu, analiza i przetwarzanie danych medycznych, a także prognozowanie szeregów czasowych w różnych dziedzinach, takich jak finanse, transport i energetyka.   |
| 2    | Sieci splotowe (CNN)                  | Często stosowane w przetwarzaniu obrazów, mogą również być wykorzystywane do analizy szeregów czasowych. W przypadku szeregów czasowych, są stosowane do ekstrakcji cech z danych wejściowych. Szczególnie przydatne są w analizie szeregów czasowych związanych z obrazowaniem biomedycznym, analizą wibracji, przewidywaniem kosztów i sprzedaży w handlu detalicznym oraz rozpoznawaniem mowy.   |
| 3    | Sieci hybrydowe                       | Łączą różne rodzaje sieci neuronowych, takie jak RNN, CNN i autoenkodery, aby uzyskać najlepsze wyniki w analizie szeregów czasowych. Pozwalają na lepsze wykorzystanie różnych typów sieci neuronowych i mają zastosowanie w dziedzinach, takich jak finanse, energetyka, produkcja, przemysł farmaceutyczny i biomedyczny.  |
| 4    | Autoenkodery                          | Uczą się reprezentacji danych wejściowych w celu zredukowania wymiarowości i wykrycia wzorców w danych z szeregów czasowych oraz do ich rekonstrukcji i analizy anomalii. Znajdują one zastosowanie w dziedzinach takich jak przetwarzanie sygnałów, rozpoznawanie wzorców, analiza sekwencji DNA oraz analiza obrazów medycznych. Są łatwe w implementacji i działają dobrze na danych z różnych dziedzin.   |
| 5    | Sieci generatywne                     | Mogą generować nowe szeregi czasowe na podstawie danych wejściowych, a także nauczyć się modelowania i przewidywania przyszłych wartości w szeregach czasowych. Sieci generatywne znajdują zastosowanie w dziedzinach takich jak analiza sekwencji DNA, analiza wibracji, finanse, detekcja oszustw, generowanie muzyki i sztucznej inteligencji w grach.   |
| 6    | Sieci uczące się ze wzmocnieniem (RL) | Stosowane do podejmowania decyzji na podstawie aktualnego stanu szeregu czasowego w celu optymalizacji celów. Mogą być stosowane w połączeniu z innymi rodzajami sieci neuronowych, takimi jak RNN lub CNN, aby lepiej modelować szeregi czasowe. Wykorzystywane są w dziedzinach takich jak sterowanie procesami przemysłowymi, zarządzanie energią czy handlowanie na rynkach finansowych. Modele RL są trudne do nauczenia i wymagają dłuższego czasu szkolenia w porównaniu do innych rodzajów sieci neuronowych. |
| 7    | Sieci transformers                    | Wykorzystują mechanizm uwagi, aby nauczyć się reprezentacji szeregów czasowych. Są stosowane w dziedzinach takich jak przetwarzanie języka naturalnego, generowanie tekstu oraz przetwarzanie danych finansowych.   |
| 8    | Sieci echo state                      | Rodzaj sieci rekurencyjnych, w których wagi są ustalane na początku szkolenia, a następnie nie są już zmieniane. Są łatwe w implementacji i znajdują zastosowanie w analizie i przetwarzaniu sygnałów, finansach oraz inżynierii.   |
| 9    | Sieci neuromorficzne                  | Inspirowane architekturą mózgu i działają w czasie rzeczywistym. Neuromorphic networks znajdują zastosowanie w dziedzinach takich jak robotyka, iot, systemy wizyjne oraz analiza sygnałów.   |



| L.p. | Nazwa           | Opis   |
|------|-----------------|--|
| 10   | Sieci attention | Wykorzystują mechanizm uwagi do nauki reprezentacji szeregów czasowych. Znajdują zastosowanie w dziedzinach takich jak przetwarzanie języka naturalnego, przetwarzanie wideo oraz rozpoznawanie obrazów. |
| 11   | Sieci ResNet    | Szczególnie skuteczne w przetwarzaniu sygnałów. Stosowane w analizie i przetwarzaniu sygnałów dźwiękowych oraz wibracji i drgań.   |

Przykładem takiej techniki jest kodowanie GAF RGB (ang. *Gramian Angular Field*) [118,119]. Pozwala ona na przygotowanie graficznych plików .png, których zastosowanie przyspiesza proces uczenia sieci neuronowych, wyspecjalizowanych w rozpoznawaniu obrazów.

