

AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA
W KRAKOWIE

WYDZIAŁ FIZYKI I INFORMATYKI STOSOWANEJ
METODY INTELIGENCJI OBLICZENIOWEJ

Budowa modelu predykcyjnego dla wskaźnika inflacji

Authors:

Damian KOPERSTYŃSKI

Łukasz WAJDA

Tomasz MADEJ

15 czerwca 2022



Spis treści

1	Opis zagadnienia	2
2	Opis Algorytmu	3
2.1	Analiza danych wejściowych	3
2.2	Model sieci oraz proces uczenia	5
2.2.1	Model stworzony na podstawie biblioteki tensorflow(Sequential())	5
2.2.2	Model stworzony na podstawie biblioteki scikit-learn(LinearRegression())	5
2.2.3	Model stworzony na podstawie biblioteki scikit-learn(SVR())	5
3	Wyniki	6
3.1	Wynik na podstawie modelu stworzonego z pomocą tensor- flow(Sequential())	6
3.2	Wynik na podstawie modelu stworzonego z pomocą scikit- learn(LinearRegression())	7
3.3	Wynik na podstawie modelu stworzonego z pomocą scikit- learn(SVR())	8
4	Wnioski	9

1 Opis zagadnienia

Celem zadania było zbudowanie modelu predykcyjnego służącego do określenia stopy inflacji.

Ceny towarów i usług w gospodarce rynkowej podlegają nieustannym zmianom - jedne rosną, a inne spadają. Jednakże z inflacją mamy do czynienia dopiero wówczas, gdy następuje powszechny wzrost cen, a nie podwyżki jedynie niektórych dóbr konsumpcyjnych. W sytuacji, gdy dochodzi do inflacji, spada wartość pieniądza, czyli za tę samą kwotę możemy kupić coraz mniej.

Istnieje kilka interesujących miar inflacji. Jedną z nich, a zarazem najczęściej używaną przez ekonomistów, jest tzw. *inflacja bazowa*, czyli wskaźnik pomijający ceny żywności, energii, paliw, koszty administracyjne itd. Jeśli miara ta jest wysoka, oznacza to, że ceny rosną z przyczyn wewnętrznych takich jak np. niskie stopy procentowe, czy zbyt duża ilość pieniędzy *bez pokrycia* w gospodarce.

Kolejną ciekawą miarą inflacji jest *CPI* (ang. *Consumer price index*) - wskaźnik cen towarów i usług konsumpcyjnych. Ta z kolei jest najpopularniejszym wskaźnikiem podawanym do informacji publicznej, ponieważ obliczana jest jako średnia ważona cen towarów i nabywanych przez przeciętne gospodarstwo domowe usług.

Analogicznie do miar, istnieje kilka rodzajów inflacji. W przypadku gdy mamy do czynienia z inflacją ujemną, mówimy o deflacji. Natomiast pozostałe rodzaje zależą od tempa jej rocznego wzrostu i tak mamy odpowiednio: inflację pełzającą, gdy nie przekracza ona 5% r/r, kroczącą 5-10% r/r, galopującą 10-150% r/r (obecnie w Polsce), a powyżej tej wartości hiperinflację.

W ramach projektu wykonane zostały trzy modele - model głębokiego uczenia, model wykorzystujący regresję liniową oraz SVM.

2 Opis Algorytmu

Projekt został zrealizowany przy użyciu języka programistycznego Python. Podczas realizowania zadania korzystaliśmy z kilku bibliotek są to między innymi:

- *pandas* - została wykorzystana do pobrania danych z pliku csv, oraz do obróbki tych danych,
- *matplotlib* - do wizualizacji wprowadzanych i wynikowych danych,
- *tenserflow* - jest odpowiedzialna za utworzenie modelu po przez przeprowadzenie algorytmu uczenia maszynowego,
- *scikit-learn* - wykorzystana w celu takim jak tenserflow, jest to proste i wydajne narzędzie do predykcyjnej analizy danych,
- *pickle* - moduł pickle implementuje protokoły binarne do serializacji i deserializacji struktury obiektu Pythona, została wykorzystana do zapisu przeszkolonych modeli.

2.1 Analiza danych wejściowych

Do wyznaczenia wskaźnika inflacji oraz analizy danych zostały użyte wskaźnik cen towarów i usług konsumpcyjnych. Do uczenia sieci zostały użyte wartości wskaźników z poprzednich miesięcy (nie z miesiąca, dla którego wyznaczamy wartość wskaźnika), a dokładniej z przedziału od 1 stycznia 1999, ponieważ wtedy nastąpiła zmiana klasyfikacji obliczania wskaźników na obliczenia według klasyfikacji *COICOP* (klasyfikacja spożycia indywidualnego według celu), co zmieniło zachowanie danych w stosunku do lat poprzednich.

