**POLITECHNIKA BIAŁOSTOCKA**

**WYDZIAŁ** **Informatyki**

**Katedra Oprogramowania**

PRACA DYPLOMOWA MAGISTERSKA

TEMAT: **Krótkoterminowa prognoza kierunku zmiany kursów walut z wykorzystaniem uczenia maszynowego.**

WYKONAWCA: Maciej Ziniewicz

*Imię i nazwisko*

PODPIS: .................................

PROMOTOR: dr inż. Jerzy Krawczuk

*Imię i nazwisko*

PODPIS: .................................

**BIAŁYSTOK 2017 ROK**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| POLITECHNIKA BIAŁOSTOCKA | | |  | | | Nr albumu studenta  90563. | |
| Wydział Informatyki | | | Studia stacjonarne  II stopnia inżynierskie | | | Rok akademicki  2015/2016 | |
| Kierunek studiów  Informatyka | |
| Katedra Oprogramowania | | |  | | | Specjalność  Inteligentne technologie internetowe | |
| **Maciej Ziniewicz**  **TEMAT PRACY DYPLOMOWEJ:**  **Krótkoterminowa prognoza kierunku zmiany kursów walut z wykorzystaniem uczenia maszynowego.**  **Zakres pracy:**   1. Przegląd podobnych rozwiązań 2. Projekt systemu 3. Implementacja modułu administracji, zarządzania zawodami, galerii zdjęć. 4. Implementacja systemu z użyciem technologii JavaScript, Java, Spring Boot, Thymeleaf, Boostrap, MySQL   **Słowa kluczowe (max 5):** serwis internetowy, Java | | | | | | | |
| .........................................................................  *Imię i nazwisko, stopień/ tytuł promotora - podpis* | | | | ......................................................................................  *Imię i nazwisko kierownika katedry - podpis* | | | |
| .....................................................*Data wydania tematu pracy dyplomowej*  *- podpis promotora* | | ...................................................  *Regulaminowy termin złożenia pracy dyplomowej* | | | ........................................................  *Data złożenia pracy dyplomowej*  *- potwierdzenie dziekanatu* | | |
|  | .....................................  *Ocena promotora* | | | ..................................  *Podpis promotora* | | |  |
| ...................................................  *Imię i nazwisko, stopień/ tytuł recenzenta* | | ...........................................  *Ocena recenzenta* | | | | ................................................  *Podpis recenzenta* | |

**Karta dyplomowa**

Thesis topic:

Web application for supporting sport club activity.

SUMMARY

The aim of the project was creating a web application for supporting sport club activity. The motivation behind choosing the thesis was my membership in Parkour Białystok sport club to which the application is dedicated. This project allows to manage various articles, members, users, photo gallery and competitions organized by sport club.

This application was created using modern technologies, these are most important: Spring Framework and based-on child projects like Spring Boot, Thymeleaf, JQuery, Bootstrap.

This thesis consists of six chapters. The first chapter is an introduction. The second presents similar solutions. The next one presents and describes technologies which were used to create application. All diagrams and schemas created during planning process are described in chapter four. Chapter five is most important as it contains the technical and implementation details. These include a description of URL mapping, graphic interface, forms, server implementation, configuration and security aspect. The last chapter is a summary of work.

Załącznik nr 4 do „Zasad postępowania przy przygotowaniu i obronie pracy dyplomowej w PB”

Załącznik nr 4 do „Zasad postępowania przy przygotowaniu i obronie pracy dyplomowej w PB”

Maciej Ziniewicz Białystok, dnia………………………

imię i nazwisko studenta

90563

nr albumu

Informatyka stacjonarne

kierunek i forma studiów

dr inż. Jerzy Krawczuk

promotor pracy dyplomowej

**OŚWIADCZENIE**

Przedkładając w roku akademickim 2015./2016. Promotorowi dr inż. Jackowi Grekowowi pracę dyplomową pt.: Aplikacja webowa wspomagająca działanie klubu sportowego dalej zwaną pracą dyplomową,

**oświadczam, że:**

1. praca dyplomowa stanowi wynik samodzielnej pracy twórczej,
2. wykorzystując w pracy dyplomowej materiały źródłowe, w tym w szczególności: monografie, artykuły naukowe, zestawienia zawierające wyniki badań (opublikowane, jak i nieopublikowane), materiały ze stron internetowych, w przypisach wskazywałem ich autora, tytuł, miejsce i rok publikacji oraz stronę, z której pochodzą powoływane fragmenty, ponadto w pracy dyplomowej zamieściłem bibliografię,
3. praca dyplomowa nie zawiera żadnych danych, informacji i materiałów, których publikacja nie jest prawnie dozwolona,
4. praca dyplomowa dotychczas nie stanowiła podstawy nadania tytułu zawodowego, stopnia naukowego, tytułu naukowego oraz uzyskania innych kwalifikacji,
5. treść pracy dyplomowej przekazanej do dziekanatu Wydziału Informatyki jest jednakowa w wersji drukowanej oraz w formie elektronicznej,
6. jestem świadomy/a, że naruszenie praw autorskich podlega odpowiedzialności na podstawie przepisów ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t. j.: Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.), jednocześnie na podstawie przepisów ustawy z dnia 27 lipca 2005 r. Prawo   
   o szkolnictwie wyższym (t. j.: Dz. U. z 2012 r. poz. 572 z późn. zm.) stanowi przesłankę wszczęcia postępowania dyscyplinarnego oraz stwierdzenia nieważności postępowania w sprawie nadania tytułu zawodowego,
7. udzielam Politechnice Białostockiej nieodpłatnie licencji na korzystanie z pracy dyplomowej w celu realizacji przeprowadzenia procedury antyplagiatowej przyjętej w Uczelni oraz na przekazanie pracy do Ogólnopolskiego Repozytorium Prac Dyplomowych, jak również udostępnianie i przechowywanie jej w Bibliotece Politechniki Białostockiej przez okres 50 lat od obrony pracy dyplomowej.

….……………………………………….

czytelny podpis studenta

Na podstawie art. 24 ust. 1 ustawy z dnia 29 sierpnia 1997 roku o ochronie danych osobowych (j.t. z 2014 r., poz. 1182 z późn. zm.) informuję, że administratorem danych jest Politechnika Białostocka, ul. Wiejska 45A, 15-351 Białystok. Dane będą przetwarzane w celach realizacji procedury antyplagiatowej przyjętej w Politechnice Białostockiej i nie będą udostępniane odbiorcom danych w rozumieniu art. 7 pkt 6 ustawy o ochronie danych osobowych. Osobie, której dane dotyczą, przysługuje prawo dostępu do treści swoich danych oraz ich poprawiania. Podanie danych jest obowiązkowe (art. 167b ustawy z dnia 27 lipca 2005 roku Prawo o szkolnictwie wyższym j.t. Dz.U. z 2012 r., poz. 572 z późn. zm.).

Białystok, dnia………………………

imię i nazwisko studenta

................................................

nr albumu

...............................................

kierunek i forma studiów

...............................................

promotor pracy dyplomowej

**OŚWIADCZENIE**

Przedkładając w roku akademickim 2015./2016. Promotorowi ………………………….……………………….. pracę dyplomową pt.: ...............…………………………………………………….……………………………………….. ………………….………………………..………………………………………..……………………………...………….…….…………………………………….........................…………………………………, dalej zwaną pracą dyplomową,

**oświadczam, że:**

1. praca dyplomowa stanowi wynik samodzielnej pracy twórczej,
2. wykorzystując w pracy dyplomowej materiały źródłowe, w tym w szczególności: monografie, artykuły naukowe, zestawienia zawierające wyniki badań (opublikowane, jak i nieopublikowane), materiały ze stron internetowych, w przypisach wskazywałem/am ich autora, tytuł, miejsce i rok publikacji oraz stronę, z której pochodzą powoływane fragmenty, ponadto w pracy dyplomowej zamieściłem/am bibliografię,
3. praca dyplomowa nie zawiera żadnych danych, informacji i materiałów, których publikacja nie jest prawnie dozwolona,
4. praca dyplomowa dotychczas nie stanowiła podstawy nadania tytułu zawodowego, stopnia naukowego, tytułu naukowego oraz uzyskania innych kwalifikacji,
5. treść pracy dyplomowej przekazanej do dziekanatu Wydziału Informatyki jest jednakowa w wersji drukowanej oraz w formie elektronicznej,
6. jestem świadomy/a, że naruszenie praw autorskich podlega odpowiedzialności na podstawie przepisów ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t. j.: Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.), jednocześnie na podstawie przepisów ustawy z dnia 27 lipca 2005 r. Prawo   
   o szkolnictwie wyższym (t. j.: Dz. U. z 2012 r. poz. 572 z późn. zm.) stanowi przesłankę wszczęcia postępowania dyscyplinarnego oraz stwierdzenia nieważności postępowania w sprawie nadania tytułu zawodowego,
7. udzielam Politechnice Białostockiej nieodpłatnie licencji na korzystanie z pracy dyplomowej w celu realizacji przeprowadzenia procedury antyplagiatowej przyjętej w Uczelni oraz na przekazanie pracy do Ogólnopolskiego Repozytorium Prac Dyplomowych, jak również udostępnianie i przechowywanie jej w Bibliotece Politechniki Białostockiej przez okres 50 lat od obrony pracy dyplomowej.

Spis treści

[1. Wstęp 7](#_Toc489995388)

[2. Predykcja rynków 7](#_Toc489995389)

[2.1 Rynek Forex 7](#_Toc489995390)

[2.1.1 Pary walutowe 8](#_Toc489995391)

[2.2 Analiza fundamentalna 9](#_Toc489995392)

[2.3 Analiza techniczna 11](#_Toc489995393)

[2.3.1 Świece japońskie 12](#_Toc489995397)

[2.3.2 Analiza trendu 15](#_Toc489995398)

[2.3.3 Wskaźniki 16](#_Toc489995399)

[2.4 Uczenie maszynowe 23](#_Toc489995400)

[2.4.1 Pojęcie uczenia się w kontekscie maszynowym 23](#_Toc489995401)

[2.4.2 Podstawowe pojęcia 24](#_Toc489995402)

[2.4.3 Przedstawienie badanych algorytmów 25](#_Toc489995403)

[3. Opis wykorzystanych narzędzi 28](#_Toc489995404)

[3.1 Technical Analysis for Java (Ta4j) 28](#_Toc489995405)

[3.1.1 Podstawy korzystania z TA4j 28](#_Toc489995406)

[3.2 Weka 29](#_Toc489995407)

[3.2.1 Wstępne przetworzenie danych 30](#_Toc489995408)

[3.2.2 Klasyfikacja danych 31](#_Toc489995409)

[3.2.3 Pozostałe funkcjonalności 32](#_Toc489995410)

[4. Metodologia 34](#_Toc489995411)

[4.1 Ogólny przebieg badań 34](#_Toc489995412)

[4.2 Omówienie użytych danych 35](#_Toc489995413)

[4.3 Omówienie procesu obróbki danych do sygnałow 37](#_Toc489995414)

[4.3.1 Tworzenie świec japońskich 37](#_Toc489995415)

[4.3.2 Generowanie sygnałów 38](#_Toc489995416)

[4.3.3 Wygenerowane dane 40](#_Toc489995417)

[4.4 Omówienie procesu badań 41](#_Toc489995418)

[4.4.1 Wznaczenie atrybutów 42](#_Toc489995419)

[4.4.2 Filtrowanie danych 42](#_Toc489995420)

[4.4.3 Klasyfikacja 42](#_Toc489995421)

[4.4.4 Testowanie 42](#_Toc489995422)

[5. Model 42](#_Toc489995423)

[6. Wnioski 43](#_Toc489995424)

[7. Bibliografia 43](#_Toc489995425)

# Wstęp

# Predykcja rynków

Poniższy podrozdział stanowi wprowadzenie teoretyczne do pojęć i technik jakie wiążą się z predykcją rynków giełdowych. Pierwszy podrozdział 2.1 jest wprowadzeniem teoretycznym do rynku forex, przedstawia jego krótki opis oraz pis par walutowych.Kolejny podrozdział 2.2 przedstawia pojęcia analizy fundamentalnej, natomiast podrozdział 2.3 opisuje analizę techniczną oraz techniki z nią związane. Ostatni podrozdział przedstawia dziedzinę jaką jest uczenie maszynowe, podstawowe pojęcia jakie są niezbędne przy pracy z uczeniem maszynowym oraz wybrane algorytmy.

