3.4, 228884, Mateusz Guściora

3.4.1

a) przykładowe procesy, dla zbioru danych Labor-Negotiations, Titanic.

Zapoznajemy się z podanymi blokami replace missing values- zastępujący wartości wartościami podanymi w parametrach, operator impute missing values- operator szacujący brakujące wartości na podstawie nadanego modelu, zawierający podproces. Replace all missings- Który automatycznie uzupełnia brakujące dane w całym zbiorze.

Uruchomiono procesy przykładowe dla bloku impute missing values dla Labour-Negotioations, dla replace all missing values dla Labor\_Negotations oraz dla replace all missings dla Titanic.

Stworzono proces z wykorzystaniem multiply dla dwóch strumieni z użyciem Replace Missing Values oraz Impute missing values. Poniżej porównanie performance vector dla algorytmu rule induction z operatorem Replace missing values i bez („z wykorzystaniem danych oryginalnych i zbiorów z uzupełnionymi brakami danych”)

Poniżej z Replace Missing values Labor Negotiations

Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

Po próbie uruchomienia bez operatora Replace missing values stwierdzono, że proces trwa zbyt długo i nie uzyskano wyników.

b)

Został stworzony plik csv bank6missingvalues który różni się od bank6 tym że nie ma wartości w miejscach unknown i 999.0 w pdays.

Proces dla danych bankowych-bank6missing values

Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

Wyniki dla danych bank6 tzn. danych gdzie znajduje się unknonw i 999 w pdays

Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

c)Proces przykładowy - churn modeling

Dla rule induction z użyciem operatora Replace missing values gdzie jako w parametrach w liscie wybrano unknown i pdays(w obu jako average) performance vector pokazano poniżej gdzie 89,18 Accuracy daje zadowalający wynik, kappa natomiast wynosi 0,382.Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

d) Proces przykładowy Churn modeling z operatorem Cost-Sensitive Scoring i operatorem SMOTE Upsampling.

Cost-Sensitive Scoring- jest to operator/algorytm wykorzystujący jeden model uczenia maszynowego. Tylko jeden model jest uczony. Operator SMOTE Upsampling, najpierw trzeba było pobrać poprzez dodanie toolbox do extensions. Jest to operator służacy dla niezbalansowanych zbiorów, korzysta także z algorytmu k-najbliższych sąsiadów.

Dla danych bank6

Przy strumieniu z cross validation zawierającej algorytm decission tree i operator Cost-sensitive Scoring

Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

Kappa 0?(nie wiem dlaczego)

e)

Dla danych niekompletnych

Używanie operatorów replace missing czy impute missing values nie zmieniało zbytnio dokładności predykcji oraz wartości kappa wyniki oscylowały w granicach 89% a w przypadku danych z unknown i z 999 także 89% dokładności.