Sposób prezentacji jaki wybrano to porównanie z poprzednim miesiącem, dla którego wynik wskaźnika był równy 100pkt. Pozwoliło to na zachowanie lepszej spójności danych oraz braku skoków, które następowały w innych sposobach prezentacji. Dane pobrano z pliku *data/goodsprices.csv* by następnie je wyselekcjonować

```
new_df = df[df['Sposob prezentacji'] == 'Poprzedni miesiac = 100']
```

i utworzyć nowy kontener na dane do przechowywania informacji o wskaźniku

```
new_df = new_df[['Rok', 'Miesiac', 'Wartosc']]
```

Następnie wartości zostały posortowane w sposób malejący ze względu na rok i miesiąc. Miało to zapewnić modelowi pewne strukturalne ułożenie danych i możliwość łatwiejszego ich porównywania miesiąc po miesiącu. Konieczna była również zamiana *floating point* na kropki zamiast przecinków, co było wymagane, aby model rozpoznawał wartości typu *float*. Dodatkowo indeks ustawiony został jako wartości *datetime*, aby model mógł posługiwać się indeksem przy porównywaniu wartości. Po posortowaniu dane prezentują się w następujący sposób (Rysunek 1):

	Rok	Miesiąc	Wartość
Data			
2022-04-01	2022	4	102.0
2022-03-01	2022	3	103.3
2022-02-01	2022	2	99.7
2022-01-01	2022	1	101.9
2021-12-01	2021	12	100.9
...
1982-05-01	1982	5	102.7
1982-04-01	1982	4	103.1
1982-03-01	1982	3	105.8
1982-02-01	1982	2	136.3
1982-01-01	1982	1	106.6
484 rows × 3 columns			

Rysunek 1: Zebrane dane po wykonaniu operacji.

2.2 Model sieci oraz proces uczenia

Analizy predykcyjnej użyliśmy na dwa sposoby korzystając z biblioteki tensorflow i scikit-learn.

2.2.1 Model stworzony na podstawie biblioteki tensorflow(Sequential())

Sequential() jest to model sekwencyjny, pozwala na tworzenie modeli warstwa po warstwie, krok po kroku. Model głębokiego uczenia (*deep learning*) wykorzystuje warstwy Dense z funkcją aktywacji 'relu', natomiast wynik interpretowany jest przez funkcję liniową. Model został dopasowany w 50 epokach (nasyca się już po około 20), przy rozmiarze partii (batch'a) 32.

2.2.2 Model stworzony na podstawie biblioteki scikit-learn(LinearRegression())

Model został oparty o obiekt tworzony przez LinearRegression(), Regresja liniowa jest prostym i powszechnym rodzajem analizy predykcyjnej. Regresja liniowa próbuje modelować związek między dwiema (lub więcej) zmiennymi przez dopasowanie linii prostej do danych. Mówiąc prościej, regresja liniowa próbuje przewidzieć wartość jednej zmiennej na podstawie wartości innej (lub wielu innych zmiennych). Model został stworzony za pomocą wartości uczących, następnie korzystając z metody predict(), dla wartości dat testujących znaleźliśmy odpowiadające im wartości wskaźnika inflacyjnego.

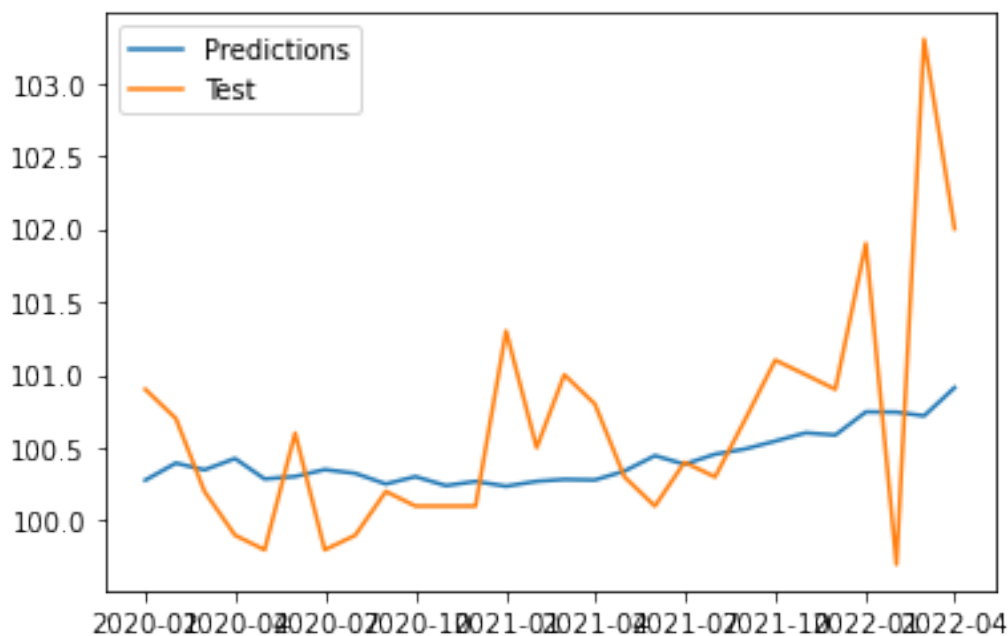
2.2.3 Model stworzony na podstawie biblioteki scikit-learn(SVR())