## Rynek Forex

Foregin Exchange Market częściej nazywany jak Forex (ang. Foregin Exchange) jest to największy na świecie rynek walutowy gdzie rządy, banki centralne, maklerzy i inni inwestorzy wykonują operacje wymiany walut. Początkowo tylko wielkie i bogate podmioty handlowe miały możliwość użytkowania rynku Forex jednak z przyjściem I popularyzacją internetu możliwość tą otrzymali również przeciętni ludzie, którzy mogą teraz podejmować decyzje inwestycyjne nie wstając od komputera. Rynek ten będzie funkcjonował tak długo jak ludzie będą używali walut. Otwarty 24 godziny na dobę, a większość platform inwestycyjnych umożliwiających wykonywanie inwestycji dla zwykłych osób jest czynna od 22:00 GMT w niedzielę do 22:00 GMT piątek. Handel odbywa się tam w różnych sesjach handlowych : azjatyckiej, amerykańskiej I europejskiej przez cały czas pomijając weekendy oraz podczas ważnych świąt. Każda waluta na forex jest oznaczona spejcalnym trzyliterowym kodem według norm ISO przykładowo euro to EUR a dolar amerykański to USD. Notowania są mierzone w dokładności do 1 pip (ang. Price Interest Point) I jest to najmniejsza wartość zmiany jaka może nastąpić w cenie danje waluty. Dla większości walut pip wynosi 0,0001 część całości np. 3,5212[[1]](#footnote-1).

### Pary walutowe

Aby wykonywać transkacje handlowe z użyciem walut na rynku forex należy znać ich wartość, którą określa się w porównaniu do innej waluty. Przykładowo wartość euro do złotkówki 4.23 oznacza że 1 euro jest warty 4.23 złotego, czy wartość euro do dolara 1.2 oznacza że jeden euro jest warty 1.2 dolara amerykańskiego. Porównania w taki sposób różnych walut tworzą pary walutowe EUR/PLN o kursie 4.23 czy jedna z najpopularniejszych EUR/USD o kursie 1.2. Walutę pierwszą w parze walutowej nazywamy walutą bazową, natomiast drugą walutą kwotowaną.

Wyróżniane są trzy podstawowe typt walut[[2]](#footnote-2):

1. **Głowne** – są wszystkie pary odnoszące się do dolata amerykańskiego po jednej ze strony kwotoweanej lub bazowej. Głowne pary z racji że są najczęsciej używane wyróżniają się wysoką płynnością oraz niskimi kosztami transakcji. Transakcje dotyczące par z tej kategorii stanowią większość transakcji na rynku forex.

|  |  |  |
| --- | --- | --- |
| Para | Kraje | Udział % w forex |
| USD/EUR | Stany Zjednoczone/Strefa Euro | 24.1 |
| USD/JPY | Stany Zjednoczone/Japonia | 18.3 |
| USD/GBP | Stany Zjednoczone/Wielka Brytania | 8.8 |
| USD/AUD | Stany Zjednoczone/Australia | 6.8 |
| USD/CAD | Stany Zjednoczone/Kanada | 3.7 |
| USD/CHF | Stany Zjednoczone/ Szwajcaria | 3.4 |

Tabela 2.1 Głowne pary walutowe dane z 2013  
 Źródłó: opracowanie własne na podstawie <http://www.bis.org/publ/rpfx13fx.pdf> (stan na 27.06.2017)

1. **Krzyżowe** – są to pary które nie odnoszą się do dolara amerykańskiego. Pary krzyżowe nie są tak popularne jak główne przez co koszty handlu nimi są wieksze.

|  |  |
| --- | --- |
| Para | Kraje |
| EUR/AUD | Strefa Euro / Kanada |
| EUR/JPY | Strefa Euro / Japonia |
| EUR/NZD | Strefa Euro / Nowa Zelandia |
| CHF/JPY | Szajcaria / Japonia |
| GBP/AUD | Wielka Brytania / Kanada |

Tabela 2.2 Krzyżowe pary walutowe   
Źródłó: Opracowanie własne

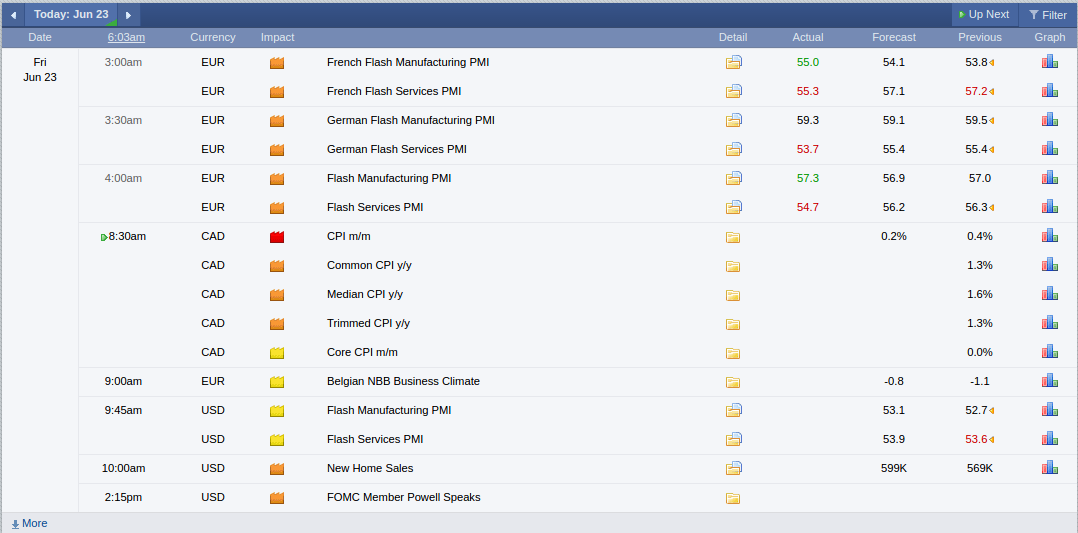
1. **Egzotyczne** - są to pary walutowe składające się z jednej z głownych bardziej znaczących walut do walut które są mniej znaczące ekonomicznie w skali globalnej. Handel z użyciem tych par walutowych stanowi najmniejszą część rynku forex.

|  |  |
| --- | --- |
| Para | Kraje |
| EUR/PLN | Strefa Euro / Polska |
| EUR/HKD | Strefa Euro / Hongkong |
| USD/SGD | Stany Zjednocznoe / Singapur |
| USD/ZAR | Stany Zjednocznone / Afryka Południowa |

Tabela 2.3Egzotyczne pary walutowe   
Źródłó: Opracowanie własne

## Analiza fundamentalna

Jest to jedna z technik analizy rynków. Głównym zadaniem analizy fundamentalnej jest śledzenie i analiza czynników zewnętrznych mających wpływ na popyt i podaż takich jak decyzje banków, przemowy ważnch polityków, prezesów dużych firm, publikacje raportów. Celem jest by na podstawie takich analizy takich infomracji zareagować i wykonać odpowiednią transkacje zanim informacje wpłyną na rynek. Wszelkiego rodzaju nowe wiadomości są ważne z punktu widzenia analizy fundamentalnej ponieważ mają potencjalny wpływ na rynek[[3]](#footnote-3). Tego rodzaju analizy są stosowane głownie w długoterminowych inwestycjach oraz częściej to pojęcie jest spotykane przy rynkach papierów wartościowych gdy określa się wartość spółek. Nie znaczy to jednak że analiza fundamentalna nie jest stosowana na rynku forex. Podstawowym narzędziem jakie jest używane podczas inwestycji na giełdzie walutowej jest kalendarz ekonomiczny gdzie można odnaleźć informacje na temat wydarzeń mających wpływ na poszczególne waluty.



Rysunek 2.1 Kalenadrz ekonomiczny   
Źródło: www.forexfactory.com

Na powyższym obrazku widać przykładowy kalendarz ekonomiczny dla dnia 23 czerwca 2017r. Wydarzenia podzielone są w nim początkowo ze względu na godzinę o której występują. Dla każdej godziny przyporządkowana jest lista wydarzeń które wtedy będą miały miejsce. Każde wydażenie ma podaną walutę na jaką będzie miało wpływ, określoną siłę wpływu w tym przypadku trzema różnymi kolorami ( w innych kalendarzach może być to inaczej prezentowane ale zazwyczaj jest to prezentacja w trzech stopniach wpływu):

* Żółty – najmniejszy wpływ
* Pomarańczowy – średni wpływ
* Czerwony – duży wpływ

rodzaj wydarzenia, dodatkowe szczegóły, realny wpływ jeżeli wydarzenie już miało miejsce, prognozę , oraz poprzedni wpływ.

## Analiza techniczna

Analiza techniczna jest to kolejna z podstawowych technik analizy rynków, a jej definicja opisuje że jest to „badanie zachowań rynku, przede wszystkim przy użyciu wykresów, którego celem, jest przewidywanie przyszłych trendów cenowych”[[4]](#footnote-4). Przedwszystkim w analiza techniczna polega na odczytywaniu informacji z wykresów instrumentów giełodwych, w tym przypadku będzie mowa o walutach. Na podstawie obserwacji zachowań ceny m.in: nagłych skoków, utworzonych formacji możliwe jest by określić przyszły trend. Cała koncepcja analizy technicznej opiera się o trzy przesłanki:

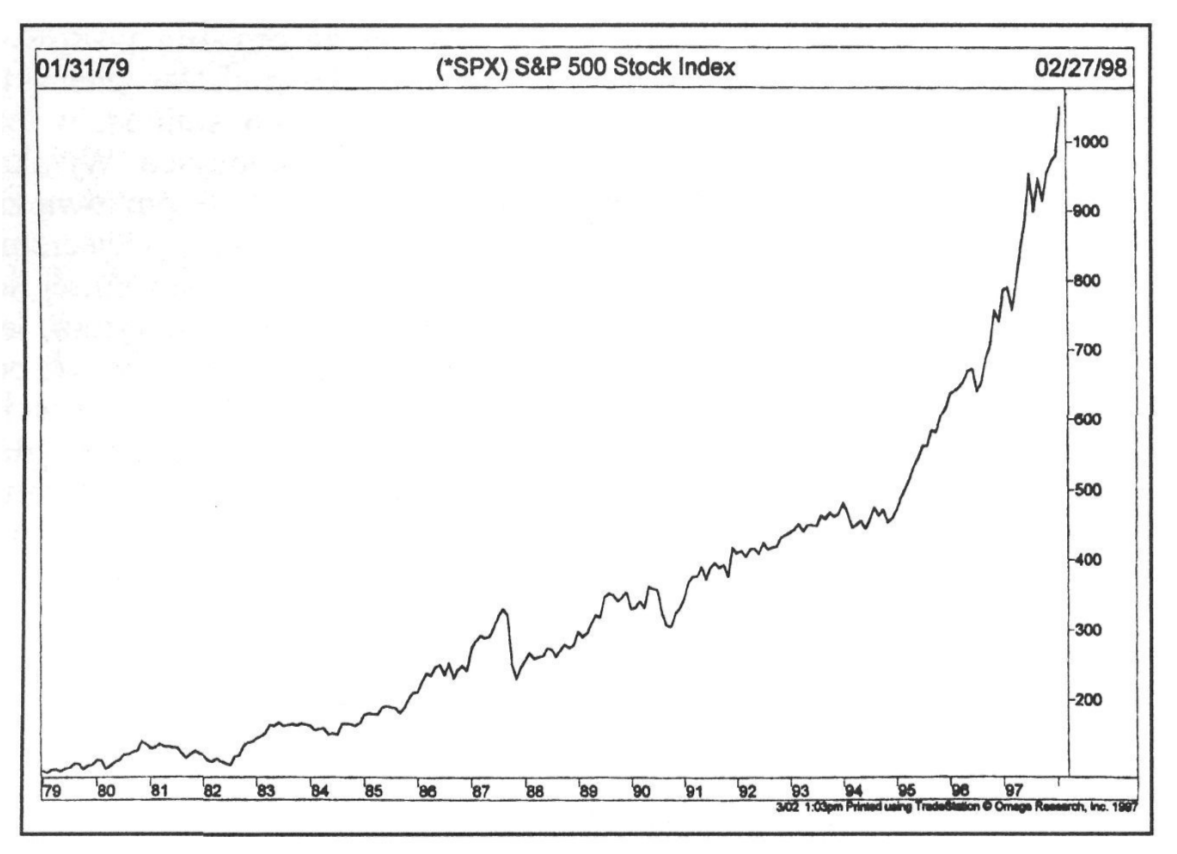
* Rynek dyskontuje wszystko
* Ceny podlegają trendom
* Historia się powtarza[[5]](#footnote-5)

### **Rynek dyskontuje wszystko**

Ta przeslanka jest podstawą analizy technicznej z której wynikają kolejne[[6]](#footnote-6). Zakłada ona że wszystkie czynniki, od fundamentalnych po psychologię rynku, mające wpływ na cenę znajdują odbicie w jej zachowaniu. Wynika z tego że do określenia przyszłego zachowania rynku wystarczy badanie zachowania cen, i nie trzeba badać każdego z czynników oddzielnie ponieważ wykres odzwierciedla je wszystnie na raz. Zgodnie z tą tezą jeżeli cena rosnie to popyt przewyższa podaż a sytuacja na rynku musi być sprzyjająca wzrostom, natomiast jeżeli cena spada to analitych wychodzi z założenia że podaż przewyższa popyt a sytacja rynkowa sprzyja spadkom.

