

D E B I U T Y S T U D E N C K I E

2 0 2 3

ZASTOSOWANIE METOD ILOŚCIOWYCH W EKONOMII I FINANSACH

pod redakcją
Alicji Grześkowiak
i Piotra Peterneka



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2023

Recenzja

Katarzyna Ostasiewicz

Redakcja wydawnicza

Elżbieta Żurawska-Łuczyńska

Korekta

Katarzyna Gwizda

Skład i łamanie

Beata Mazur

Projekt okładki

Beata Dębska

Na okładce wykorzystano zdjęcia z zasobów 123 Royalty Free

Praca opublikowana na licencji Creative Commons Uznanie autorstwa

Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0).

Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>



ISBN 978-83-67899-08-6 (wersja papierowa)

ISBN 978-83-67899-09-3 (wersja elektroniczna)

DOI: 10.15611/2023.09.3

Druk i oprawa: TOTEM

Streszczenie: W artykule został podjęty temat budowy i testowania wstecznego w pełni automatycznej strategii inwestycyjnej opartej na uczeniu maszynowym. Podczas pisania artykułu zostały przeprowadzone obszerne studia literaturowe. Zostały przedstawione zarówno najważniejsze założenia i metody stosowane w budowie strategii, jak i narzędzia do testowania wstecznego. Przedstawiona została analiza wskaźnikowa i metody uczenia maszynowego z wyróżnieniem drzew decyzyjnych i lasu losowego. Zbudowany algorytm wykazał się zyskownością w prawie każdych warunkach panujących od 2008 do 2022 roku. System wykazywał dodatni wynik zarówno w trendzie wzrostowym, bocznym, jak i spadkowym. Wyniki w trendzie spadkowym i bocznym w ujęciu procentowym były znacząco wyższe od benchmarku. Wyniki w ujęciu procentowym w trendzie wzrostowym były poniżej benchmarku. Na całym okresie w stosunku do maksymalnego użycia kapitału zysk procentowy wyniósł 82%. Daje to roczny CAGR na poziomie 4,39%.

Słowa kluczowe: handel algorytmiczny, uczenie maszynowe, testowanie wsteczne

1. Wstęp

W niniejszym artykule poruszono tematykę związaną z zastosowaniem metody uczenia maszynowego do konstrukcji systemu przewidującego optymalny moment zakupu spółek notowanych na giełdzie papierów wartościowych. Przewidywanie rynku finansowego stanowi niezwykle wyzwanie, a przez dziesięciolecia powstało wiele badań dotyczących możliwości przewidywania kierunku zmian cen instrumentów finansowych notowanych na giełdzie. Inwestorzy, badacze i spekulanci opracowali liczne techniki mające na celu trafne określenie optymalnego momentu zakupu lub sprzedaży instrumentów finansowych.

Głównym celem przeprowadzonego badania jest stworzenie systemu, który będzie w stanie przewidzieć optymalny moment zakupu instrumentów finansowych przy wykorzystaniu lasu losowego. Następnie przeprowadzone zostaną testy wsteczne strategii oraz porównanie wyników z benchmarkiem oraz rentownością 10-letnich obligacji. Badanie opiera się na spółkach z indeksu WIG20 i obejmuje okres od 2008 do 2022 roku.

2. Metodyka badań własnych

2.1. Analiza wskaźnikowa – wskaźniki techniczne

Wskaźniki techniczne to narzędzia analizy, które służą do badania trendów rynkowych oraz sygnałów kupna i sprzedaży. Wskaźniki techniczne opierają się na analizie historycznych danych rynkowych, takich jak ceny, wolumeny czy zmienność. Ich głównym celem jest pomoc inwestorom w podejmowaniu decyzji inwestycyjnych i przewidywanie przyszłego ruchu cen. Jak podaje Chen (2021) można wyróżnić 2 typy wskaźników:

- oscylatory,
- nakładki.

Oscylatory

Oscylatory – czyli wskaźniki, które oscylują między lokalnym minimum a maksimum. Są nakładane na wykres powyżej lub poniżej niego. Przykładowymi wskaźnikami są RSI i MACD.

Wskaźnik RSI

Jednym z popularnych wskaźników technicznych jest RSI, który mierzy siłę ruchu cenowego w określonym okresie. RSI pokazuje, czy na rynku panuje nadmierny optymizm czy pesymizm. Jak podaje Wilder (1978) oblicza się go następująco:

$$RSI = 100 - \frac{100}{1 + RS}$$

gdzie

$$RS = \left(\frac{a}{b} \right)$$

a – średnia wartość wzrostu cen zamknięcia z y dni,

b – średnia wartość spadku cen zamknięcia z y dni.