Model został oparty o obiekt tworzony przez SVR(), "Support Vector Regression" (SVR) to funkcja regresji, która jest uogólniana przez "Support Vector Machines" — model uczenia maszynowego używany do klasyfikacji danych na danych ciągłych. Dalszy proces postępowania dla tej metody jest taki sam jak w punkcie powyższym.

3 Wyniki

Korzystając z stworzonych modeli stworzyliśmy wykres przewidywanych wskaźników inflacyjnych w latach 01.2020 - 04.2022, dzięki czemu możliwe było porównanie ich z rzeczywistymi wartościami.

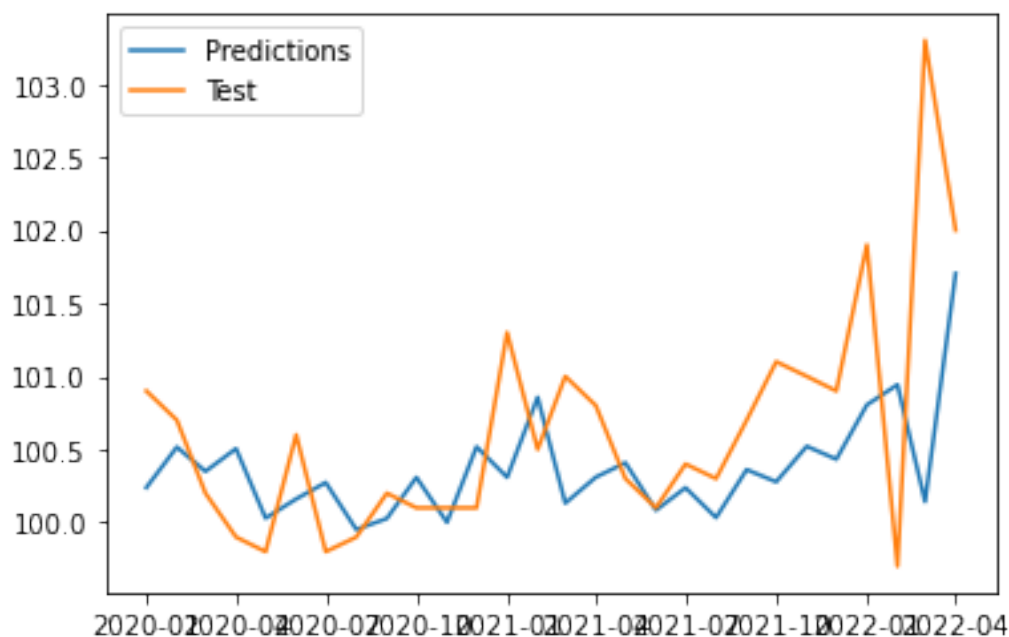
3.1 Wynik na podstawie modelu stworzonego z pomocą tensorflow(Sequential())



Rysunek 2: Wykres pokazujący różnicę między wskaźnikiem inflacyjnym rzeczywistym a tym stworzonym na podstawie modelu stworzonego za pomocą tensorflow.

Na podstawie Rysunek 2. możemy zobaczyć, że wartości z naszego modelu przewidują wzrost wskaźnika inflacyjnego, jednakże jest on bardzo delikatny, bez gwałtownych zmian.

