

Instytut Automatyki i Informatyki Stosowanej



na kierunku Informatyka

w specjalności Systemy Informacyjno-Decyzyjne

System optymalizacji spłaty kredytów walutowych

Łukasz Komorowski

Numer albumu 276373

promotor

Dr inż. Sebastian Plamowski

WARSZAWA 2021

Streszczenie:

[TEKST]

Słowa kluczowe – [SŁOWA]

Summary:

[TEKST]

Keywords – [SŁOWA]

Zusammenfassung:

[TEKST]

Stichworte – [SŁOWA]

Oświadczenie autora pracy - z systemu USOS oświadczenie autora pracy dyplomowej (Zarządzenie nr 28/2016 Rektora Politechniki Warszawskiej)

Spis treści

[1 Wstęp 8](#_Toc74491804)

[2 Rynek kredytowy 9](#_Toc74491805)

[2.1 Sektor bankowy 9](#_Toc74491806)

[2.2 Sektor pożyczkowy (pozabankowy) 10](#_Toc74491807)

[2.3 Kredyty mieszkaniowe 11](#_Toc74491808)

[2.4 Kredyt walutowy 13](#_Toc74491809)

[3 Rynek walutowy 16](#_Toc74491810)

[3.1 Źródła danych 19](#_Toc74491811)

[3.1.1 Dane z NBP Web API 20](#_Toc74491812)

[3.1.2 Inne serwisy walutowe 20](#_Toc74491813)

[4 Metody analizy danych 22](#_Toc74491814)

[4.1 Metody statystyczne 22](#_Toc74491815)

[4.2 Modele parametryczne 22](#_Toc74491816)

[4.2.1 Model ARIMA 22](#_Toc74491817)

[4.3 Wskaźniki techniczne 24](#_Toc74491818)

[4.3.1 SMA - Simple moving average 25](#_Toc74491819)

[4.3.2 EMA - Exponential Moving Average 27](#_Toc74491820)

[4.3.3 DEMA - Double Exponential Moving Average 27](#_Toc74491821)

[4.3.4 TEMA - Triple Exponential Moving Average 27](#_Toc74491822)

[4.3.5 MACD - Moving Average Convergence Divergence 27](#_Toc74491823)

[4.3.6 RSI - Relative Strength Index 28](#_Toc74491824)

[4.3.7 CMO - Chande Momentum Oscillator 29](#_Toc74491825)

[4.3.8 UO - Ultimate Oscillator 30](#_Toc74491826)

[4.4 Uczenie maszynowe 30](#_Toc74491827)

[4.4.1 SVR - Support Vector Regression 31](#_Toc74491828)

[4.4.2 KRR - Kernel Ridge Regression 32](#_Toc74491829)

[4.4.3 GPR - Gaussian Process Regression 32](#_Toc74491830)

[4.4.4 MLPRegressor - Multi-layer Perceptron regressor 32](#_Toc74491831)

[5 Opis systemu 33](#_Toc74491832)

[6 Wnioski 34](#_Toc74491833)

[7 Bibliografia 35](#_Toc74491834)

[8 Wykaz symboli i skrótów 36](#_Toc74491835)

[9 Spis rysunków 37](#_Toc74491836)

[10 Spis tabel 38](#_Toc74491837)

[11 Spis załączników 39](#_Toc74491838)

# Wstęp

Tekst

# Rynek kredytowy

Rynek kredytowy jest ogromny. Tylko w Polsce na koniec 2019 było 15 444 tys. klientów sektorów bankowego i pożyczkowego, którzy posiadali czynne zobowiązania, a łączna kwota zadłużenia gospodarstw domowych wyniosła 671 mld zł. i wzrosła o 7% w porównaniu z końcem 2018 roku. Zauważalny jest silny i trwały trend wzrostowy. W porównaniu z początkiem 2012 roku zadłużenie wzrosło aż o 46,6% (213 mld).

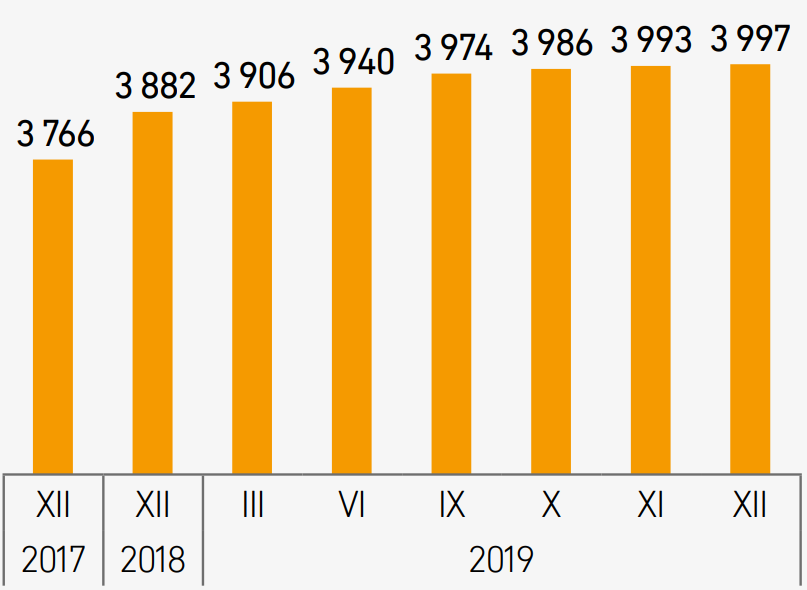
Podziału rynku można dokonać ze względu na sektory:

* bankowy,
* pożyczkowy (pozabankowy).

## Sektor bankowy

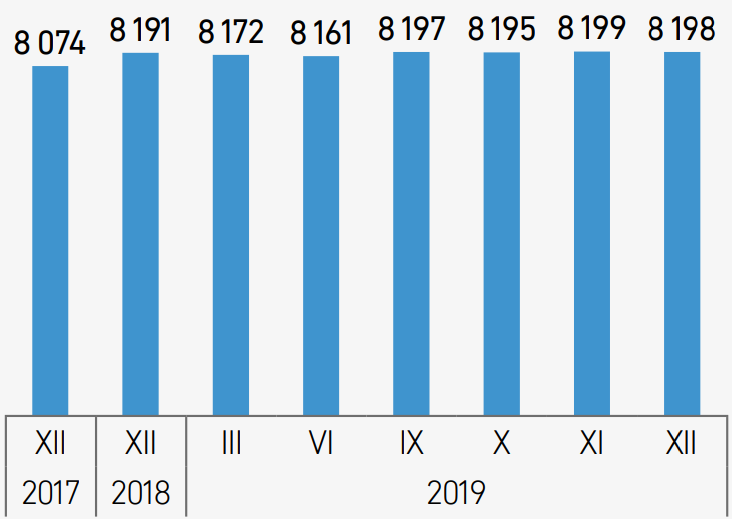
W sektorze bankowym kredytobiorców podzielić można ze względu na posiadany produkt:

* kredyt mieszkaniowy
* kredyt ratalny i gotówkowy
* linie i karty kredytowe



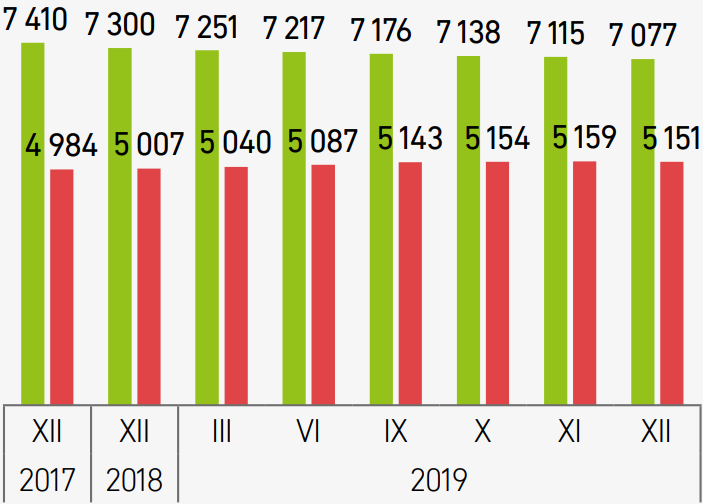
Wykres .Kredytobiorcy w tys. – kredyty mieszkaniowe[1]

Kredytobiorców mieszkaniowych w latach 2017-2019 przybywało około 3% rocznie, a na koniec roku 2019 było ich 3 997 tys. W porównaniu z początkiem roku 2012 liczba osób z czynnym kredytem mieszkaniowym wzrosła o 1 mln.



Wykres .Kredytobiorcy w tys. – kredyty ratalne i gotówkowe[1]

Liczba osób posiadająca kredyt konsumpcyjny jest dość stabilna. W porównaniu z początkiem roku 2012 zmniejszyła się ona jednak aż o 321 tys. czego częściowej przyczyny można dopatrywać się w rozwoju sektora pożyczkowego i migracji klientów.

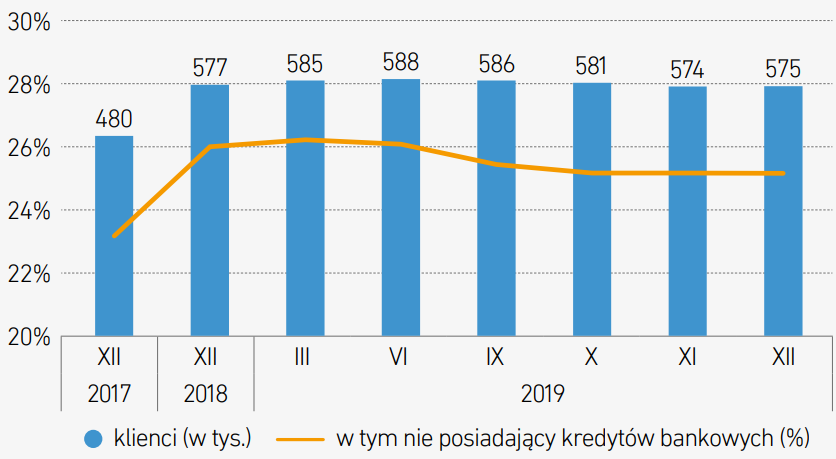


Wykres .Kredytobiorcy w tys. – linie i karty kredytowe[1]

Wzrasta liczba osób posiadających karty kredytowe, w 2018 roku jedynie o 23 tys, ale już rok później przybyło 144 tys. osób. Wyraźny spadek widać w liczbie osób, które posiadają limity kredytowe do rachunków bankowych. W roku 2018 był to spadek o 110 tys, natomiast w roku 2019 już o 223 tys.