Popularna interpretacja tego wskaźnika mówi, że kiedy spada on poniżej 30, oznacza to, że inwestorzy są bardziej skłonni do sprzedaży niż kupna, co może sugerować, że cena jest zbyt niska i może wkrótce wzrosnąć. Natomiast kiedy RSI wzrasta powyżej 70, inwestorzy są bardziej skłonni do kupna niż sprzedaży, co może sugerować, że cena jest zbyt wysoka i może wkrótce spaść. Wskaźnik RSI na rysunku 1 przedstawiającym cenę zamknięcia bitcoina prezentuje się następująco.



Rysunek 1. Wskaźnik RSI

Źródło: opracowanie własne.

Nakładki

Nakładki – wskaźniki, które wykorzystują tę samą skalę co ceny instrumentów finansowych i są nakładane na ich ceny na wykresie giełdowym. Ich przykładami są wstęgi Bollingera i SMA.

Wskaźnik wstęgi Bollingera

Wstęga Bollingera to narzędzie analizy technicznej, które zostało opracowane przez Johna Bollingera w latach 80. XX wieku. Narzędzie to, jak podaje Bollinger (2001), składa się z 3 linii na wykresie, które opisują przedział zmienności cenowej wokół średniej ruchomej. Górna linia przedziału zmienności cenowej jest oznaczona przez wartość, która jest dwukrotnością odchylenia standardowego od średniej ruchomej w górę. Dolna linia przedziału zmienności cenowej jest oznaczona przez wartość, która jest dwukrotnością odchylenia standardowego od średniej ruchomej w dół. Ceny instrumentu finansowego, które przekraczają górną lub dolną granicę przedziału zmienności cenowej, mogą wskazywać na zmianę trendu lub większą zmienność cen. Jak pisze Hayes (2023) wstęgi Bollingera są często stosowane przez inwestorów w celu analizy rynku i identyfikacji punktów zwrotu. Na przykład, gdy ceny instrumentu przekraczają górną granicę przedziału zmienności cenowej, może to

sugerować, że rynek jest przewartościowany i może dojść do spadku cen. Z kolei, gdy ceny instrumentu spadają poniżej dolnej granicy przedziału zmienności cenowej, może to sugerować, że rynek jest niedowartościowany i może dojść do wzrostu cen. Na rysunku 2 zaprezentowano ten wskaźnik.



Rysunek 2. Wskaźnik wstęgi Bollingera

Źródło: opracowanie własne.

2.2. Uczenie maszynowe

Uczenie maszynowe, jak podaje Mitchell (1997), to dziedzina sztucznej inteligencji, która umożliwi algorytmom uczenie się z danych i podejmowanie decyzji na ich podstawie bez wyraźnej instrukcji ze strony człowieka. Algorytmy zbudowane według reguł matematycznych uczą się na podstawie znalezionych wzorców w danych wejściowych, zwanych zbiorem uczącym, a następnie stosują te wzorce do podejmowania decyzji, prognozowania na nowych danych, zwanych zbiorem treningowym. Proces uczenia maszynowego zazwyczaj składa się z kilku kroków: zbierania i przetwarzania danych, wyboru odpowiedniego algorytmu uczenia maszynowego, trenowania modelu na danych treningowych, a następnie testowania i walidacji modelu na danych testowych. Algorytmy uczenia maszynowego, jak podaje Microsoft Azure (2023), są podzielone na 3 główne kategorie uczenia:

- nadzorowane,
- nienadzorowane
- ze wzmocnieniem.

Drzewa decyzyjne

Drzewa decyzyjne, jak podają Breiman, Friedman, Olshen i Stone (1984), są przykładem uczenia nadzorowanego. Są używane zarówno do klasyfikacji, jak i do regresji. Proces uczenia drzewa decyzyjnego polega na budowie hierarchii węzłów decyzyjnych, z których każdy reprezentuje jedną cechę wejściową. Algorytm budowy drzewa polega na wyborze najlepszej cechy, która najlepiej dzieli zbiór danych na podzbiory jednorodne. Następnie, dla każdego podzbioru danych, tworzone są kolejne węzły, a proces powtarzany jest rekurencyjnie, aż do osiągnięcia maksymalnej głębokości drzewa lub uzyskania liści jednorodnych, co oznacza, że dla danego liścia przypisana jest jedna klasa. Najczęściej stosowanymi metodami do wyboru cechy rozdzielającej w drzewach decyzyjnych są:

- entropia,
- zysk informacyjny,
- indeks Giniego.

