Analiza predykcyjna danych giełdowych przy użyciu Microsoft Azure

Spis treści

[1. Wprowadzenie 2](#_Toc138269260)

[2. Baza danych 2](#_Toc138269261)

[2.1. Opisanie istoty problemu 4](#_Toc138269262)

[2.2. Schemat bazy danych w Microsoft Azure 4](#_Toc138269263)

[3. Algorytm prognostyczny 4](#_Toc138269264)

[3.1. Wbudowane możliwości analizy predykcyjnej w Microsoft Azure 4](#_Toc138269265)

[3.1.1. Model ARIMA 5](#_Toc138269266)

[3.1.2. Model GARCH 5](#_Toc138269267)

[3.1.3. Rekurencyjne sieci neuronowe 6](#_Toc138269268)

[3.2. Algorytm predykcyjny w języku Python 7](#_Toc138269269)

[3.3. Porównanie wyników poszczególnych metod prognostycznych 7](#_Toc138269270)

[4. Interfejs 7](#_Toc138269271)

[4.1. Jupiter Notebook (?) 7](#_Toc138269272)

[4.2. Przykład użycia 7](#_Toc138269273)

[5. Literatura 7](#_Toc138269274)

# 1. Wprowadzenie

# 2. Baza danych

Resource groups



Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Opis wygenerowany automatycznie

Upload Danych

Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Ikona komputerowa

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, numer, oprogramowanie

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, oprogramowanie

Opis wygenerowany automatycznie

## 2.1. Opisanie istoty problemu

## 2.2. Schemat bazy danych w Microsoft Azure

# 3. Algorytm prognostyczny

## 3.1. Wbudowane możliwości analizy predykcyjnej w Microsoft Azure

Microsoft Azure oferuje szereg wbudowanych możliwości analizy predykcyjnej, które umożliwiają budowanie, wdrażanie i zarządzanie modelami uczenia maszynowego oraz przeprowadzanie zaawansowanej analizy danych.

Azure Machine Learning to pełna usługa do tworzenia, trenowania i wdrażania modeli uczenia maszynowego. Zapewnia środowisko do eksploracji danych, automatycznego uczenia maszynowego, zarządzania eksperymentami, tworzenia i wdrożenia modeli, a także monitorowania i skalowania wdrożonych usług modeli.

Automatyczne uczenie maszynowe, Usługa oferuje funkcjonalność automatycznego uczenia maszynowego (AutoML), która umożliwia automatyczne wyszukiwanie i dobieranie optymalnych modeli oraz hiperparametrów na podstawie danych wejściowych. Automatyczne uczenie maszynowe upraszcza proces budowy modeli, szczególnie dla osób bez głębokiej wiedzy w dziedzinie uczenia maszynowego.

Azure Machine Learning zapewnia narzędzia do zarządzania eksperymentami, co umożliwia rejestrowanie różnych wersji modeli, śledzenie metryk i porównywanie wyników eksperymentów. Można przeglądać, analizować i udostępniać eksperymenty zespołom i współpracownikom.

Wdrażanie modeli - usługa umożliwia wdrażanie modeli jako usług w chmurze, co pozwala na udostępnianie modeli za pomocą interfejsów API. Można łatwo wdrażać modele jako usługi sieciowe, które można wywoływać z aplikacji lub innymi usługami.

Skalowanie i zarządzanie zasobami, w Azure Machine Learning zapewnia skalowalne i elastyczne środowisko obliczeniowe, które umożliwia zarządzanie zasobami potrzebnymi do treningu i wdrożenia modeli. Można dostosować zasoby obliczeniowe do wymagań i rozszerzać je w miarę potrzeb.

Integracja z innymi usługami Azure, usługa jest częścią ekosystemu Azure, co oznacza, że ​​można ją integrować z innymi usługami Azure, takimi jak Azure Data Lake Storage, Azure SQL Database, Azure Cognitive Services, Azure Synapse Analytics i wiele innych. To zapewnia kompleksowe rozwiązania dla analizy danych i uczenia maszynowego.

By modelować zmienność finansową, istnieje kilka popularnych podejść i technik. Oto kilka modeli, które często są stosowane w celu modelowania zmienności finansowej:

- Model ARIMA

- Model GARCH

- Rekurencyjne sieci neuronowe

### 3.1.1. Model ARIMA

Model ARIMA (Autoregressive Integrated Moving Average) jest jednym z popularnych modeli używanych do analizy szeregów czasowych, w tym do prognozowania zmienności finansowej.

Autoregresja (AR) to składowa autoregresyjna opiera się na założeniu, że wartość zmiennej zależnej w danym czasie zależy od poprzednich wartości tej samej zmiennej. Model AR wykorzystuje poprzednie wartości do prognozowania przyszłych wartości.

Różnicowanie (I) to składowa różnicująca odnosi się do różnicy między kolejnymi obserwacjami w celu uzyskania szeregów czasowych o stacjonarnych właściwościach.

Średnia ruchoma (MA) to składowa średniej ruchomej opiera się na założeniu, że wartość zmiennej zależnej w danym czasie zależy od losowych szoków lub błędów.

Aby zastosować model ARIMA, najpierw należy dostosować parametry modelu do danych poprzez analizę szeregów czasowych, taką jak identyfikacja trendów, sezonowości czy stacjonarności. Następnie można oszacować parametry modelu i przeprowadzić prognozy na podstawie znanych wartości.