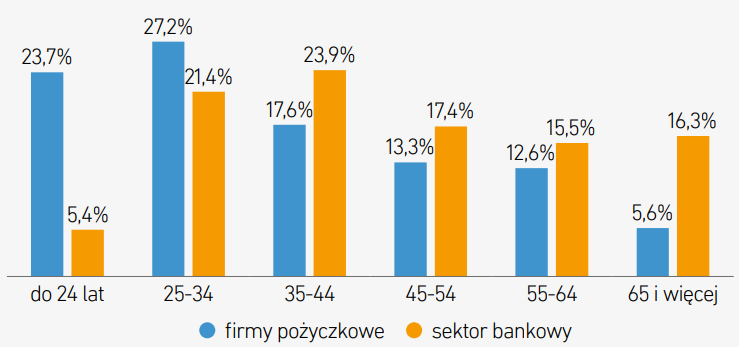
## Sektor pożyczkowy (pozabankowy)

W latach 2016-2017 widoczny był znaczny i szybki wzrost liczby pożyczkobiorców pozabankowych, jednak znaczna część tego wzrostu wynikała z coraz bliższej współpracy pomiędzy firmami pożyczkowymi a Biurem Informacji Kredytowej (BIK).



Wykres . Klienci firm pożyczkowych[1]

W 2019 roku liczba klientów firm pożyczkowych spadła w Polsce o 2,3 tys. osób.

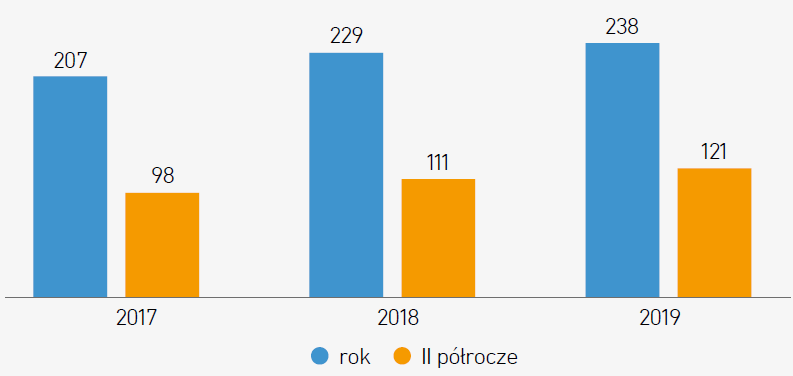


Wykres . Klienci sektora pożyczkowego bez zobowiązań wobec banków oraz posiadacze kredytów konsumpcyjnych. Struktura wiekowa[1]

Wyłącznie z pożyczek pozabankowych w Polsce na koniec 2019 roku korzystało 144,5 tys. osób. Głównie były to osoby młode (poniżej 35 roku życia).

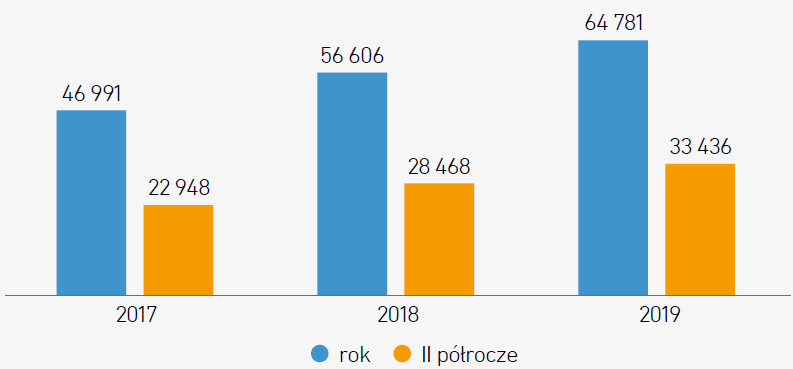
## Kredyty mieszkaniowe

Kredyt mieszkaniowy często uznawany jest za to samo co kredyt hipoteczny, jednak pojęcie kredytu mieszkaniowego jest węższe. Kredyt hipoteczny jest to kredyt którego zabezpieczenie stanowi hipoteka, ale cel kredytu może być dowolny. Kredyt mieszkaniowy jest to kredy którego zabezpieczenie również stanowi hipoteka, ale służy on do finansowania zakupu mieszkania. Ogólnie kredyt mieszkaniowy zawsze jest hipoteczny, natomiast kredyt hipoteczny nie musi być mieszkaniowy.



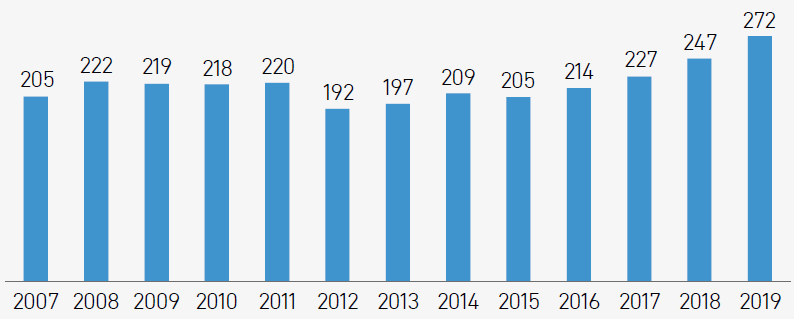
Wykres . Kredyty mieszkaniowe (w tys. szt.) udzielone w latach 2017–2019[1]

W 2019 roku liczba udzielonych kredytów wzrosła o 4% w porównaniu z 2018 (z 229 tys. do 238 tys.) i utrzymuje trend wzrostowy, pomimo wycofania wsparcia z budżetu (program MdM).



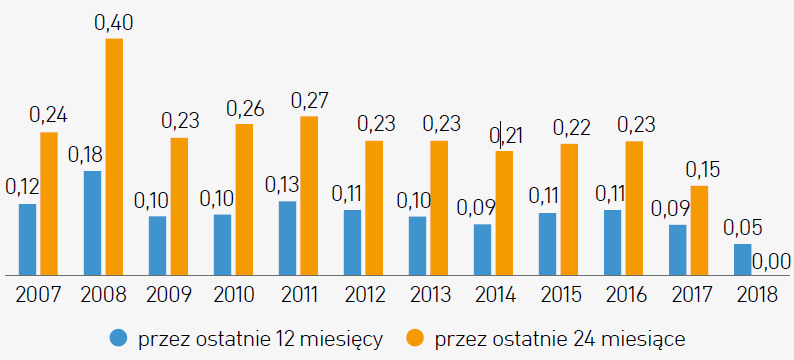
Wykres . Kwota kredytów mieszkaniowych (w mln zł) udzielonych w latach 2017–2019[1]

Stale rośnie również wartość udzielanych kredytów mieszkaniowych. Porównując do 2018 roku kwota udzielonych kredytów wzrosła o ponad 14% (z 56,61 mld do 64,78 mld).



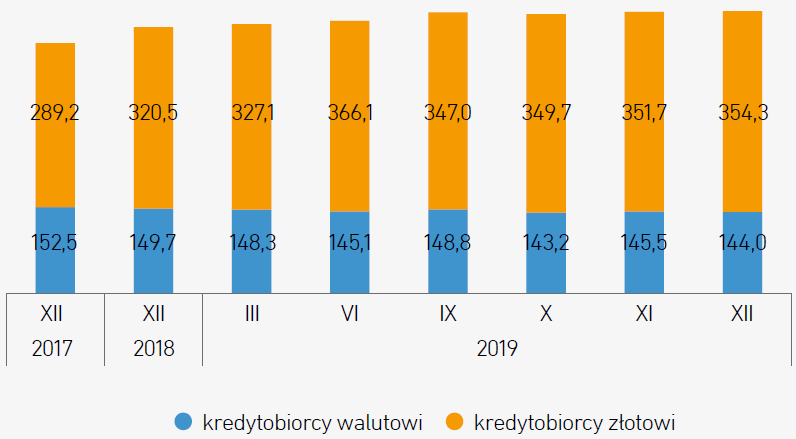
Wykres . Średnie kwoty kredytów mieszkaniowych (w tys. zł) udzielonych w latach 2007–2019[1]

Wzrost wartości udzielonych kredytów spowodowany jest nie tylko wzrostem ich liczby ale również średniej kwoty kredytu, która od 2012 ma trend wzrostowy. Porównując do roku poprzedniego, w 2019 roku kwota ta wzrosła o 10% (z 247 tys. do 272 tys.).



Wykres . Wzrosty (w pkt. proc.) udziałów rachunków opóźnionych w liczbie otwartych w rocznikach 2007–2018[1]

Bardzo powolny jest proces wzrostu wraz z wiekiem udziału kredytów opóźnionych. Zdecydowanie najwyższy wzrost notuje się w kryzysowym roczniku 2008, natomiast pozostałe roczniki kredytów psują się znacznie wolniej (około 0,1% r/r).



Wykres . Zadłużenie w (mld zł) w kredytach mieszkaniowych[1]

Stale rośnie całkowite zadłużenie na kredytach mieszkaniowych. Łączna kwota zadłużenia, w porównaniu z rokiem 2018, wzrosła w roku 2019 o prawie 6% (z 470,2 mld do 498,3 mld). Wśród kredytobiorców złotowych zadłużenie wzrosło o 33,8 mld (10,5%), a wśród kredytobiorców walutowych spadło o 5,7 mld (3,8%).

## Kredyt walutowy

Kredyty walutowe, a zwłaszcza kredyty frankowe, były w Polsce bardzo popularne w latach 2004-2009. Dawane były ze znacznie niższym oprocentowaniem niż kredyty złotowe, dlatego tak dużo osób decydowało się na nie, nie zawsze będą świadomymi ryzyka jakie podejmują.