Zestawienie w tabeli 4 daje obraz zaledwie jednego z obszarów zastosowań sieci neuronowych. Wybór odpowiedniego modelu sieci neuronowej zależy od charakteru danych, podlegających analizie oraz celu tej analizy. W przypadku predykcji notowań giełdowych liczba użytecznych architektur sieci, które można wykorzystać do realizacji takich zadań, jest jednak ograniczona, jednakże otrzymywane rezultaty świadczą o dużym potencjale tych struktur.

## 5. Podsumowanie

W pracy przedstawiono przegląd zagadnień dotyczących modelowania i predykcji szeregów czasowych. W przeglądzie wyodrębniono trzy części istotne z punktu widzenia badania szeregów czasowych. Każda z tych części dotyczy różnych aspektów opisywanego zagadnienia:

- W pierwszej części przedstawiono charakterystykę szeregów czasowych w tym danych zawierających notowania rynkowe. Pokazane zostały cztery grupy metod analizy danych giełdowych. Szerzej opisano metody analizy technicznej będące elementem wielu strategii inwestycyjnych. Dokonano przeglądu wybranych narzędzi analizy technicznej stosowanych do automatyzacji prac analitycznych.
- Z kolei w drugiej części omówiono metody algorytmiczne pomocne w analizie szeregów czasowych. Na wstępie pokazana została grupa algorytmów i metod używanych na etapie wstępnego przygotowania danych. Dalej opisano kluczowe metody algorytmiczne, należące do obszaru uczenia maszynowego, których zakres stosowania obejmuje również predykcję. W tej

części przedstawiono również informacje o predykcji i sposobach jej oceny.

- W części trzeciej zaprezentowano sieci neuronowe przydatne w analizie oraz predykcji szeregów czasowych. Opisano historię powstania i rozwoju sieci neuronowych oraz pokazano ich ogólną taksonomię. Omówiono wybrane sieci reprezentujące grupę tzw. uczenia nadzorowanego jak również sieci uczące się w trybie nienadzorowanym. Zestawiono również własności sieci neuronowych w odniesieniu do analizy szeregów czasowych.

Zebrane w artykule informacje przeglądowe dotyczące struktur przetwarzających dane zawarte w szeregach czasowych, rysują pewien obszar do prowadzenia badań, który może być rozszerzony i objąć sieci hybrydowe oraz rozbudowane zastosowanie metod uczenia zespołowego. Stwarza to perspektywę na zwiększenie efektywności przetwarzania danych oraz na uzyskiwanie coraz lepszych wyników predykcji. Dynamiczny rozwój badań teoretycznych oraz obserwowane obecnie efekty wdrożeń sztucznej inteligencji (AI), przynoszą wiele optymizmu w odniesieniu do osiągania znaczących rezultatów z zastosowaniem takich technologii.

