

Projekt Hurtownie Danych – funkcjonalności aplikacji

Dla projektów 1-osobowych:

funkcjonalności	ocena		
	3	4	5
Odczyt danych z pliku CSV	TAK	TAK	TAK
Obliczanie miar statystycznych (min, max, odchylenie std., mediana, moda, inne) – dla danych numerycznych i nienumerycznych	TAK	TAK	TAK
Wyznaczanie korelacji cech / atrybutów – 1 wybrana metoda liczenia korelacji	TAK	TAK	TAK
Ekstrakcja podtablic poprzez podanie numerów wierszy lub nazw kolumn. Podajemy numery/nazwy, które mają zostać usunięte.	TAK	TAK	TAK
Zastępowanie wartości w tablicach danych (zamiana wartości z danej kolumny na inną) – zarówno ‘ręcznie’ jak i w sposób zautomatyzowany np. zamiana wszystkich wartości ‘a’ w kolumnie A na wartość ‘b’ (użytkownik decyduje która wartość w której kolumnie ma być zmieniona).	TAK	TAK	TAK
Skalowanie i standaryzacja kolumn w tabelach danych (MinMaxScaler, StandardScaler)		TAK	TAK
Usuwanie wierszy z brakującymi wartościami i wypełnianie pustych miejsc. Chodzi o mechanizm, który pozwoli automatycznie usuwać wiersze lub kolumny (decyduje użytkownik) z brakującymi wartościami oraz mechanizm, który pozwoli automatycznie zamienić wartości brakujące na inne, np. wartość średnią, medianę itp. (decyduje użytkownik).		TAK	TAK
Usuwanie powtarzających się miejsc (wierszy) w tablicach danych.		TAK	TAK
Kodowanie kolumn symbolicznych (1 metoda do wyboru: One-Hot Encoding lub Binary Encoding)		TAK	TAK
Proste wykresy dot. danych (np. wykres zależności wybranych 2 atrybutów, histogram wartości danej cechy, ...). Min. 2 rodzaje wykresów, np. słupkowy, liniowy, kołowy, punktowy, mapa cieplna, inne		TAK	TAK
Do wyboru jedno z: <ul style="list-style-type: none"> Klasyfikacja danych wybranym klasyfikatorem (zadawanie parametrów budowy modelu, wybór rodzaju/metody przeprowadzania eksperymentów na danych) Grupowanie danych K-Means (z różnymi parametrami) Reguły asocjacyjne apriori (z różnymi parametrami) 			TAK
Intuicyjne GUI z możliwością zadawania parametrów i prezentacją wyników	TAK	TAK	TAK
Dokumentacja projektu wg. wzoru dokumentacjaHD.docx	TAK	TAK	TAK

Dla projektów 2-osobowych:

funkcjonalności	ocena		
	3	4	5
Odczyt danych z pliku CSV	TAK	TAK	TAK
Obliczanie miar statystycznych (min, max, odchylenie std., mediana, moda, inne) – dla danych numerycznych i nienumerycznych	TAK	TAK	TAK
Wyznaczanie korelacji cech / atrybutów – 2 metody liczenia korelacji spośród: 'pearson', 'kendall', 'spearman'	TAK	TAK	TAK
Ekstrakcja podtablic poprzez podanie numerów wierszy lub nazw kolumn. 2 opcje: (1) podajemy numery/nazwy, które mają zostać usunięte z danych, (2) podajemy numery/nazwy, które mają pozostać w danych wynikowych	TAK	TAK	TAK
Zastępowanie wartości w tablicach danych (zamiana wartości z danej kolumny na inną) – zarówno ‘ręcznie’ jak i w sposób zautomatyzowany np. zamiana wszystkich wartości ‘a’ w kolumnie A na wartość ‘b’ (użytkownik decyduje która wartość w której kolumnie ma być zmieniona).	TAK	TAK	TAK
Skalowanie i standaryzacja kolumn w tabelach danych (MinMaxScaler, StandardScaler)		TAK	TAK
Usuwanie wierszy z brakującymi wartościami i wypełnianie pustych miejsc. Chodzi o mechanizm, który pozwoli automatycznie usuwać wiersze lub kolumny (decyduje użytkownik) z brakującymi wartościami oraz mechanizm, który pozwoli automatycznie zamienić wartości brakujące na inne, np. wartość średnią, medianę itp. (decyduje użytkownik).		TAK	TAK

Usuwanie powtarzających się miejsc (wierszy) w tablicach danych.		TAK	TAK
Kodowanie kolumn symbolicznych (3 metody: One-Hot Encoding, Binary Encoding, Target Encoding)		TAK	TAK
Proste wykresy dot. danych (np. wykres zależności wybranych 2 atrybutów, histogram wartości danej cechy, ...). Min. 4 rodzaje wykresów, np. słupkowy, liniowy, kołowy, punktowy, mapa cieplna, inne		TAK	TAK
Do wyboru: <ul style="list-style-type: none"> Klasyfikacja danych wybranym klasyfikatorem (zadawanie parametrów budowy modelu, wybór rodzaju/metody przeprowadzania eksperymentów na danych) Grupowanie danych (z różnymi parametrami) Reguły asocjacyjne (z różnymi parametrami) Możliwości zastosowania, do wyboru: <ul style="list-style-type: none"> 3 różne metody klasyfikacji (np. w celu porównywania wyników uzyskanych różnymi metodami) 2 różne metody klasyfikacji + grupowanie K-Means 2 różne metody klasyfikacji + reguły asocjacyjne apriori 1 metoda klasyfikacji + grupowanie K-Means + reguły asocjacyjne apriori 			TAK
Intuicyjne GUI z możliwością zadawania parametrów i prezentacją wyników	TAK	TAK	TAK
Dokumentacja projektu (wspólna) wg. wzoru dokumentacjaHD_2.docx	TAK	TAK	TAK