

## **WPROWADZENIE**

Sztuczna inteligencja (AI) to badanie, w którym systemom komputerowym nadaje się zdolność do wykonywania zadań, które zwykle zależą od ludzkiej inteligencji. AI pozwala zaawansowanej technologii zastąpić człowieka w procesach gromadzenia i analizy danych, a nawet w procesie podejmowania ostatecznych decyzji. W sektorze bankowym drastycznie wzrosło wykorzystanie nowoczesnych technologii. Dlatego sztuczna inteligencja jest obecnie wdrażana w wielu bankach na całym świecie. Wcześniej banki wymagały do swojego funkcjonowania zaangażowania człowieka. Teraz, wraz z wdrożeniem różnych technologii w bankach, zaowocowało to maksymalną wydajnością. Stąd sztuczną inteligencję można znaleźć w bankach w usługach związanych z spersonalizowanymi finansami, gdzie AI pomaga w badaniu giełd i oferuje rekomendacje zgodnie z celami monetarnymi klientów. Analizowanie ogromnych ilości danych oraz przeprowadzanie różnych obliczeń i przewidywań pomogło w automatycznych decyzjach o pożyczkach podejmowanych przez sztuczną inteligencję w celu osiągnięcia lepszych zwrotów. Unikalne zdolności AI w przetwarzaniu języka pomogły zwiększyć efektywność procesów w bankach, co w konsekwencji pomogło w ograniczeniu błędów ludzkich. Moc przetwarzania mowy za pomocą sztucznej inteligencji pomogła zautomatyzować różne zgłoszenia do obsługi klienta, a to pomogło skrócić czas oczekiwania klientów, którzy czekają na przedstawiciela klienta, który przyjmie ich telefon. Klienci branży ubezpieczeniowej i bankowej oczekują teraz szybkiej i sprawnej obsługi. Milenialsi stawiają na wygodę. W ubezpieczeniach obszary takie jak rozpatrywanie roszczeń, wykrywanie oszustw, underwriting, obsługa klienta itp. ewoluowały w kierunku technologii usprawniającej te procesy. Głównym powodem takiego rozprzestrzeniania się sztucznej inteligencji jest lepsza analiza i wgląd w dane. Branża usług bankowych i finansowych również znajduje kilka zastosowań sztucznej inteligencji, takich jak doradcy kredytowi, planowanie portfela obligacji i uzyskiwanie wglądu w transakcje, diagnozowanie błędów itp. Obliczenia, takie jak stopy zwrotu, analizy szeregów czasowych, indeksy aktualizacji itp. są specyficzne dla dziedziny bankowości i finansów, i są one przeprowadzane za pomocą technik obliczeniowych rankingu ryzyka, algorytmów programowania liniowego, wieloatrybutowego scoringu, weryfikacji protokołów itp. Banki inwestycyjne stosują sztuczną inteligencję, aby zapobiegać nieuczciwym działaniom oraz zapewniać standardową i ulepszoną obsługę klienta oraz rozwiązania w czasie rzeczywistym, do dokumentacji cyfrowej itp. Banki wykorzystują inteligencję maszynową i Big Data do zbierania informacji o klientach w celu oceny ich zdolności kredytowej i oferowania spersonalizowanych usług. Algorytmy AI umożliwiają opracowywanie wysoce wyrafinowanych strategii inwestycyjnych i ograniczanie prania pieniędzy poprzez analizę danych wewnętrznych, które są publicznie dostępne w sieci klientów. Sztuczna inteligencja pozwala bankom oceniać ryzyko przy użyciu dużych ilości danych, rozumieć i dostosowywać się do zmieniającego się krajobrazu finansowego, być bardziej spójnym w swoich operacjach, dodawać wartość i podejmować wysokiej jakości decyzje. AI jest również wykorzystywana w sektorze ubezpieczeniowym. Konsument zazwyczaj kupuje polisę ubezpieczeniową i zgłasza roszczenia, i to jest podstawowy związek pomiędzy konsumentem a ubezpieczycielem. Obecnie sztuczna inteligencja w sektorze ubezpieczeniowym jest wykorzystywana do wzmacniania zdolności i wiedzy ubezpieczyciela, a nie konsumenta. Jeżeli ubezpieczyciel ma większą wiedzę na temat przeszłych roszczeń klienta i jeżeli roszczeń jest bardzo mniej lub wcale, to produkt ubezpieczeniowy może być sprzedany po niższej cenie, co może być konkurencyjne na rynku. W tym rozdziale omówiono niektóre z kluczowych zastosowań sztucznej inteligencji w ubezpieczeniach, bankowości i finansach, takie jak analiza predykcyjna, algorytmy genetyczne (GA) i wykrywanie anomalii.

## **ANALIZA PREDYKTYWNA I JEJ ZASTOSOWANIA**

Eksploatacja danych i analiza predykcyjna mają szerokie zastosowanie w biznesie i są znane pod różnymi nazwami, takimi jak odkrywanie wiedzy, tworzenie sensów itp. Jednak najnowsza nazwa to data

science. Eksploracja danych polega na wykorzystywaniu ogromnych zbiorów danych do identyfikowania trendów lub wzorców w celu rozwiązywania problemów. Sztuczna inteligencja polega na wykorzystaniu zautomatyzowanych metod do systematycznego analizowania ogromnych porcji danych i wyciągania rozsądnych i rozsądnych wniosków. Ten rodzaj analizy danych pomaga nam udowodnić lub obalić hipotezę, odkryć nieznane informacje, zidentyfikować relacje między danymi oraz klasyfikować i przewidywać dodatkowe dane dotyczące przyszłych zdarzeń. Jednym z najczęstszych przykładów analizy predykcyjnej w sektorze e-commerce jest plan reagowania kryzysowego Walmart. Walmart skutecznie wykorzystuje dane historyczne dotyczące punktu sprzedaży, aby zwiększyć satysfakcję klientów. Główne zdarzenia pogodowe wpływają na zachowania zakupowe klientów, takie jak hurtowe zakupy artykułów takich jak woda butelkowana, pop-tart, taśma izolacyjna itp. przed burzą. Woda butelkowana i taśma klejąca zostały kupione, ponieważ były częścią zaleceń udzielonych przez rząd, więc to jest potwierdzenie; jednak wzrost sprzedaży pop-tart jest zaskakujący, a więc jest odkryciem nowego związku. Ta wiedza na temat zachowań zakupowych klientów może pomóc Walmart spełnić ich wymagania, zapewniając gotowość dostaw. Przewidywanie takich potrzeb klientów pozwoliło Walmartowi zareagować na czas podczas huraganu Katrina. Analiza predykcyjna jest rozszerzeniem procesu eksploracji danych; można ją uznać za jej uzupełnienie. Polega na tworzeniu modeli wykorzystywanych do przewidywania przyszłych wydarzeń poprzez wykorzystanie przeszłych trendów lub wzorców. Eksploracja danych i analiza predykcyjna są tak samo analityczne, jak uczenie maszynowe i matematyka. Proces analizy to w 80% przygotowanie danych i 20% analiza. Narzędzia data science, analiza predykcyjna i Big Data znacząco wpłynęły na sposób zarządzania łańcuchami dostaw, tworząc nowe wyzwania i możliwości. Nowe środowisko bogate w dane sprawiło, że niektóre standardowe praktyki w SCM (zarządzanie łańcuchem dostaw) stały się przestarzałe. Analiza predykcyjna to podkategoria nauki o danych. Wiąże się to z kilkoma podejściami ilościowymi i jakościowymi, w przeciwieństwie do statystyki, która jest ilościowa. Analityka predykcyjna to nie tylko prognozowanie; odpowiada na pytania, co by się stało w przeszłości, gdyby warunki były inne. W przeciwieństwie do optymalizacji, która polega na znalezieniu maksymalnej lub minimalnej funkcji z pewnymi ograniczeniami, analityka predykcyjna ma również na celu ustalenie, jakie są cechy systemu, który nie jest optymalny. Analityka predykcyjna pozwala firmom optymalizować istniejące procesy, uzyskiwać lepszy wgląd w zachowania klientów, identyfikować szanse oraz przewidywać wyzwania i je łagodzić. Obejmuje duży zestaw technik i metodologii, na które składają się matematyka, statystyka i sztuczna inteligencja. Dodaje do tego dużą ilość zarządzania danymi. Analityka predykcyjna ma charakter indukcyjny. Zamiast zakładać rzeczy, pozwala danym przejąć inicjatywę i bada te dane za pomocą obliczeń neuronowych, uczenia maszynowego, technik statystycznych, sztucznej inteligencji, matematyki obliczeniowej itp. w celu zidentyfikowania relacji lub wzorców w tych danych, które mogą ujawnić się w przyszłości. Analiza predykcyjna odeszła daleko od statystyki, a dziś, z pomocą zaawansowanych komputerów i technologii, stosuje się techniki, takie jak drzewa decyzyjne, sieci neuronowe, maszyny wektorów nośnych (SVM), zaawansowane algorytmy matematyczne i GA, które obejmują złożone obliczenia i iteracje, aby zrozumieć ogromne ilości danych. Nadzorowane i nienadzorowane analizy predykcyjne: nadzorowane uczenie się polega na wykorzystaniu danych historycznych, które składają się z wyników, które próbujesz przewidzieć. Obejmuje techniki, takie jak regresja, klasyfikacja, analiza szeregów czasowych itp. Uczenie nienadzorowane nie wykorzystuje danych historycznych ani wyników w modelowaniu predykcyjnym, ale wykorzystuje statystyki opisowe i nie przewiduje wartości docelowej. Istotnym zastosowaniem analizy predykcyjnej jest prognozowanie cen akcji i ich zmienności. Relacja między zwrotem a ryzykiem to jeden z najważniejszych aspektów nowoczesnych finansów. Właśnie dlatego pomiar i prognozowanie zmienności jest podstawowym obszarem wyceny aktywów i zarządzania ryzykiem. Narzędzia do modelowania dynamiki zmienności są wykorzystywane także w innych dziedzinach ekonomii, a także naukach społecznych, przyrodniczych, a nawet medycynie. Znajduje również