Model ARIMA (Autoregressive Integrated Moving Average) jest popularnym modelem do analizy i prognozowania szeregów czasowych. Chociaż nie jest dostępny jako wbudowany model w Azure Machine Learning, można go zaimplementować, korzystając z bibliotek

### 3.1.2. Model GARCH

Model GARCH (Generalized Autoregressive Conditional Heteroskedasticity) jest narzędziem stosowanym w analizie finansowej do modelowania zmienności cenowych, szczególnie na rynkach finansowych, gdzie występują niestabilne warunki i wahania cen.

GARCH został wprowadzony przez Roberta Engle'a w 1982 roku jako rozwinięcie modelu ARCH (Autoregressive Conditional Heteroskedasticity), który zakłada, że wariancja błędów jest heteroskedastyczna i zależy od poprzednich błędów. Model GARCH dodaje do tego autoregresję dla wariancji, co umożliwia uwzględnienie wpływu wcześniejszych wartości wariancji na jej aktualną wartość.

Podstawowe założenie modelu GARCH jest to, że zmienność jest dynamicznym procesem, który ma tendencję do utrzymywania się w czasie. Model GARCH uwzględnia to, przewidując przyszłą zmienność na podstawie wcześniejszych wartości zmienności.

Model GARCH może być zapisany jako GARCH(p, q), gdzie p oznacza liczbę opóźnień (poprzednich wartości błędów ARCH), a q oznacza liczbę opóźnień (poprzednich wartości zmienności GARCH).

Model GARCH jest użyteczny w analizie finansowej, ponieważ pozwala uwzględnić niestabilność i zmienność w danych finansowych. Jest stosowany do prognozowania ryzyka finansowego, zarządzania portfelem, modelowania wahań cen aktywów, oceny ryzyka inwestycji i tworzenia scenariuszy symulacyjnych.

Ważne jest, aby zrozumieć, że model GARCH ma pewne założenia, takie jak normalność rozkładu błędów i stacjonarność szeregów czasowych. W praktyce może być konieczne dostosowanie modelu do uwzględnienia specyficznych cech danych finansowych i zastosowanie technik walidacji modelu w celu oceny jego jakości i trafności.

Azure Machine Learning nie posiada wbudowanego modelu GARCH (Generalized Autoregressive Conditional Heteroskedasticity). Jednakże, można zaimplementować model GARCH w Azure Machine Learning, korzystając z bibliotek.

### 3.1.3. Rekurencyjne sieci neuronowe

Rekurencyjne sieci neuronowe (RNN, Recursive Neural Networks) są rodzajem modelu uczenia maszynowego, które są stosowane w analizie sekwencyjnych danych, takich jak szeregi czasowe, teksty, mowy itp. RNN są szczególnie przydatne w przypadkach, gdy kolejność i kontekst informacji są istotne dla analizy i prognozowania.

Podstawową różnicą między RNN a innymi modelami, takimi jak standardowe sieci neuronowe (feedforward networks), jest to, że RNN mają pamięć wewnętrzną, która pozwala na przechowywanie informacji o poprzednich stanach i wykorzystywanie ich przy analizie kolejnych wejść.

Kluczowym elementem RNN jest komórka rekurencyjna, która jest powtarzana dla każdego kroku czasowego sekwencji danych. W każdym kroku czasowym komórka rekurencyjna pobiera aktualne wejście oraz informację o poprzednim stanie, a następnie generuje aktualny stan wyjściowy oraz przekazuje go do kolejnego kroku czasowego. W ten sposób RNN są w stanie uwzględniać kontekst i zależności czasowe w danych sekwencyjnych.

Najpopularniejszym typem komórki rekurencyjnej w RNN jest komórka LSTM (Long Short-Term Memory). Komórka LSTM ma dodatkowe mechanizmy, które pozwalają na efektywne przechowywanie i usuwanie informacji z pamięci wewnętrznej, co jest szczególnie przydatne w przypadku długotrwałych zależności w danych sekwencyjnych.

RNN są powszechnie stosowane w różnych dziedzinach, takich jak analiza szeregów czasowych, rozpoznawanie mowy, tłumaczenie maszynowe, generowanie tekstu i wiele innych.

W Microsoft Azure Machine Learning istnieje wiele możliwości korzystania z rekurencyjnych sieci neuronowych (RNN) do analizy danych sekwencyjnych.

Azure Machine Learning SDK: Azure Machine Learning SDK to zestaw narzędzi i bibliotek, które umożliwiają tworzenie i zarządzanie eksperymentami uczenia maszynowego w Azure. Możesz użyć SDK w połączeniu z bibliotekami Python do tworzenia i trenowania własnych modeli RNN,

## 3.2. Algorytm predykcyjny w języku Python

## 3.3. Porównanie wyników poszczególnych metod prognostycznych

# 4. Interfejs

## 4.1. Jupiter Notebook (?)

## 4.2. Przykład użycia

# 5. Literatura

1. „Wprowadzenie do algorytmów”

Druk: Warszawa, 2022

Wydanie/Copyright: 2012

Autor: Cormen Thomas H., Leiserson Charles E., Rivest Ronald L, Clifford Stein

Tłumacz: Krzysztof Marian Diks, Adam Malinowski, Daria Roszkowska

Wydawca: Wydawnictwo Naukowe PWN

1. <http://www.home.umk.pl/~piter/Modele_klasy_Garch.pdf>
2. <http://www2.im.uj.edu.pl/katedry/seminaria/inda/wp-content/uploads/szeregi-wstep.pdf>
3. <https://learn.microsoft.com/pl-pl/azure/architecture/data-guide/big-data/non-relational-data>
4. <https://learn.microsoft.com/en-us/training/modules/introduction-to-machine-learning/>