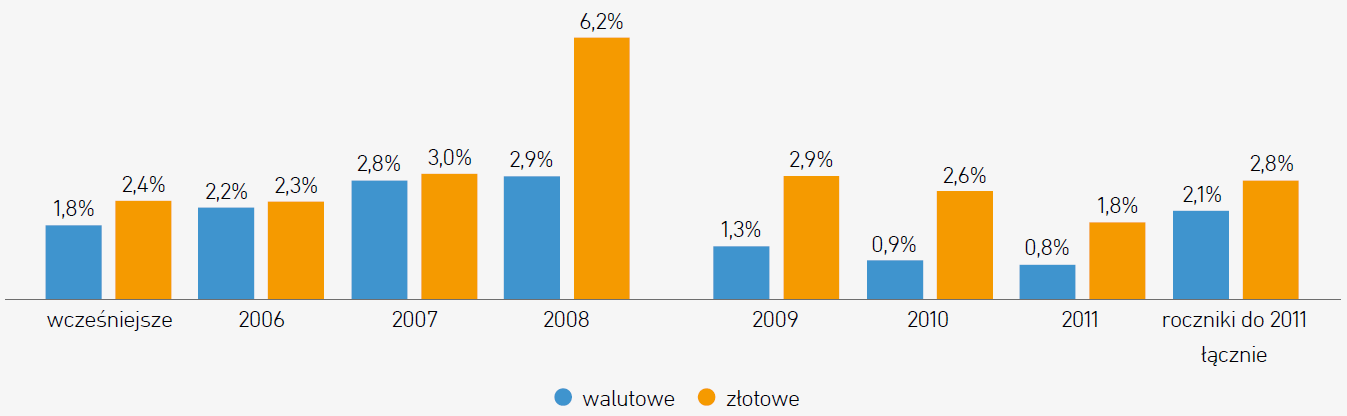
Wykres . Kurs franka szwajcarskiego[2]

Znaczny wzrost kursu franka szwajcarskiego (CHF) spowodował, że sytuacja wielu osób pogorszyła się. Kwota ich zadłużenia przewyższała kwotę zaciągniętego kredytu, a także wartość kupionej nieruchomości. W celu uregulowania udzielania kredytów w obcej walucie wprowadzono rekomendację S. Zgodnie z rekomendacją S wydaną przez KNF w czerwcu 2013 roku od dnia 1 lipca 2014 roku banki mogą udzielać kredytów tylko w walucie zarobków:

***Rekomendacja 6***

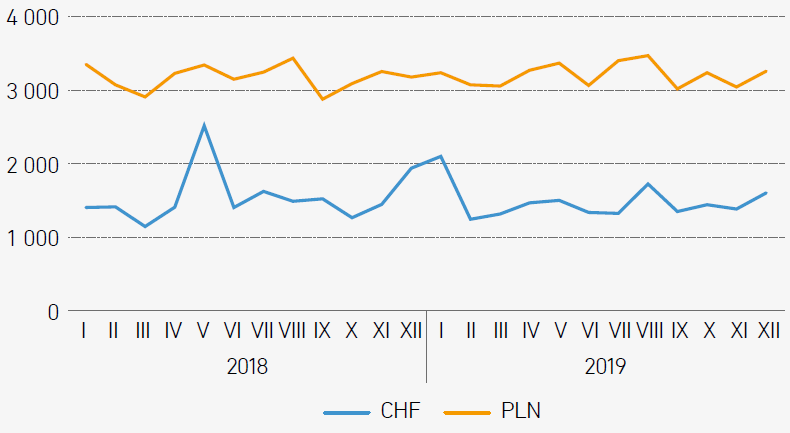
*Bank powinien udzielać klientom detalicznym kredytów zabezpieczonych hipotecznie wyłącznie w walucie, w jakiej uzyskują oni dochód, także w przypadku klientów o wysokich dochodach.[3]*

Ma to na celu wyeliminowanie ryzyka walutowego dla klientów indywidualnych.

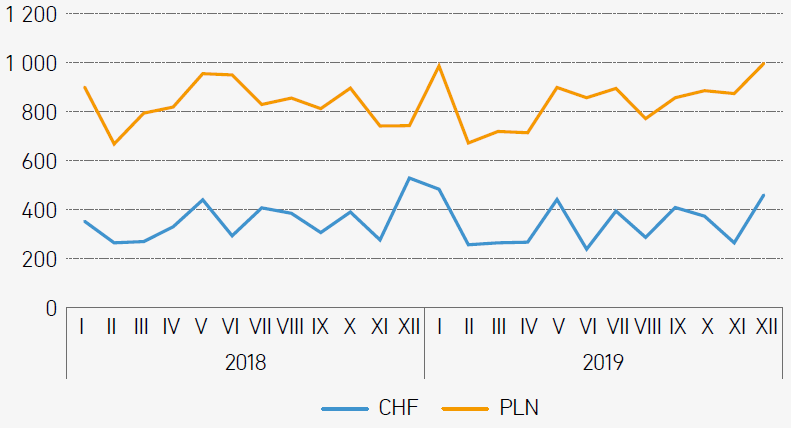


Wykres . Udziały rachunków opóźnionych w obsłudze >90 dni w liczbie kredytów mieszkaniowych udzielonych w rocznikach do 2011 r. Stan bazy BIK na koniec 2019 r. po korekcie o przewalutowania[1]

Pomimo zawirowań na rynku walutowym, szkodowość kredytów walutowych, mierzona udziałem kredytów opóźnionych o ponad 90 dni, jest niższa niż kredytów złotowych. Wysoka szkodowość rocznika 2008 może być spowodowana tym, że było w nim najwięcej kredytów walutowych i jest w nim najwięcej przewalutowań.



Wykres . Liczba wejść do statusu >30 dni opóźnienia w portfelu złotowym i frankowym. Dane miesięczne z lat 2018–2019[1]



Wykres . Liczba wejść do statusu >90 dni opóźnienia w portfelu złotowym i frankowym. Dane miesięczne z lat 2018–2019[1]

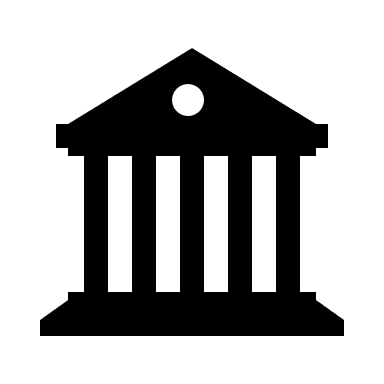
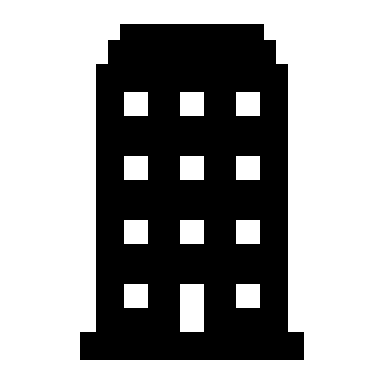
Stabilna jest jakoś kredytów (zarówno złotowych, jak i frankowych) mierzona odsetkiem wejść w status pierwszych opóźnień (ponad 30 dni) oraz poważnych opóźnień (ponad 90 dni). W 2019 roku liczba rachunków wchodzących w status pierwszych opóźnień spadła minimalnie dla kredytów frankowych i wzrosła minimalnie dla kredytów złotowych. Wartym zauważenia jest również stosunek wejść w status pierwszych opóźnień do wejść w status poważnych opóźnień. Dla kredytów złotowych wejść w poważne opóźnienia jest 3,8 razy mniej niż wejść w pierwsze opóźnienia. Dla kredytów frankowych wejść w poważne opóźnienia jest 4,3 razy mniej niż wejść w pierwsze opóźnienia. Oznacza to, że większość wczesnych zaległość regulowana jest przed przejściem w opóźnienia długoterminowe.

# Rynek walutowy

Rynek finansowy można podzielić na 3 zasadnicze części:

* rynek kapitałowy – tworzony przez giełdy papierów wartościowych, na nim inwestorzy kupują oraz sprzedają akcje, obligacje i instrumenty pochodne (termin zapadalności dłuższy niż 1 rok),
* rynek pieniężny – zawiera instrumenty finansowe o terminie zapadalności krótszym niż 1 rok, np. bony skarbowe,
* rynek walutowy – przedmiotem wymiany są tu waluty

Rynek walutowy powszechnie nazywany jest FOREX albo FX (akronimy pochodzące od angielskiego Foreign Exchange). Jest rynkiem nieregulowanym (OTC – Over The Counter) na którym obrót odbywa się poprzez wiele pojedynczych transakcji pomiędzy uczestnikami. W rzeczywistości na rynku handel odbywa się za pomocą stałego mechanizmu:

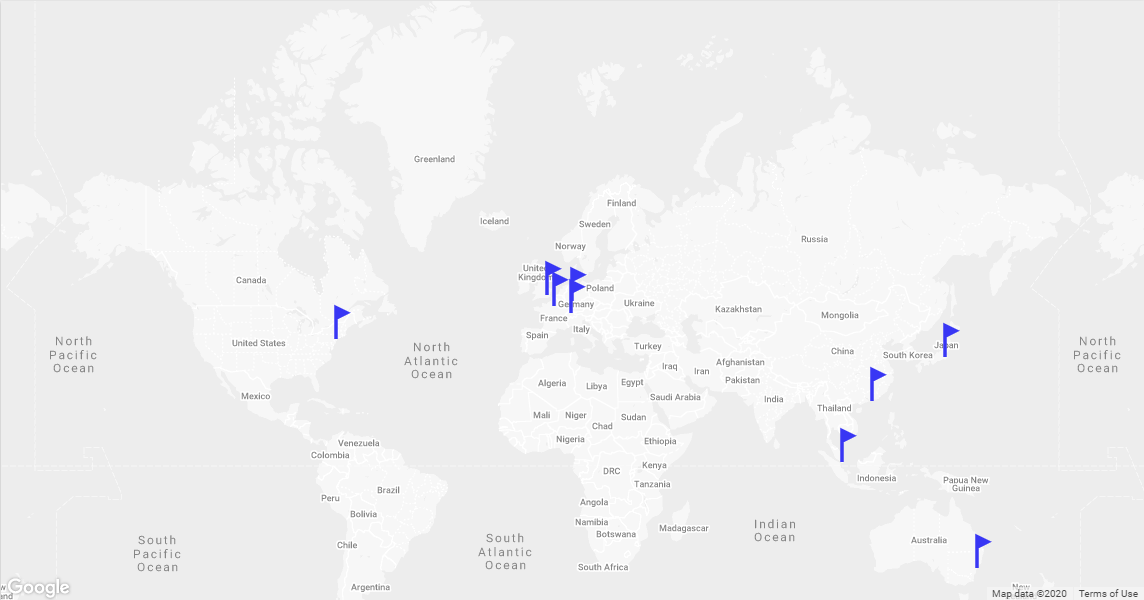


Zlecenie

Zlecenie

Rysunek . Diagram przedstawiający mechanizm działania rynku walutowego

Klienci indywidualni zlecają zakup/sprzedaż waluty firmom brokerskim, które następnie zlecają to międzynarodowym podmiotom bankowym, które dokonują między sobą transakcji o ogromnym wolumenie.