zastosowanie w ekonomice rolnictwa. Deregulacje w sektorze użyteczności publicznej doprowadziły również do kilku nowych zastosowań w modelowaniu zmienności cen gazu i energii. Engle (1982) zaproponował model ARCH (autoregresywny warunkowy heteroskedastyczność), który był kluczowym fundamentem w modelowaniu i prognozowaniu zmienności jako zmiennej w czasie funkcji obecnych informacji. Bollerslev (1986) zaproponował klasę modeli GARCH (uogólniona autoregresyjna warunkowa heteroskedastyczność), która jest niesławna ze względu na jej zastosowania w prognozowaniu zmienności. Podstawowy model GARCH zakłada, że przyszłe warunkowe wariancje są pod wpływem szoków symetrycznych, podczas gdy niektóre aktywa reagują asymetrycznie na przeszłe stopy zwrotu, co może prowadzić do większych zmienności w przyszłości. Modele T-GARCH, A-GARCH i wykładnicze GARCH są powszechnie używanymi modelami GARCH do wyjaśnienia tego typu asymetrii. Modele te sugerują, że szoki związane z zmiennością wykazują wykładniczy spadek. Prognozy korelacyjne to kolejna dziedzina zastosowania analizy predykcyjnej. W przeciwieństwie do zmienności, korelacje uwzględniają zwroty z dwóch lub więcej aktywów. Jeśli portfel składa się ze średnich korelacji, wskazuje to na symetrię. Oznacza to, że wyniki inwestycji nie będą zbyt dobre na spadkowych rynkach, ponieważ te korelacje będą wzrastać. Korelacje są powiązane z cyklem koniunkturalnym, a zmiana sytuacji gospodarczej może wpłynąć na oczekiwane zwroty, a tym samym na strukturę korelacji. Stwierdzono, że korelacja nie jest stała na podstawie badania Longina i Solnika, w którym oszacowali model GARCH, który dał miarę warunkowej korelacji miesięcznej i przyjęto, że jest ona stała. Model był przydatny do testowania stałości korelacji, ale nie jest jasne, w jaki sposób można go wykorzystać do prognozowania korelacji. Prognozowanie korelacji jest niezbędne przy wycenie pochodnych papierów wartościowych.

Ceny papierów wartościowych często wykazują duże skoki, a zastosowanie standardowego modelu GARCH może prowadzić do przeszacowania zmienności. W przeciwieństwie do tego, bezwarunkowe prognozy zmienności pokazują prognozy większe niż w rzeczywistości powinny być. Podobnie w przypadku szacunków korelacji, jeśli tylko jedna z akcji wykazuje tendencję do dużego skoku cen, to oszacowanie jest zorientowane w kierunku zera. Jeśli ekoskoki mają ten sam znak, to powoduje przesunięcie oszacowań w kierunku (minus) 1. Estymatory DCC (dynamicznej korelacji warunkowej) mają elastyczność jednowymiarowych modeli GARCH, a jednocześnie nie są tak skomplikowane, jako wielowymiarowe modele GARCH. Mają one znaczną przewagę obliczeniową nad MV-GARCH, ponieważ liczba parametrów, które należy oszacować, nie zależy od liczby szeregów, które należy skorelować. DCC są uogólnieniem CCC (stała korelacja warunkowa) zaproponowanej przez Bollersleva w 1990 roku. Model DCC oparty na zakresie łączy model CARR (warunkowy zakres autoregresji) z ramą DCC i jest w stanie przewyższyć standardowy model DCC, który jest oparty na zwrocie pod względem zdolności prognozowania. Dzieje się tak, ponieważ dane dotyczące zakresu są bardziej wydajne niż dane dotyczące zwrotów w szacowaniu zmienności. Dotyczy to zarówno analizy w próbie, jak i poza próbą. ARIMA (autoregresywna zintegrowana średnia ruchoma) to kolejna ważna metoda prognozowania w celu przewidywania wartości szeregów czasowych na podstawie danych wprowadzanych do modelu. ARIMA to solidna metoda uchwycenia dynamicznego zachowania szeregów czasowych w przypadku krótkich okresów czasu. Zmieniające się trendy lub składnik sezonowości nie wpływają negatywnie na skuteczność tej metody prognozowania. ARIMA działa znacznie lepiej niż technika prognozowania HoltWinters, a także prognozowanie przy użyciu agregatu wartości trendów. Podejście do analizy sentymentu połączone z analizą predykcyjną służy do zrozumienia związku między przyszłymi ruchami cen akcji a nastrojami giełdowymi na mikroblogach. Krótki, ale zwięzły charakter mikroblogów giełdowych wraz z ich dużą liczbą, częstotliwością i publikowaniem w czasie rzeczywistym sprawia, że mikroblogi giełdowe są kluczowym czynnikiem napędzającym wydarzenia na giełdzie, a tym samym przyczyniają się do ich zdolności predykcyjnej w tej dziedzinie. Zakłada się również, że sentymenty niedźwiedzi mają lepszą dokładność predykcyjną niż

sentymenty bycze, które są bardziej myśleniem życzeniowym, ponieważ przyciągają większą uwagę inwestorów. Badania pokazują również, że kanały Twittera są dobrym źródłem danych do predykcyjnej analizy nastrojów. Pozytywne nastroje w tweetach zapowiadają podobny wzrost cen akcji. Klasyfikowanie tweetów za pomocą strefy neutralnej SVM pomaga nam poprawić korelację między tweetami a ceną zamknięcia akcji. Odchylenie trendu w bogactwie ma znaczną zdolność predykcyjną w prognozowaniu zwrotów dużych portfeli. Bez rozważania decyzji dotyczących portfela wielookresowego, jeśli skupimy się na przewidywalności w zwrotach z aktywów, może to być bardziej wiarygodna miara, ponieważ możemy nie być zmuszeni polegać na technikach symulacji, które mogą znacznie zmniejszyć liczbę wstępnie przypisanych zmiennych, które można uwzględnić w regresji predykcyjnej. Regresja predykcyjna może obejmować kilka zmiennych, takich jak płynność, makroekonomiczne i opóźnione zwroty. Kolejnym istotnym zastosowaniem jest przewidywanie upadłości banków. Korzystanie z komputerowych systemów EWS (systemów wczesnego ostrzegania) do tworzenia tych prognoz może mieć ograniczenia w próbkowaniu. Badania, w których stosowano analizę logitową (podejście parametryczne) i rozpoznawanie cech (podejście nieparametryczne) wykazały, że obie z nich wypadły dobrze w wynikach klasyfikacji, ale trafność predykcyjna i rozpoznawanie cech stabilności okazały się bardziej skuteczne. Nieparametryczny system EWS dostarcza istotnych informacji dotyczących zbliżającej się rentowności banków, a rozpoznawanie cech jest potencjalnie przydatne w tej dziedzinie. Sieci neuronowe i analiza dyskryminacyjna to inne techniki, które pozwalają nam przewidywać rentowność banków. Sieci neuronowe wykazują lepszą skuteczność pod względem zdolności predykcyjnych niż te pierwsze, zwłaszcza że bardziej rygorystycznie wazą błędy typu I. Modele predykcyjne z powodzeniem uzupełniają badania przeprowadzane przez banki na miejscu, ponieważ są mniej inwazyjne i mogą być od czasu do czasu aktualizowane. Ale nie zapewniają bezpośredniego sposobu oceny zarządzania i indywidualnych praktyk pożyczkowych i ubezpieczeniowych.

## ANALIZA PREDYKCYJNA CEN MAGAZYNOWYCH Z WYKORZYSTANIEM MODELU DCC GARCH W R

Przeprowadziliśmy modelowanie DCC GARCH trzech spółek giełdowych, a mianowicie Netflix (NFLX), Disney (DIS) i Google (GOOGL). Oprogramowanie statystyczne jest używane do tej analizy w R Studio. Rysunek.1 przedstawia funkcję multifit GARCH wygenerowaną w R.

```
> uspec.n<- multispec(replicate(3,ugarchspec(mean.model = list(armaOrder= c(1,0))))))
> multf= multifit(uspec.n, rx)
> multf

*-----*
*   GARCH Multi-Fit   *
*-----*
No. Assets.:3
GARCH Multi-Spec Type.: Equal
GARCH Model Spec
-----
Model.: sGARCH
Exogenous Regressors in variance equation: none

Mean Equation.:
Include Mean.: 1
AR(FI)MA Model.: (1,d,0)
GARCH-in-Mean.: FALSE
Exogenous Regressors in mean equation: none
Conditional Distribution: norm

GARCH Model Fit
-----
Optimal Parameters:
      rNFLX      rGOOGL      rDIS
mm      0.00210      0.00071      0.00062
ar1      0.04352      0.01183     -0.02745
omega    0.00000      0.00003      0.00002
alpha1   0.00788      0.08780      0.10523
beta1    0.99031      0.77735      0.81020
Log-Lik 5130.91387 6976.94039 7425.93499
```

Rysunek 2 wskazuje kryteria informacyjne, takie jak Akaike, Bayes itp.

```

-----*
*          DCC GARCH Fit          *
-----*

Distribution      : mvnorm
Model            : DCC(1,1)
No. Parameters   : 20
[VAR GARCH DCC UncQ] : [0+15+2+3]
No. Series       : 3
No. Obs.         : 2515
Log-Likelihood   : 19912.28
Av. Log-Likelihood : 7.92

Optimal Parameters
-----*

