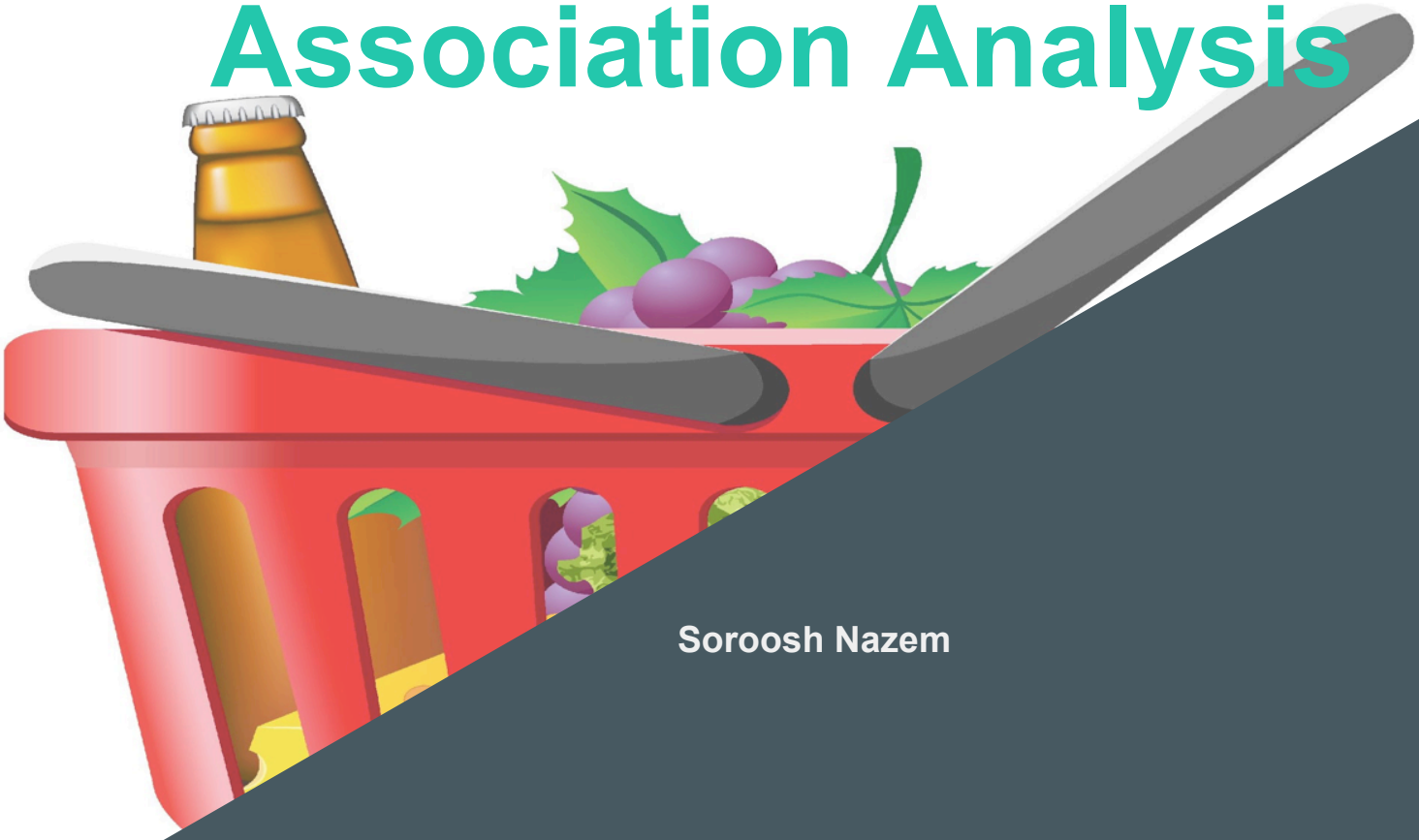
Association Analysis



Soroosh Nazem

Introduction

- **One of the most importants areas in data mining**
- **Talks about correlations between items on a set in transactions**
- **Application Areas:**
  - Market Basket Analysis
  - Recommendation Systems
  - Customer Relationship Management (CRM)
  - Medical Diagnosis
  - Census Data
  - ……