### **Ceny podlegają trendom**

Kolejne podstawowe założenie technicznej mówi że ceny podlegają trendom krótko, średnio, długo terminowym. Rynek wykazuje tendencje do kontynuowania trendu niż do jego zmiany[[7]](#footnote-7). Innymi słowy duża część analizy technicznej sprowada się do tego by badać trendy, oraz wykrywać ich zmiany wystarczająco wczesnej fazie co pozwoli dokonywać zyskowne transakcje.



Rysunek 2.2 Trend wzrostowy  
Źródło: Murphy, Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s 4

### **Historia się powtarza**

Analiza techniczna w dużym stopniu powiązana jest z badaniem ludzkiej psychiki, ponieważ ludzie mają tendencję do powtarzalności swoich zachowań. Formacje swiecowe znane sa od lat, odzwierciedlają one zachowania ludzkie. Powtarzalność zachowań ceny jest podstawą tej tezy, ponieważ jeżeli formacje cenowe sprawdziły się w przeszłości to zakłada się że będą skuteczne również w przyszłości, dlatego analiza techniczna używa narzędzi do badań przeszłości aby określić zachowanie ceny w przyszłości.

### Świece japońskie

Świece japońskie jest to jedna z trzech najpopularniejszych metod przedstawiania wykresów rynków. Jak sama nazwa wskazuje świece te pochodzą z Japonii a pierwsze informacje o ich użyciu sięgają XVIII w, gdzie były używane przez japońskiego kupca ryżu nazywającego się Munehisa Homma. Który w póznijeszym czasie został konsultantem finansowym rządu japońskiego oraz honorowo mianowany samurajem, a jego wskazówki handlowe i książka wpłyneły na pózniejsze ukształtowanie się obecnie używanych świec[[8]](#footnote-8).

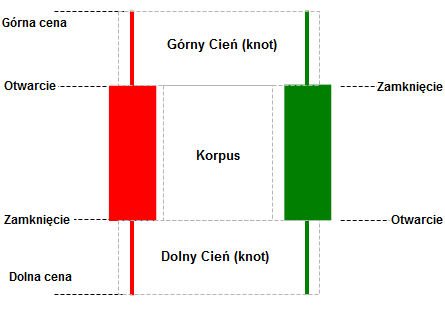
Wykres świecowy daje największą ilość informacji, w porównaniu do dwóch innych również często używanych czyli liniowego oraz słupkowego.



Rysunek 2.3 Rodzaje wykresów  
Źródło: opracowanie własne

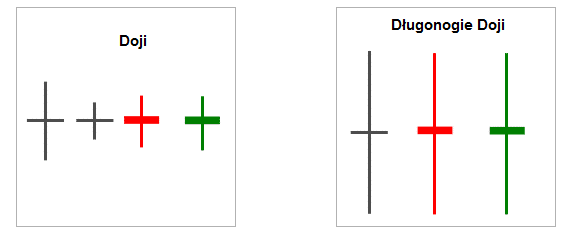
Tak jak widać na powyższym rysunku (rys. 2.3) wykres liniowy dostarcza najmnieszą ilość informacji dla inwestora ponieważ łączy on ceny zamknięcia z każdego przedziału czasowego, to znaczy że inwestor otrzymuje informacje o cenie otwarcia czyli cenie początkowej w określonym przedziale czasowym oraz cenie zamknięcia, czyli cenie końcowej. Wykres słupkowy dostarcza nam dodatkowo informacje o maksymach jakie wystąpiły podczas ruchu ceny w określonym przedziale czasowym, czyli wartość maksymalną i minimalną jakie wystąpiły w tym czasie. Podobnie jak wykresy słupkowe wykresy świecowe dostarczają informacje o cenie otwarcia, zamknięcia oraz maksymalnej i minimalnej wartości ceny w przedziale czasowym.

Świece japońskie składają się z korpusu czyli zakresu od ceny otwarcia do ceny zamknięcia. Kolor świecy rownież ma znaczenie, standardowo są to kolory biały zastąpiony zielonym, oraz czarny zastąpiony czerwonym. Jeżeli kolor świecy jest czerwony to cena zamknięcia jest niższa niż cena otawrcia i jest to tak zwana świeca spadkowa, natomiast jeżeli świeca jest zielona to cena zamknięcia jest wyższa niż cena otwarcia i jest to świeca wzrostowa. Dodatkowymi informacjami które dostarczane są przez świece są jej cienie, a informują one o ekstremach ceny występujących w czasie przedziału czasowego prezentowanego przez świecę. Cień górny informuje o maksymalnej cenie jak wystąpiła podczas przedziału, a cień dolny informuje o cenie minimalnej[[9]](#footnote-9).



Rysunek 2.4 Świeca spadkowa oraz wzrostowa  
Źródło: https://comparic.pl/swiece-japonskie-podstawy/ (dostęp na 08.07.2017)

Wygląd pojedyńczych świec może również stanowić dodatkową informację, ale również sposób w jaki ustawiają się świece kolejno po sobie jest wartościową informacją dla inwestora, nazwane jest to formacjami. Najbardziej podstawowym oraz niemniej ważnym i znaczącym wzorcem jest tzw „doji” czyli świeca o tej samej cenie otwarcia i zamknięcia lub o bardzo małej różnicy między tymi cieniami, sygnalizuje ona niezdecydowanie rynku i możliwą zmianę trendu[[10]](#footnote-10). Doji mogą różnić się długością cienia.



Rysunek 2.5 Wzorzec doji  
Źródło: https://comparic.pl/swiece-japonskie-podstawy/ (dostęp na 08.07.2017)

### Analiza trendu

Pojęcie trendu jest absolutną podstawą przy analizie technicznej, ze względu na to że wszystkie narzędzia którymi posługuje się inwestor służą do badania trendów. Przy inwestowaniu można często spotkać się z powiedzeniami że walka z trendem czy inwestycje w kierunku przeciwnym niż idzie trend nie maja sensu, dlatego zrozumienie pojęcia trendu jest tak ważne. Trend można zdefiniować jako „kierunek, jaki

przyjmują szczyty i dołki”[[11]](#footnote-11), co znaczy że trend wzrostowy można zdefiniować jako szczyty i dołki kolejno występujące po sobie coraz wyżej o coraz wyższych wartościach, natomiast trend spadkowy byłyby to kolejno występujące po sobie szczyty i dołki położone coraz niżej.



Rysunek 2.6 Trend wzrostowy i spadkowy  
Źródło: opracowanie własne

Istnieje jeszcze trend boczny czyli szczyty i dołki które następują po sobie horyzontalnie. W taki sposób że różnice pomiędzy ich wartościami są niezaczne, i cena nie posiada szczególnego kierunku zmiany. Taka sytuacja również określana jest czasem jako brak trendu ze względu na to że cena nie podąża ani do góry ani do dołu.



Rysunek 2.7 Trend boczny  
Źródło: opracowanie własne

W trendach wyróżnia sięnie tylko kierunek ale również na rodzaj. Mianowicie trend określając ze względu na rodzaj może być głowny, średnio lub krótko okresowy. Trend głowny jest to najdłuższy trend, i żeby określić trend jako głowny według teroii Dowa trwać on musi conajmniej rok. Trend średnio okresowy jest częscią trendu głownego, tak jak trend krótko okresowy jest częścią trendu średniookresowego i głownego. Generalnie w większych trendach wsytępują mniejsze trendy korygujące jednak wyróżniane są najczęściej trzy powyższe. Przykładowo gdy trend głowny jest wzrostowy, może wystąpić krótszy trend średnio okresowy boczny bądź spadkowy który obniży cenę następnie jak się skończy cena będzie nadal się poruszała zgodnie z trendem głownym[[12]](#footnote-12).

### Wskaźniki

Wskaźniki używane są podstawowym narzędziem anaizy technicznej obok interpretacji wykresów, świec czy występujących formacji. Jest to statystyczne podejście do analizy technicznej, oraz znacznie wspomaga proces decyzyjny podczas klasycznej subiektywnej analizy technicznej czyli interpretacji tego co inwestor widzi na wykresach. Wskazniki generują dodatkowe informacje odnosnie sytuacji na rynku oraz pomagają w identyfikacji stanów w jakich znajduje się rynek. Są to narzędzia analizujące dane trudno widoczne na pierwszy rzut oka, czyli przepływy pienięzne, trendy, zmienność i dynamikę, oraz reprezentują je za pomocą wykresów na podstawie których możlwe jest generowanie sygnałów. Wskaźniki mimo to że mają wsperać inwestora przy podejmowaniu decyzji, mogą przynieść odwrotny efekt dlatego nie nalezy traktować ich jako podstawę do decyzji oraz korzystanie ze zbyt dużej ilości wskazników może wprowadzić w błąd.

Wskaźniki możemy podzielić na dwa typy[[13]](#footnote-13):

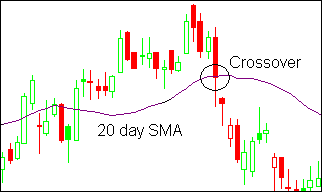
1. Wiodące (ang. leading) – wskaźniki poprzedzające ruchy ceny oraz próbujące je przewidzieć. Przydają się podczas trendów bocznych, gdy rynek nie ma widocznych trendów.
2. Opóźnienia (ang. lagging) – wskaźniki podążające za ruchem ceny, śledzące ją. Służą najczęsciej jako potwierdzenia wspomagająe decyzje. Przydatne podczas trendów.

Poza ogolną klasyfikacją wskaźniki można grupować na podstawie tego co wskazują, np.: zmienność ceny, siła ceny, średnie ruchome, popyt, wolumen, itd. Na podstawie wskazników można wywnioskować lub wygenerować sygnał zakupu/sprzedaży.

#### Oscylaotry

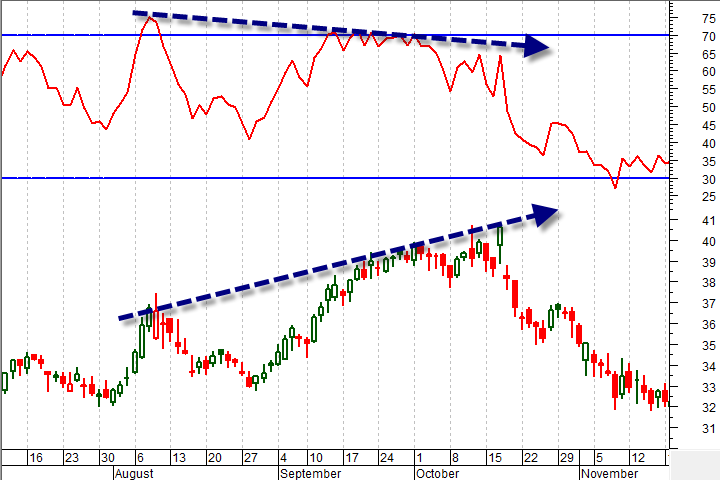
Szczególnej uwagi wymagają oscylatory. Wskazniki należące do tej rodziny są pomocniczne przy analizie trendu, którego skuteczność jest największa gdy nie ma dużych ruchów cen. Konstrukcja wszystkich oscylatorów jest podobna, jest to rysowany na bierząco wykres o ośi na poziomie zero która dzieli wykreś na górną i dolną połowke lub o maksymalnej wartości 100 i minimalnej 0. Oscylatory sygnalizują możliwość zmiany kierunku ruchu ceny w momencie gdy wykres zbliża się do ekstremalnych wartości oscylatora. Gdy wartość oscylatora jest blisko maksimum mowa jest o wykupieniu rynku, natomiast gdy jego wartość zbliża się do minimum mówimy o wysprzedaniu rynku[[14]](#footnote-14).