```

	Estimate	Std. Error	t value
[rNFX].mu	0.002105	0.000626	3.35956
[rNFX].ar1	0.043524	0.025576	1.70172
[rNFX].omega	0.000002	0.000001	2.26327
[rNFX].alpha1	0.007876	0.000837	9.41195
[rNFX].beta1	0.990312	0.000122	8106.17963
[rGOOGL].mu	0.000715	0.000309	2.31082
[rGOOGL].ar1	0.011828	0.022993	0.51444
[rGOOGL].omega	0.000034	0.000027	1.25624
[rGOOGL].alpha1	0.087797	0.071559	1.22692
[rGOOGL].beta1	0.777349	0.153274	5.07163
[rDIS].mu	0.000616	0.000306	2.01423
[rDIS].ar1	-0.027447	0.024512	-1.11973
[rDIS].omega	0.000015	0.000003	4.63766
[rDIS].alpha1	0.105235	0.015309	6.87408
[rDIS].beta1	0.810198	0.032688	24.78547
[Joint]dcca1	0.041001	0.016444	2.49343
[Joint]dccb1	0.860692	0.042913	20.05691

```


```

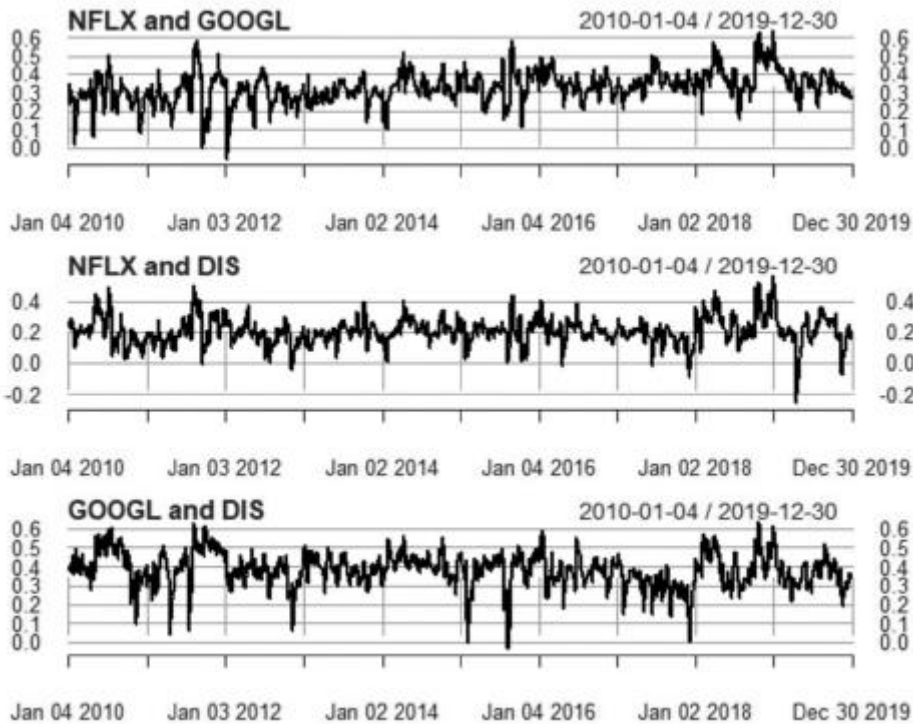
	Pr(> t )
[rNFX].mu	0.000781
[rNFX].ar1	0.088808
[rNFX].omega	0.023619
[rNFX].alpha1	0.000000
[rNFX].beta1	0.000000
[rGOOGL].mu	0.020843
[rGOOGL].ar1	0.606943
[rGOOGL].omega	0.209027
[rGOOGL].alpha1	0.219853
[rGOOGL].beta1	0.000000
[rDIS].mu	0.043985
[rDIS].ar1	0.262828
[rDIS].omega	0.000004
[rDIS].alpha1	0.000000
[rDIS].beta1	0.000000
[Joint]dcca1	0.012652
[Joint]dccb1	0.000000

```

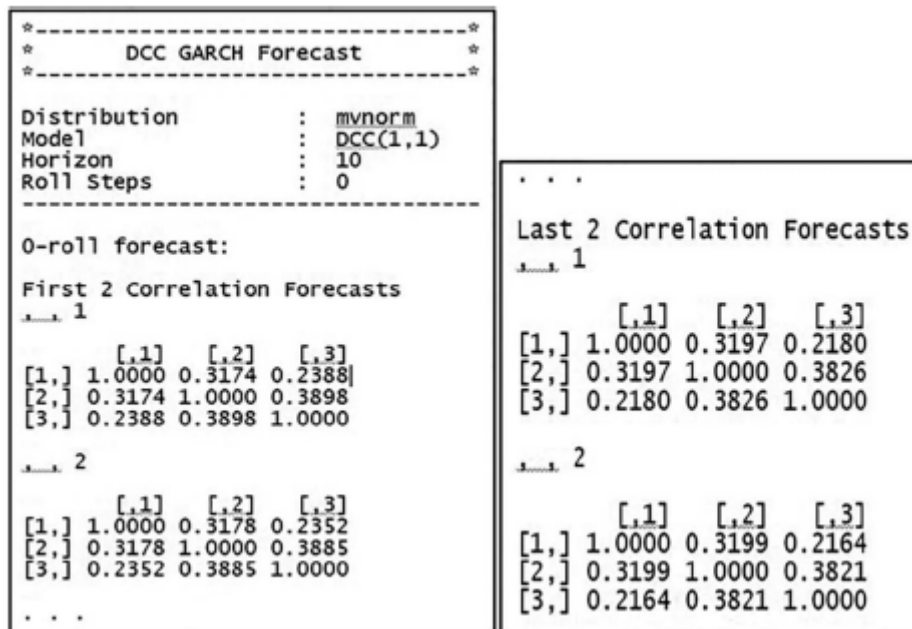
Information Criteria
-----*
Akaike      -15.819
Bayes       -15.773
Shibata     -15.819
Hannan-Quinn -15.802

```

Kryterium Akaike mierzy prawdopodobieństwo estymowanych modeli statystycznych. Mierzy jakość modelu w stosunku do innych modeli. Kryterium informacyjne Bayesa jest również wykorzystywane do oceny zestawu modeli. Niskie wartości tych kryteriów informacyjnych dla tego modelu gwarantują wykonalność tego modelu i wskazują, że jest on przydatny do predykcji]. Rysunek 3 przedstawia modelowanie DCC dla tych trzech spółek dla okresu poprzednich 10 lat, począwszy od 2010 do 2019, które zostało przeprowadzone w celu analizy ich korelacji w tym czasie.



DCC między trzema firmami wykazał pozytywną korelację między NFLX i GOOGL, a także między GOOGL i DIS. Można to przypisać temu, że są to firmy z różnych branż. Z kolei NFLX i DIS, które są konkurentami z tej samej branży, wykazują nieco ujemną korelację w 2018 r. Za pomocą tych danych historycznych wygenerowano model prognozy korelacji, jak pokazano na rysunku 4.



W kontekście analizy predykcyjnej wykonano prognozy korelacji na okres 10 najbliższych lat. Na rysunku tym przedstawiono prognozy korelacji dla dwóch pierwszych i dwóch nadchodzących lat. Te same firmy przedstawiają korelację 1, co wskazuje na doskonałą korelację. Model wykazuje dodatnią korelację (ale mniej niż 0,5) – jest to korelacja bardzo niskiego stopnia. Przewiduje się więc, że w nadchodzących latach wybrane firmy wykażą się nieznacznie dodatnią korelacją. W celu zmniejszenia

ryzyka sugeruje się dywersyfikację portfela inwestycyjnego, wybierając akcje, które mają niską lub ujemną korelację, tak aby w przypadku, gdy jeden przyniesie złe zyski, możesz polegać na innych akcjach. W ten sposób banki mogą korzystać z analizy predykcyjnej, aby uzyskać wgląd w transakcje, aby poprowadzić swoich klientów do kupowania, sprzedawania lub handlu akcjami. Analiza predykcyjna połączona z wykrywaniem anomalii, które wyjaśniono w dalszej części tego rozdziału, może być również wykorzystana do wykrywania przyszłych oszustw lub niewykonania zobowiązań kredytowych.

## **ALGORYTMY GENETYCZNE**

Techniki uczenia maszynowego okazały się bardzo popularne ze względu na szybki rozwój technologii informatycznych. Techniki takie jak sieci neuronowe i GA są coraz częściej stosowane w dziedzinie finansów i inwestycji. Ludzie mają bardzo ograniczone zdolności poznawcze w porównaniu z komputerami. Z pomocą technologii i komputerów można opracować bardziej efektywne strategie handlowe, scenariusze finansowe i portfele. GA są jedną z takich metod efektywnej i szybkiej oceny możliwości finansowych i inwestycyjnych w czasie rzeczywistym. GA to skuteczna technika optymalizacji. Ta technika jest znana jako technika optymalizacji stochastycznej, opracowana przez Johna Hollanda. GA jest iteracją służącą do znalezienia optymalnego rozmiaru i manipuluje populacją o stałym rozmiarze. Populacja składa się z chromosomów, a chromosomy składają się z genów. Te chromosomy konkurują ze sobą o przetrwanie, a w zastosowanym scenariuszu reprezentują potencjalne rozwiązania problemu. Najbardziej optymistyczne z tych rozwiązań zostaną pozostawione jako ocaleni. Każda iteracja daje początek nowej populacji o tej samej liczbie chromosomów, które są lepiej przystosowane do funkcji selekcyjnej. Każda nowa generacja coraz bardziej zbliża się do optymalnego rozwiązania. Selekcja, krzyżowanie i mutacja to operatory stosowane do tworzenia nowej populacji. Operacje selekcji, krzyżowania i mutacji pomagają zbliżyć się lub zredukować populację do jednego optymalnego chromosomu. Były używane w kilku aplikacjach finansowych i są uważane za jedno z najskuteczniejszych i najczęściej używanych podejść w finansach, takich jak generowanie reguł handlowych i formułowanie portfela, optymalizacja strategii crossover itp. W optymalizacji portfela GA jest wykorzystywane do zoptymalizowania wagi każdego papieru wartościowego lub zdecydować o optymalnej proporcji środków, które należy zainwestować w każdy papier wartościowy. Każdy chromosom, który wskazuje tę wagę, jest zoptymalizowany. GA wykonuje pięć kroków: inicjalizacja, ocena, selekcja, crossover i mutacja. GA generuje pierwszy zestaw rozwiązań, który jest początkową populacją chromosomów. Geny w chromosomie wskazują na identyfikację stada. Wartość tych genów wskazuje na wagę każdego stada. Ocena dotyczy procesu definiowania funkcji fitness dla portfela. Aby stworzyć nowy chromosom, chromosomy rodzicielskie są wybierane według proporcji do wartości dopasowania i jest to proces selekcji. Krzyżowanie jest unikalną cechą GA, w której chromosomy rodzicielskie są krzyżowane w celu uzyskania potomstwa. Może być jednolita lub tradycyjna. Krzyżowanie jednolite polega na losowym doborze genu i zachowuje schemat w porównaniu z tradycyjnym. Mutacja zmienia geny w chromosomie, aby zapewnić, że roztwór nie zostanie wychwycony w lokalnym optimum. Proces jest powtarzany wielokrotnie, aby uzyskać optymalne rozwiązanie lub optymalny portfel. Proces iteracyjny jest powtarzany do momentu, gdy nie jest możliwa dalsza optymalizacja lub poprawa. Wykorzystanie GA w optymalizacji portfela inwestycyjnego zostało szczegółowo omówione poniżej.