## Literatura

- [1] J. Kłobukowska. Laureaci Nagrody Nobla w dziedzinie nauk ekonomicznych. Historia i wkład laureatów do współczesnej myśli ekonomicznej, *Optimum. Studia Ekonomiczne* tom 64, nr 4, ss. 144-155, 2013. [Online]. [https://repozytorium.uwb.edu.pl/jspui/bitstream/11320/981/1/14\\_Justyna%20K%C5%81OBUKOWSKA.pdf](https://repozytorium.uwb.edu.pl/jspui/bitstream/11320/981/1/14_Justyna%20K%C5%81OBUKOWSKA.pdf).
- [2] *Predykcja*. [Online]. <https://sjp.pwn.pl/slowniki/predykcja.html>.
- [3] A. Nielsen. *Szeregi czasowe. Praktyczna analiza i predykcja z wykorzystaniem statystyki*, wyd. 1. Gliwice: Helion SA, 2020.
- [4] G. Box. *Analiza szeregów czasowych: prognozowanie i sterowanie*. Warszawa: PWN, 1983.
- [5] *Analiza szeregów czasowych*. [Online]. [https://el.us.edu.pl/ekonofizyka/index.php/Analiza\\_Szereg%C3%B3w\\_Czasowych/Wst%C4%99p#Definicja\\_szeregu\\_czasowego](https://el.us.edu.pl/ekonofizyka/index.php/Analiza_Szereg%C3%B3w_Czasowych/Wst%C4%99p#Definicja_szeregu_czasowego).
- [6] *Szeregi czasowe*. [Online]. <https://stat.gov.pl/metainformacje/szeregi-czasowe-4712/>.

- [7] Szereg czasowy. [Online]. [https://pl.wikipedia.org/wiki/Szereg\\_czasowy](https://pl.wikipedia.org/wiki/Szereg_czasowy).
- [8] R. Ciesielski. *Model Blacka-Scholesa*. [Online]. <https://www.fuw.edu.pl/~ciesiel/essey.pdf>.
- [9] *Modele szeregów czasowych*. [Online]. [http://www.ekonometria.4me.pl/modele\\_szeregow\\_czasowych.htm](http://www.ekonometria.4me.pl/modele_szeregow_czasowych.htm).
- [10] N. Ferguson, Potęga pieniądza. Finansowa historia świata, wyd. 1. Kraków: Wydawnictwo Literackie, 2022.
- [11] A. Grün, T. Kohrs, *Jak etycznie inwestować*. Kraków: Salwator, 2010.
- [12] K. van Tharp, D. Barton, S. Sjuggerud. *Bezpieczne strategie inwestycyjne. Jak osiągnąć wolność finansową*, wyd. 2. Kraków: Oficyna Ekonomiczna, 2006.
- [13] J.D. Schwager. *Analiza techniczna rynków terminowych*. Warszawa: WIG-Press, 2002.
- [14] *Analiza sentymentu*. [Online]. <https://www.udemy.com/course/applied-sentiment-analysis-commitments-of-traders-report/learn/lecture/11802872#notes>.
- [15] S. Nison. *Świece japońskie i inne techniki analizowania wykresów*, wyd. 1. Warszawa: WIG-Press, 1996.
- [16] J.J. Murphy. *Analiza techniczna rynków finansowych*. Warszawa: WIG-Press, 1999.
- [17] K. Oziemczuk. *Ichimoku japońska strategia inwestycyjna*. Warszawa: Wyd. Bullet Books, 2012.
- [18] *Teoria Dowa*. [Online]. <https://www.edukacjagieldowa.pl/gieldowe-abc/analiza-techniczna/teoria-dowa/>.
- [19] B. Zawadzki. *Teoria Dowa – założenia*. [Online]. <https://comparic.pl/teoria-dowa-glowne-zalozenia-twierdzenia/>.
- [20] *Wprowadzenie do Teorii Dowa*. [Online]. <https://academy.binance.com/pl/articles/an-introduction-to-the-dow-theory>.
- [21] A.J. Frost, R.R. Prechter. *Teoria fal Eliota*. Warszawa: WIG-Press, 1995.
- [22] *Metoda Wyckoffa – dogłębne wytłumaczenie*. [Online]. <https://academy.binance.com/pl/articles/the-wyckoff-method-explained>.
- [23] M. Kosior. *Czym jest metoda Wyckoffa*. [Online]. <https://bithub.pl/artykuly/czym-jest-metoda-wyckoffa/>.
- [24] R.C. Miner. *Strategie inwestycyjne o wysokim prawdopodobieństwie sukcesu*. Warszawa: infoin-westor.pl, 2012.