Najczęsciej poszukiwane na oscylatorach są przecięcia linii (ang. crossovers) oraz dywergencje (ang. divergence). Przecięcia linii używane są często w odniesienu do średnich ruchomych czy wstęg bollingera, wskazniki te można zaliczyć do wskazników opóznienia ponieważ sledzą one ruch ceny. Cena częściej wacha się i idzie w jakimś kierunku rysując wykres w kształcie sinusoidy niż po linni prostej. Przy takim ruchu gdy cena przetnie linie rysowaną przez wskaznik opóznienia może być to interpretowane jako sygnał odwrócenia trendu.[[15]](#footnote-15) Poniżej przykład przecięcia linii dwudziestodniowej średniej kroczącej.



Rysunek 2.8 Przecięcie linii średniej kroczącej  
Źródło: http://www.investopedia.com/terms/c/crossover.asp (dosęp 11.07.2017)

Dywergencje są to sytuacje kiedy wskaznik i wykres ceny podążają w przeciwnych kierunkach. Przykładowo gdy cena rosnie, a wartość wskaznika spada może być to interpretowane jako sygnał odwrócenia trendu. Dywergencja jest silniejsza im wieksza jest rozbierzność między wskaznikiem a wykresem[[16]](#footnote-16).

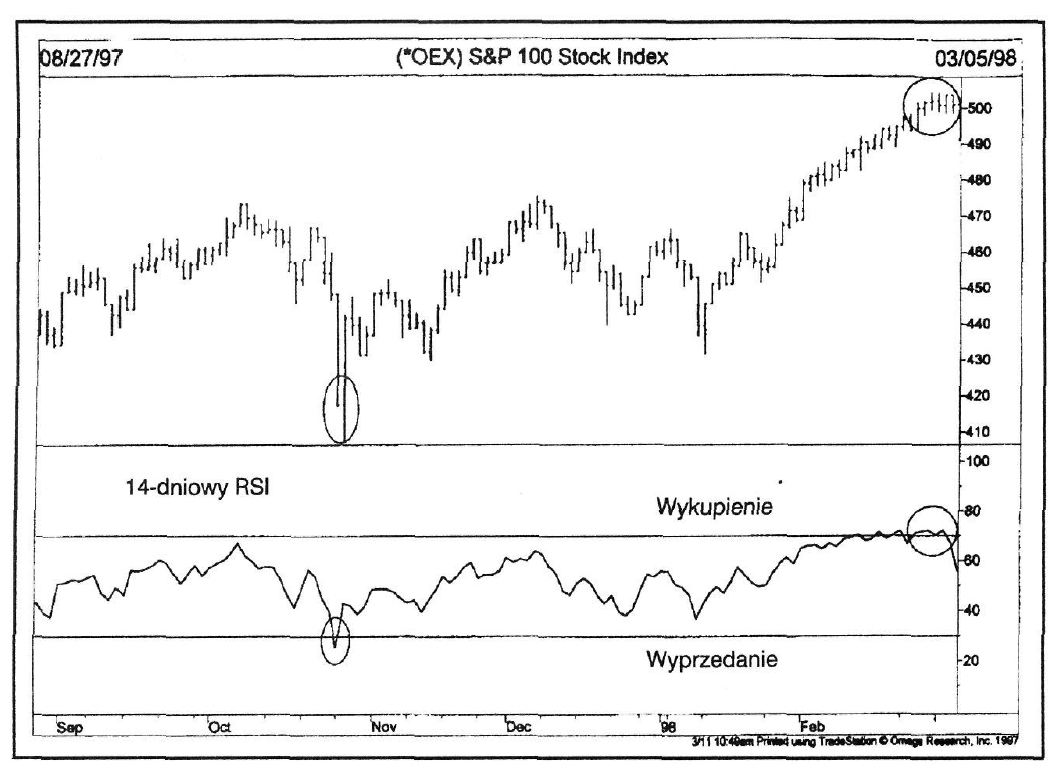


Rysunek 2.9 Dywergencja  
Źródło: https://www.learningmarkets.com/how-to-trade-bullish-and-bearish-technical-divergences/ (dostęp 11.07.2017)

Poniżej zostaną opisane wskaźniki używane do stworzenia w prezentowanej pracy:

#### RSI – Relative Strength Index

Jest to jeden z najpopularniejszych wskaźników siły względnej należący do oscylatorów. Zakres wartości wskaznika to od 0 do 100, stały zakres daje możliwość dokonywania porównań. Wadą tego wskaźnika jest to że podczs silnych trendów i gwałtownych zmian ceny jego skuteczność mocno spada. Zdefiniowane są w nim poziomy domyślne 70 i 30 przekroczenie których uważane jest za sygnał wykupienia bądz wyprzedania rynku.



Rysunek 2.10 Wskaznik RSI  
Źródłó: Murphy, Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s 210

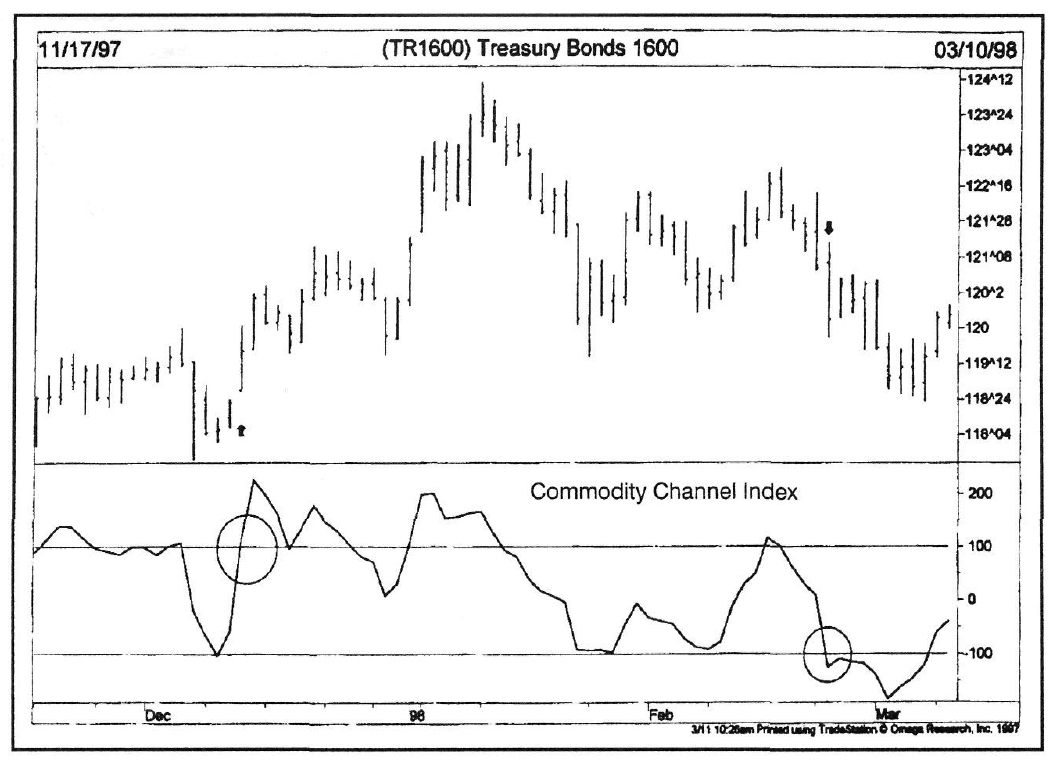
Do inicjalizacji wskaznika podajemy okres w postaci cyfry, przykładowo 14 to okres domyślny oraz najpopularniejszy i oznacz on że do obliczeń przyjmowany jest okres 14 dni. Do obliczania wartości wskaźnika używany jest następujący wzór:

Do otrzymania średniej wartości wzostu cen trzeba dodać wszystkie punkty zyskane podczas wzrostów, pod czas przyjętego okresu (w tym przypadku 14 dni) i podzielić ich sumę przez przyjęty okres czyli 14. Aby potrzymać średnią wartość spadku analogicznie należy zsumować wszystkie punkty straty podczas dni spadkowych i podzielić je przez 14. Siłą względna czyki RS obliczana jest poprzez podzielenie średniej wzrostu przez średnią spadku, a następnie ta wartość jest podstawiana pod wzór na obliczenie RSI.

Oscylator ten jest tym bardziej czuły i większa jest jego amplituda im jego okres jest krótszy. Dzięki temu do analizy krótkoterminowej można obniżyć okres by zwiększyć czułość oscylatora, natomiast aby uniknąć zakłóceń przy analizie długoterminowej można zwiększyć okres wskaźnika[[17]](#footnote-17).

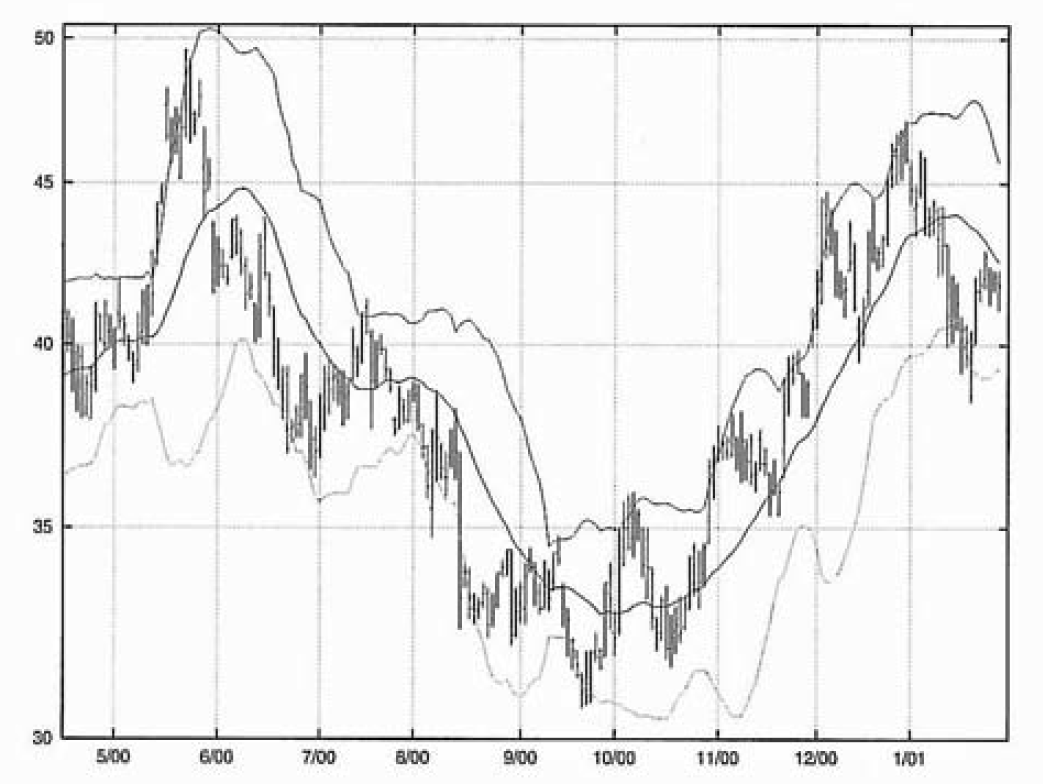
#### CCI - Commodity Channel Index

Jest to skonstruowany przez Donalda R. Lamberta wskaźnik stworzony do wykrywania początków i końców trendów na giełdach towarowych ale jest stosowany również na innych. Porównuje on aktualną cenę średnią ze średnią dla podanego okresu, domyślnie jest to 20 dni. Po tym otrzymane wartości oscylatora są normalizowane za pomocą dzielnika bazującego na przeciętnym odchyleniu. Wynikiem tego jest wartość wskaznika CCI oscylująca zazwyczaj od od +100 do -100. Twórca wskaźnia zaleca przyjmowanie długich pozycji kupna gdy CCI wskazuje wartość powyżej +100 aż do momentu gdy wróci do zakresu, oraz krótkich sprzedaży gdy spada poniżej -100[[18]](#footnote-18).