## **ALGORYTMY GENETYCZNE W OPTYMALIZACJI PORTFELA**

W zależności od tolerancji inwestora na ryzyko można zastosować aktywne lub pasywne strategie portfelowe. Strategia pasywna opiera się na teorii spaceru losowego i obejmuje śledzenie indeksu rynkowego i jest bardziej strategią pasywną lub strategią kup i trzymaj. Strategia aktywna jest wykorzystywana głównie przez risk takers i ma na celu przewyższenie standardowego indeksu

rynkowego. Optymalizacja portfela zasadniczo oznacza maksymalizację zwrotów i minimalizację ryzyka portfela. Poziom ryzyka portfela jest często wybierany na podstawie tolerancji ryzyka inwestora. To jest sedno Modern Portfolio Theory Harry'ego Markowitza. Teoria średniej-wariancji to propozycja, że albo inwestor powinien minimalizować ryzyko dla określonego poziomu zwrotu, albo odwrotnie. Otwiera drogę do granicy efektywności, która jest wykresem przedstawiającym wszystkie możliwe do osiągnięcia portfele, które są osiągalne, przy różnych poziomach ryzyka i zwrotu. Oczekiwany zwrot aktywów to średnia ważona historycznych zwrotów:

$$E[r_p] = \sum_{i=1}^n \omega_i E[r_i]$$

gdzie  $E[r_i]$  odnosi się do oczekiwanego zwrotu z poszczególnych akcji, a  $\omega_i$  jest wagą każdej akcji w portfelu lub proporcją środków zainwestowanych w każdą akcję. Ryzyko składnika aktywów lub papieru wartościowego to niepewność jego wartości w przyszłości. Mierzy się je za pomocą wariancji, odchylenia standardowego, semiwariancji, VaR, CVaR itp. Ryzyko mierzone jest w następujący sposób:

$$\sigma_i^2 = \text{Var}(r_i) = E[(r_i - E[r_i])^2]$$

Kowariancję mierzy się jako:

$$\sigma_{ij} = \text{Covar}(r_{ij}) = E[(r_i - E[r_i])(r_j - E[r_j])]$$

Ponadto średnia korelacja między akcjami w portfelu wynosi:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

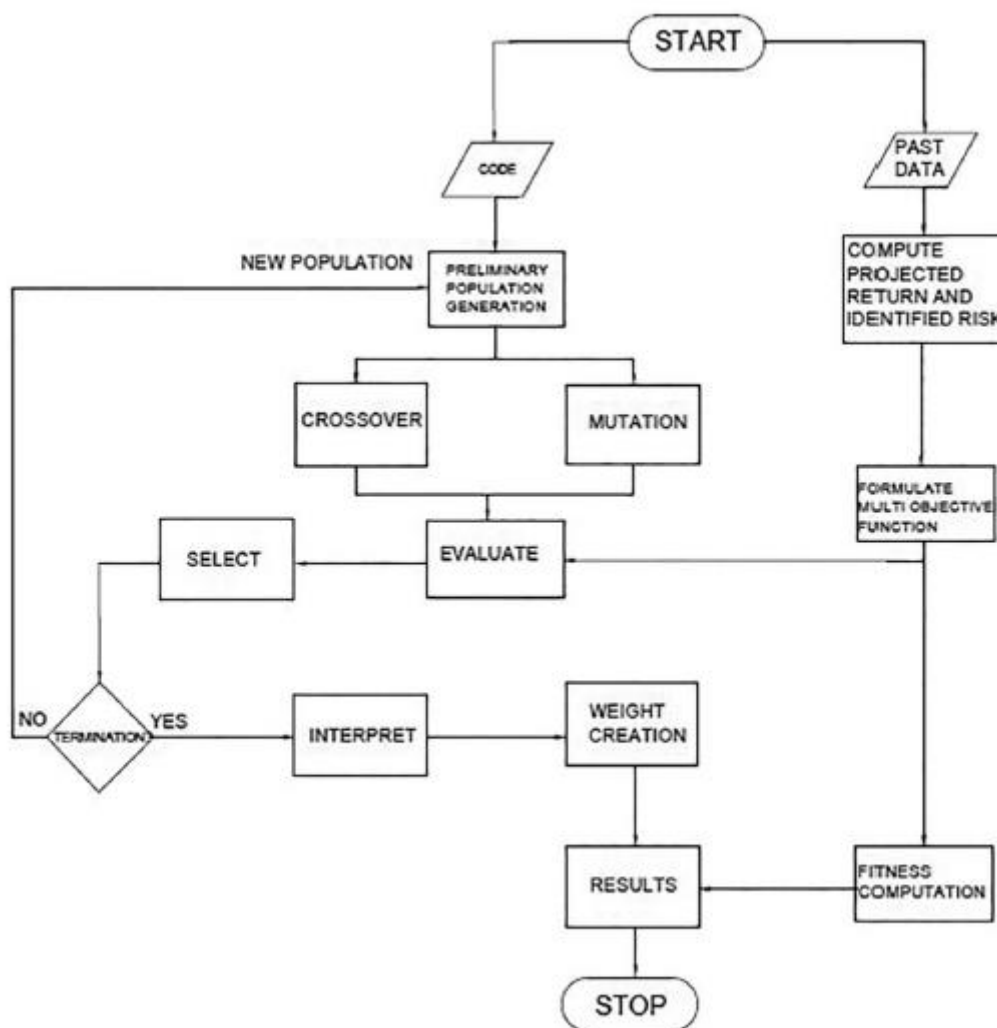
Większość inwestorów wybiera portfele o niskim lub średnim ryzyku, a ludzie na ogół mają awersję do ryzyka. Aby obniżyć ryzyko, portfel powinien mieć ujemną lub niską korelację. Pomaga to zdywersyfikować ryzyko. Wskaźnik Sharpe'a mierzy premię za ryzyko dla inwestycji i jest obliczany jako:

$$S_p = \frac{R_p - R_f}{\sigma_p}$$

gdzie  $R_f$  jest zwrotem standardowego papieru wartościowego,  $R_p$  jest zwrotem portfela, a  $\sigma_p$  jest ryzykiem tego portfela. Wysoki wskaźnik Sharpe'a wskazuje, że portfel daje wyższą stopę zwrotu z określonego ryzyka. Wagi akcji, które pozwalają na maksymalizację współczynnika Sharpe'a, są znane jako wagi optymalne i są obliczane dla każdej akcji za pomocą GA. Rysunek 5 przedstawia schemat blokowy optymalizacji portfela przy użyciu GA. W optymalizacji portfela funkcji celu bierzemy pod uwagę całkowity oczekiwany zwrot i ryzyko, które należy odpowiednio maksymalizować i minimalizować. Funkcja sprawności jest zdefiniowana jako:



$$\text{fitness} = \frac{f(\text{return})}{f(\text{risk})} = \frac{\sum_{i=0}^n r_i \cdot w_i}{\sum_{i=0}^n \sum_{j=1}^n w_i \cdot w_j \cdot \sigma_{ij}}$$



GA kończy się testowaniem konwergencji; to znaczy daje wyniki, które zbiegają się do optymalnego lub lepszego rozwiązania. Zaletą korzystania z GA do optymalizacji wielocelowej jest to, że pozwala nam pracować z kilkoma portfelami i znajdować wiele niezdominowanych rozwiązań w jednym przebiegu. GA są również mniej podatne na ataki w porównaniu z niewypukłą przestrzenią poszukiwań. Ale z drugiej strony są one czasochłonne i może być konieczne zatrzymanie algorytmu przed wprowadzeniem jakichkolwiek ulepszeń na efektywnej granicy, jeśli optymalne rozwiązanie nie jest znane. Nowatorską metodę, w której próbki są dzielone według kwartyli ryzyka, omawiają J. Samuel i in. w 2012 roku. Takie podejście pozwala na generowanie nowej przestrzeni rozwiązań bez wydłużania czasu obliczeń.