W przypadku stosowania drzew decyzyjnych należy uważać na możliwość przetrenowania. Trzeba stosownie dobierać głębokość drzewa do posiadanych danych. Zaletą drzew decyzyjnych jest łatwa interpretacja i możliwość wizualizacji.

Las losowy

Las losowy, jak pisze Yiu (2019), to popularna metoda uczenia maszynowego, która jest rozszerzeniem drzewa decyzyjnego i polega na budowaniu wielu drzew decyzyjnych. Każde indywidualne drzewo w lesie losowym wydaje prognozę klasy i klasa z największą liczbą głosów staje się prognozą naszego modelu. Las losowy to metoda, której podstawową ideą jest wykorzystanie „mądrości tłumu”. W uproszczeniu oznacza to, że zbiór drzew decyzyjnych działających w grupie może osiągnąć lepszą dokładność predykcji niż pojedyncze drzewo decyzyjne. Kluczową rolę odgrywa tu niski poziom korelacji między poszczególnymi drzewami.

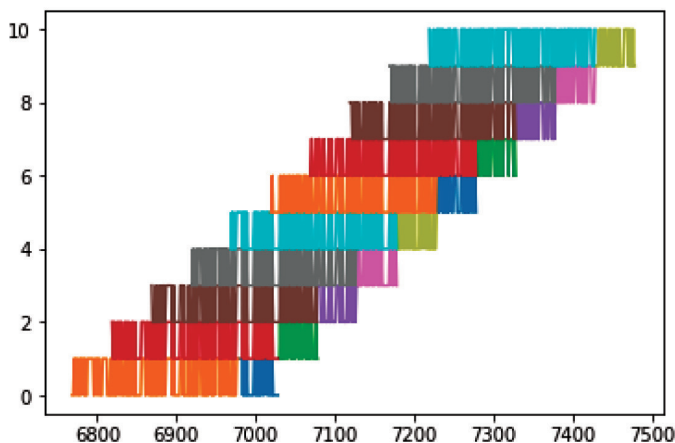
3. Przygotowanie danych i budowa modelu

Do wykonania analizy zostały wybrane dane o interwale dziennym dla spółek z indeksu WIG20. Dane obejmowały lata 2008-2022. Pobrane dane zawierały dzienną cenę zamknięcia, otwarcia, najwyższą cenę danego dnia, najniższą cenę danego dnia i wolumen dla każdej spółki.

Do przewidzenia momentu kupna akcji został wykorzystany model uczenia maszynowego w postaci lasu losowego. Został użyty algorytm klasyfikacyjny, który klasyfikuje sygnał kupna z etykietą „1”, a inną sytuację z etykietą „0”. Zmienna przewidywana została ustalona jako zmiana procentowa w ciągu 21 dni handlu na giełdzie. Algorytm przewiduje czy cena zamknięcia danej akcji wzrośnie o 10% w ciągu następujących 21 dni. Zmiennymi objaśniającymi zostały:

- cena zamknięcia,
- wolumen,
- wskaźnik RSI,
- wstęgi Bollingera,
- wskaźnik ATR,
- wskaźnik MACD,
- wskaźnik ADX,
- odchylenie standardowe.

Zbiory testowe i treningowe



Rysunek 3. Przykładowy podział danych na zbiór treningowy i testowy

Źródło: opracowanie własne.

Algorytm został tak zbudowany, żeby był wrażliwy na zmieniające się otoczenie. Model powinien się dostosowywać do trendów panujących na giełdzie. Zbiór treningowy został ustawiony na 21 dni handlu na giełdzie, a zbiór testowy na 5 dni. Czyli algorytm uczy się cały miesiąc, żeby przewidzieć następny tydzień. Program dodaje nowe dane co tydzień. Rysunek 3 obrazuje przykładowe działanie modelu w czasie dla 10 ram czasowych.

3.1. Założenia backtestera

Podczas tworzenia backtestera przyjęto założenie, że algorytm inwestuje 1000 zł w każdą transakcję. System nie ma ograniczeń kapitałowych i wykorzystuje cały dostępny sygnał do zakupu akcji. Zgodnie z założeniem nie jest możliwe ponowne zakupienie akcji tej samej spółki przed wcześniejszą sprzedażą. Analizowane wyniki odnoszą się do maksymalnego wykorzystania kapitału w danym okresie. Moment sprzedaży akcji może nastąpić w dwóch momentach, jeżeli cena akcji wzrośnie o 10% albo jeżeli spadnie o 5%.