Rysunek 2.11 Wskaźnik CCI  
Źródłó: Murphy, Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s 208

#### Wstęgi Bollingera

Wstęgi Bollingera (ang. Bollinger Bands) to wskaźnik rysujący wstęgi zgodnie ze strukturą ceny. Jego celem jest zdefiniowanie względnych definicji cen wysokich i niskich. Co oznacza że gdy cena znajduje sie w pobliżu górnej wstęgi jest wysoka i można spodziewać się jej spadku, natomiast gdy cena znajduje się w pobliżu dolenj wstęgi jej wartość jest niska i można spodziewać się jej wzrostów. 

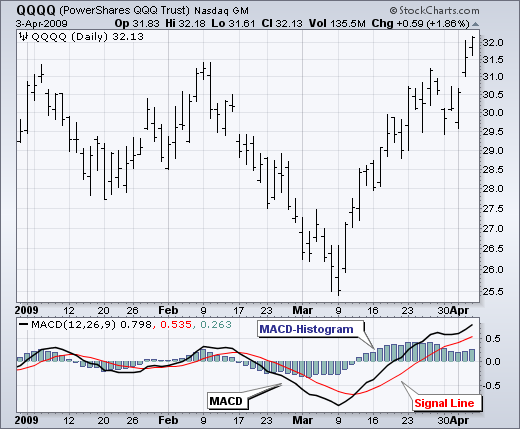
Rysunek 2.12 Bollinger Bands  
Źródło: Bollinger, Bollinger on Bollinger Bands, McGraw-Hill New York 2001, s xxi

Podstawą wstęg Bollingera jest wskaźnik średniej ruchomej, stanowi on również środkową wstęgę opisywanego wskaźnika. Domyślna wrartość śrendniej ruchomej użytej w wstęgach Bollingera to 20 okresów. Serokość wstęg definiowana jest przez miarę zmienności określanej jako odchylenie standardowe. Dane do obliczenia zmienności to te same dane które były użyte do obliczeń średniej ruchomej. Wstęga górna i dolna rysowane są w odległości dwóch odchyleń standardowych od średniej.

Analiza z pomocą wstęg bollingera sprowadza się do odczytywania formacji występujących na rynku w okolicy wstęg , analizy przecięć ceny z wstęgami czy wybić ceny poza wstęgi[[19]](#footnote-19).

#### MACD - Moving Average Convergence/Divergence

Wskaźnik wynalazł Gerald Appel pod koniec lat siedemdziesiątych. Jest to bardzo popularny wskaźnik, zamieniający dwie średnie ruchome (wskaźniki śledzące trend) w oscylator mierzący zmiany ceny. W rezultacie otrzymany został wskaźnik dający informację o kierunku trendu oraz zmienności ceny. MACD oscyluje powyżej i poniżej poziomu 0, wskazując dywergencje, konwergencje oraz przecięcia linii. MACD nie ma ograniczeń co do wartości, więc nie jest wartościowym wskaźnikiem do oceny wykupienia wyprzedania raynku.



Rysunek 2.13 Wskaźnik MACD  
Źródło: http://stockcharts.com/school/doku.php?id=chart\_school:technical\_indicators:moving\_average\_convergence \_divergence\_macd (dostęp na 17.07.2017)

Wskaźnik MACD składa się z linii MACD która jest różnicą między dwoma wskaźnikami wykładniczej średniej kroczącej (ang. EMA – Exponential Moving Average), od 12 dniowej jest odejmowania 26 dniowa. Linia MACD przedstawia konwergencję i dywengercję tych dwóch linii. Linia sygnałowa jest średnią linii MACD z 9 okresów, natomiast histogram rpzedstawia różnicę między linią MACD a linią sygnałową.

Dywergencja zachodzi gdy linie średnich kroczących na podstawie których linia MACD jest obliczna się rozchodzą wtedy wartość linii MACD rośnie gdy 12 dniowa średnia jest większa od 26 dniowej, i maleje gdy sytuacja jest odwrotna. W obu przypadkach podczas dywergencji linia MACD oddala się od poziomu zero, natomiast podczas konwergencji gdy linie średnich kroczących podążają w tym samym kierunku linia MACD zbliża się do poziomu 0.

MACD daje sygnały w momencie przecięcia linii MACD z poziomem 0, gdy przecięcie następuje od góry jest to sygnał do sprzedaży, a gdy od dołu to do kupna. Przecięcia linii MACD z linią sygnałową w ekstremach również sa sygnałami które mogą zwiastować zmianę trendu. Dodatkowymi informacjami również sa dwyergencje gdy linie MACD i sygnałowa poruszają się w przeciwnym kierunku niż wykres[[20]](#footnote-20).

## Uczenie maszynowe

Uczenie maszynowe to dziedzina powiązana z statystyką, informatyką, logiką, teorią informacji, teorią prawdopodobieństwa i nie tylko. Jest częścią większego pojęcia jakim jest sztuczna inteligencja. Celem uczenia maszynowego jest automatyczna detekcja znaczących danych z dużych źródeł danych. Automatyzacja niesie za soba konieczność uczenia się tak aby dostarczane dane były jak najbardziej precyzyjne. Narzędzia uczenia maszynowego uczą się i adoptują w taki sposób by zwracać celne wyniki[[21]](#footnote-21).

### Pojęcie uczenia się w kontekscie maszynowym

Proces uczenia się zawiera kilka ważnych elementów które spełniać musi również system określany jako uczący się. Wymagane zmian o charakterze uczenia się, to oznacza że system powinien wprowadzać zmiany w interpretacji danych w taki sposób by poprawić swoje działanie. Poprawa za pośrednictwem zmian powinna wynikać z działania systemu, czyli to on musi zdecydować że dana zmiana wpłynie pozytywnie na jego działanie i ją zaastosować. Nie mogą być to zmiany funkcjonalności czy inne zmiany zewnętrzne. Takie zmiany wynikające z wnętrza systemu można określić jako doświadczenie[[22]](#footnote-22).

W kontekscie maszyn proces uczenia w dużym uproszczeniu polega na analizie zbioru danych treningowych na których algorytm uczy się czyli wprowadza zmiany w interpretacji danych w celu poprawienia wyników, i to za pomocą odpowiedniego algorytmu zmiany wynikają z wnętrza systemu na podstawie analizy danych wprowadzonych z zewnątrz po czym otrzymany jest wynik czyli informacje na bazie wprowadzonego zbioru danych[[23]](#footnote-23).

### Podstawowe pojęcia

Poniżej zostaną opisane podstawowe pojęcia związane z uczeniem maszynowym użyte w pracy:

* Atrybuty – rodzaj wartości jaki przyjmuje określona kolumna w zbiorze danych, atrybuty można określić jako symboliczne czyli określona lista wartości jakie może przyjmować, numerczny, tekstowy, format daty,
* Filtry – narzędzia modyfikujące w określony sposób zestawy danych, mogą m.in usuwać lub dodawać atrybuty, usuwać instance, dyskretyzować,
* Instancja – reprezentuje pojedyńczy wiersza w zestawie danych, składa się z określonej liczby atrybutów,
* Klasa – atrybut decyzyjny do ktorego przypisywana jest instancja po klasyfikacji,
* Klsyfikator – każdy algorytm uczenia maszynowego dostarczony przez wekę. Z użyciem dostarczonego zestawu danych tworzy model na podstawie którego mogą być klasyfikowane nowe przykłady,
* Macierz pomyłek – macierz prezentująca wynik klasyfikacji, wiersze odpowiadaja poprawnym klasom decyzyjnym, a kolumny decyzjom podjętym przez klasyfikator,
* Model – jest sposobem klasyfikacji danych wytrenowany przez klasyfikator, gdy otrzyma instancję na wejsciu, na wyjsciu otrzymana zostanie informacja do jakiej klasy pasuje. Modele mogą być dynamiczne co oznacza że mogą trenować się w miarę otrzymywania nowych danych,
* Precyzja – jest to wartość poprawnie sklasyfikowanych instancji w relacji do wszystkich danych z zestawu,
* Skuteczność (ang. Recall) – wartość poprawnie sklasyfikowanych instancji ze wszystkich instancji danej klasy,
* Walidacja krzyżowa – metoda testowania która dzieli zestaw danych na określoną ilość podzbiorów losowych oraz dla każdego buduje model i go testuje,
* Zbiór treningowy – zbiór danych wykorzystywany przez klasyfikator do treningu modelu,
* Zbiór testowy – zbiór danych wykorzystywany przez klasyfikator do testowania modelu,
* Zestaw danych – wszystkie dane wprowadzone do weki na podstawie których wykonywane są działania, składa się z instancji.

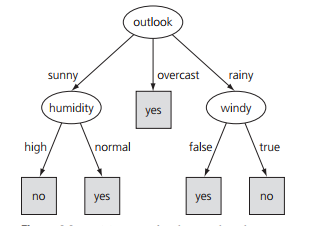
### Przedstawienie badanych algorytmów

Podczas badań użyta została wiekszość algorytmów uczenia maszynowego jakie dostarcza biblioteka Weka (opisana w 3.2). Poniżej zostaną przedstawione skrótowo algorytmy wybrane jako najlepsze.

#### Lasy losowe

Lasy losowe (ang. Random Forests) jest to metoda klasyfikacji z użyciem wielu algorytmów drzew decyzyjnych. Polega na tworzeniu wielu drzew decyzyjnych bazując na losowym zestawie danych. Każde drzewo dezyzyjne klasyfikje problem a następnie końcowa decyzja jest podejmowana w drodze głosowania większościowego na najbardziej popularna klase.

Drzewo decyzyjne jest to skierowany graf acykliczny, oparty na strukturze drzewiastej. Każde drzewo posiada węzły, a może być nimi decyzja lub stan i gałęzi które odpowiadają wariantom. Z węzła możę wychodzić tyle gałęzi ile możliwości jest wariantów decyzji, a każdy liść oznacza decyzję. Klasyfikacja rozpoczyna się od korzenia drzewa, następnie podąża po gałęziach a kończy na lisciach czyli ostatnim elemencie drzewa.



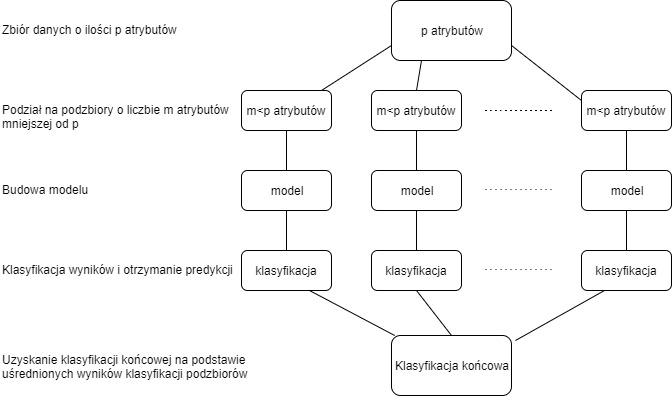
Rysunek 2.14 Przykładowe drzewo  
Źródło: Witten,Eibe,Mark., Data Mining: Practical Machine Learning Tools and Techniques

Użytkownik ma możliwość zdefinowania ilości drzew klasyfikujących. Tworzone są różne drzewa z użyciem róznych algorytmów, za pomocą algorytmu bagging który generuje zróżnicowany zestaw klasyfikatorów użytych w lasach.