- [25] P. Danielewicz, *Geometria Fibonacciego – nowe ujęcie*. Poznań: maklerska.pl, 2014.
- [26] M. Żelazny. *Harmonic Trading. Precyzyjna strategia inwestycyjna od otwarcia do zamknięcia pozycji na rynku Forex*, wyd. 1. Sieradz: Harmonic Traders, 2013.
- [27] A. Brooks. *Analiza Price Action Trendy*. Poznań: maklerska.pl, 2014.
- [28] Lukasz. *Ile można stracić na Forex? 96% przegrywa – Prawda czy mit?* [Online]. <https://www.opcje24h.pl/ile-mozna-stracic-na-forex-96-przegrywa-prawda-czy-mit/>.
- [29] *Algorithmic trading*. [Online]. [https://en.wikipedia.org/wiki/Algorithmic\\_trading](https://en.wikipedia.org/wiki/Algorithmic_trading).
- [30] K. Jajuga. *Podstawy inwestowania na Gieldzie Papierów Wartościowych*, wyd. II uaktualnione. Warszawa: Giełda Papierów Wartościowych w Warszawie, 2006.
- [31] J.J. Murphy. *Międzyrynkowa analiza techniczna, strategie inwestycyjne na rynkach akcji, obligacji, towarów i walut*. Warszawa: WIG-Press, 1998.
- [32] *Platforma cTrader na stronie IC Markets*. [Online]. <https://ct.icmarkets.com/>
- [33] M. Wiśniewski. *Elementarz giełdowy, czyli pierwsze kroki do fortuny*. Kraków: Wyd. DWN, 1993.
- [34] Z. Komar. *Sztuka spekulacji*. Warszawa: Wydawnictwo PRET S.A., 1993.
- [35] J. Bernstein. *Inwestor Jednosesyjny*. Kraków: Oficyna Ekonomiczna Wolters Kluwer, 2006.
- [36] J. DiNapoli. *Poziomy DiNapolego*. Warszawa: WIG-Press, 2004.
- [37] C. Carolan. *Kalendarz spiralny*. Warszawa: WIG-Press, 1996.
- [38] *Średnia ruchoma*. [Online]. [https://pl.wikipedia.org/wiki/%C5%9Arednia\\_ruchoma](https://pl.wikipedia.org/wiki/%C5%9Arednia_ruchoma).
- [39] *Wskaźnik Moving Average*. [Online]. [https://www.instaforex.com/pl/forex\\_technical\\_indicators/moving\\_average](https://www.instaforex.com/pl/forex_technical_indicators/moving_average).
- [40] *Time Series Moving Average*. [Online]. <https://ctrader.com/forum/calgo-support/11309>.
- [41] *Zero lag exponential moving average*. [Online]. [https://en.wikipedia.org/wiki/Zero\\_lag\\_exponential\\_moving\\_average](https://en.wikipedia.org/wiki/Zero_lag_exponential_moving_average).
- [42] J. Ehlers, R. Way. *Zero lag (well, almost)*. [Online]. <https://www.mesa-software.com/papers/ZeroLag.pdf>.

