Tehnike i metode analize podataka

**Primena asocijativne analize i apriori algoritma za analizu transakcija u prodavnicama**

Danilo Milošević 1732

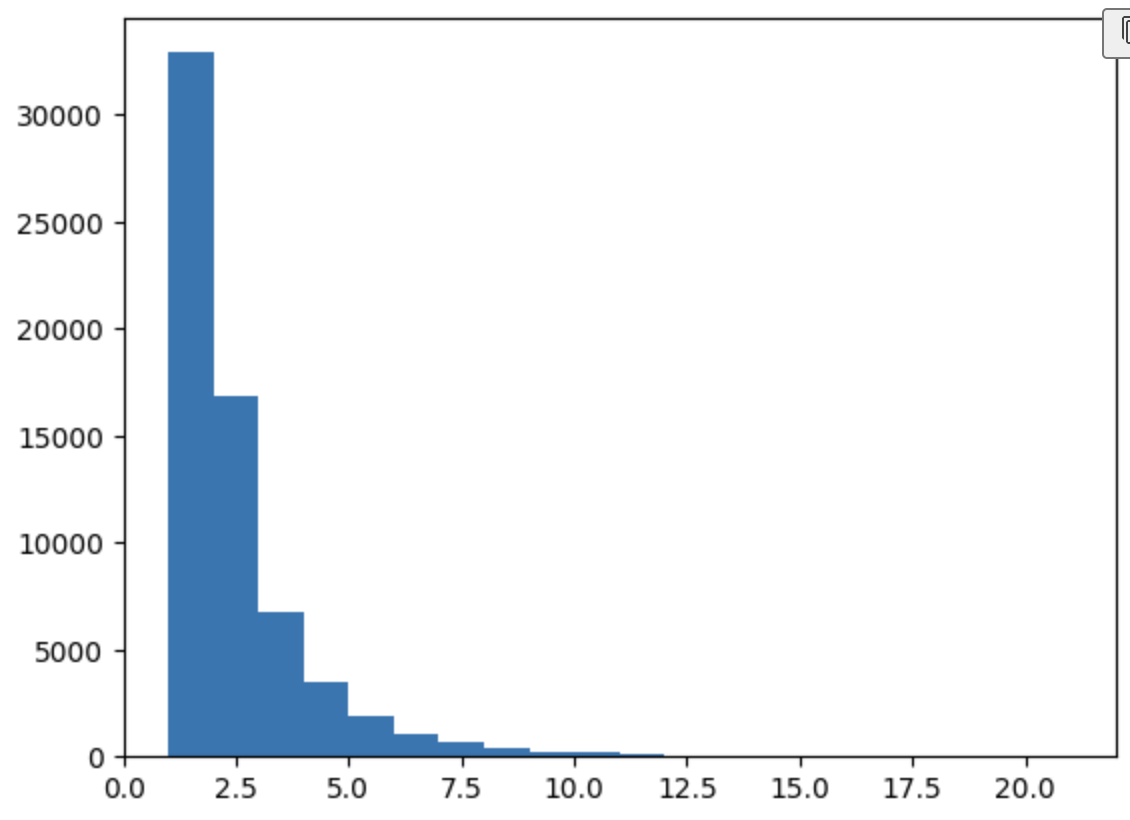
**Opis podataka i problema**

Vrši se analiza skupa podatka o transakcijama u prodavnicama sa ciljem određivanja asocijativnih pravila transakcija i kupaca. Želimo da odredimo da li postoje veze između proizvoda - da li kupovina nekih proizvoda uslovljava kupovinu drugih kao i koje proizvode kupci često kupuju zajedno. Atributi u skupu podataka su sledeći

* Unnamed:0 - ID svakog reda i suvišan je atribut, zato ga izbacujemo
* Date$- datum kada je izvršena prodaja
* Customer\_ID - ID kupca
* Transaction\_ID - ID transakcije, odnosno kupovine
* SKU\_Category - kategorija prodatog artikla
* SKU - ID prodatog artikla
* Quantity - količina prodatog artikla
* Sales\_amount - zarada od prodaje (cena \* kolicina)

Analizom podataka možemo sledeće činjenice da zaključimo

* Ukupno imamo 131706 ulaza u skupu podataka
* Imamo 8 atributa
* Postoje 187 jedinstvenih kategorija artikala
* Imamo 5242 jedinstvenih artikala
* Većina transakcija sadrže samo 1 artikal



*Broj transakcija po broju artikala*

**Obrada atributa**

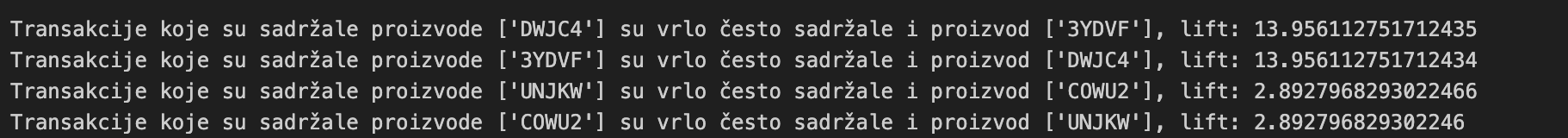
Prvo ćemo izbaciti atribut Unnamed:0. Kasnije vršimo grupaciju transakcija po transakciji odnosno kupcu i ID-u kupljenog artikla, pri čemu se dobija matrica koja predstavlja činjenicu da li je artikal kupljen u nekoj transakciji ili da li ga je neki kupac kupio.



**Određivanje pravila**

Primenom apirori i association\_rules funkcija biblioteke mlxtend dobijamo pravila kao i support i lift vrednosti.

Analizom transakcija dobijamo sledeća pravila sa visokim liftom



Analizom kupaca dobijamo sledeća pravila sa veoma velikim vrednostima lift-a