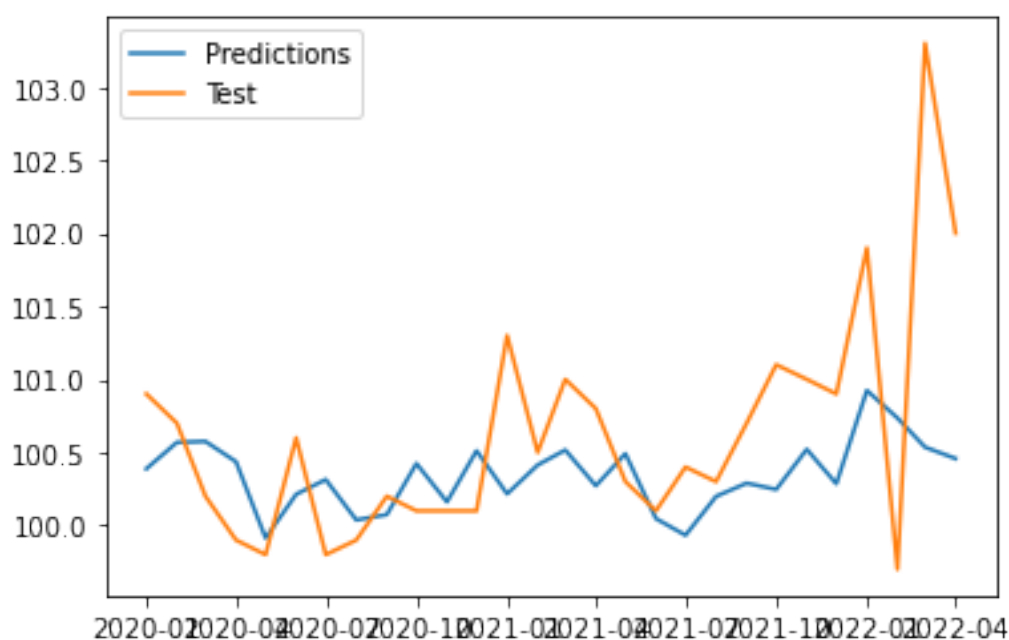
3.2 Wynik na podstawie modelu stworzonego z pomocą `scikit-learn(LinearRegression())`



Rysunek 3: Wykres pokazujący różnicę między wskaźnikiem inflacyjnym rzeczywistym a tym stworzonym na podstawie modelu stworzonego za pomocą `scikit-learn(LinearRegression())`.

Na podstawie Rysunek 3. możemy zobaczyć, że wartości z naszego modelu przewidują wzrost wskaźnika inflacyjnego, jest on bardzo zbliżony do rzeczywistego, uwzględnia gwałtowne zmiany, jednakże w naszym modelu wzrosty krytyczne pojawiają się nieco później.

3.3 Wynik na podstawie modelu stworzonego z pomocą scikit-learn(SVR())



Rysunek 4: Wykres pokazujący różnicę między wskaźnikiem inflacyjnym rzeczywistym a tym stworzonym na podstawie modelu stworzonego za pomocą tensorflow.

Na podstawie Rysunek 4. możemy zobaczyć, że wartości z naszego modelu przewidują wzrost wskaźnika inflacyjnego, widzimy na nim gwałtowne zmiany, ale nie są one aż tak gwałtowne i duże jak rzeczywiste.

4 Wnioski

Stworzone przez nas modele predykcyjne działają dobrze o czym świadczą stosunkowo niewielkie różnice pomiędzy wartościami testowymi, a tymi przewidzianymi, co jest szczególnie widoczne w przypadku modelu wykorzystującego regresję liniową. Warto zwrócić tutaj również uwagę na fakt, iż nasze modele nie biorą pod uwagę czynników takich jak zmiany stóp procentowych, czy dodruk gotówki, który w 2020r. wyniósł 238mld zł, a które również wpływają na poziom inflacji.

Dodatkowym powodem, aby twierdzić, że są to dobre modele jest fakt, że nasze modele nie prognozowały deflacji, jak te wykorzystywane przez NBP, o których w kwietniu 2020r. na konferencji prasowej prezes Adam Glapiński mówił: *„mamy najlepsze na świecie modele prognostyczne”* i dalej cytując: *„w oczy zagląda nam groźba deflacji - inflacja jest teraz ostatnim problemem. To będzie trudny i ciężki kryzys, ale mam nadzieję, że wyjście z niego będzie energiczne”*.^[2]

Podsumowując projekt oraz przygotowywanie modeli predykcyjnych można powiedzieć, że jest to dość prosta metoda do prognozowania np. wskaźnika inflacji, ale równie dobrze może sprawdzić się przy przewidywaniu kursu walut lub pogody.

Literatura

- [1] <https://stat.gov.pl/obszary-tematyczne/ceny-handel/wskazniki-cen/wskazniki-cen-towarow-i-uslug-konsumpcyjnych-pot-inflacja-/>, 14 06 2022.
- [2] <https://finanse.gazetaprawna.pl/artykuly/1468038,glapinski-grozba-deflacji.html>, 14 06 2022.