- [43] F. Butin. (2020) *A bounded operator approach to technical indicators without lag*. [Online]. <https://arxiv.org/pdf/2009.08821.pdf>.
- [44] *Kanały Keltnera*. [Online]. <https://pl.tradingview.com/scripts/keltnerchannels/>.
- [45] *Natural Language Processing with Disaster Tweets*. [Online]. <https://www.kaggle.com/competitions/nlp-getting-started>.
- [46] *JPX Tokyo Stock Exchange Prediction*. [Online]. <https://www.kaggle.com/competitions/jpx-tokyo-stock-exchange-prediction>.
- [47] *Co określamy jako finanse behawioralne?* [Online]. <https://businessinsider.com.pl/gospodarka/przepisy/co-to-sa-finance-behawioralne/wgccj90>.
- [48] J. Matejka, G. Fitzmaurice. *Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing*, Proceedings of the 2017 CHI conference on human factors in computing systems, 2017, ss. 1290-1294. [Online]. <https://damassets.autodesk.net/content/dam/autodesk/www/autodesk-research/pdf/publications/Same%20Stats,%20Different%20Graphs%20Generating%20Datasets%20with%20Varied%20Appearance%20and%20Identical%20Statistics%20through%20Simulated%20Annealing.pdf>.
- [49] M. Szeliga. *Data science i uczenie maszynowe*, wyd. 1. Warszawa: PWN, 2017.
- [50] D.T. Larose. *Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych*. Warszawa: PWN, 2009.
- [51] F.J. Anscombe. *Graphs in statistical analysis, The American Statistician* Vol. 27, No. 1, ss. 17-21, 1973.
- [52] J. Stefanowski., *Analiza Szeregów Czasowych*. [Online]. <http://www.cs.put.poznan.pl/jstefanowski/aed/TPtimeseries.pdf>.
- [53] *Godziny handlu Forex i godziny otwarcia giełd na świecie*. [Online]. <https://admiralmarkets.com/pl/education/articles/forex-basics/forex-godziny-handlu>.
- [54] P. Kościelniak. *Statystyka I i II*. [Online]. <https://im.uj.edu.pl/documents/65632/144164127/Statystyka/f6386896-9211-4807-818b-382509b6d995>.
- [55] T. Górecki. *Analiza danych*. [Online]. [http://drizzt.home.amu.edu.pl/images/DADA\\_AIPD/W4.pdf](http://drizzt.home.amu.edu.pl/images/DADA_AIPD/W4.pdf).

- [56] *Standaryzacja (statystyka)*. [Online]. [https://pl.wikipedia.org/wiki/Standaryzacja\\_\(statystyka\)](https://pl.wikipedia.org/wiki/Standaryzacja_(statystyka)).
- [57] *Standaryzacja zmiennych – opis*. [Online]. [https://www.naukowiec.org/wiedza/statystyka/standaryzacja-zmiennych\\_713.html](https://www.naukowiec.org/wiedza/statystyka/standaryzacja-zmiennych_713.html).
- [58] *Normalization (statistics)*. [Online]. [https://en.wikipedia.org/wiki/Normalization\\_\(statistics\)](https://en.wikipedia.org/wiki/Normalization_(statistics)).
- [59] *Regresja liniowa*. [Online]. <https://pl.wikipedia.org/wiki/Regresja liniowa>.
- [60] *Regresja (statystyka)*. [Online]. [https://pl.wikipedia.org/wiki/Regresja\\_\(statystyka\)](https://pl.wikipedia.org/wiki/Regresja_(statystyka)).
- [61] *Regression analysis*. [Online]. [https://en.wikipedia.org/wiki/Regression\\_analysis](https://en.wikipedia.org/wiki/Regression_analysis).
- [62] *Homoscedasticity and heteroscedasticity*. [Online]. [https://en.wikipedia.org/wiki/Homoscedasticity\\_and\\_heteroscedasticity](https://en.wikipedia.org/wiki/Homoscedasticity_and_heteroscedasticity).
- [63] D.J. Hand, H. Mannila, P. Smyth. *Eksploracja danych*. Warszawa: WNT, 2005.
- [64] D.T. Larose. *Metody i modele eksploracji danych*. Warszawa: PWN, 2012.
- [65] Robust regression. [Online]. [https://en.wikipedia.org/wiki/Robust\\_regression](https://en.wikipedia.org/wiki/Robust_regression)
- [66] W. Pieriegudow. *Metoda najmniejszych kwadratów i jej zastosowanie*. Warszawa: Państwowe Wyd. Ekonomiczne, 1967.
- [67] C. Tofallis, Least Squares Percentage Regression, *Journal of Modern Applied Statistical Methods* Vol. 7, No. 2, ss. 526-534, 2008. [Online]. <https://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=1466&context=jmasm>.
- [68] Z. Gniazdowski., *Prywatna korespondencja*, październik 2023.
- [69] W. Niemirow. *Statystyka*. [Online]. <https://www.mimuw.edu.pl/~adamm/statystyka.pdf>.
- [70] R. Waśko. *Rangowanie danych. Metody i przykłady*. [Online]. <https://predictivesolutions.pl/rangowanie-danych-metody-i-przyklady>.
- [71] Z. Gniazdowski, M. Grabowski. Numerical Coding of Nominal Data, *Zeszyty Naukowe WWSI* tom 9, nr 12, ss. 53-61, 2015. [Online]. <https://doi.org/10.26348/znwysi.12.53>.
- [72] J. Trzęsiok. Wykorzystanie regresji nieparametrycznej do modelowania wielkości oszczędności gospodarstw domowych, *Studia Ekonomiczne* nr 159, ss. 99-108, 2013. [Online]. <https://bibliotekanauki.pl/articles/591136.pdf>.