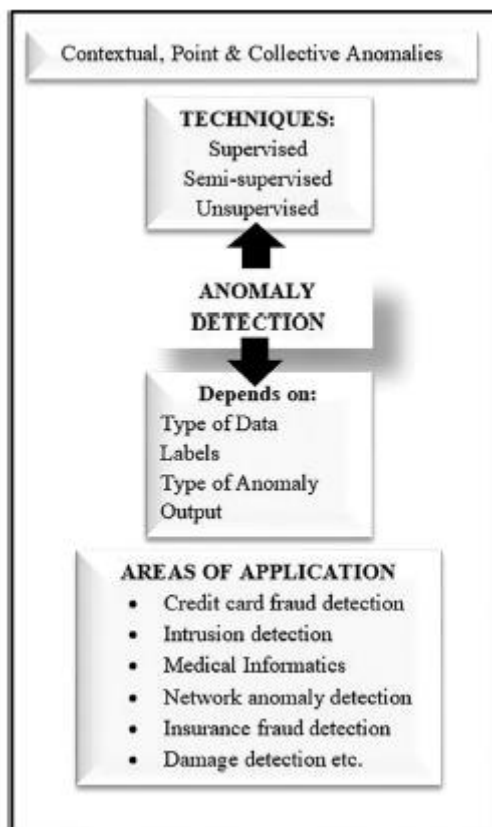
## ALGORYTMY GENETYCZNE W MAKSYMALIZACJI ZYSKU BANKOWEGO

GAMCC (algorytm genetyczny oparty na wielopopulacyjnej konkurencyjnej koewolucji) może być skutecznie wykorzystany do optymalizacji decyzji kredytowych banku w celu maksymalizacji zysku w warunkach kryzysu kredytowego. Tradycyjne metody mogą nie być w tym przypadku skuteczne, ponieważ funkcje celu i ograniczenia w tym problemie są nieliniowe i niewypukłe, co sprawia, że portfel kredytowy jest złożonym problemem do oszacowania. Zmiennymi, od których zależy decyzja kredytowa, są okres kredytowania, limit kredytowy, wielkość i rodzaj kredytu, jego oprocentowanie oraz ocena kredytowa klienta. Funkcja dopasowania, którą należy zmaksymalizować, składa się z całkowitego kosztu transakcji, przychodów z kredytu, kosztu depozytu na żądanie, kosztu kredytu itp. W tym przypadku GA generuje początkową populację, generując zbiór decyzji kredytowych na zasadzie losowej, z których każda jest oceniana za pomocą funkcji fitness. Wartość tej funkcji przystosowania decyduje o tym, czy można ją wyselekcjonować jako chromosom w nowej populacji dla następnego pokolenia. Następne pokolenie zaczyna się od reprodukcji: najlepsza reprezentacja chromosomu uzyskuje więcej kopii, podczas gdy najgorsze reprezentacje są wykluczane w przyszłych pokoleniach. Algorytm zapewnia, że łączna kwota pożyczonej pożyczki nie przekroczy stopy rezerwy depozytu. Jeżeli banki nie będą w stanie efektywnie zarządzać swoimi portfelami kredytowymi, może to doprowadzić do krachu kredytowego, będącego konsekwencją nieostrożnego udzielania kredytów, powodującego straty i złe długi. W takich sytuacjach banki uciekają się do podwyższania stóp procentowych w celu kontroli kredytu. Banki mogą wykorzystać GA do podjęcia optymalnej decyzji, gdy muszą ograniczyć podaż kredytów; w tym czasie doświadczają negatywnego szoku płynnościowego, koncentrując się na maksymalizacji zysku i minimalizacji kosztów kredytowania. Modele GA można również wykorzystać do optymalizacji portfela inwestycyjnego, wykorzystując wzorce behawioralne, takie jak nastroje inwestorów i nadmierna pewność siebie. Model ten pomaga zracjonalizować decyzję inwestora na giełdzie. Wdrażający bankomat ma trudności z utrzymaniem się na rynku, ponieważ popyt w tej dziedzinie jest mniejszy niż podaż bankomatów, a marże zysku są pod presją. Aby sprostać wymaganiom klientów, banki powinny zmaksymalizować użyteczność poprzez efektywne wdrażanie bankomatów. Jest to kolejny problem maksymalizacji, w którym można zastosować GA. Algorytm genetyczny oparty na rangach wykorzystujący konwolucję (RGAC) może być wykorzystany do rozwiązania problemu lokalizacji banków w tym bankomacie, który okazuje się być bardziej wydajny niż algorytm heurystyczny oparty na konwolucji (HAC), ponieważ ten pierwszy pozwala na szybszą zbieżność jednocześnie uzyskując możliwy do zrealizowania wynik. RGAC obejmuje minimalizowanie różnicy między narzędziem klienta (CU) a narzędziem usług (SU) oraz obliczanie komputera (procent pokrycia) dla każdego wdrożenia w celu uzyskania maksymalnej liczby komputerów. W tym konkretnym problemie gen wskazuje lokalizację ATM, która może wynosić 1 lub 0. Równanie sprawności opracowano tak, aby uzyskać najwyższą PC CU. W ten sposób RGAC pozwala bankom uzyskać najwyższą wartość CU, a także poprawia ich efektywność kosztową. Algorytm genetyczny selekcji wieloobiektywnej (MSGA) jest również wykorzystywany do identyfikacji zmiennych, które są najbardziej przydatne dla banków i innych instytucji finansowych w przewidywaniu zaufania użytkowników do systemów bankowości elektronicznej. Niezdominowany genetyczny algorytm sortowania (NSGA-II) jest dość kosztowny pod względem czasu obliczeń, a ponadto generuje zmienne w bardzo niewielu liczbach, co utrudnia zastosowanie w prawdziwym życiu. MSGA oferuje pełniejsze rozwiązania z istotnymi zmiennymi umożliwiającymi podejmowanie lepszych decyzji zarządczych. Dlatego zmienne wybrane przez MSGA są bardziej niezawodne w ulepszaniu bankowości elektronicznej w porównaniu z innymi metodami. GA są używane do automatycznego przypisywania żądań klientów do odpowiednich analityków helpdesku. GA określają, który analityk jest lepiej przystosowany do obsługi konkretnego żądania klienta za pomocą pewnych słów kluczowych powiązanych z analitykiem (które są regularnie aktualizowane). Algorytm ma na celu znalezienie najlepszego analityka rozwiązania, do którego automatycznie przypisywana jest prośba klienta. Instytucje kredytowe również potrzebują lepszych technik oceny ryzyka kredytowego, ponieważ

branża wymaga obecnie szybszych i mniej ryzykownych operacji na precyzyjnych danych wraz ze zmniejszoną marżą zysku. Innym ważnym aspektem w tym zakresie jest selekcja cech, wykonywana przez hybrydowy algorytm genetyczny z sieciami neuronowymi (HGA-NN). Heurystyka HGA-NN może szybko usunąć nieistotne funkcje i jest solidną metodą dodaną do obecnych technik eksploracji danych.

## WYKRYWANIE ANOMALII

Wykrywanie anomalii to proces znajdowania wzorców w zestawie danych, które nie zachowują się w normalny lub oczekiwany sposób. Wzorce te są określane jako wartości odstające lub anomalie, wyjątki, aberracje, osobliwości itp. w zależności od obszaru zastosowania. Proces ten jest szeroko stosowany w wykrywaniu oszustw w transakcjach kartami kredytowymi, opiece zdrowotnej, ubezpieczeniach, wykrywaniu włamań dla cyberbezpieczeństwa i systemów bezpieczeństwa, a także w nadzorze wojskowym. Ta dziedzina nauki o danych jest istotna, ponieważ anomalie lub wartości odstające w danych często prowadzą do krytycznych informacji, które należy natychmiast zbadać i podjąć odpowiednie działania. Wykrywanie anomalii różni się od usuwania szumu i rekomendacji szumu, które obejmują radzenie sobie z szumem, który stanowi przeszkodę w analizie danych. Wykrywanie nowości to kolejny pokrewny obszar pracy, w którym analitycy dążą do wykrycia nowych wzorców w danych, które wcześniej nie zostały wykryte. Różni się to od anomalii, ponieważ w przeciwieństwie do anomalii, nowości są zwykle włączane do modelu po ich wykryciu. Wykrywanie anomalii nie jest łatwym problemem do rozwiązania w jego ogólnej postaci, dlatego bardziej odpowiednie jest rozwiązanie konkretnego sformułowania problemu. Rysunek 6 przedstawia kluczowe koncepcje związane z techniką wykrywania anomalii.



Anomalie są klasyfikowane jako anomalie punktowe, kontekstowe i zbiorowe. Anomalia punktowa jest najprostsza; dzieje się tak, gdy dane indywidualne lub punktowe są odbiegające od pozostałych danych. Na przykład, jeśli dysponujemy zbiorem danych dotyczących transakcji kartą kredytową danej

osoby, zakładamy, że dane są określone tylko przez wydaną kwotę. Tak więc transakcja, która ma dość wysoką kwotę w porównaniu z normalnym zakresem wydatków, będzie anomalią punktową, a konkretna transakcja będzie musiała zostać zbadana, aby sprawdzić, czy nie jest oszustwem. Anomalia warunkowa lub kontekstowa ma miejsce, gdy wystąpienie danych jest anomalne w określonym kontekście; w innym kontekście te same dane mogą nie być nietypowe. Na przykład, w przypadku transakcji kartą kredytową, dana osoba ma zwykły rachunek za zakupy w wysokości 100 USD, ale w okresie Bożego Narodzenia wzrasta on do 1000 USD, co jest uważane za normalne; ale zakup tej samej wysokiej kwoty w czerwcu będzie anomalią kontekstową. Anomalia zbiorcza odnosi się do sytuacji, w której zbiór wystąpień danych jest nieprawidłowy w porównaniu z resztą danych. Ich zbiorowe występowanie jest powodem, dla którego uważa się je za anomalne. Techniki stosowane do wykrywania różnych typów anomalii są różne; mogą być nadzorowane, częściowo nadzorowane lub nienadzorowane.

