**Data Mining**

**HW 4**

**Atieh Kashani**

**RED ID: 817365647**

1. What is the confidence for the rules A 🡪 ∅ ? Here, ∅ indicates an empty set.

Conf ( A 🡪 ∅ ) = Support (A∪∅) = Support(A) = 100%

Support (A) Support (A)

1. Let c1, c2, and c3 be the confidence values of the rules {p} 🡪{q}, {p} 🡪{q, r}, and {p, r} 🡪{q}, respectively. If we assume that c1, c2, and c3 have different values, which rule has the lowest confidence? Explain the reason.

{p} 🡪{q}: c1 = Support (p∪q)

Support (p)

{p} 🡪{q, r}: c2 = Support (p∪q∪r)

Support (p)

{p, r} 🡪{q}: c3 = Support (p∪q∪r)

Support (p∪r)

Assume: Support (p) ≥ Support (p∪q) ≥ Support (p∪q∪r)

So c1 ≥ c2 and c3 ≥ c2 , Therefore : c2 has the lowest confidence

3. trafficMain matlab file contains the code for creating and storing the table for traffic accident data and binarize each categorical attributes: I have used two variables ‘Yes’ and ‘No” for the following attributes and “Major” and “Minor” for Crash Severity’:

% Binarize categorical attributes

Weather\_Condition\_Good={'Yes';'No';'Yes';'Yes';'No';'Yes';'No';'Yes';'Yes';'No';'Yes';'No'};

Driver\_Condition\_Sober={'No';'Yes';'Yes';'Yes';'Yes';'No';'No';'Yes';'No';'Yes';'No';'Yes'};

Trafic\_Violation\_Exceed\_Speed\_Limit={'Yes';'No';'No';'Yes';'No';'No';'No';'No';'No';'No';'Yes';'No'};

Trafic\_Violation\_Disobey\_traffic\_Signal={'No';'No';'No';'No';'Yes';'No';'No';'Yes';'No';'Yes';'No';'No'};

Trafic\_Violation\_Disobey\_Stop\_Sign= {'No';'No';'Yes';'No';'No';'Yes';'No';'No';'No';'No';'No';'Yes'};

Seat\_Belt={'No';'Yes';'Yes';'Yes';'No';'Yes';'Yes';'Yes';'No';'No';'Yes';'Yes'};

%Crash Severity classification

Crash\_Severity={'Major';'Minor';'Minor';'Major';'Major';'Minor';'Major';'Major';'Major';'Major';'Major';'Minor'};

% Create table

T=table(Crash\_Severity,Weather\_Condition\_Good,Driver\_Condition\_Sober,Trafic\_Violation\_Exceed\_Speed\_Limit,Trafic\_Violation\_Disobey\_traffic\_Signal,Trafic\_Violation\_Disobey\_Stop\_Sign,Seat\_Belt);

header={'Crash\_Severity','Weather\_Condition\_Good ','Driver\_Condition\_Sober','Trafic\_Violation\_Exceed\_Speed\_Limit','Trafic\_Violation\_Disobey\_traffic\_Signal','Trafic\_Violation\_Disobey\_Stop\_Sign','Seat\_Belt'};

T.Properties.VariableNames = header



Also, this file contains the related codes for displaying the frequent itemsets and the association rules after generating them.

Function findFrequentItemsets get a nested cell array of items that each nested array is one of the records of the traffic accident data table. After calculating the support of each itemset, the result is stored in a Map object that is used in rule generation and it is part of computing the confidence. Also findFrequentOneItemsets function is used to find frequent one itemset and generateItemsets function to find the candidate k-itemsets. The candidate items that their support is below the threshold are pruned. For minimum support threshold I used 0.3. Following are the frequent itemsets:

Min support: 0.3

Number of Frequent Itemsets: 70

{Major, }

{Minor, }

{Weather\_Condition\_Good = Yes, }

{Driver\_Condition\_Sober = Yes, }

{Seat\_Belt = Yes, }

{Weather\_Condition\_Good =No, }

{Driver\_Condition\_Sober = No, }

{Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Seat\_Belt = No, }

{Major, Weather\_Condition\_Good = Yes, }

{Major, Driver\_Condition\_Sober = Yes, }

{Major, Seat\_Belt = Yes, }

{Major, Driver\_Condition\_Sober = No, }

{Major, Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Major, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Major, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Major, Seat\_Belt = No, }