- [73] *Modele autoregresyjne*. [Online]. <http://www.ekonometria.4me.pl/modele-autoregresyjne.htm>.
- [74] A. Torój. *Zaawansowane metody odsezonowywania szeregów czasowych*. [Online]. [https://web.sgh.waw.pl/~atoroj/eszcz/wyklad1\\_odsezonowanie.pdf](https://web.sgh.waw.pl/~atoroj/eszcz/wyklad1_odsezonowanie.pdf).
- [75] V. Mehandzhiyski. (2023) *What Is an ARIMAX Model?* [Online]. <https://365datascience.com/tutorials/python-tutorials/arimax/>.
- [76] J. Mielniczuk. *Procesy nieliniowe ARCH, GARCH*. [Online]. <https://home.ipipan.waw.pl/j.mielniczuk/SC14-2020.pdf>.
- [77] *Autoregressive conditional heteroskedasticity*. [Online]. [https://en.wikipedia.org/wiki/Autoregressive\\_conditional\\_heteroskedasticity](https://en.wikipedia.org/wiki/Autoregressive_conditional_heteroskedasticity).
- [78] M. Rubaszek. *Modele klasy ARCH*. [Online]. <https://web.sgh.waw.pl/~mrubas/EFzR/pdf/R7Prezentacja.pdf>.
- [79] *Modele klasy ARCH*. [Online]. <https://gieldowyracjonalista.blogspot.com/2010/08/modele-klasy-arch.html>.
- [80] J.R. Hyndman, G. Athanasopoulos. (2021) *Seasonal plots*. In: [w:] *Idem, Forecasting: Principles and Practice* (3rd ed.). Melbourne: Monash University, [Online]. <https://otexts.com/fpp3/seasonal-plots.html>.
- [81] R.J. Hyndman, G. Athanasopoulos. (2021) *Forecasting: principles and practice*, 3rd edition. Melbourne: Monash University, [Online]. <https://otexts.com/fpp3/>.
- [82] J.R. Hyndman, G. Athanasopoulos. (2018) *Seasonal plots*. [w:] *Idem, Forecasting: Principles and Practice* (2nd ed). Melbourne: Monash University, [Online]. <https://otexts.com/fpp2/seasonal-plots.html>.
- [83] J.R. Hyndman, G. Athanasopoulos. (2018) *Seasonal subseries plots*. [w:] *Idem, Forecasting: Principles and Practice* (2nd ed). Melbourne: Monash University, [Online]. <https://otexts.com/fpp2/seasonal-subseries-plots.html>.
- [84] *Spectral density estimation*. [Online]. [https://en.wikipedia.org/wiki/Spectral\\_density\\_estimation](https://en.wikipedia.org/wiki/Spectral_density_estimation).
- [85] *Autocorrelation*. [Online]. <https://en.wikipedia.org/wiki/Autocorrelation>.
- [86] J.R. Hyndman, G. Athanasopoulos. (2018) *Moving averages*. [w:] *Idem, Forecasting: Principles and Practice* (2nd ed). Melbourne: Monash University, [Online]. <https://otexts.com/fpp2/moving-averages.html>.
- [87] E.B. Dagum, S. Bianconcini. *Seasonal adjustment methods and real time trend-cycle estimation*. Cham: Springer International Publishing, 2016.