Rysunek . Mapa najistotniejszych centrów finansowych rynku walutowego

Handel na rynku walutowym odbywa się 24 godziny na dobę, od poniedziałku do piątku. Najważniejszymi centrami finansowymi z punktu widzenia wymiany walut pod względem liczby i wielkości zawieranych transakcji są:[4]

* w Europie
  + Frankfurt,
  + Londyn,
  + Paryż,
  + Zurych,
* w Azji
  + Hong Kong,
  + Singapur,
  + Tokio,
* w Australii
  + Sydney,
* w Amerykach
  + Nowy Jork.

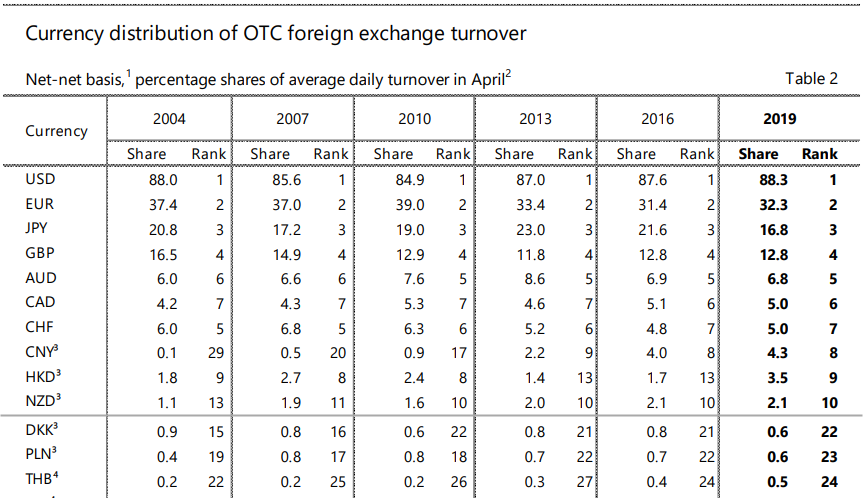


Tabela . Rozkład walut obrotu walutowego na FOREX[5]

Według ukazującego się co 3 lata raportu Banku Rozrachunków Międzynarodowych (BIS – Bank for International Settlements) najpopularniejszymi walutami na FOREX w kwietniu 2019 roku były:

* USD – dolar amerykański
* EUR – euro
* JPY – jen japoński
* GBP – funt brytyjski
* AUD – dolar australijski

Tylko te 5 walut brało udział w 78,5% transakcji zawieranych na rynku walutowym. Polski złoty znalazł się na 23 miejscu zestawienia, odpowiadając za 0,3% transakcji.

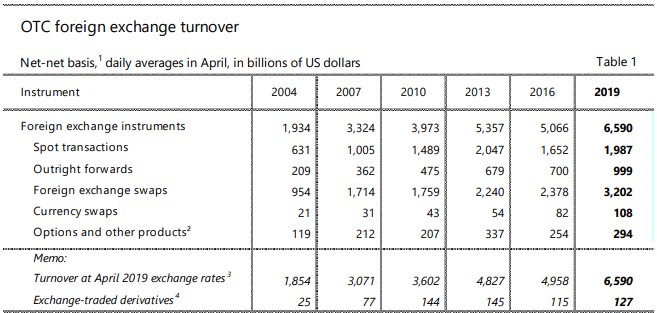


Tabela . Średnie dzienne obrotu na rynku FOREX[5]

Wartość transakcji zawieranych na rynku walutowym jest ogromna. W kwietniu 2019 roku przeciętna dzienna wartość obrotów na rynku FOREX wyniosła 6,59 bln dolarów amerykańskich. Widać trend rosnący dla wartości, która w ciągu 15 lat (2004-2019) powiększyła się 3,4 razy.

Na rynku walutowym zawiera się transakcje o różnym charakterze. Można podzielić je ze względu na motywację stron zawierających i wtedy wyróżnia się następujące typy operacji:

* inwestycyjne
* spekulacyjne
* hedgingowe
* arbitrażowe

Transakcje inwestycyjne zawierane są z długim horyzontem czasowym, a wybrana waluta związana jest najczęściej z przekonaniem co do pozycji gospodarki danego państwa, a co za tym idzie również jego waluty. Zalicza się tu również transakcje, gdzie wymiana waluty jest tylko środkiem do realizacji celu (np. zagraniczny przedsiębiorca wymienia swoją narodową walutę na PLN w celu otwarcia siedziby w Polsce).

Transakcje spekulacyjne zawierane są z dużo krótszym horyzontem czasowym, dochodzącym nawet do kilku minut lub sekund. Zawierane są na podstawie przekonania co do kształtowania się krótkookresowego kursu danej waluty.

Transakcje hedgingowe (zabezpieczające) mają na celu zabezpieczenie przed ryzykiem kursu walutowego. Wykorzystywane są najczęściej przez eksporterów którzy narażeni są na zmianę kursu waluty w której otrzymują zapłatę do waluty ich kraju. Aby się przed tym uchronić zawierają odpowiednie transakcje na rynku Forex dzięki czemu mogą rekompensować straty w przypadku osłabienia się obcej waluty w stosunku do waluty ich kraju.

Transakcje arbitrażowe są to transakcje pozbawione ryzyka. Polegają na wyszukiwaniu i wykorzystywaniu czasowej nieefektywności rynku do zawarcia konkretnych kontraktów kupna/sprzedaży. Zysk z pojedynczej transakcji arbitrażowej jest zazwyczaj niewielki, przez co staje się opłacalny tylko w przypadku dysponowania dużymi zasobami finansowymi. Klasycznym przykładem jest taki układ trzech kursów który gwarantuje zysk, np.:

* waluta2/waluta1 = 4
* waluta3//waluta2 = 2,5
* waluta3/waluta1 = 10,01

Mając 1000 j. waluty 1, inwestor (pomijając koszty zawarcia transakcji) wymienia je na walutę 2 (250 j.), następnie wymienia to na walutę 3 (100 j.) które ponownie wymienia na walutę 1 (1001 j.) co daje mu zysk w wysokości 1 j. waluty 1.

Rynek walutowy podzielić można również ze względu na termin realizacji transakcji. Wyróżnia się następujące grupy:

* transakcje SPOT
* transakcje terminowe.

Transakcje SPOT to transakcje natychmiastowe. Ustalenie kursu walut oraz rozliczenie odbywa się bardzo szybko, a waluta dostarczana jest już po 2 dniach roboczych.

Transakcje terminowe mają czas rozliczenia dłuższy niż 2 dni robocze.

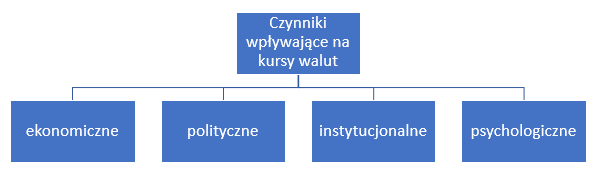
## Źródła danych

Kursy walut są bardzo ważną informacją dla wielu osób. Dotyczą one nie tylko eksporterów i importerów, ale również osób, które posiadają zobowiązania finansowe w innej walucie (np. kredyt walutowy) oraz osób, które wyjeżdżają na wakacje za granicę.

Mechanizm kształtowania się kursów walut jest podobny do analogicznych mechanizmów dla innych towarów. Głównym czynnikiem są popyt i podaż, a wyceną jest wartość jednej waluty względem innej. Wszystkie podmioty dokonujące wymiany walut mają wpływ na kształtowanie ceny.

Wszystkie czynniki wpływające na kursy walut możemy podzielić w następujący sposób:

* Czynniki ekonomiczne – zmiany wzrostu gospodarczego, inflacja, stopy procentowe oraz poziom bezrobocia,
* Czynniki polityczne – zmiany systemowe, stopień stabilności politycznej, programy poszczególnych partii politycznych oraz wydarzenia losowe lub jednorazowe,
* Czynniki instytucjonalne – inwestycje państwa na rynku walutowym,
* Czynniki psychologiczne – emocje ludzi inwestujących na Forexie.



Rysunek . Czynniki wpływające na kursy walut[6]

Bardzo często istotna jest znajomość kursu waluty w czasie rzeczywistym. Są one płynne i potrafią zmieniać się w krótkich okresach, jeżeli tylko dochodzi w nich do wymiany walut. W praktyce kursy aktualizowane są w następującym oknie czasowym:

* Od północy w niedzielę (rusza wtedy rynek Azjatycki)
* Do północy w piątek (zamyka się wtedy rynek Amerykański)

Kursy średnie walut są wyliczane na podstawie informacji o wymianie waluty przekazywanej przez banki oraz inne duże instytucje finansowe do agencji informacyjnych, takich jak Bloomberg czy Thompson Reuters. W odpowiedzi, banki otrzymują informację o kursach średnich i generują wykresy wahań kursowych. Następnie z tych informacji korzystają kantory (stacjonarne oraz internetowe), które do kursów średnich dokładają swoją marżę, a następnie udostępniają te wartości swoim klientom.

### Dane z NBP Web API

Kurs średni podawany przez NBP opiera się na danych z banków. W celu rzetelnego wyliczenia kursu średniego, powinny zostać zebrane informacje o kursach ze wszystkich banków i kantorów, jednak byłoby to zadanie niezwykle trudne w realizacji oraz nieopłacalne. Dlatego NBP swoje wyliczenia opiera na kursach EUR/PLN i USD/PLN otrzymanych w godzinach 10:55-11:00 każdego dnia roboczego od 10 największych banków w Polsce. Odrzucone zostają dwa najwyższe i dwa najniższe kursy, aby następnie, wyciągając średnią z kursów sprzedaży i kupna, obliczyć kurs średni.