4. Wyniki badań własnych

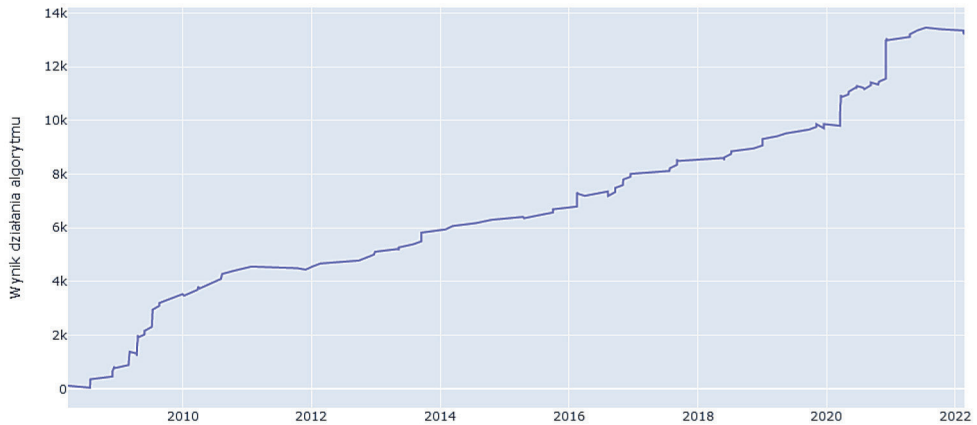
System został poddany testowaniu wstecznemu na okresie od 01.01.2008 roku do 04.07.2022 roku, który charakteryzował się dynamicznym i zróżnicowanym otoczeniem rynkowym. W tym czasie indeks WIG20 odnotował spadek o 51,01%. Przez cały badany okres miały miejsce różne wydarzenia i sytuacje, takie jak światowy kryzys finansowy w 2008 roku, wywołany kryzysem na rynku kredytów hipotecznych (subprime) w Stanach Zjednoczonych oraz upadkiem banku Lehman Brothers. Kolejnymi wyzwaniami była zmiana władzy oraz różne wizje polityki gospodarczej w Polsce, a także nowa polityka banków centralnych na całym świecie, polegająca na stosowaniu masowego luzowania ilościowego oraz niskich stóp procentowych. W 2020 roku wybuchła pandemia koronawirusa, która spowodowała paraliż wielu krajów, w tym Polski, a wiele osób zaczęło pracować zdalnie. Na koniec badanego okresu na rynki finansowe wpływały wydarzenia związane z wojną w Ukrainie oraz masową inflacją, której poziom nie był obserwowany od dziesięcioleci. Wszystkie te wydarzenia można zobaczyć na rysunku 4.



Rysunek 4. WIG20 cały badany zakres

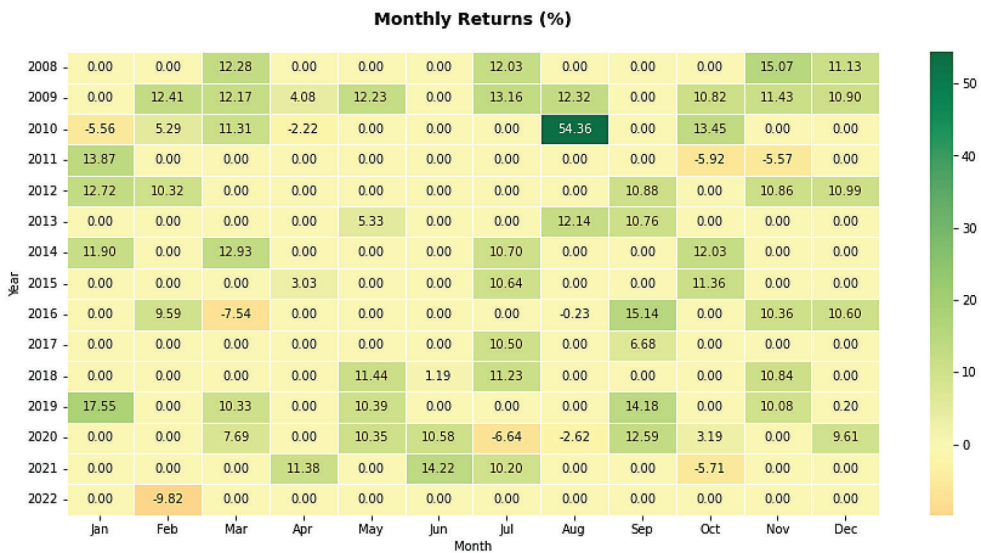
Źródło: (Bankier, 2023).

