***Linear regression model with KERAS.***

Nim przystąpiłem do wykonania w Colab (wstępnie):

Mam trochę żal z powodu nie otrzymania odpowiedzi na moje wiadomości, przez to straciłem 2 dni.

Samo wykonanie modelu w Python było dla mnie sprawą dalszorzędną, liczą się wnioski (model GRETL). Uważam, że zrealizowałem zadanie na poziomie 75% z racji niewielkich braków w praktycznym zastosowaniu Python, a Keras to dla mnie zupełna nowość. Myślę, że dzięki krótkim konsultacjom i/lub intensywnemu kursowi wykonanie modeli ekonometrycznych w Python niestanowiłoby dla mnie problemu.

Kompletnie nie wiem jakie kwestie są dla Was istotne, stworzyłem **standardowy model regresji liniowej z wieloma zmiennymi. Przewidujący wysokość przyznawanych napiwków** na skutek zmiany zmiennych objaśniających o zadaną wielkość.

Nie wszystkie szczegółowe wyniki przeprowadzonych badań umieszczałem w raporcie (Bezproblemowo mogę uzupełnić).

\*przy pomocy poleceń SQL na platformie GCP ustaliłem strukturę danych. Wnioski: udokumentowane dane o ilości i wysokości przyznawanych napiwków są w odniesieniu do transakcji płaconych bezgotówkowo (płatności kartą). Dalsze badania wykluczyły wpływ ilości pasażerów na wysokość napiwków.

Z bazy liczącej ponad 1 800 000 rekordów pobrałem do badania plik ‘DANE.CSV’ zawierający 16 000 rekordów. Uznałem to za reprezentatywną próbę na podstawie, której można wnioskować o całym zbiorze.

\*w GRETLU, wykonałem statystyki opisowe każdej ze zmiennych (poza współrzędnymi i data time), dokonałem oceny istotności zmiennych(tłumaczących zmienną objaśnianą [y = wysokość napiwków{tip\_amount}], ***o jaką wielkość zmieni się y w wyniku wzrostu o 1% zmiennej z uwzględnieniem błędu szacunku***. Wyprowadziłem równanie modelu ze zmiennymi, model w wysokim stopniu dopasowany do zjawiska (predykcja i dane empiryczne [podział danych na estymację i weryfikację prognoz]. Jedynie 4/10 klientów płacących kartą gratyfikowała przejazd napiwkiem. Zestawiłem dane na wykresach, zbadałem relacje między zmiennymi, histogramy, rozkłady, stopień korelacji. A także zweryfikowałem reszty modelu, pozytywnie. KMNK



Zmienne objaśniające: (po odrzuceniu nieistotnych, na podstawie uzyskanych parametrów)

**X1-*fare\_amount* {wysokość opłaty taryfowej}**

**X2-*trip\_distance* {dystans kursu}**

**X3-*total\_amount* {kwota do zapłaty [z uwzględnieniem podatków, opłat dodatkowych, m.in. napiwków etc.]**

**========================================================================================================== coef std err t P>|t| [95.0%Conf.Int.]**

**==========================================================================================================**

Całe badanie zostało podzielone na następujące etapy:

\*Identyfikacja,

\*Estymacja,

\*Prognoza,

\*Weryfikacja,

Klienci płacący kartą są skłonni………….



Colab: (większość etapów uruchamia się prawidłowo)

1.W nowym notatniku zaimportowałem wszystkie niezbędne biblioteki, paczki, moduły niezbędne do zaimportowania i przetwarzania danych. Zaimportowane dane Przystosowałem do wykonywania operacji matematycznych.

2. Statystki opisowe zmiennych, wykresy zmiennych [x, y] rozkłady częstości(density). Współczynniki korelacji, błąd szacunku

3. Prognoza, weryfikacja (m.in. badanie reszt etc.).

Nie wszystkie wyniki są prezentowane,

Dane zostały podzielone na train:test w proporcjach 7:3

Layer, density, epoch (machine learning)

Zał. 1 [**Zawartość tabeli do badań(Nazwy pól)]**

vendor\_id,

pickup\_datetime,

dropoff\_datetime,

passenger\_count,

trip\_distance,

pickup\_longitude,

pickup\_latitude,

rate\_code,

store\_and\_fwd\_flag,

dropoff\_longitude,

dropoff\_latitude,

payment\_type,

fare\_amount,

extra,

mta\_tax,

tip\_amount

,tolls\_amount,

imp\_surcharge,

total\_amount