Między godziną 11:45, a 12:15 w każdy dzień roboczy NBP publikuje tabelę A kursów średnich walut obcych. Zawiera ona informacje o kursie średnim dla 35 najpopularniejszych walut. Dodatkowo w każdą środę, w tych samych godzinach publikowana jest tabela B kursów średnich walut obcych w której znajdują się informacje o kursach średnich dla 115, bardziej „egzotycznych” walut.

Łatwy dostęp do danych z rynku walutowego oraz najbardziej aktualnych kursów walut sprawia, że w dzisiejszych czasach kursy średnie prezentowane przez NBP służą głownie do obliczeń statystycznych oraz są pomocne przy wyliczeniach księgowych.

Narodowy Bank Polski udostępnia publiczne Web API umożliwiające wykonywanie zapytań na następujących zbiorach:

* Aktualne i archiwalne kursy walut obcych
* Aktualne i archiwalne ceny złota

Odpowiedź, zależnie od wymagania, zwracana jest w formacie JSON lub XML (domyślnie – JSON). Pojedyncze zapytanie nie może obejmować więcej niż 93 dni, natomiast dane archiwalne dostępne są dla następujących okresów:

* Dla kursów walut – od 2 stycznia 2002 r.
* Dla cen złota – od 2 stycznia 2013 r.

### Inne serwisy walutowe

W Internecie istnieje wiele serwisów oferujących dostęp do danych z rynku Forex w czasie rzeczywistym. Dane, udostępniane poprzez API pozwalają na wykorzystanie ich w komputerowym przetwarzaniu danych oraz, dzięki odpowiedniej analizie, na inwestowanie na rynku walutowym. Przykładowe serwisy udostępniające takie dane to:

* FCS API
* AvaTrade
* IG
* FXCM
* FIXER

Dostęp do API w takich serwisach jest zazwyczaj darmowy, ale również ograniczony (ograniczona dostępna liczba zapytań, ograniczona liczba dni dostępu albo opóźnienie w danych). Najczęściej za dane w czasie rzeczywistym trzeba zapłacić wykupując odpowiedni abonament.

Z uwagi na charakter tej pracy oraz na brak zapotrzebowania na najświeższe dane, zdecydowano się wykorzystać API udostępnione przez NBP z uwagi na jego wiarygodność, wystarczający poziom aktualności danych oraz darmowy dostęp.

# Metody analizy danych

W kolejnych rozdziałach opisane zostaną wykorzystane w tej pracy metody wyboru optymalnego dnia do wymiany waluty. Można podzielić je na cztery kategorie:

* Metody statystyczne
* Modele parametryczne
* Wskaźniki techniczne
* Uczenie maszynowe

Każdy z rozdziałów zawierać będzie opis danej metody, sposób doboru odpowiednich parametrów oraz ocenę uzyskiwanych wyników.

## Metody statystyczne

Pierwszą wykorzystaną w tej pracy metodą wyboru optymalnego dnia do wymiany waluty jest statystyka. Celem tego badania było sprawdzenie czy istnieje dzień, tydzień lub miesiąc który statystycznie najczęściej jest najtańszy lub najdroższy. Przebadano następujące statystyki:

* Ranking dni w tygodniu
* Najtańszy dzień w miesiącu
* Najtańszy tydzień w roku

**UZUPEŁNIĆ DATY DLA DANYCH**

Pobrane uprzednio dane z API NBP dla okresu od \_\_\_.\_\_\_.\_\_\_\_\_\_ do \_\_\_.\_\_\_.\_\_\_\_\_\_ zostały podzielone na:

* Dane treningowe (od … do …)
* Dane testowe (od … do …)

Dane treningowe posłużyły do wyliczenia wymienionych wcześniej statystyk, natomiast dane testowe wykorzystano do oceny przydatność tych statystyk w opisywanym w tej pracy procesie decyzyjnym.

## Modele parametryczne

Moduł statsmodels udostępnia klasy i funkcję służące estymacji wielu różnych modeli statystycznych, a także przeprowadzaniu testów statystycznych oraz eksploracji danych statystycznych. Oferuje on szeroką listę parametrów statystycznych dla każdego estymatora. Wyniki obliczeń są porównywane z istniejącymi modelami statystycznymi w celu zapewnienia ich poprawności. Moduł statsmodels jest udostępniany na licencji open source.



Rysunek . Moduł statsmodels w Pythonie[7]

### Model ARIMA

Jednym z elementów modułu statsmodels jest autoregresyjny zintegrowany model średniej ruchomej - ARIMA (ang. Autoregressive Integrated Moving Average). Model ten, zgodnie z nazwą, złożony jest z trzech komponentów:

* Procesu autoregresyjnego,
* Procesu średniej ruchomej,
* Stopnia integracji.

Proces autoregresyjny korzysta z modelu AR (ang. Autoregressive Model). Każda z wartości jest liniową kombinacją p wartości poprzednich, gdzie p jest rzędem modelu autoregresyjnego. Model taki definiuje się w następujący sposób:

gdzie:

* wartość szeregu w chwili t,
* parametry modelu – wskazują na wpływ poprzednich wartości szeregu na wartość bieżącą,
* pewna stała (pomijana w celu uproszczenia),
* biały szum.

Proces średniej ruchomej korzysta z modelu MA (ang. Moving Average). Podobieństwo do modelu AR polega na korzystaniu z danych z chwil poprzednich, jednak model MA korzysta z wartości zaburzeń w poszczególnych momentach z przeszłości. Rząd modelu q określa z ilu wcześniejszych zaburzeń korzystamy. Model MA definiuje się w następujący sposób:

gdzie:

* wartość szeregu w chwili t,
* wartość oczekiwana ,
* parametry modelu – wskazują na wpływ poprzednich wartości białego szumu na wartość bieżącą szeregu,
* biały szum.

Model ARIMA, w odróżnieniu od modelu ARMA, można stosować do procesów które są niestacjonarne. Aby możliwe było takie zastosowanie, procesy te muszą być sprowadzalne do stacjonarnej postaci. Procesem stacjonarnym nazywamy taki proces, który zachowuje niezmienne w czasie średnią i wariancję, natomiast jednym ze sposobów sprowadzenia procesu niestacjonarnego do postaci stacjonarnej jest różnicowanie szeregu (ang. Differencing). Polega ono na obliczaniu przyrostów (różnic) jakie dany szereg osiągnął w czasie. Nie zawsze jednokrotne różnicowanie pozwala na osiągnięcie zamierzonego efektu. W takim wypadku korzysta się z wielokrotnego różnicowania, a za określenie stopnia różnicowania odpowiada parametr d.

Połączenie trzech opisanych powyżej metod generuje model ARIMA(p, d, q), gdzie:

* p – rząd autoregresji,
* d – stopień integracji szeregu,
* q – rząd średniej ruchomej.[8]

W 1970 roku George Box wraz z Gwilymem Jenkinsem w książce pod tytułem „*Time series analysis: Forecasting and control”* opisali swoje podejście do modelowania procesów ARIMA. Według nich modelowanie to obejmuje identyfikację odpowiedniego procesu ARIMA, dopasowanie go do danych, a następnie wykorzystanie dopasowanego modelu do prognozowania dalszych wartości. Oryginalnie procedura modelowania Boxa-Jenkinsa obejmowała iteracyjny, trzyetapowy proces wyboru modelu, jednak nowsze źródła (Makridakis, Wheelwright i Hyndman: *„Forecasting: methods and applications”,* 1998 rok) dodają na początku etap przygotowania danych oraz na końcu etap stosowania wybranego modelu. Poniżej opisano każdy z tych etapów.

* Przygotowanie danych – etap ten obejmuje transformację i różnicowanie danych. Transformacja pozwala na ustabilizowanie wariancji szeregu, natomiast proces różnicowania został już opisany powyżej.
* Identyfikacja modelu ARIMA – w tym etapie na podstawie wykresów autokorelacji oraz częściowej autokorelacji danych przekształconych i zróżnicowanych dobiera się odpowiednie procesy ARIMA. Powstały różne narzędzie wyboru, np. kryterium informacyjne Akaikego AIC (ang. Akaike Information Criterion).
* Wybór parametrów – etap ten polega na wyszukaniu takich współczynników modelu, które zapewniają najlepsze dopasowanie do danych.
* Sprawdzenie modelu – w tym etapie wykonywane są testy dopasowania modelu do danych w celu zidentyfikowania obszarów, w których model jest nieodpowiedni. W przypadku nieodpowiedniego dopasowania modelu należy wrócić do etapu identyfikacji modelu ARIMA i wybrać lepszy model.
* Prognozowanie – jest to etap w którym wykorzystuje się wybrany model do obliczenia prognozy.[9]

WYKORZYSTANIE ARIMA DO PRZEWIDYWANIA KURSU

## Wskaźniki techniczne

Wskaźniki techniczne służą do analizy sygnałów płynących z rynku. Są to wzory matematyczne wykorzystujące dane historyczne i używane przez analityków oraz inwestorów w celu przewidywania przyszłych trendów cenowych, a następnie podejmowania odpowiednich decyzji na tej podstawie.

Wskaźniki techniczne prezentowane są zazwyczaj w formie wykresu i porównywane z odpowiednim wykresem cenowym w celu analizy. Różnorodne mechaniki wskaźników technicznych pozwalają na wychwycenie zachowania i psychologii inwestorów, aby na tej podstawie zapewnić wskazówki o przyszłych trendach cenowych. Niektóre wskaźniki generują sygnały samodzielnie, inne natomiast stanowią jedynie uzupełnienie informacji. W celu lepszego zrozumienia sygnałów płynących z rynku w analizie techniczne wykorzystuje się zazwyczaj więcej niż jeden wskaźnik techniczny. Większość wskaźników technicznych jest niespecyficznych, co oznacza, że można wykorzystać je na dowolnym rynku, natomiast niektóre wskaźniki są przeznaczone do stosowania dla określonego rynku finansowego.

Wyróżnić można dwa typy wskaźników technicznych:

* Oscylatory – koncentrują się na dynamice rynku. Inwestorzy określają ceny wejścia i wyjścia za pomocą oscylatorów wahających się w określonym zakresie.
* „Nakładki” – używają tej samej skali co cena, dlatego z powodzeniem można je „nakładać” na wykres ceny. Najczęściej stosowane są średnie ruchome różnego rodzaju.