- [88] J.R. Hyndman, G. Athanasopoulos. (2018) *STL decomposition*. [w:] *Idem, Forecasting: Principles and Practice* (2nd ed). Melbourne: Monash University, [Online]. <https://otexts.com/fpp2/stl.html>.
- [89] R.B. Cleveland, W.S. Cleveland, J.E. McRae, I. Terpenning. STL: A seasonal-trend decomposition procedure based on loess, *Journal of Official Statistics* Vol. 6, No. 1, ss. 3-73, 1990. [Online]. <http://bit.ly/stl1990>.
- [90] I. Goodfellow, Y. Bengio, A. Courville. *Deep Learning: Systemy uczące się*. Warszawa: PWN, 2018.
- [91] *Support vector machine*. [Online]. [https://en.wikipedia.org/wiki/Support\\_vector\\_machine](https://en.wikipedia.org/wiki/Support_vector_machine).
- [92] *Naive Bayes classifier*. [Online]. [https://en.wikipedia.org/wiki/Naive\\_Bayes\\_classifier](https://en.wikipedia.org/wiki/Naive_Bayes_classifier).
- [93] *Analiza głównych składowych*. [Online]. [https://pl.wikipedia.org/wiki/Analiza\\_g%C5%82%C3%B3wnych\\_sk%C5%82adowych](https://pl.wikipedia.org/wiki/Analiza_g%C5%82%C3%B3wnych_sk%C5%82adowych).
- [94] *Odkrywanie asocjacji*. [Online]. <https://wazniak.mimuw.edu.pl/images/d/dd/ED-4.2-m02-1.0.pdf>.
- [95] *Metoda asocjacyjna*. [Online]. [https://pl.wikipedia.org/wiki/Metoda\\_asocjacyjna](https://pl.wikipedia.org/wiki/Metoda_asocjacyjna).
- [96] *Weighted majority algorithm (machine learning)*. [Online]. [https://en.wikipedia.org/wiki/Weighted\\_majority\\_algorithm\\_\(machine\\_learning\)](https://en.wikipedia.org/wiki/Weighted_majority_algorithm_(machine_learning)).
- [97] *Ensemble learning*. [Online]. [https://en.wikipedia.org/wiki/Ensemble\\_learning](https://en.wikipedia.org/wiki/Ensemble_learning).
- [98] J. Grochowska. *Miary dopasowania*. [Online]. <https://www.etrapez.pl/wp-content/uploads/domowe/ke/Miary%20dopasowania.pdf>.
- [99] V. Yathish. *Loss Functions and Their Use In Neural Networks*. [Online]. <https://towardsdatascience.com/loss-functions-and-their-use-in-neural-networks-a470e703f1e9>.
- [100] M. Bryant. *The Analogy of System Optimization to Least Squares Regression*. [Online]. <http://www.adaptrade.com/BreakoutFutures/Newsletters/Newsletter0503.htm>.
- [101] M. Bryant. *Strategy Quality Metrics Improve Out-of-Sample Results*. [Online]. <http://www.adaptrade.com/Newsletter/NL-QualityMetrics.htm>.
- [102] R. Tadeusiewicz, B. Leper, B. Borowik, T. Gąciarz. *Odkrywanie właściwości sieci neuronowych przy użyciu programów w języku C#*. Kraków: PAU, 2007.