Istnieją różne podejścia do wykrywania anomalii. Jednym z nich jest wykrywanie anomalnych podstruktur, w ramach którego analizowany jest wykres w celu zidentyfikowania w nim anomalnych danych lub wartości odstających. Ta metoda nie jest zbyt realistyczna w przypadku bardzo małych podbudów i jest skuteczna przy wykrywaniu dużych podbudów. A duże podkonstrukcje są rzadsze niż małe podkonstrukcje. Druga metoda to wykrywanie anomalii podwykresów, w której wykres jest podzielony na unikalne podwykresy i jest oceniany, aby zobaczyć, jak anomalny jest każdy podwykres w porównaniu z innymi. Podgrafy zawierające kilka wspólnych podstruktur są zwykle bardziej anomalne niż podgrafy zawierające kilka wspólnych podstruktur. Pojęcie entropii warunkowej oznacza ilość informacji, która jest niezbędna do opisanie zdarzenia, jeśli miało miejsce inne szczególne zdarzenie. Jest to dobra miara do obliczania regularności lub przewidywalności wykresu. Regularność wykresu znacząco wpływa na możliwość wykrycia na nim anomalii; wraz ze wzrostem regularności maleje średni ranking anomalii. O zdarzeniu lub obiekcie mówi się, że jest anomalny, jeśli znacznie odbiega od tego, co jest definiowane jako normalne lub tego, co jest określone przez model normalności. Jeśli  $S$  jest podejściem nadzorowanym do wykrywania anomalii, to zależy od dwóch czynników  $M$  i  $D$ , które są odpowiednio modelem normalności i stopniem odchylenia. Dlatego każdy system będzie posiadał moduł detekcji i moduł modelowania. Wykrywanie anomalii jest szczególnie przydatne przy wykrywaniu włamań, które odnoszą się do działań mogących naruszyć bezpieczeństwo komputerów i sieci. Anomalie w sieciach można przypisać wydajności lub bezpieczeństwu. Wykrywanie włamań do sieci na podstawie anomalii (ANID) jest głównym obszarem badań, a więcej systemów z ANID możliwości są już dostępne, ale wiele zasadniczych obszarów wymaga większej przejrzystości, zanim takie systemy będą mogły być zastosowane na dużą skalę.

Anomalie sieciowe można przypisać problemom takim jak wada urządzenia, przeciążenie sieci, złośliwe ataki typu „odmowa usługi” itp. Jednym z ważnych wymagań wykrywania anomalii sieciowych jest uzyskanie odpowiedniego rodzaju danych o wydajności sieci. Większa synergia między obszarami przetwarzania sygnału a siecią pozwoli na opracowanie jeszcze bardziej zaawansowanych narzędzi do wykrywania anomalii sieciowych. Narzędzia do przetwarzania sygnałów sprawiają, że istniejące narzędzia do zarządzania siecią są bardziej efektywne. Nienadzorowane podejście do wykrywania anomalii obejmuje duże zbiory danych, w których większość elementów jest normalna, ale w danych istnieją włamania. Zakładają, że normalne zdarzenia są bardzo duże w porównaniu z anomaliami, które również różnią się jakościowo. Są ograniczone w tym sensie, że nie będą w stanie zidentyfikować tych anomalii, które nie różnią się jakościowo od normalnych wydarzeń lub przypadków. Geometryczne ramy do nienadzorowanego wykrywania anomalii zasugerowali Eskin i inni, w którym elementy danych są mapowane do przestrzeni cech, a anomalie są identyfikowane poprzez ocenę, które elementy występują w rzadkich obszarach przestrzeni cech. Przestrzeń funkcji w przypadku nienadzorowanego wykrywania anomalii jest specyficzna dla jej zastosowania; dlatego najlepiej jest wybrać przestrzeń

funkcji po przeanalizowaniu aplikacji pod kątem optymalnych wyników. Do wykrywania punktów leżących w rzadkich regionach można zastosować algorytmy oparte na klastrach, podejście SVM lub podejście k-najbliższego sąsiada. ANID ma duże znaczenie, ponieważ ataki internetowe stają się coraz częstsze i powinny być rozwiązywane za pomocą technik, które charakteryzują się zarówno precyzją, jak i elastycznością. Wykrywanie anomalii można również przeprowadzić za pomocą wejściowych zapytań HTTP, aby opracować system specjalnie stworzony do wykrywania ataków internetowych. Model ten wykorzystuje korelację między programami po stronie serwera i ich parametrami, które są specyficzne dla obszaru aplikacji. Rozszerzenie tego podejścia może umożliwić wykrywanie anomalii w czasie rzeczywistym w przypadku witryn, które codziennie odpowiadają na ogromne ilości zapytań. W przypadku różnych typów ataków sieciowych można zastosować różne techniki wykrywania, a ich sukces w dużej mierze zależy od obszaru zastosowania. Nienadzorowane maszyny SVM są bardzo skuteczne, ponieważ mają bardzo wysoki współczynnik wykrywalności, ale mimo to mają również wysoki współczynnik fałszywych alarmów. Rozwój technik wykrywania anomalii jako taki jest dziedziną poszukiwaną, ponieważ algorytmy eksploracji danych szeregowych okazują się mniej skuteczne i niedopuszczalne w przypadku dużych wolumenów danych heterogenicznych. Tradycyjne metody wykrywania anomalii obejmują podejście oparte na modelu GARCH i model OC-SVM. GARCH, ARIMA, ARMA (autoregresywna średnia ruchoma) i EWMA (wykładniczo ważona średnia ruchoma) to tylko niektóre z metod wczesnego wykrywania anomalii parametrycznych. W tej technice albo zakładamy rozkład anomalii, albo stosujemy model oparty na regresji. W tym drugim przypadku do określenia wyników anomalii wykorzystywana jest wartość rezydualna, która jest częścią, której nie uwzględnia model regresji. Z drugiej strony maszyny SVM są modelami nieparametrycznymi i wykorzystują lokalne modele jądra w przeciwieństwie do jednego globalnego modelu dystrybucji danych. Modele te są szybsze, mniej skomplikowane, a jednocześnie elastyczne. Ten model SVM jest podejściem nadzorowanym i wykorzystuje funkcję jądra, dlatego należy go odpowiednio dostroić, aby uzyskać dokładność klasyfikacji. Głębokie uczenie jest wykorzystywane do analizy zbiorów danych w sektorach bankowym i ubezpieczeniowym w celu wykrywania oszustw. Algorytmy, takie jak AE (autoenkoder), SVM i głębokie sieci konwolucyjne, opierają się na głębokim uczeniu. Chociaż dostępnych jest kilka algorytmów opartych na głębokim uczeniu, deweloper wdrażający technologię powinien być szczególnie świadomy problemu, a także wiedzieć, jak działa każdy algorytm [42]. Modele SVM są powiązane z sieciami neuronowymi. Ich celem jest znalezienie optymalnego hiperpasu do rozdzielenia próbek. Próbkę zgrupowane w pobliżu hiperpasu są wektorami pomocniczymi. Modele OC-SVM różnią się od tych modeli, ponieważ dane szkoleniowe należą tylko do jednej klasy. Nadużywanie kart kredytowych to istotny obszar, w którym wykrywanie anomalii może być bardzo przydatne. Rozwój technologiczny ułatwił obsługę pieniędzy, ale jednocześnie ułatwił dokonywanie oszukańczych transakcji i produkcję podrobionych kart. Można tego dokonać, zdobywając dane karty przez ukryte urządzenie w bankomacie lub przez ramię surfing itp. Anomalie w transakcjach kartą kredytową można zidentyfikować, wybierając transakcje, które wiążą się z bardzo wysokimi płatnościami lub zakupem nietypowych przedmiotów, a także częstotliwość zakupów. Wykrywanie anomalii jest również stosowane w identyfikowaniu oszustw związanych z telefonami komórkowymi, poprzez monitorowanie wzorców użytkownika i tworzenie profilu dla każdego konta klienta, tak aby każde odchylenie od normalnego wzorca użytkownika klienta mogło zostać wykryte jako anomalia, a następnie można było wydać ostrzeżenie klientowi, aby zapobiec wszelkim naruszeniom. Oszustwa ubezpieczeniowe, takie jak oszustwa samochodowe, zwykle mają miejsce w wyniku fałszowania dokumentów. Są one zwykle sprawdzane ręcznie przez inspektorów ubezpieczeniowych, ale wykrywanie tych oszustw w sposób zautomatyzowany będzie szybsze i jeszcze bardziej efektywne. Wczesne jest również kluczowe w przypadku wykorzystywania informacji poufnych, w przypadku których informacje poufne są nielegalnie uzyskiwane, zanim zostaną one upublicznione. Jest to działalność przestępcza, która może sztucznie wpływać na ceny akcji. Głównym wyzwaniem w

wykrywaniu anomalii jest fakt, że nie jest ona uogólniona; oznacza to, że technika stosowana w jednym obszarze lub domenie może nie być odpowiednia dla innej domeny, nawet jeśli typ danych jest podobny. Ponadto dane zawierają szum, który może być postrzegany jako anomalia, co utrudnia ich rozróżnienie. Ponadto to, co jest obecnie uważane za normalne zachowanie, może ulec zmianie w przyszłości, więc obecne techniki wykrywania anomalii mogą nie być prawidłowe lub przydatne w przyszłości. Wraz z rozwojem tych technik oszuści stają się mądrzejsi, skutecznie naśladowując normalne zachowanie, aby nieuczciwe działania stały się normalne. Wykrywanie anomalii jest szczególnie ważne, ponieważ najeźdźcy stają się niezwykle inteligentni w swoich atakach, a sektor finansowy na całym świecie stoi w obliczu ogromnych naruszeń danych przy użyciu złośliwych narzędzi cyberataków. Niezwykle konieczne staje się wykorzystanie Big Data Analytics do zapobiegania takim wyzwaniom i ich łagodzenia. Umożliwi to instytucjom finansowym wywiązanie się z zobowiązania do ochrony wrażliwych danych swoich klientów i interesariuszy.