W 2018 roku powstał projekt którego wynikiem było powstanie paczki FinTA dla Pythona. Zawiera ona implementację wielu popularnych wskaźników technicznych z użyciem Pandas. Wskaźniki wykorzystywane w tej pracy znajdują się na liście dostępnych w FinTA, co pozwoliło na pominięcie etapu implementacji i wcześniejsze zajęcie się testowaniem oraz doborem parametrów.

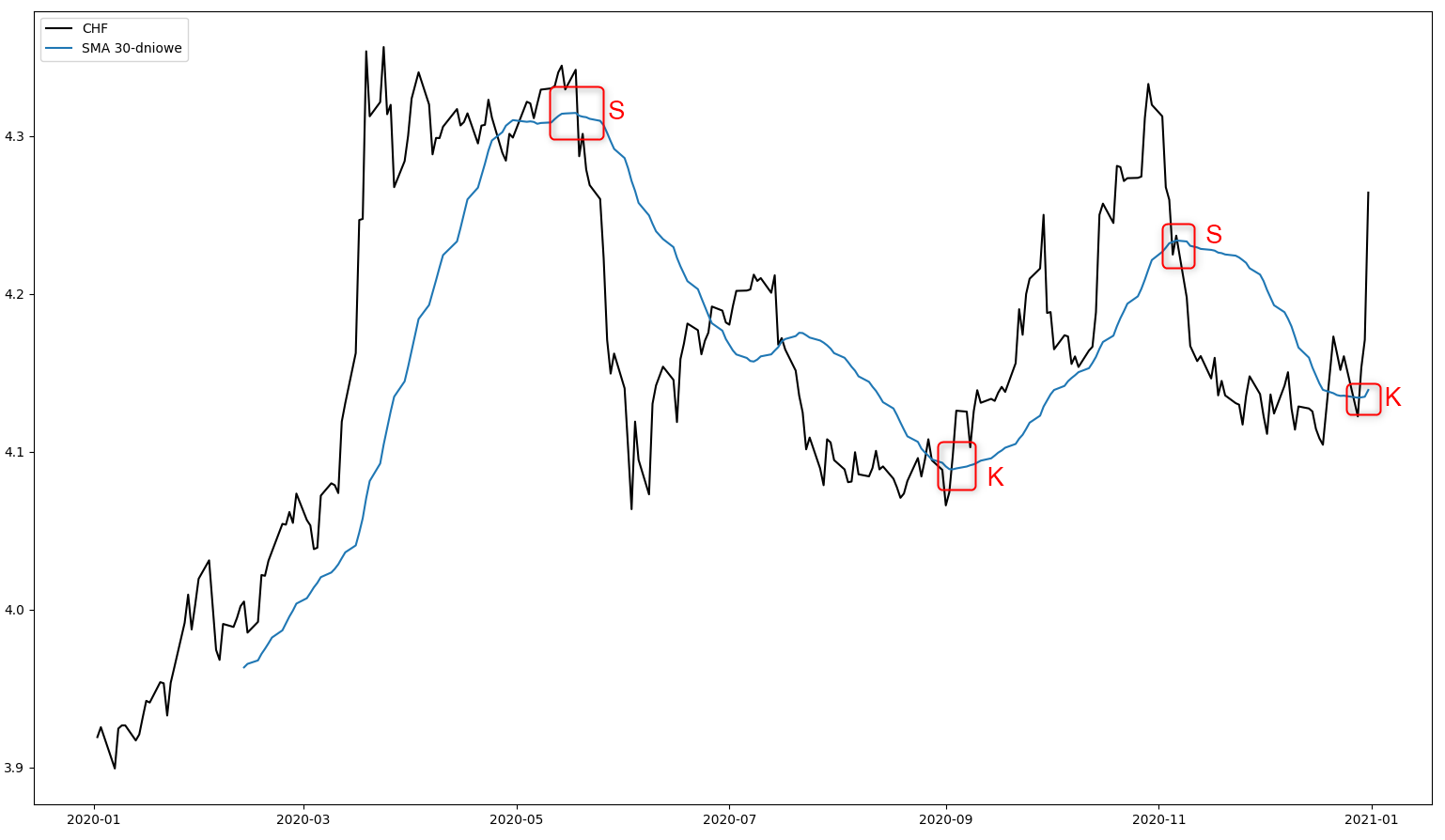
W kolejnych podrozdziałach opisane zostały wskaźniki techniczne wykorzystane w tej pracy oraz poddano ocenie ich użyteczność przy optymalizacji spłaty kredytu walutowego.

### SMA - Simple moving average

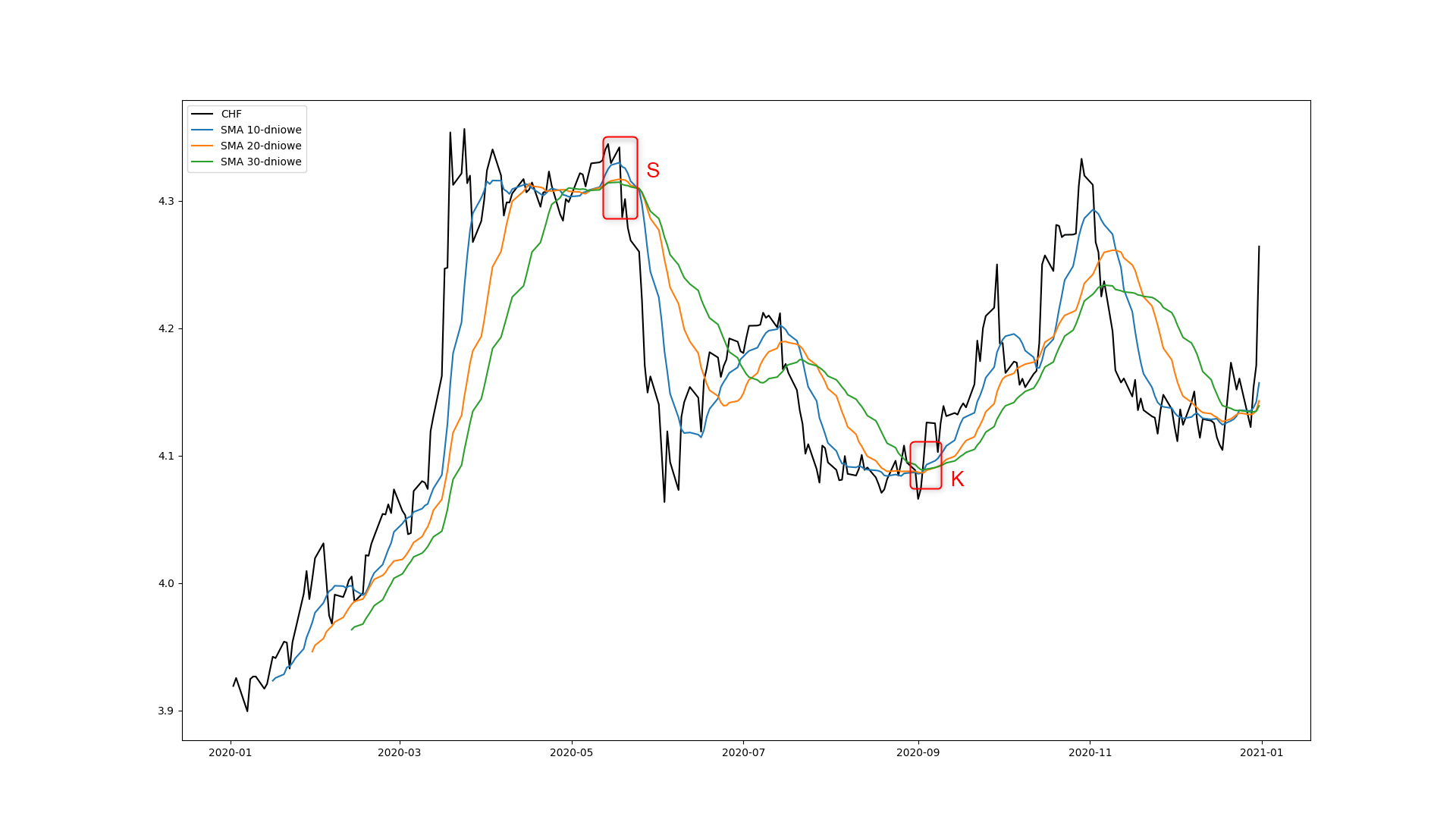
Prosta średnia krocząca (SMA – Simple Moving Average) jest wskaźnikiem opartym na średniej arytmetycznej cen z pewnej liczby ostatnich okresów. Wzór na obliczenie jego wartości wygląda następująco:

WZÓR

Wyróżnia się trzy sposoby wykorzystania informacji jakie niesie SMA:

* Obserwacja punktów przecięcia się SMA z wykresem ceny – w przypadku gdy wykres ceny przecina nierosnący wykres SMA od góry, oznacza to sygnał sprzedaży (oznaczony jako S). W odwrotnym przypadku, czyli, gdy wykres ceny przecina niemalejący wykres SMA od dołu, oznacza to sygnał kupna (oznaczony jako K).  
  

Rysunek . Wykorzystanie SMA - wariant 1

* Obserwacja punktów przecięcia się kilku SMA o różnym okresie z wykresem ceny - w tym przypadku gdy wykres ceny przecina wszystkie nierosnący wykresy SMA od góry, oznacza to sygnał sprzedaży (oznaczony jako S). W odwrotnym przypadku, czyli, gdy wykres ceny przecina wszystkie niemalejący wykresy SMA od dołu, oznacza to sygnał kupna (oznaczony jako K).   
  

Rysunek . Wykorzystanie SMA - wariant 2

* Obserwacja punktów przecięcia się dwóch SMA o różnym okresie ze sobą – w tym wypadku gdy wykres średniej o krótszym okresie przecina od dołu niemalejącą średnią o dłuższym okresie, oznacza to sygnał kupna (oznaczony jako K). W przypadku gdy wykres „krótszej” SMA przecina od góry nierosnący wykres „dłuższej” SMA, oznacza to sygnał sprzedaży (oznaczony jako S).   
  