Algorytm na całym okresie osiągnął 82% zysku w stosunku do maksymalnego użycia kapitału. Benchmark, jakim jest WIG20, w tym samym okresie spadł o 51,01%. Wynik systemu można uznać za bardzo dobry. Przez cały okres zarobił 13 205,93 zł. Maksymalne użycie kapitału wynosiło 16 000 zł. Trafność podejmowanych decyzji wyniosła 78%. Zwrot w czasie w stosunku do wykorzystanego kapitału przedstawiono na rysunku 5.



Rysunek 5. Zwrot do zainwestowanego kapitału – cały okres

Źródło: opracowanie własne.



Rysunek 6. Miesięczna zmiana procentowa – cały okres

Źródło: opracowanie własne.

Jak można zauważyć na, rysunku 6 algorytm działa bardzo powtarzalnie i stabilnie. Miesięczne zwroty w stosunku do wykorzystanego kapitału są podobne. Jedynie w sierpniu 2010 roku algorytm wykazał się dużym miesięcznym zwrotem. Było to w silnym trendzie wzrostowym. Algorytm dostosowuje się do panujących warunków i nie ulega znacznym wahaniom.

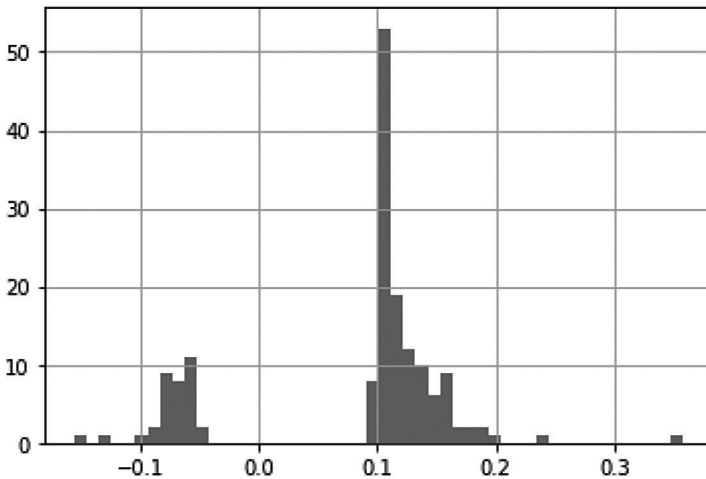
Tabela 1. Wyniki roczne – cały okres

Data	Zysk (zł)
31.12.2008	776,08
31.12.2009	2756,59
31.12.2010	888,05
31.12.2011	23,78
31.12.2012	667,56
31.12.2013	711,65
31.12.2014	475,67
31.12.2015	394,08
31.12.2016	1319,09
31.12.2017	477,18
31.12.2018	471,17
31.12.2019	905,74
31.12.2020	3120,77
31.12.2021	414,87
04.07.2022	-196,34

Źródło: opracowanie własne.

Wynik systemu w polskich złotych z podziałem na każdy rok można zobaczyć w tabeli 1. Jak można zaobserwować, największy zysk wyrażony w polskich złotych osiągnięty został w 2009 roku, 2016 i 2020 roku. Wszystkie te lata charakteryzowały się końcówką bardzo dużych spadków i mocnym po nich odbiciem. Potwierdza to dobrą zdolność do przewidywania tak zwanych dołków. Najłabszy wynik odnotowano w latach 2022 i 2011. W 2022 roku system odniósł stratę w wysokości -196,34 zł, jednak należy podkreślić, że okres badawczy zakończył się w lipcu tego roku. W roku 2011 system zarobił tylko 23,78 zł. W tym roku mogliśmy zobaczyć szczyt poprzednich wzrostów i mocne odreagowanie w dół.

Rozkład otwieranych i zamykanych pozycji można zobaczyć na rysunku 7.



Rysunek 7. Rozkład pozycji

Źródło: opracowanie własne.