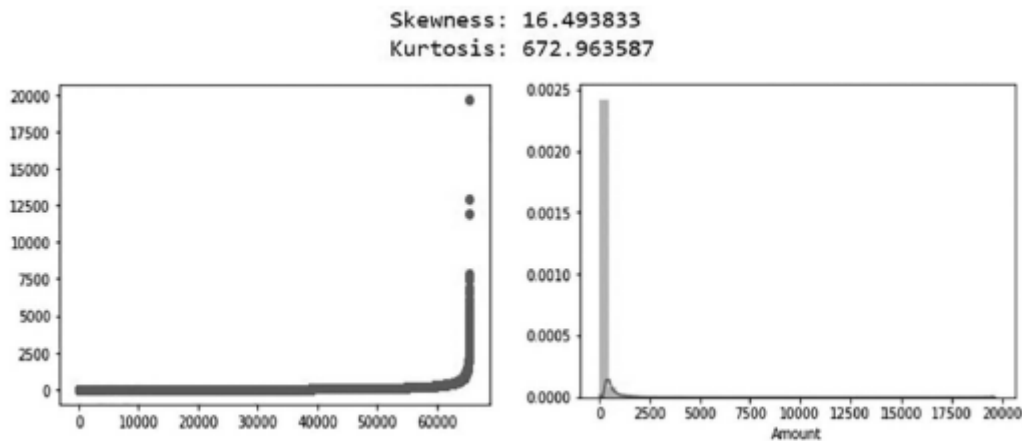
## **WYKRYWANIE ANOMALII W CELU IDENTYFIKACJI OSZUSTW NA KARCIE KREDYTOWEJ ZA POMOCĄ PYTHONA**

### **Biblioteki Pythona**

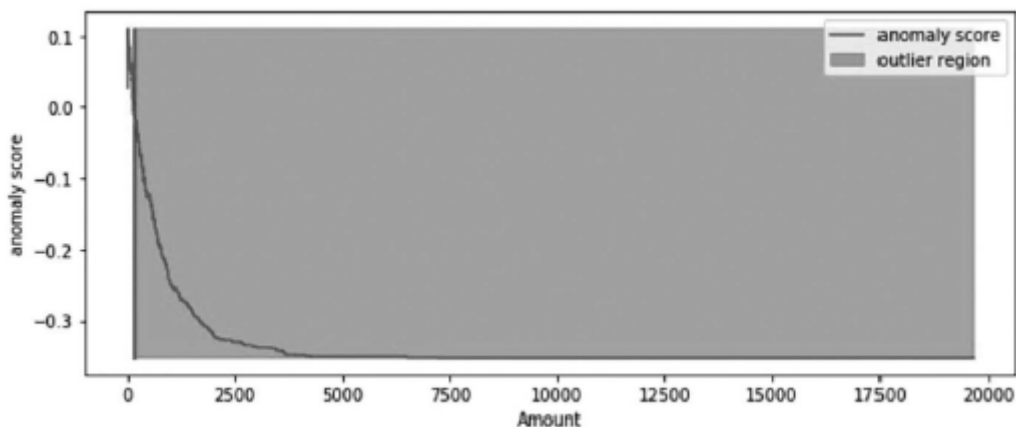
NumPy to moduł Pythona o otwartym kodzie źródłowym, powszechnie używany do obliczeń związanych z tablicami i macierzami. Obliczenia wysokiego poziomu związane z funkcjami matematycznymi są wykonywane przy użyciu numpy. Pandas jest podobny do numpy, jest używany w Pythonie i obsługuje narzędzia do analizy danych. Pandas udostępnia arkusz kalkulacyjny, taki jak wiersze i kolumny; dlatego nazywa się to ramką danych. Wykreślanie wykresów i tworzenie tabel przestawnych było możliwe z pomocą Pand. Matplotlib w pythonie to dwuwymiarowa biblioteka do kreślenia, dlatego jest używana w skryptach Pythona i serwerach aplikacji internetowych. Python używa biblioteki wizualizacji danych znanej jako Seaborn, która pomaga w zapewnieniu interfejsu do rysowania zrozumiałych i użytecznych ilustracji statystycznych. Scikit-learn to biblioteka Pythona, która zapewnia funkcje takie jak regresja, klasyfikacja, grupowanie, wybór modeli i wstępne przetwarzanie, wspierając w ten sposób uczenie nadzorowane i nienadzorowane.

### **Wykrywanie anomalii w zestawie danych karty kredytowej**

Przeprowadziliśmy wykrywanie anomalii w zestawie danych dotyczących transakcji kartami kredytowymi, które uzyskaliśmy od Kaggle.com. Dane zostały poddane jednowymiarowemu wykryciu anomalii w Pythonie, aby zidentyfikować wzorce, które nie odpowiadają normalnemu zachowaniu. Otrzymany przez nas rozkład zmiennej „Kwota” bardzo różni się od rozkładu normalnego, ponieważ większość rozkładu koncentruje się po lewej stronie wykresu. Dane mają bardzo małe prawdopodobieństwo pojawienia się po prawej stronie rozkładu. Rozkład jest mocno wypaczony; obliczono, że jest to wartość dodatnia 16,49, co wskazuje, że prawy ogon rozkładu jest dłuższy niż lewy, jak pokazano na rysunku 7. Dystrybucja jest leptokurtyczna; kurtoza jest obliczana jako 672,963, co oznacza skrajne wartości w rozkładzie



Algorytm, którego użyliśmy tutaj do wykrywania anomalii, to Isolation Forest, model oparty na drzewie. Rysunek 8 pokazuje region, w którym spadają wartości odstające. Zgodnie z wykresem transakcje, które przekraczają 500, są uważane za wartości odstające i należy przeprowadzić dalsze dochodzenie w sprawie takich transakcji, aby ocenić, czy są one oszustwem, czy nie. Za pomocą tego modelu jesteśmy również w stanie wizualnie zbadać każdą anomalię.



## DEMONSTRACJA WYKRYWANIA ANOMALII W CENACH ETHERU Z WYKORZYSTANIEM R

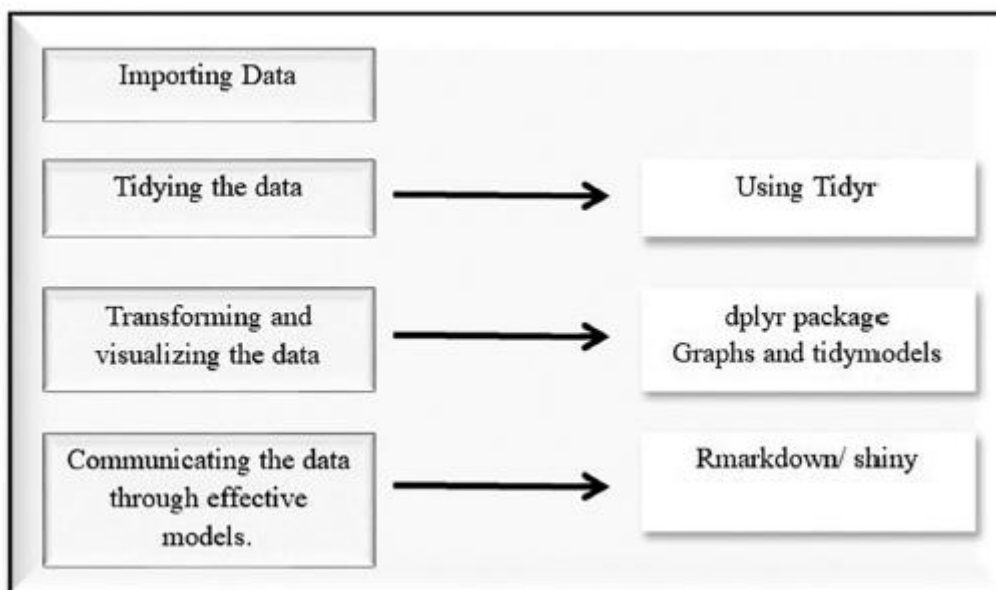
### Ethereum

Ethereum można uznać za maszynę stanu opartą na transakcjach, która zawiera dane, takie jak reputacje, salda kont, dane dotyczące świata fizycznego itp. Transakcje są zestawiane w bloki, które są łączone ze sobą za pomocą kryptografii. Bloki działają jako dziennik lub księga, łącząc te transakcje z poprzednim blokiem. Zdecentralizowany system umożliwia wszystkim zaangażowanym stronom tworzenie nowych bloków. Bitcoin i Ethereum to obecnie najpopularniejsze kryptowaluty. Ethereum ma na celu pokonanie kilku ograniczeń Bitcoina. Obsługuje wszystkie rodzaje obliczeń i zawiera pętle; obsługuje również stan transakcji i ma wiele dodatkowych ulepszeń w porównaniu ze strukturą Bitcoina. Pozwala każdemu tworzyć własne reguły własności za pomocą inteligentnych kontraktów, które można wykonać tylko po spełnieniu określonych warunków. Ethereum i inne kryptowaluty pozostają szybką, bezpieczną warstwą nad istniejącym systemem walutowym i jest to wynalazek, który może zastąpić obecne scentralizowane systemy płatności. Ethereum nadaje się do konstruowania systemów ekonomicznych wyłącznie w oprogramowaniu. Umożliwia ludziom przenoszenie pieniędzy z łatwością i szybkością, z jaką dziś przynosisz dane. Eliminuje zmienne okresy i opłaty związane z dostawcami, takimi jak Mastercard, Visa itp. Ethereum zachowuje kilka podstawowych koncepcji

Bitcoin, ale jest zupełnie nowy, ponieważ jego kluczowe elementy są różne. Może być replikowany do kompatybilnych systemów; jest to darmowy system o otwartym kodzie źródłowym.

