

Programowanie w języku Python

dr Przemysław Juszczuk

Zestaw zaliczeniowy 1

Projekt zaliczeniowy cz. I

Rzeczywiste dane finansowe zawierają informację o cenie instrumentu oraz wartościach wskaźników technicznych. Decyzja o zakupie (BUY lub STRONGBUY), sprzedaży (SELL lub STRONGSELL), lub czekaniu (WAIT) podejmowana jest w oparciu o poniższe wskaźniki, a także na podstawie relacji pomiędzy wskaźnikami, a ceną. Poniżej szczegółowy opis danych:

- Close – wartość instrumentu na zamknięcie danej sesji;
- SMA14 i SMA50 – wartości prostych średnich kroczących wyznaczonych w oparciu o 14 (SMA14) i 50 (SMA50) ostatnich odczytów;
- SMA14IND i SMA50IND - wartości zmiany dla średnich kroczących obserwowane pomiędzy sąsiednimi odczytami;
- Bulls – wskaźnik określający siłę byków (siła trendu wzrostowego na rynku).

Projekt zaliczeniowy cz. I

- CCI,DM, RSI – Commodity Channel Index, DeMarker oraz Relative Strength Index – oscylatory szacujące poziom wykupienia (wyprzedania na rynku);
- OSMA – Oscillator of Moving Average – oscylator średniej kroczącej;
- Stoch – oscylator stochastyczny;
- Decision – wartość decyzji podejmowanej na podstawie ustalonych wartości wskaźników (SELL, STRONGSELL, BUY, STRONGBUY, WAIT).

Projekt zaliczeniowy cz. I

- 1 Wczytaj dane do formatu DataFrame wybierając tylko 2500 obiektów z pliku;
- 2 Usuń kolumny oznaczone jako SMA14IND oraz SMA50IND;
- 3 Dla kolumny Close policz liczbę wystąpień danych pustych. Napraw dane w taki sposób, że pusta wartość zastępowana jest wartością uśrednioną dwóch sąsiednich elementów;
- 4 W przypadku danych pustych w kolumnach SMA14 i SMA50 – napraw wartości puste dowolną metodą;
- 5 Dla wszystkich pozostałych atrybutów wypełnij wartości puste zerami;
- 6 Wyznacz korelację pomiędzy SMA14 i SMA50;
- 7 Wyznacz korelację pomiędzy Close oraz SMA14 a także pomiędzy Close oraz SMA50. Usuń kolumnę, dla której wartość korelacji była większa;
- 8 Podaj liczbę elementów ujemnych dla atrybutu CCI;
- 9 Podaj informację o wartości maksymalnej i minimalnej dla każdego atrybutu;

Projekt zaliczeniowy cz. I

- 1 Przeprowadź normalizację dwóch wybranych atrybutów;
- 2 Przeprowadź dyskretyzację dwóch wybranych atrybutów (podział odpowiednio na 2 i 4 kategorie);
- 3 Na wykresie kołowym przedstaw rozkład wartości decyzji (atrybut Decision);
- 4 Na wykresie liniowym przedstaw przebieg zmienności atrybutu Close;
- 5 Dane po preprocessingu zapisz do pliku w formacie JSON.

Wszystkie powyższe kroki należy wykonać w ramach jednego lub kilku skryptów. Do kodu należy dołączyć komentarze.

Projekt zaliczeniowy cz. II

Należy wygenerować sztuczny zestaw danych o takiej samej wielkości, jak zestaw bazowy. Jednocześnie należy kierować się następującymi założeniami:

- w pierwszym wierszu dla danych znajdują się elementy losowe z zakresu $\langle min_i, max_i \rangle$, gdzie min_i jest wartością minimalną i-tego atrybutu, a wartość max_i jest wartością maksymalną i-tego atrybutu;
- losowe wartości w poszczególnych kolumnach nie mogą wyjść poza wskazany zakres $\langle min_i, max_i \rangle$, dla $i = 1, 2, \dots, k$ gdzie k jest liczbą atrybutów z wyłączeniem atrybutu decyzyjnego;
- zmiana wartości dla poszczególnych atrybutów w kolejnych wierszach musi być w zakresie $\langle prev_{val} - prev_{val} \cdot 1\%; prev_{val} + prev_{val} \cdot 1\% \rangle$;
- w przypadku, gdy wynikiem będzie wartość poza zakresem max (lub min) należy jako nową wartość przyjąć właśnie dany zakres;
- po wygenerowaniu tablicy należy policzyć korelację pomiędzy każdą z kolumn z pierwszego zestawu oraz odpowiadającym kolumnom z drugiego zestawu (tj. korelacja pomiędzy pierwszymi kolumnami, korelacja pomiędzy drugimi kolumnami i tak dalej) – 45 minut (etap 8).

Dziękuję za uwagę.