{Minor, Seat\_Belt = Yes, }

{Minor, Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Minor, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Weather\_Condition\_Good = Yes, Seat\_Belt = Yes, }

{Weather\_Condition\_Good = Yes, Driver\_Condition\_Sober = No, }

{Weather\_Condition\_Good = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Weather\_Condition\_Good = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Weather\_Condition\_Good = Yes, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Driver\_Condition\_Sober = Yes, Seat\_Belt = Yes, }

{Driver\_Condition\_Sober = Yes, Weather\_Condition\_Good =No, }

{Driver\_Condition\_Sober = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Driver\_Condition\_Sober = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Driver\_Condition\_Sober = Yes, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Seat\_Belt = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Seat\_Belt = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Seat\_Belt = Yes, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Weather\_Condition\_Good =No, Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Weather\_Condition\_Good =No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Trafic\_Violation\_Disobey\_traffic\_Signal = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Trafic\_Violation\_Disobey\_Stop\_Sign = No, Seat\_Belt = No, }

{Major, Weather\_Condition\_Good = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Major, Weather\_Condition\_Good = Yes, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Major, Driver\_Condition\_Sober = Yes, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Major, Seat\_Belt = Yes, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Major, Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Major, Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Major, Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Major, Trafic\_Violation\_Disobey\_traffic\_Signal = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Major, Trafic\_Violation\_Disobey\_Stop\_Sign = No, Seat\_Belt = No, }

{Minor, Seat\_Belt = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Minor, Seat\_Belt = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Minor, Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Weather\_Condition\_Good = Yes, Seat\_Belt = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Weather\_Condition\_Good = Yes, Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Weather\_Condition\_Good = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Driver\_Condition\_Sober = Yes, Seat\_Belt = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Driver\_Condition\_Sober = Yes, Seat\_Belt = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Driver\_Condition\_Sober = Yes, Weather\_Condition\_Good =No, Trafic\_Violation\_Exceed\_Speed\_Limit = No, }

{Driver\_Condition\_Sober = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Seat\_Belt = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

{Seat\_Belt = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Weather\_Condition\_Good =No, Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Major, Weather\_Condition\_Good = Yes, Trafic\_Violation\_Disobey\_traffic\_Signal = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Major, Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, }

{Minor, Seat\_Belt = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, }

4. The association rules are generated from the frequent

itemsets based on the minimum confidence threshold and the rules are presented as {antecedent} => {consequent}. For minimum confidence threshold I used 0.6 and those rules that don’t meet the minimum confidence are pruned. The following rules are generated:

{Driver\_Condition\_Sober = No, } => {Major}

{Trafic\_Violation\_Disobey\_Stop\_Sign = No, } => {Major}

{Seat\_Belt = No, } => {Major}

{Weather\_Condition\_Good = Yes, Trafic\_Violation\_Disobey\_Stop\_Sign = No, } => {Major}

{Driver\_Condition\_Sober = Yes, Trafic\_Violation\_Disobey\_Stop\_Sign = No, } => {Major}

{Seat\_Belt = Yes, Trafic\_Violation\_Disobey\_Stop\_Sign = No, } => {Major}

{Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, } => {Major}

{Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, } => {Major}

{Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, } => {Major}

{Trafic\_Violation\_Disobey\_traffic\_Signal = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No, } => {Major}

{Trafic\_Violation\_Disobey\_Stop\_Sign = No, Seat\_Belt = No, } => {Major}

{Driver\_Condition\_Sober = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, Trafic\_Violation\_Disobey\_Stop\_Sign = No,} => {Major}

{Seat\_Belt = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, } => {Minor}

{Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No, } => {Minor}

{Seat\_Belt = Yes, Trafic\_Violation\_Exceed\_Speed\_Limit = No, Trafic\_Violation\_Disobey\_traffic\_Signal = No,} => {Minor}

References:

https://www-users.cs.umn.edu/~kumar001/dmbook/ch6.pdf

https://blogs.mathworks.com/loren/2015/01/29/introduction-to-market-basket-analysis/

https://github.com/toshiakit/apriori