Rysunek . Wykorzystanie SMA - wariant 3

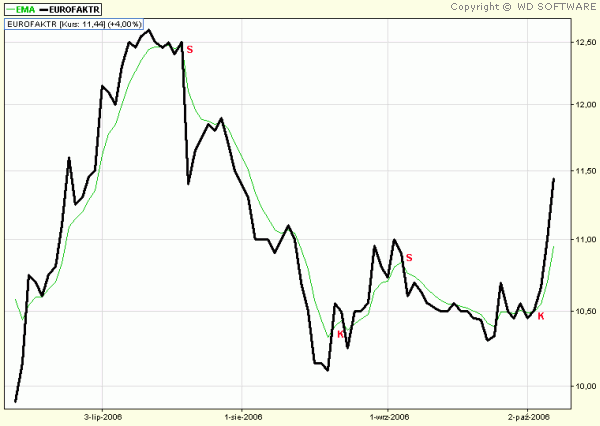
Jak widać na wykresach powyżej, korzystanie z prostej średniej kroczącej może stanowić dobre źródło informacji o nadchodzących trendach kursu. Należy jednak pamiętać, że dużo zależy od przyjętego okresu dla SMA, dlatego uzyskane w ten sposób sygnały trzeba skonfrontować z innymi wskaźnikami technicznymi.

### EMA - Exponential Moving Average

Wykładnicza średnia krocząca (EMA - Exponential Moving Average) jest wskaźnikiem który dużo większą wagę przywiązuje do najnowszych wartość niż do starszych danych, dzięki czemu jest bardziej czuła na zmiany i pozwala na dokładniejsze przewidywanie trendu. Wzór na obliczenie EMA wygląda następująco:

WZÓR

Wykorzystanie informacji jakie niesie wykładnicza średnia krocząca jest identyczne jak w przypadku SMA omówionym w poprzednim podrozdziale. Poniższy wykres przedstawia wykorzystanie EMA w sposób analogiczny jak opisany jako wariant 1 wykorzystania SMA.



Rysunek . Wykorzystanie EMA - wariant 1[10]

EMA liczona z CHF i ocena ewentualnego wykorzystania

### DEMA - Double Exponential Moving Average

Tekst

### TEMA - Triple Exponential Moving Average

Tekst

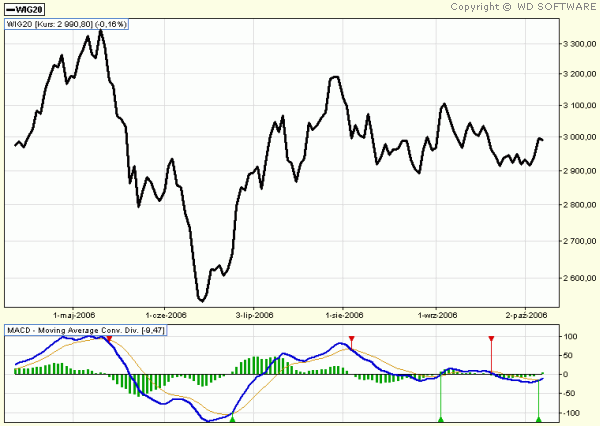
### MACD - Moving Average Convergence Divergence

Wskaźnik MACD jest jednym z najpopularniejszych i najchętniej używanych wskaźników technicznych. Dzieje się tak, ponieważ jest on niezwykle łatwy do interpretacji, daje wiarygodne sygnały i pozwala uzyskiwać wiarygodne wyniki zarówno w okresach stabilizacji, jak i podczas spadków lub wzrostów cen. Wadą tego wskaźnika jest stosunkowo wysoki poziom opóźnienia generowanych przez niego sygnałów, dlatego lepiej sprawdza się w inwestycjach z długim horyzontem czasowym.

Wzór na obliczenie MACD wygląda następująco:

WZÓR

Najczęściej wykorzystuje się EMA z 26 oraz 12 dni. Dodatkowo na wykresie rysuje się linię prezentującą EMA (najczęściej z 9 dni) obliczoną ze wskaźnika MACD. Linia ta jest nazywana linią sygnałową.



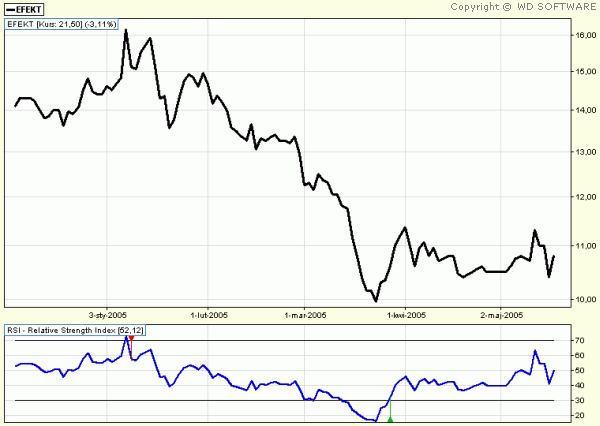
Rysunek . Wskaźnik MACD[10]

Powyższy wykres przedstawia wykorzystanie wskaźnika MACD. Gdy linia MACD przecina od dołu niemalejącą linię sygnałową oznacza to sygnał kupna (zielona strzałka na wykresie). W przypadku gdy linia MACD przecina od góry nierosnącą linię sygnałową oznacza to sygnał sprzedaży (czerwona strzałka na wykresie

MACD liczona z CHF i ocena ewentualnego wykorzystania

### RSI - Relative Strength Index

Indeks siły względnej (RSI - Relative Strength Index) jest obok MACD jednym z najczęściej wykorzystywanych wskaźników technicznych. Autor wskaźnika, John Welles Wilder Jr., proponował wykorzystanie do obliczeń okres 14 dni, jednak spotyka się wykorzystania okresów 7 i 9 dniowych (analizy krótkoterminowe) oraz 21 i 25 dniowych (analizy długoterminowe). Wartości wskaźnika wahają się w zakresie od 0 do 100.



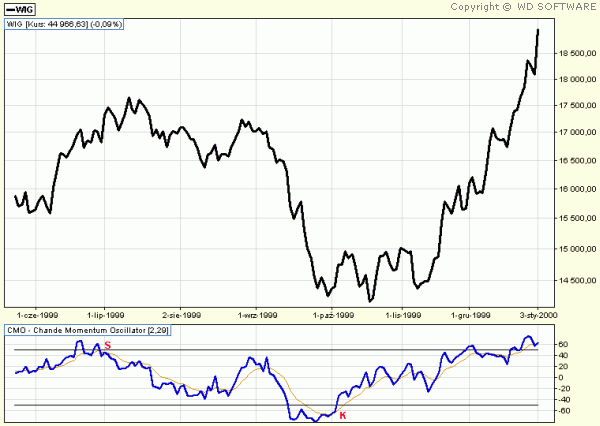
Rysunek . Wskaźnik RSI[10]

Powyższy wykres przedstawia wykres kursu oraz obliczonego z niego wskaźnika RSI. Wykorzystując ten wskaźnik określa się poziom wykupienia (RSI = 70-80) oraz poziom wyprzedania (RSI = 20-30). Gdy wykres przetnie od dołu ustalony poziom wyprzedania, oznacza to sygnał kupna (zielona strzałka na wykresie), a gdy przetnie od góry ustalony poziom wykupienia, oznacza to sygnał sprzedaży (czerwona strzałka na wykresie).

RSI liczone z CHF i ocena ewentualnego wykorzystania

### CMO - Chande Momentum Oscillator

Wskaźnik CMO generuje sygnały podobne do RSI, jednak mocniej reaguje na zachowania kursu i częściej generuje sygnały. Osiąga on również inne wartości – od -100 do 100, a linie wykupienia i wyprzedania oznacza się w nim odpowiednio na poziomie 50 i -50.



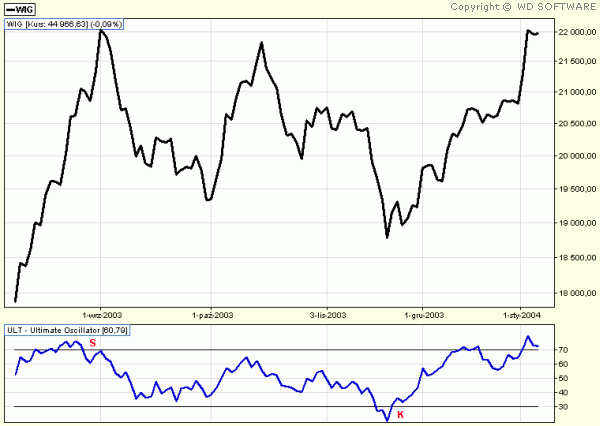
Rysunek . Wskaźnik CMO

Powyższy wykres przedstawia wykres kursu oraz obliczony z niego wskaźnik CMO (niebieska linia). Wykorzystanie tego wskaźnika jest identyczne jak w przypadku RSI: gdy wykres przetnie od dołu poziom wyprzedania, oznacza to sygnał kupna (oznaczony jako K), a gdy przetnie od góry poziom wykupienia, oznacza to sygnał sprzedaży (oznaczony jako S).

CMO liczone z CHF i ocena ewentualnego wykorzystania

### UO - Ultimate Oscillator

Wskaźnik Ultimate Oscillator nazywany jest często po prostu Ultimate i oznaczany skrótem ULT. Od RSI oraz CMO odróżnia go to, że do jego wyliczenia używa się ważonej sumy 3 oscylatorów opartych na okresach o różnym czasie trwania. Najczęściej wykorzystuje się oscylatory 7, 14 i 28 dniowe oraz wagi o wartości odpowiednio 4, 2 i 1. Problemem wielu różnych oscylatorów jest to, że w sytuacji, gdy cena szybko rośnie potrafią one generować fałszywe sygnały. Oscylator Ultimate ma za zadanie skorygować taki sygnał poprzez uwzględnienie kilku przedziałów czasowych w wyliczeniach, co odróżnia go od wielu innych wskaźników impetu, które wyliczane są na podstawie wybranego przedziału czasowego.



Rysunek . Wskaźnik UO

Powyższy wykres przedstawia wykres kursu oraz obliczony z niego wskaźnik UO. Wskaźnik osiąga wartości od 0 do 100. Wykorzystanie informacji płynących ze wskaźnika jest identyczne jak w przypadku wskaźnika RSI. Dla tego wskaźnika określa się poziom wykupienia (RSI = 70-80) oraz poziom wyprzedania (RSI = 20-30). Gdy wykres przetnie od dołu ustalony poziom wyprzedania, oznacza to sygnał kupna (oznaczony jako K), a gdy przetnie od góry ustalony poziom wykupienia, oznacza to sygnał sprzedaży (oznaczony jako S).