Zasada działania algorytmu bagging polega na wygenerowaniu określonej liczby klasyfikatorów ekspertów dla podzbioru zadań. Z całego zestawu danch jest losowany podzbiór poprzez losowanie ze zwracaniem a następnie dla niego jest tworzony ekspert czyli drzewo, po ty losowany jest kolejny pozbiór i tworzony jest kolejne drzewo. Proces ten powtarzany jest określoną ilośc razy, a na koniec stworzone drzewa są używane do głosowania. Decyzja posiadająca najwięcej głosów jest przyjmowana.

#### Metoda podprzestrzeni losowych

Metoda podprzestrzeni losowych(ang. Random Subspace Method, RSM) wykorzystywana przy danych z dużą ilością atrybutów, kiedy wyszukiwanie pośród takiej ilości zmiennych może być długie i trudne aby zidentyfikować istotne zmienne.



Rysunek 2.15 Wykres etapów algorytmów RSM  
Źródło: opracowanie własne

Metoda operuje na losowo generowanych podzbiorach atrybutów która należą do całości atrybutów dostępnych w zestawie danych. Na danych o mniejszej ilości atrybutów algorytm uczący może poradzić sobie lepiej niż przy ich dużej ilości a tymsamym uzyskać uzyskać celniejsze informacje. Ze względu na dużą ilość atrubutów losowanie podzbiorów odbywa sie do momemntu gdy wszystkie zostaną pokryte. W rezultacie tyle otrzymujemy taką samą ilość podzbiorów ile było losowań a każdy podzbiór. Dla każdego podzbioru generowany jest model zgodnie z wybranym algorytmem. Każdy podzbiór jest klasyfikowany a wyniki predykcji wszystkich klasyfikatorów są sumowane i otrzymywana jest ich średnia[[24]](#footnote-24).

# Opis wykorzystanych narzędzi

Poniższy dział opisuje narzędzia jakie zostały wykorzystane do badań. Podrodział 3.1 opisuje narzędzie w języku Java dedykowane do przeprowadzania analizy technicznej oraz budoawania strategii. Kolejny podrodział 3.2 opisuje narzędzie stanowiące zestaw algorytmów uczenia maszynowego które zostało wykorzystane do budowy i testowania modelu.

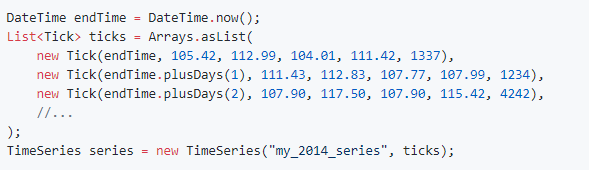
## Technical Analysis for Java (Ta4j)

Ta4j jest to biblioteka napisana przeprowadzania analizy technicznej w języku Java typu open source na licencji MIT. Dostarcza podstawowe komponenty do tworzenia, oceny oraz stosowania strategi inwestowania. Jest to biblioteka napisana w całości w języku Java działająca na wersjach 1.6 lub wyższych. Zawiera ponad 100 gotowych oraz przetestowanych wskaźników analizy technicznej, silnik do tworzenia strategii inwestycyjnych.

### Podstawy korzystania z TA4j

Aby używać biblioteki wystarczy dołączyć plik .jar do projektu. Podstawowymi obiektami w bibliotece są TimeSeries czyli obiekty zawierające zagregowane dane o określonych przedziałach czasowych, a każde zakończenie przedziału czasowego jest reprezentowane przez obiekt typu Tick. TimeSeries można porównać do wykresu giełdowego o określonym przedziale czasowym.

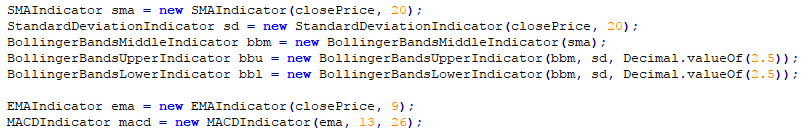
Obiekt Tick jest obiektem zawierającym informacje o pojedyńczym zakończonym przedziale czasowym, jeste reprezentacją świecy japońskiej i zawiera informacje o cenie otwarcia, cenie zamknięcia, najwyższej i najniższej cenie oraz wolumenie ceny.



Rysunek 3.1 Tworzenie obiektu TimeSeries  
Źródłó: <https://github.com/mdeverdelhan/ta4j/wiki/Time%20series%20and%20ticks> (dostęp na 24.07.2017)

Jednym ze sposobów utworzenia obiektu TimeSeries jest utworzenie listy obiektów Tick reprezentujących świece i na podstawie tej listy utworzenie obiektu TimeSeries za pomocą konstruktora.

Kolejnym ważnym elementem biblioteki są dostarczone wskaźniki analizy technicznej. Aby swtorzyć wskaźnik zazwyczaj potrzeba obiektu TimeSeries dla wskaźników podstawowych oraz dla wskaźników bardziej złożonych wymagane są inne wskaźniki na podstawie których są generowane wartości.



Rysunek 3.2 Wskaźniki  
Źródło:opracowaie własne

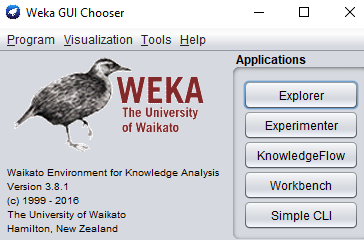
Dodatkowo konstruktowy wskaźników przyjmują wartości liczbowe które mogą oznaczać odchylenia,czy okresy czasowe jakie przyjmuje wskaźnik, wszystko zależnie od rodzaju wskaźnika. Wartości wskaźników są obliczane dynamnicznie w momencie wywoałani metody get(int index) na obiekcie wskaźnika.

Innymi możliwościami biblioteki jest tworzenie strategii czyli zestawu reguł według którego miały by być podejmowane decyzje kupna/sprzedaży, oraz tester strategii za pomocą którego można przetestować stworzoną strategię na historycznych danych[[25]](#footnote-25).

## Weka

Weka jest to zbiór zawierający wiele algorytmów uczenia maszynowego do eksploracji danych. Z weki można kożystać za pomocą przygotowanego interfejsu graficznego, lub jako biblioteki Java którą można załączyć do projektu i prosto z kodu używać zawartych algorytmów. Częścią weki są narzędzia do wstępnego przetwarzania danych, klasyfikacji, regresji, klastrowania, relacji oraz wizualizacji. Jest również przystosowana do rozwijania nowych schematów uczenia maszynowego. Rozwijana przez Uniwersytet Waikato w Nowej Zelandii jako projekt open source pod licencją GNU (GNU General Public License, GPL). Weka dostarcza również gotowe aplikacje które wybiera się podczas uruchamiania programu, są to:

* + Explorer - środowisko do badania danych z użyciem weki,
  + Experimenter – środowisko do wykonywania eksperymentów, oraz testów statystycznych pomiędzy schematami uczenia,
  + KnowledgeFlow – ta sama funkcjonalość co Explorer dodatkowo obsługująca interfejs „drag and drop” oraz obsługująca uczenie inkrementacyjne,
  + Workbench – aplikacja zawierająca wszystkie powyższe funkcjonalności
  + SimpleCLI – wierz poleceń umożliwiający bezpośrednie wykonywanie komend weki, przeznaczony dla systemów nie dostarczających własnego wiersza poleceń.



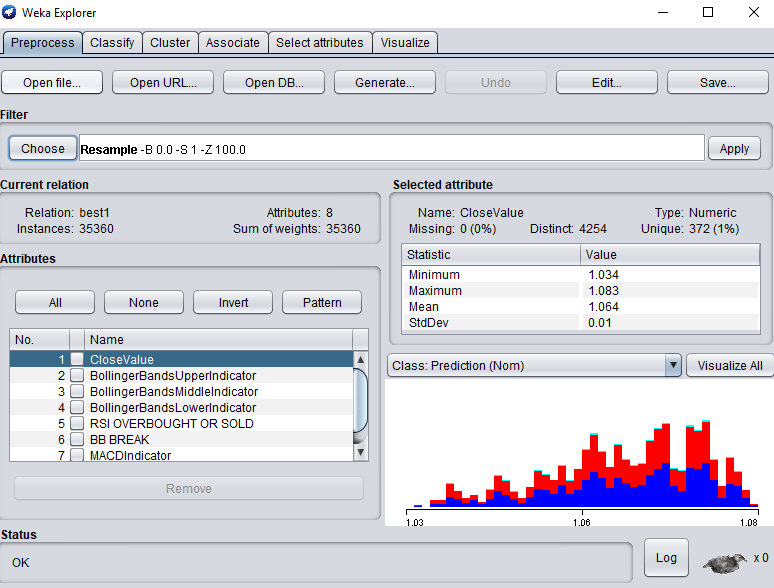
Rysunek 3.3 Weka – aplikacje  
Źródłó: opracowanie własne

Funkcjonalnośći weki opisywane w kolejnych podrozdziałach będą na podstawie aplikacji Explorer, z uwagi na największe wykożystanie własnie tej aplikacji podczas badań.

### Wstępne przetworzenie danych

Do wstępnego przetwarzania danych służy sekcja „Preproscess”, umożliwia ona prwoadzenie danych w postaci różnych rozszerzeń plików takich jak .arff czy .csv, również dane mogą zostać wprowadzone z odnośnika URL, bazy danych lub wygenerowane za pomocą wbudowanego generatora.W sekcji wstępnego przetwarzania danch zawiera się kilka innych podsekcji:

* **Filter** - daje możliwość filtrowania danych za pomocą dostarczanych przez wekę gotowych filtrów,
* **Current Relation**  - prezentuje podstawowe dane wprowadzonego zbioru danych czyli nazwę, ilość arybutów, ilość wierszy danych oraz sumę wag,
* **Attributes** – prezentuje atrybuty jakie posiada dostarczony zbiór danych oraz umożliwia usuwanie wybranych,
* **Selected attribute** – prezentuje parametry wybranego atrybutu, jego ekstrema, ilość różnych powtarzających oraz uniklanych wartości i dodatkowo graf reprezentujący zbierzność wartości z wartościami wybranego innego atrybutu.

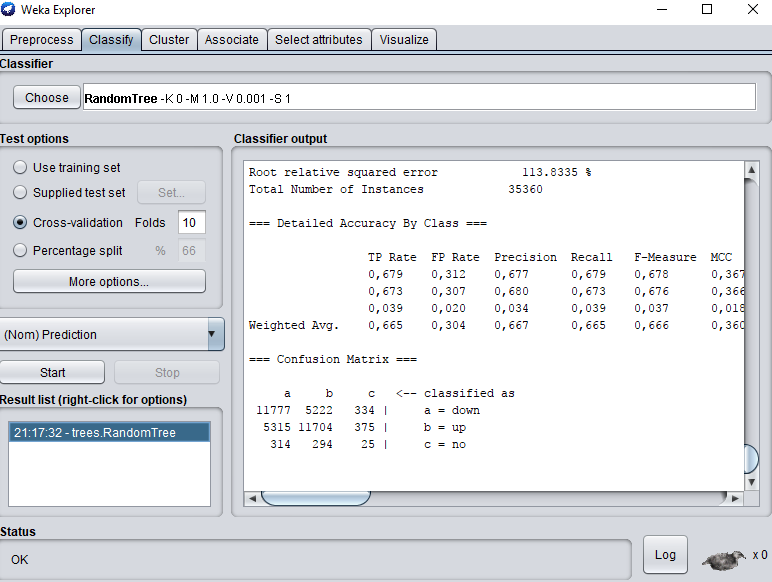


Rysunek 3.4 Sekcja preprocess  
Źródło:opracowanie własne

### Klasyfikacja danych

Do klasyfikacji danych służy sekcja „Classify”, która składa się z następujących podsekcji:

* **Classifier** – umożliwia wybór klasyfikatora oraz ustawienie jego parametrów po kliknięciu w nazwę,
* **Test options** – wybór opcji testowania stworzonego modelu. Wybrać można jedną opcję, a do wybrania jest użycie tylko zbioru treningowego, użycie dodatkowego zbioru testowego, walidacja krzyżowa z możliwością określenia ilości zbiorów testowych na które zostaną podzielone dane treningowe oraz podział procentowy,
* **Result list** – lista wyników klasyfikacji, zapisywana po każdym wykonaniu dzięki czemu jest możliwość sprawdzenia poprzednich klasyfikacji kliknięcie prawym przyciskiem myszy na wynik wywoła listę czynności jakie można wykonać z stworzonym modelem m.in zapisanie modelu, wizualzacja modelu, analiza kosztów,
* **Classifier output** – dokładny wynik klasyfikacji z m.in opisem modelu, macierzy pomyłek, dokładności względem klas.