- [103] J. Robinson. *From Basic Gates to Deep Neural Networks: The Definitive Perceptron Tutorial*. [Online]. <https://towardsdatascience.com/the-definitive-perceptron-guide-fd384eb93382#3a73>.
- [104] J. Kepner, V. Gadepally. *Mathematics of Big Data and Machine Learning*. [Online]. <https://www.youtube.com/watch?v=t4K6lney7Zw>.
- [105] M. Grochowski. *Sztuczne sieci neuronowe*. [Online]. [https://www.is.umk.pl/~grochu/wiki/lib/exe/fetch.php?media=zajecia:nn\\_2021\\_1:nn-wyklad.pdf](https://www.is.umk.pl/~grochu/wiki/lib/exe/fetch.php?media=zajecia:nn_2021_1:nn-wyklad.pdf).
- [106] M. Flasiński. *Wstęp do sztucznej inteligencji*. Warszawa: PWN, 2011.
- [107] E. Gately. *Sieci neuronowe, prognozowanie finansowe i projektowanie systemów transakcyjnych*. Warszawa: WIG-Press, 1999.
- [108] J. Krutsinger. *Systemy transakcyjne. Sekrety mistrzów*. Warszawa: WIG-Press, 1998.
- [109] S. Sharma. *Activation Functions in Neural Networks*. [Online]. <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>.
- [110] *Activation function*. [Online]. [https://en.wikipedia.org/wiki/Activation\\_function](https://en.wikipedia.org/wiki/Activation_function).
- [111] P. Baheti. *Activation Functions in Neural Networks*. [Online]. <https://www.v7labs.com/blog/neural-networks-activation-functions>.
- [112] J. Brownlee. *How to Choose an Activation Function for Deep Learning*. [Online]. <https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/>.
- [113] *Metody Heurystyczne – wykład 4*. [Online]. <http://www.imio.polsl.pl/Dopobrania/MH%20ME3%20wyklad%204%20do%20druku.pdf>.
- [114] F. van Veen, S. Leijnen. *The Neural Network ZOO*. [Online]. <https://www.asimovinstitute.org/neural-network-zoo/>.
- [115] A. Pacut. *Sieci neuronowe*. [Online]. <http://zbum.ia.pw.edu.pl/PL/dydaktyka/SNR/PUBLIC/SNR-czI.pdf>.
- [116] K. Piczak. *Klasyfikacja dźwięku za pomocą splotowych sieci neuronowych*. Warszawa: Politechnika Warszawska, Wydz. Elektroniki i Technik Informacyjnych, 2018. [Online]. [https://repo.pw.edu.pl/docstore/download/WUT54648a8a84d1476abf2a16173a0a9e2f/Rozprawa-doktorska\\_Karol-Piczak.pdf](https://repo.pw.edu.pl/docstore/download/WUT54648a8a84d1476abf2a16173a0a9e2f/Rozprawa-doktorska_Karol-Piczak.pdf).
- [117] K. Chandrakant. *Reinforcement Learning with Neural Network*. [Online]. <https://www.baeldung.com/cs/reinforcement-learning-neural-network>.

- [118] S. Xiang. *RGB GAF image: A possible solution to one weak point of Gramian Angular Field Imaging*. [Online]. <https://towardsdatascience.com/rgb-gaf-image-a-possible-solution-to-one-weak-point-of-gramian-angular-field-imaging-ffc6b31edfbc>.
- [119] J. Faouzi. *Data set of Gramian angular fields*. [Online]. [http://pyts.readthedocs.io/en/stable/auto\\_examples/image/plot\\_dataset\\_gaf.html](http://pyts.readthedocs.io/en/stable/auto_examples/image/plot_dataset_gaf.html).
- 

## Selected problems of time series analysis

### Abstract

The article provides an overview of selected problems related to the analysis and prediction of time series containing stock market data. The taxonomy of time series is presented, and the main directions encountered in the analysis of financial data are characterized. Selected tools of technical analysis are described in more detail. Subsequent chapters provide a review of selected machine learning methods, divided into a section on algorithmic methods and a section on neural networks useful in the analysis and prediction of time series.

**Keywords:** *Time Series, Technical Analysis, Prediction, Machine Learning, Neural Networks*