### Porządek w porządku

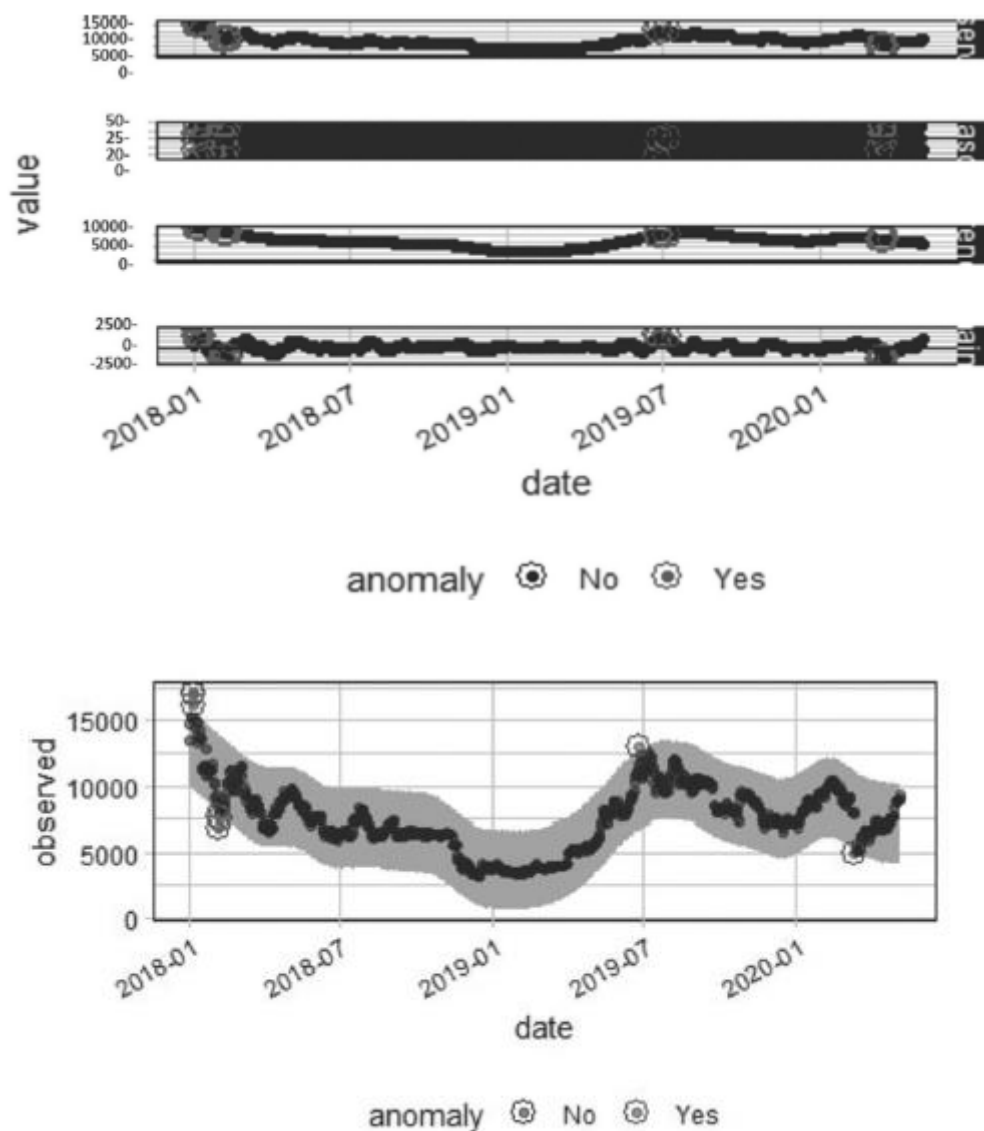
Tidyverse to zbiór pakietów języka R, które charakteryzują się niskopoziomowymi konfiguracjami gramatyki i danych oraz projektami wysokiego poziomu. Jest to łatwiejszy sposób na ułatwienie powtarzających się zadań, takich jak importowanie dużych ilości danych, porządkowanie ich, manipulowanie, wizualizacja i programowanie tych danych. Pozwala nam pobrać i zainstalować wszystkie pakiety tidyverse za pomocą jednego polecenia; ten pojedynczy metapakiet, który to umożliwia, nazywa się tidyverse. Tidyverse pracuje z uporządkowanymi danymi, w których każdy wiersz jest obserwacją, a każda kolumna jest zmienną. Każda komórka ma jedną wartość. Tidyverse umożliwia konwersję danych do tego uporządkowanego formatu, co ułatwia pracę. Przekształcenie tych danych odbywa się za pomocą pakietu dplyr. Wizualizacja tych danych odbywa się za pomocą wykresów, a modelowanie odbywa się za pomocą tidymod-els. Komunikacja tych danych z innymi nie jest zawarta w pakiecie tidyverse i odbywa się za pomocą innych pakietów w R, takich jak rmarkdown i shiny. Programowanie obejmuje wszystkie te wyżej wymienione aspekty i obejmuje każdy aspekt nauki o danych. Purr, tibble itp. to narzędzia programistyczne w pakiecie tidyverse. Elementy pakietu Tidyverse pokazano na rysunku 9.



### Wykrywanie anomalii

Zebraliśmy historyczne dane cen Ethereum z 2018 roku z Coindesk.com i przeprowadziliśmy wykrywanie anomalii za pomocą tidyverse w R. Najpierw rozkładamy dane szeregów czasowych cen Ethereum na składniki sezonowe, trendy i pozostałe. Następnie używamy biblioteki „Anomalizacja” w R do wykrywania i oznaczania anomalii. Białe punkty na wykresie wskazują anomalie, które należy dokładniej zbadać, podczas gdy czarne punkty na wykresie wskazują na normalne zachowanie. Możemy również użyć R do wyodrębnienia rzeczywistych punktów, które są anomaliami. Rozkład szeregów czasowych i wykryte anomalie pokazano na rysunkach 10 i 11.





## WNIOSEK

Sztuczna inteligencja to obecnie największa rewolucja technologiczna, która wpływa i oczekuje się, że nadal będzie wpływać na sposób, w jaki firmy realizują swoje procesy. Sztuczna inteligencja umożliwia wykorzystanie ogromnych ilości danych, które są generowane z dnia na dzień, aby przedsiębiorstwa mogły działać w bardziej wydajny i efektywny sposób. Sztuczna inteligencja ma istotne zastosowania w dziedzinie ubezpieczeń i bankowości, a niektóre z nich omówione w tym rozdziale obejmują analizę predykcyjną, GA i wykrywanie anomalii. Istnieje również wiele innych przydatnych i solidnych aplikacji. Techniki te pozwalają bankom i innym instytucjom finansowym zoptymalizować swoje procedury i skuteczniej reagować na wymagania klientów. Analiza predykcyjna umożliwia organizacjom uzyskanie informacji o trendach i wzorcach w danych historycznych w celu zrozumienia przewidywanej przyszłości. Wykorzystywana jest przez organizacje do identyfikacji trendów w zachowaniach klientów i wzorców zakupowych, prognozowania sprzedaży, zarządzania jakością itp. W obszarze ubezpieczeń i bankowości analiza predykcyjna jest szczególnie wykorzystywana do przewidywania odpływu klientów i analityki, oceny ryzyka kredytowego, w celu uzyskania wglądu handlowego itp. Zastosowanie analizy predykcyjnej we wglądzie handlowym jest omówione w tym rozdziale; analiza predykcyjna pozwala bankom na wskazywanie swoim klientom, czy sprzedać, zatrzymać lub kupić akcje. Demonstracja modelu prognostycznego DDC GARCH na okres najbliższych 10 lat została dokonana pomiędzy NFLX,

DIS i GOOGL przy użyciu oprogramowania statystycznego R. Wyniki wskazują na nieco pozytywną korelację w nadchodzących latach dla wybranych spółek. Korelacja jest istotną miarą, która umożliwia inwestorom dywersyfikację portfela inwestycyjnego. Banki wykorzystują również analitykę predykcyjną do przewidywania ryzyka kredytowego i niespłacania kredytów za pomocą SVM, modeli regresji logistycznej itp., które są łatwiejsze dzięki sztucznej inteligencji i pakietom oprogramowania obliczeniowego. GA pomagają generować wysokiej jakości rozwiązania dla złożonych problemów optymalizacyjnych. Powtarzalny proces pozwala bankom wybrać najbardziej optymalny sposób przeprowadzenia procesu lub operacji. GA pozwalają bankom maksymalizować zysk poprzez optymalizację różnych obszarów, takich jak wdrażanie bankomatów, optymalizacja portfela kredytowego, optymalizacja portfela inwestycyjnego itp. Umożliwiają również lepszą obsługę klienta i optymalizację sposobu, w jaki analitycy odpowiadają na zapytania klientów.

Wykrywanie anomalii, omówione pod koniec, jest ważnym aspektem dotyczącym transakcji kartami kredytowymi i bankowości mobilnej; umożliwia bankom wykrycie potencjalnych nadużyć. Polega na identyfikowaniu wartości odstających lub anomalii w dużych zestawach danych, które mogą być potencjalnymi oszustwami. Identyfikowanie i zapobieganie oszustwom jest ważne dla każdej organizacji, a dla banków i innych instytucji finansowych jest to kluczowe, ponieważ mają do czynienia z dużymi kwotami pieniędzy, a klienci muszą ufać instytucji, w której deponują swoje pieniądze. Potencjalne oszustwa lub naruszenia mogą zniszczyć zaufanie klienta i niekorzystnie wpłynąć na długoterminową reputację instytucji finansowej. Pokazaliśmy, w jaki sposób wykrywanie anomalii może być wykorzystywane przez banki do wykrywania oszustw w transakcjach kartami kredytowymi za pomocą Pythona. Przeprowadzana jest kolejna demonstracja, aby pokazać anomalie w cenach Ethereum z 2018 roku przy użyciu pakietu Tidyverse w R. Istnieje kilka innych zastosowań sztucznej inteligencji, które nie są omówione w tym rozdziale, takich jak przetwarzanie języka naturalnego, generowanie języka naturalnego itp.