Rysunek 3.5 Sekcja klasyfikacji danych  
Źródło: opracowanie własne

### Pozostałe funkcjonalności

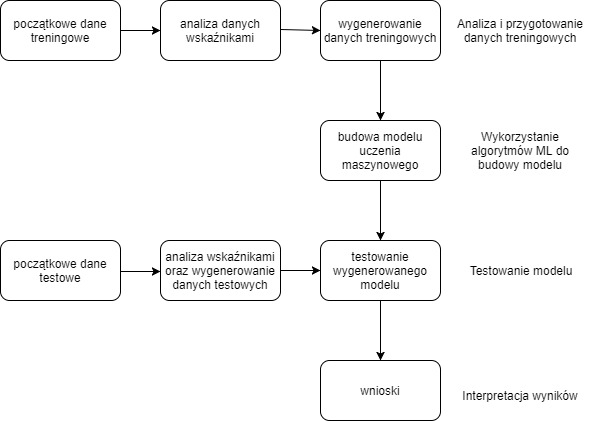
Pozostałe funkcjonalności były używane w mniejszym stopniu do badań jednak wciąż warto o nich wspomieć, są to sekcje:

* **Cluster** – umożliwia wykonanieprocesu klastrowania jest to proces grupowania obiektow w taki sposób by obiekty w tej samej grupie były jak najbardziej do siebie podobne. Interfejs jest niemal identyczny jak w sekcji „Classify” z tą różnicą że zamiast wyboru algorytmów klasyfikujących są do wyboru algorytmy klastrujące.
* **Associate** –sekcja uczenia maszynowego bazującego na regułach. Polega na poszukiwaniu silnych relacji pomiędzy zmienymi dostarczonymi w zbiorze danych. Interfejs umożliwia wybór algorytmu oraz przeglądanie listy rezultatów.
* **Select attributes** – jest to narzędzie do oceny wartości atrybutów, szczególnie przydatne gdy zbiór danych posiada wiele atrybutów i niektóre mogą wprowadzać zamęt. Sekcja ta dostarcza narzędzia odpowiednie do selekcji wartościowych atrybutów. Dostarczona jest możliwość wybrania sposobu oceny atrybutów, metody wyszukiwania, sposobu selekcji atrybutów ze zbioru oraz okno z listą rezultatów i wynikiem selekcji atrybutów.
* **Visualize –** sekcja prezentuje matrycę wykresów rozproszenia dla relacji każdego atrybutu z każdym. Atrybuty są oznaczone kolorami względem klas do których należą. Wykresy mogą być powiększane, zarówno jak i obiekty atrybutów na wykresie.

# Metodologia

## Ogólny przebieg badań

Przebieg badań składał się z pewnych etapów następujących po sobie. Dane wykorzystane do badań były początkowo w stanie ym, wymagały obróbki i interpretacji co zostało wykonane przez program napisany w języku Java. Wygenerowany nowy zestaw danych bazujący na zbiorze początkowym, służy jako zbiory treningowe i testowe dla algorytmów uczenia maszynowego, bedacym przedmiotem badań w celu ich skuteczności przy przewidywaniu przyszłych ruchów cen rynków walutowych.

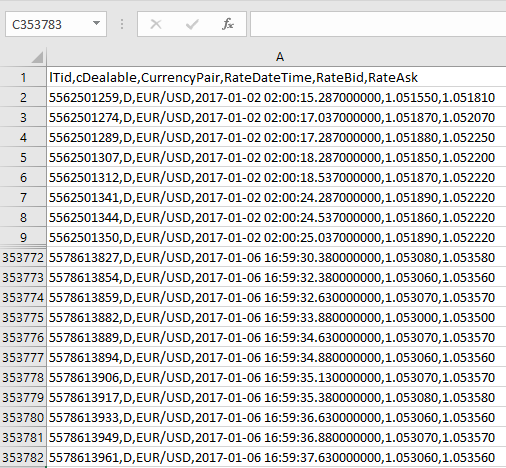


Rysunek 4.1 Etapy pracy  
Źródło: Opracowanie własne

Tak jak na powyższym rysunku etapów pracy (rys. 4.1) całość można podzielić na cztery etapy. Pierwszy to przygotowanie danych treningowych gdzie dostarczone do napisanego programu Java początkowe dane treningowe są mapowane do odpowiednich obiektów a następnie wykonywane na nich są obliczenia wskaźników analizy technicznej, które cześciowo stanowią dane treningowe. Na podstawie wskaźników generowane są sygnały usupełniające dane treningowe. Całosć jest zbierana do jednego pliku csv. Kolejnym etapem jest budowa modelu gdzie wcześniej wygenerowany plik csv jest ładowany do programu weka, a tam za pomocą filtrów modyfikowany jest zbiór treningowy oraz generowany jest model uczenia maszynowego. Następny etap to testowanie modelu które odbywa się zaraz po wygenerowaniu modelu. Początkowy zbiór danych testowych przechodzi identyczną drogę analizy i generowania sygnałów. Otrzymany gotowy zbiór danych testowych jest wskazywany podczas ustawiania procesu generowania modelu za pomocą wybranego algorytmu uczenia maszynowego. Na postawie testowania otrzymywana jest macierz pomyłek oraz inne szczegółowe informacje o wygenerowanym modelu gotowe do analizy w ostanim kroku badań.

## Omówienie użytych danych

Do badań użyte zostały dane historyczne rynków walutowych. Ze wszystkich dostępnych par walutowych wybrana została para EUR/USD ze względu na jej największą popularność i dużą zmienność, co się wiąże z największą ilością transkacji dokonywanych własnie w zakresie tej pary walut. Dostępne są dane od roku 2000 jednak ze względu na charakter badań gdzie celem jest krótkoterminowa prognoza kierunku zmiany kursu walut wykorzystane zostały dane nowe pochodzące z stycznia roku 2017.



Rysunek 4.2 Przykład danych  
Źródło: opacowanie własne

Zestaw danych pochodzący z jednego miesiąca składa się z dokładnie takiej samej liczby plików ile jest tygodni w danym miesiącu. Powyższe dane (rys. 4.2) jest to mała część danych pochodzących z pierwszego tygodnia stycznia 2017r. Pojedyńczy plik posiada w tym przypadku ponad 353 tyś. wierszy, podobna liczba wierszy występuje w pozostałych plikach. Ilość wierszy zależy od ilości ruchów ceny w tygodniu ponieważ każdy wiersz odpowiada pojedyńczej zmianie w cenie pary walutowej. Dane dostarczone są w formacie CSV, a szczegóły jakie przedstawia pojedyńczy wiersz to id wiersza, para wautowa, data zmiany ceny, cena kupna i cena sprzedaży. Podczas pracy z danymi nie wszystkie były istotne, wykorzystane zostały poszczególne kolumny odpowiadające za date zmiany ceny, rodzaj pary walutowej oraz cena kupna. Jakość danych jest wysoka ze względu na to że zanotowana jest każda minimalna zmiana ceny, nieraz jest to kilka wierszy dotyczącyh jednej sekundy.

## Omówienie procesu obróbki danych do sygnałow

Aby uzyskać odpowiednie dane do badań został napisany program w języku Java. Dane wejściowe wymagają specjalnej konwersji by pasowały do formatu użytego w bibliotece do analizy technicznej. Podrozdział przedstawia w jaki sposób w apllikacji z danych początkowych są tworzone świece japońskie, jak obliczane są wartości wskaźników na podstawie świec oraz w jaki sposób na podstawie wskaźnków są generwoane sygnały.

### Tworzenie świec japońskich

Wyżej przedstawione dane wejściowe aby zostały zamienione na świece japońskie zostają początkowo czytywane do listy obiektów StockRecord który reprezentuje pojedyńczy wiersz z otrzymanego pliku CSV. Następnie dla każdego takiego wiersza wykonywana jest akcja przyporządkowania do jakiej świecy należy dany wiersz. Tworzony jest obiekt MutableTick reprezentujący pojedynczą świecę z informacją zakresu przedziału czasowego jaki obejmuje. Po czym każdy wiersz z danych jest przyporządkowywany do świecy, a jego wartości nadpisują odpowiednie ekstrema świecy jeżeli je przekraczają. Gdy pojawi się wiersz danych należący do kolejnego przedziału czasowego, aktualna świeca jest zapisywana i tworzona jest następna do której są zapisywane wiersza należące do kolejnego okresu.

Rysunek 4.3 Pseugokod algorytmu klasyfikującego wiersz do świecy  
Źródło: opraowanie własne

**Jeżeli** wiersz należy do okresu **to:**

Pobierz i porównaj jego parametry z parametrami świecy, gdy przekracza ekstrema świecy podmień wartość.

**W przecywnym razie:**

Zapisz aktualną świecę, stwórz nową i przepisz wartości wiersza do świecy

**Koniec**

### Generowanie sygnałów

Na podstawie otrzymanego zestawu świec japońskich tworzone są obiekty wskaźników które dla każdej świecy obliczają odpowiednią wartość liczbową. Obliczone wartości już same w sobie są istotnymi informacjami do oceny przyszłego kierunku ceny, niemniej jendak na podstawie niektórych wskaźników generowane są dodatkowo sygnały kupna i sprzedaży.

#### Sygnał RSI

Wskaźnik RSI może generować sygnał prawdopodobnego spadku ceny reprezentowany przez jeden oraz wzrostu ceny reprezentowany przez minus jeden, natomiast gdy wartość sygnały wynosi zero to informacja o braku okazji do decyzji. Zakładając że wartość poziomu górnego wskaźnika RSI jest oznaczony jako ***x*** a poziom dolny jako ***y*** i wartość wskaźnika RSI to ***rsi*** to generowany sygnał jest opisany następującymi rownaniami:

Rysunek 4.4 Równania opisujące sygnał RSI  
Źródłó: opracowanie własne

Gdy wartość wskaźnika rsi jest wyższa niż poziom górny oznacza to wykupienie rynku i generowany jest sygnał spadku ceny czyli jeden, natomiast w odwrotnej sytuacji gdy wartość wskaźnika rsi jest poniżej poziomu dolnego oznacza to wysprzedanie rynku i generowany jest sygnał wzrostu ceny czyli minus jeden. Natomiast gdy wskaźnik porusza się pomiędzy poziomami generowany jest sygnał zero będący informacją o braku okazji do decyzji.

#### Sygnał CCI

Kolejny wskaźnik generujący sygnały to CCI w którym generowanie sygnału działą podobnie jak w powyższym RSI. Możliwe jest wygerenowanie trzech sygnałów. Sygnał oznaczony jako jeden oznacza sygnał do spadku ceny, minus jeden to sygnał wzrostu ceny zero to brak sygnału.

Rysunek 4.5 Równania opisujące sygnał CCI  
Źródło: opracowanie własne

W CCI również występują poziomy górny i dolny na podstawie których generowane są sygnały. Gdy wartość wskaźnika oznaczona jako ***cci*** przegracza poziom górny oznaczony jako ***x*** generowana jest jeden czyli sygnał spadku ceny, w odwrotnej sytuacji gdy ***cci*** przegracza poziom dolny oznaczony jako ***y*** oznacza to sygnał wzrostu ceny czyli minus 1, w przypadku gdy wartość wskaźnika porusza się miedzy poziomami generowane jest zero jako brak sygnału.