Jak można zauważyć na tym rysunku, rozkład zamykanych pozycji prezentuje się dobrze. Nie zauważono nieprawidłowości w działaniu backtestera. Pewne odchyły prezentowane na wykresie wynikają z wybranego interwału dziennego. Najwyżej

zamknięta pozycja to Cd Projekt na poziomie 35,8%. Było to 08.2010 roku i wtedy w jeden dzień CDR urósł, jak podaje Bankier (2023), o 26,44%. To zdarzenie tylko potwierdza poprawność wykonanych obliczeń.

5. Podsumowanie

Zbudowany algorytm wykazał się zyskowością w prawie każdych warunkach panujących od 2008 do 2022 roku. System wykazywał dodatni wynik zarówno w trendzie wzrostowym, bocznym, jak i spadkowym. Wyniki w trendzie spadkowym i bocznym w ujęciu procentowym były znacząco wyższe od benchmarku, a wyniki w ujęciu procentowym w trendzie wzrostowym były od niego niższe. Na całym okresie w stosunku do maksymalnego użycia kapitału zysk procentowy wyniósł 82%, a zysk w polskich złotych wyniósł 13,205. Daje to roczny CAGR na poziomie 4,39%. Jak podaje Investing (2023), przeciętna rentowność polskich obligacji 10-letnich na tym okresie wyniosła 3,98%. Algorytm zatem przyniósł większy zysk od benchmarku WIG20, który spadł o 51,01% i porównywalny wynik do 10-letnich obligacji. W dalszej części należałoby dodać koszty utrzymania systemu (koszt prądu i dostępu do internetu) oraz przetestować optymalną liczbę wykorzystywanego kapitału w celu zmaksymalizowania wyniku systemu. Jednak tak zbudowany system oparty na uczeniu maszynowym można uznać za bardzo dobrą bazę do dalszego rozwoju. System okazał się zyskowny przez ponad 14 lat w wielu warunkach i wykazał lepsze wyniki od benchmarków.

Literatura

- Andrzejewski, W. i Boiński, P. (b.d.). Materiały dydaktyczne. Pobrane 21.05.2023 z http://www.cs.put.poznan.pl/pboinski/files/ED/ED_slajdy_drzewa.pdf
- Bankier. (2023). *CD Projekt SA*. Pobrane 05.05.2023 z <https://www.bankier.pl/inwestowanie/profile/quote.html?symbol=CDPROJEKT>
- Bollinger, J. (2001). *Bollinger on Bollinger Bands* (p. 0). McGraw-Hill.
- Breiman, L., Friedman, J., Olshen, R. i Stone, C. (1984). *Classification and Regression Trees*. Chapman & Hall.
- Chen, J. (2021). *Technical Indicator: Definition, Analyst Uses, Types and Examples*. Pobrane 21.05.2023 z <https://www.investopedia.com/terms/t/technicalindicator.asp>
- Cichosz, P. (2000). *Systemy uczące się*. Wydawnictwa Naukowo-Techniczne.
- Hayes, A. (2023). *Bollinger Bands: What They Are, and What They Tell Investors*. Pobrane 5.03.2023 z <https://www.investopedia.com/terms/b/bollingerbands.asp>
- Investing.com (2023). *Rentowność obligacji: Polska 10-letnie*. Pobrane 02.05.2023 z <https://pl.investing.com/rates-bonds/poland-10-year-bond-yield-historical-data>
- Microsoft Azure. (2023). Pobrane 21.05.2023 z <https://azure.microsoft.com/pl-pl/resources/cloud-computing-dictionary/what-is-machine-learning-platform>
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Quinlan, J. R. (2014). *C4.5: Programs for Machine Learning*. Elsevier.

Welles Wilder, J. (1978). *New Concepts in Technical Trading Systems*. Trend Research.

Yiu, T. (2019). *Understanding Random Fores*. Pobrane 21.05.2023 z <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>

Building and Backtesting Investment Strategy Based on Machine Learning

Abstract: The article addresses the construction and testing of a fully automated, machine learning-based investment strategy. Extensive literature studies were conducted during the writing of the article. The key assumptions and methods used in strategy construction, as well as tools for backtesting, were presented. The article includes a performance analysis and machine learning methods, highlighting decision trees and random forest. The developed algorithm demonstrated profitability under almost all prevailing conditions from 2008 to 2022. The system yielded positive results in both upward, sideways, and downward trends. The percentage-based results in downward and sideways trends were significantly higher than the benchmark, while the percentage-based results in the upward trend were below the benchmark. Over the entire period, the percentage profit in relation to the maximum capital utilization amounted to 82%. This translates to an annual Compound Annual Growth Rate (CAGR) of 4.39%.

Keywords: algorithmic trading, machine learning, backtesting