UO liczone z CHF i ocena ewentualnego wykorzystania

## Uczenie maszynowe

Uczenie maszynowe (ang. Machine Learning) pozwala na przeprowadzenie na komputerze procesu uczenia się. Nie istnieje jedna definicja określająca czym dokładnie jest uczenie maszynowe. Po raz pierwszy określenia Machine Learning użył w 1959 roku Arthur Samuel i służyło ono wtedy do określenia kierunku studiowania, który daje komputerom możliwość uczenia się bez wcześniejszego wyraźnego zaprogramowania („Field of study that gives computers the ability to learn without being explicitly programmed”). Inną, równie znaną, definicję uczenia maszynowego sformułował Tom Mitchell w 1997 roku: „Mówi się, że program komputerowy uczy się na podstawie doświadczenia E w odniesieniu do pewnej klasy zadań T i pomiaru wydajności P, jeśli jego wydajność przy zadaniach w T, mierzona przez P, poprawia się wraz z doświadczeniem E” („A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E „)[11].

Istnieją różne kryteria według których można podzielić Machine Learning. Jednym z nich jest podział ze względu na dostarczane dane:

* Uczenie nadzorowane – dane dostarczane maszynie zawierają również informację o oczekiwanej odpowiedzi.
* Uczenie nienadzorowane – dane dostarczane maszynie nie zawierają informacji o oczekiwanej odpowiedzi
* Uczenie przez wzmocnienie – brak określonych danych wejściowych lub wyjściowych, maszyna otrzymuje jedynie sygnały wzmocnienia (nagrody i kary).

Uczenie maszynowe można podzielić również ze względu na typ zadanie jakie jest stawiane prze maszyną:

* Klasyfikacja – zadanie klasyfikacji polega na przypisaniu każdego obiektu do odpowiedniej kategorii na podstawie jego cech.
* Grupowanie – zadanie grupowania polega na pogrupowaniu na podstawie cech wszystkich obiektów w grupy o zbliżonych charakterystykach
* Regresja – zadanie regresji polega na przewidywaniu wartości pewnego atrybutu na podstawie cech obiektu.

Wykorzystanie uczenia maszynowego w Pythonie ułatwia paczka scikit-learn, zawierająca wiele narzędzi związanych z Machine Learningiem. Zbudowana na NumPy, SciPy oraz matplotlib, jest uważana za jedno z najlepszych rozwiązań dostępnych obecnie w licencji Open Source[12]



Rysunek . Logo projektu scikit-learn

W kolejnych podrozdziałach opisane zostały metody uczenia maszynowego wykorzystane w tej pracy oraz poddano ocenie ich użyteczność przy optymalizacji spłaty kredytu walutowego.

### SVR - Support Vector Regression

Regresja wektora nośnego została w 1996 roku opracowana przez Harrisa Druckera i jego wspólników. Jest to technika uczenia nadzorowanego u której podstaw leży koncepcja wektorów pomocniczych opracowana przez Vladimira Vapnika. Celem SVR jest określenie hiperpłaszczyzny celem minimalizacji błędu oraz zmniejszenie różnic pomiędzy przewidywaniem, a rzeczywistą obserwacją.

WYKRESY SVR DLA CHF Z WYKORZYSTANIEM RÓŻNYCH JĄDER I OCENA PRZYDATNOŚCI

### KRR - Kernel Ridge Regression

Regresja grzbietowa jądra łączy regresję grzbietową i klasyfikację z wykorzystaniem jądra. Dzięki temu KRR uczy się funkcji liniowej w przestrzeni tworzonej przez odpowiednie jądro i dane. W przypadku, gdy wykorzystuje się jądro nieliniowe, odpowiada to nieliniowej funkcji w oryginalnej przestrzeni. Forma uzyskiwanego modelu nie różni się od SVR, ale do jego uzyskania stosuje się różne funkcje straty.

WYKRESY KRR DLA CHF Z WYKORZYSTANIEM RÓŻNYCH JĄDER I OCENA PRZYDATNOŚCI

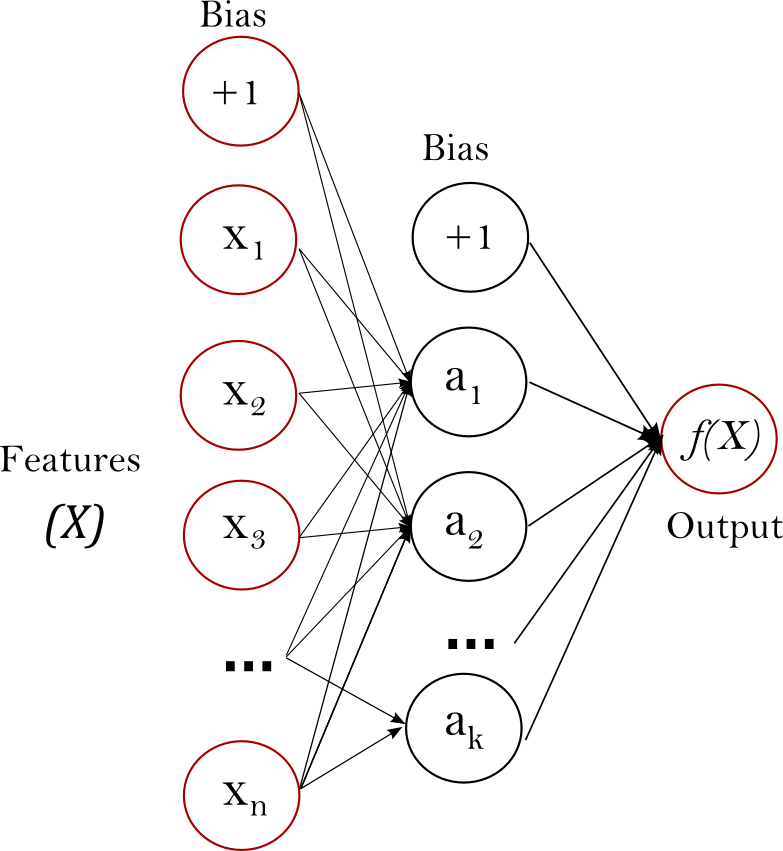
### GPR - Gaussian Process Regression

GPR jest nieparametrycznym, bayesowskim podejściem do regresji. Jest efektywne przy względnie małych zestawach danych i pozwala na predykcję z pominięciem wcześniejszego dopasowania.

WYKRESY GPR DLA CHF I OCENA PRZYDATNOŚCI

### MLPRegressor - Multi-layer Perceptron regressor

Perceptron wielowarstwowy jest najpopularniejszym typem sztucznych sieci neuronowych. Wykorzystywany w uczeniu nadzorowanym jest w stanie nauczyć się aproksymować funkcję nieliniową w celu dalszej klasyfikacji lub regresji. MLP składa się z jednej warstwy wejściowej i wyjściowej oraz kilku warstw ukrytych. W przypadku regresji, warstwa wyjściowa składa się z neuronów liniowych.



Rysunek . MLP z jedną warstwą ukrytą[12]

WYKRESY WYKORZYSTANI MLP DLA CHF I OCENA PRZYDATNOŚCI

# Opis systemu

Tekst

# Wnioski

Tekst

# Bibliografia

1. Biuro Informacji Kredytowej, Raport Biura Informacji Kredytowej KREDYT TRENDY 2019
2. Kursy średnie NBP – Frank szwajcarski (CHF) [w:] money.pl [on-line]
3. Komisja Nadzoru Finansowego, Rekomendacja S dotycząca dobrych praktyk w zakresie zarządzania ekspozycjami kredytowymi zabezpieczonymi hipotecznie, czerwiec 2013 r.
4. Rynek walutowy bez tajemnic [w:] nbportal.pl [on-line]
5. Bank Rozrachunków Międzynarodowych, Triennial Central Bank Survey – Foreign exchange turnover in April 2019, 16.09.2019
6. Czynniki wpływające na kursy walut [w:] inwestycyjnyporadnik.pl [on-line]
7. statistical models, hypothesis tests, and data exploration [w:] statsmodels.org [on-line]
8. Płonka M.: Co trzeba wiedzieć korzystając z modelu ARIMA i które parametry są kluczowe? [w:] support.predictivesolutions.pl [on-line]
9. Hyndman R.J.: Box-Jenkins modeling, 25.05.2001
10. Encyklopedia Analizy Technicznej [w:] bdm.com.pl [on-line]
11. Mitchell, T.: Machine Learning, 1997r.
12. https://scikit-learn.org/stable/index.html

# Wykaz symboli i skrótów

Tekst

# Spis rysunków

[Rysunek 1. Diagram przedstawiający mechanizm działania rynku walutowego 16](#_Toc74488008)

[Rysunek 2. Mapa najistotniejszych centrów finansowych rynku walutowego 16](#_Toc74488009)

[Rysunek 3. Czynniki wpływające na kursy walut[6] 19](#_Toc74488010)

[Rysunek 4. Moduł statsmodels w Pythonie[7] 22](#_Toc74488011)

[Rysunek 5. Wykorzystanie SMA - wariant 1 25](#_Toc74488012)

[Rysunek 6. Wykorzystanie SMA - wariant 2 26](#_Toc74488013)

[Rysunek 7. Wykorzystanie SMA - wariant 3 26](#_Toc74488014)

[Rysunek 8. Wykorzystanie EMA - wariant 1[10] 27](#_Toc74488015)

[Rysunek 9. Wskaźnik MACD[10] 28](#_Toc74488016)

[Rysunek 10. Wskaźnik RSI[10] 29](#_Toc74488017)

[Rysunek 11. Wskaźnik CMO 29](#_Toc74488018)

[Rysunek 12. Wskaźnik UO 30](#_Toc74488019)

[Rysunek 13. Logo projektu scikit-learn 31](#_Toc74488020)

[Rysunek 14. MLP z jedną warstwą ukrytą[12] 32](#_Toc74488021)

# Spis tabel

[Tabela 1. Rozkład walut obrotu walutowego na FOREX[5] 17](#_Toc62636476)

[Tabela 2. Średnie dzienne obrotu na rynku FOREX[5] 18](#_Toc62636477)

# Spis załączników

Tekst