#### Sygnał Wstęg Bollgingera

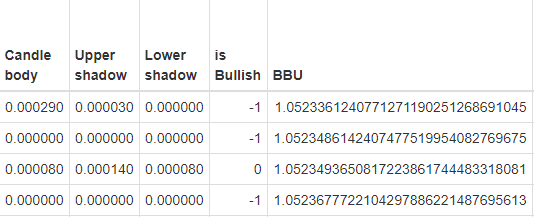
Na podstawie wskaźnika wstęg Bollingera również generowane są sygnały. Reprezentacja podobna jest jak w poprzednich. Generowane są trzy sygnały: jeden, minus jeden , zero. Wskaźnik składa się z linni górnej i dolnej które śledzą ruch ceny. Gdy wartość ceny przekroczy wartość wskaźnika wstęgi górnej generowany jest sygnał jeden czyli sygnał spadku ceny, w przeciwnym wypadku gdy wartość ceny przekroczy dolną wstęgę generowany jest sygnał minus jeden sygnaliżujący wzrost ceny. Gdy cene porusza się pomiędzy wstęgami generowane jest 0 jako brak sygnału. Metodę generowania sygnału można opisać następującym układem równań, gdzie ***p*** wyraża wartość ceny, ***bbu*** wartość górnej wstęgi bollingera a ***bbd*** wartość dolnej wstęgi bollingera.

### Wygenerowane dane

Dane wygenerowane przez apliację stworzoną na potrzeby badań składają sie z wielu elementów takich jak cechy świecy, wartości wskaźników i sygnały. Z miesięcznego zestawu ruchów ceny który zawierał ponad 350 tysięcy wierszy, po zamianie na świece zostało ponad 35 tysięcy wierszy które stanowią dane wyjściowe aplikacji generującej, a będą służyły jako zbiór wejsciowy dla algorytmu uczenia maszynowego.

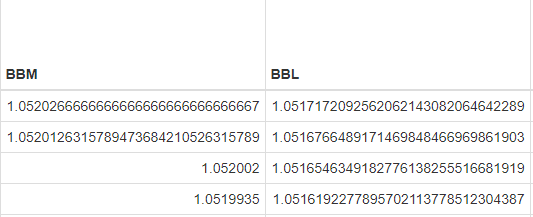
Dane wyjściowe jakie otrzymywane są po analizach to:

* Candle body - wartość ciała świecy,
* Upper shadow - wartość ciena górnego,
* Lower shadow - wartość cienia dolnego,
* Is Bullish – czy świeca jest wzrostowa, jest to jednocznoczesna informacja czy świeca jest spadkowa tj. świeca wzrostowa nie może być spadkową,
* BBU – wartość górnej wstęgi Bollingera,



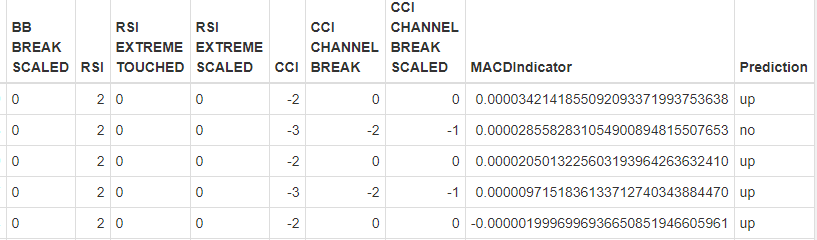
Rysunek 4.6 Wygenerowane dane 1/3  
Źródło: opracowanie własne

* BBM – wartość środkowej wstęgi Bollingera,
* BBL – wartość dolnej wstęgi Bollingera,



Rysunek 4.7 Wygenerowane dane 2/3  
Źródło: opracowanie własne

* BB Break Scaled – przebicie wstęgi bollingera skalowane w zależności od wielkości świecy ponad wstęgą. Skala reprezentuje wartości od -5 do 5,
* RSI – wartość wskaźnika RSI,
* RSI extreme touched – informacja czy RSI osiągneło ekstremum jeżeli tak to wartość 1 lub -1 w zależności od osiągniętego ekstremum, w innym przypadku 0,
* RSI extreme scaled – informacja czy RSI osiągneło ekstremum wartość reprezentowana jest w skali od -5 do 5 w zależności od przebicia poziomu,
* CCI – wartość wskaźnika CCI,
* CCI channel break – informacja czy CCI przebiło kanał, jeżeli tak to wartość 1 lub -1 w zależności od osiągniętego ekstremum, w innym przypadku 0,
* CCI channel break scaled – informacja czy CCI przebilo kanał, wartość reprezentowana jest w skali od -5 do 5 w zależności od przebicia poziomu,
* MACD Indicator – wartość wskaźnika MACD,
* Prediction – faktyczne zachowanie ceny w okresie następującym. Wartość która ma być przewidywana przez algorytm uczenia maszynowego.



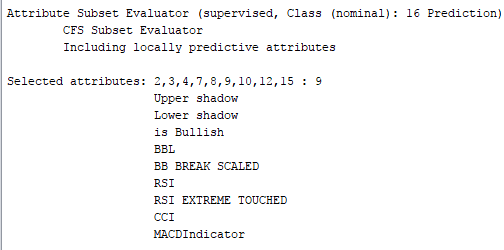
Rysunek 4.8 Wygenerowane dane 3/3  
Źródło: opracowanie własne

## Omówienie procesu badań

Proces badań składał się z czterech poszczególnych etapów. Wyznaczania atrybutów gdzie za pomocą algorytmów oceniających były wyznaczane atrybuty najbardziej znaczące dla badań, kolejno filtorwanie danych aby pozbyć się atrybutów obniżających wyniki badań, następnie klasyfikacja za pomocą wybranych metod klasyfikujących oraz testowanie modelu. Dodatkowo w procesie badań można wyróżnić dwie częsci składające się z tych samych czterech etapów. Pierwsza część gdy każdy etap był powtarzany dla tego samego zbioru danych tak by na domyślnych wartościach wyznaczyć zbiór atrybutów najbardziej znaczących które będą stanowiły bazę do badań w kolejnym etapie i wyznaczyć klasyfikatory które najlepiej radzą sobie z problemem tak by wyniki badań były najlepsze. W drugiej częsci zaczęto wprowadzać zmiany w ustawieniach wskaźników, sygnałów oraz oraz parametarach klasyfikatorów tak by poprawić wcześniejsze rezultaty.

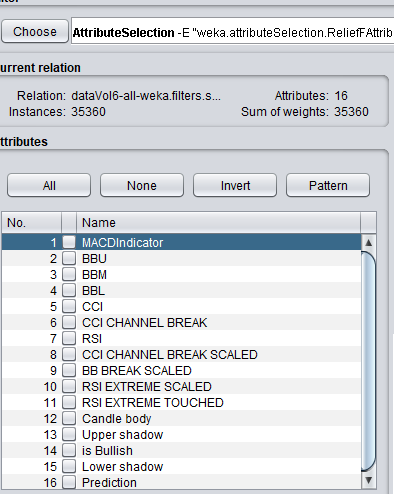
### Wznaczenie i filtrowanie atrybutów

Etap wyznaczania atrybutów polegał na wykorzystaniu do wprowadzonych danych algorytmów selekcji atrybutów. Użyte zostało narzędzie „Select attributes” programu Weka które dostarcza algorytmy umożliwiające selekcję. Zastosowany algorytm na zbiorze danych zwracał jako wynik listę potencjalnie najbardziej znaczących atrybutów bądź posegregowaną listę atrybutów względem oceny wykonanej przez algorytm. Selekcja odbywała się z wykorzystaniem wszystkich kombinacji metod oceniających i wyszukujących.



Rysunek 4.9 Przykładowy wynik selekcji atrybutów  
Źródło: opracowanie własne

Dostarczane wyniki przez algorytmy miały różną skuteczność przy klasyfikacji zbioru danych dlatego do najlepszych wyników algorytmów traktowanych jako podstawa wybierane były dodatkowe atrybuty tak by zwiększyć trafność klasyfikacji w kolejnych etapach. Na podstawie dostarczonych wyników filtrowane były atrybuty używane do klasyfikacji w kolejnym etapie.



Rysunek 4.10 Zbiór atrybutów po użyciu filtra "AttributeSelection"  
Źródło: opracowanie własne

Filtracja polegała na usunięciu atrybutów posiadających mniejszą wagę a pozostawienie tych posiadających większą wagę, lub nałożeniu filtru „AttributeSelection” który wykonuje wybrany algorytm selekcji atrybutów i automatycznie nanosi zmiany na listę atrybutów zgodnie z wynikiem algorytmu. Gdy algorytm zwraca podzbiór atrybutów jako wynik algorytmu, niezawarte w podzbiorze atrybuty były usuwane. Natomiast gdy wynikiem algorytmu była posortowana lista według oceny, kolejność była nanoszona na zbiór atrybutów.

### Klasyfikacja

### Testowanie

# Model

Omowienie najlepszych wynikow

// wydaje mi się że omówienie modelu zajmie mi cały rozdział ale nie wiem szczerze mówiąc jeszcze jak to będę opiywał.

# Wnioski

Wnioski płynące z pracy

# Bibliografia

WEKA Manual for Version 3-9-1

<http://www.cs.waikato.ac.nz/ml/weka/index.html>

<https://github.com/mdeverdelhan/ta4j>

[https://nlp.stanford.edu/IR-book/html/htmledition/text-classification-and-naive-bayes-1.html 24.07.2017](https://nlp.stanford.edu/IR-book/html/htmledition/text-classification-and-naive-bayes-1.html%2024.07.2017)

1. J. Mazurek http://www.bankier.pl/wiadomosc/Czym-jest-rynek-Forex-1797860.html (stan na 27.06.2017) [↑](#footnote-ref-1)
2. http://www.sharptrader.com/new-to-trading/forex/majors-minors-exotic-currency-pairs/ (stan na 27.06.2017) [↑](#footnote-ref-2)
3. M.C.Thomsett, *Mastering Fundametnal Analysys,* Dearborn Financial Publishing 1998, s 2-3 [↑](#footnote-ref-3)
4. J.J.Murphy, *Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s 1* [↑](#footnote-ref-4)
5. 9 10 .J.Murphy, *Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s 2* [↑](#footnote-ref-5)
6. [↑](#footnote-ref-6)
7. http://www.investopedia.com/university/technical/techanalysis1.asp (stan na 04.06.2017) [↑](#footnote-ref-7)
8. S.Nison, *Japanise candlestick charing techniques*, Prentice Hall Press USA 2001, s 16 [↑](#footnote-ref-8)
9. S.Nison, Japanise candlestick charing techniques, Prentice Hall Press USA 2001, s 21-26 [↑](#footnote-ref-9)
10. S.Nison, Japanise candlestick charing techniques, Prentice Hall Press USA 2001, s 149-150 [↑](#footnote-ref-10)
11. .J.Murphy, *Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s 42* [↑](#footnote-ref-11)
12. .J.Murphy, *Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s 42-47* [↑](#footnote-ref-12)
13. http://www.investopedia.com/university/technical/techanalysis10.asp (dosęp 10.07.2017) [↑](#footnote-ref-13)
14. J.Murphy, *Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s 197-198* [↑](#footnote-ref-14)
15. http://www.investopedia.com/terms/c/crossover.asp (dosęp 11.07.2017) [↑](#footnote-ref-15)
16. J.Murphy, *Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s 137* [↑](#footnote-ref-16)
17. J.Murphy, *Analiza techniczna rynków finansowych, WIG-PRESS Warszawa 1999, s*  [↑](#footnote-ref-17)
18. <https://www.linnsoft.com/techind/cci-commodity-channel-index> (dostęp na 14.07.2017) [↑](#footnote-ref-18)
19. J. Bollinger, *Bollinger on Bollinger Bands,* McGraw-HillNew York 2001, s xx-xxii [↑](#footnote-ref-19)
20. http://stockcharts.com/school/doku.php?id=chart\_school:technical\_indicators:moving\_average\_

    convergence\_divergence\_macd (dostęp na 17.07.2017) [↑](#footnote-ref-20)
21. S. Shalev-Shwartz, S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms,*

    Cambridge University Press 2014, s vii [↑](#footnote-ref-21)
22. P.Cichosz, *Systemy uczące się*, Wydawnictwa Naukowo-Techniczne, Warszawa, 2000 [↑](#footnote-ref-22)
23. S. Shalev-Shwartz, S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms,*

    Cambridge University Press 2014, s 19-22 [↑](#footnote-ref-23)
24. http://isplab.tudelft.nl/sites/default/files/Lai05a.pdf (dostęp na 26.07.2017) [↑](#footnote-ref-24)
25. https://github.com/mdeverdelhan/ta4j (dostęp na 25.07.2017) [↑](#footnote-ref-25)