Application:Market Basket

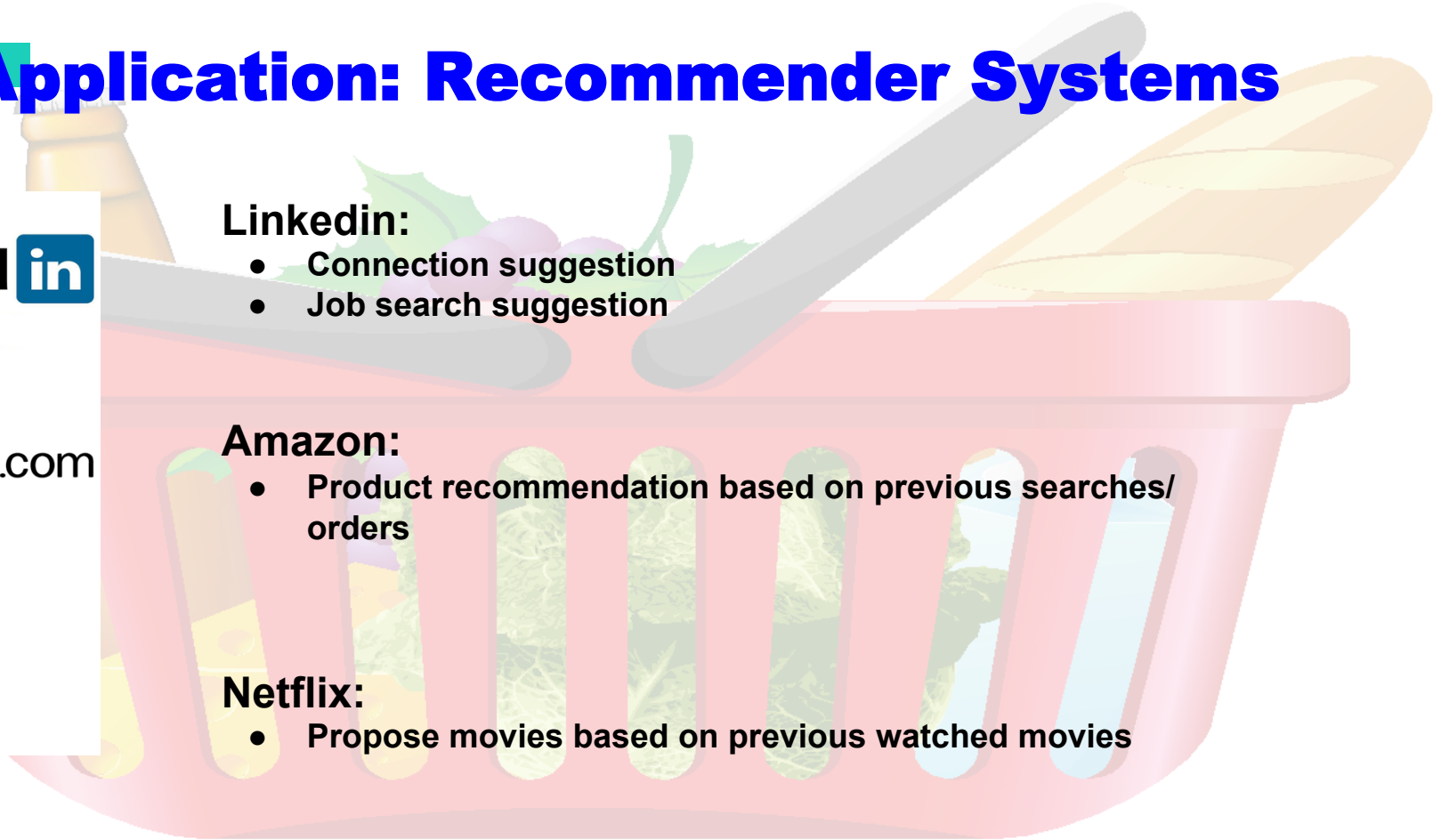
- Predicting of customers behaviors based on previous transactions.
- Example (based on table):
  - Most customers buy “whole milk”.
  - Customers with “curd” buy “whole milk” 100%.

Why Supermarket need this info?

- loyalty program management
- Location of items in supermarket
- promotions/discount management
- …….

TRANSACTION	ITEMS
T1	Tropical fruit, yogurt, coffee
T2	Whole milk
T3	Whole milk, butter, yogurt, rice
T4	Whole milk, cereals
T5	Citrus fruit, tropical fruit, whole milk, butter, curd, yogurt, flour, bottled water, dishes
T6	chicken, tropical fruit
T7	Root vegetables, other vegetables, whole mik, beverages, sugar
T8	Berries, yogurt
T9	Whole milk, curd, yogurt, pastry

Application: Recommender Systems



Linkedin:

- Connection suggestion
- Job search suggestion

Amazon:

- Product recommendation based on previous searches/ orders

Netflix:

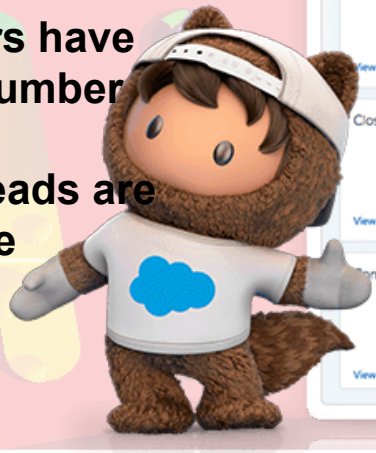
- Propose movies based on previous watched movies

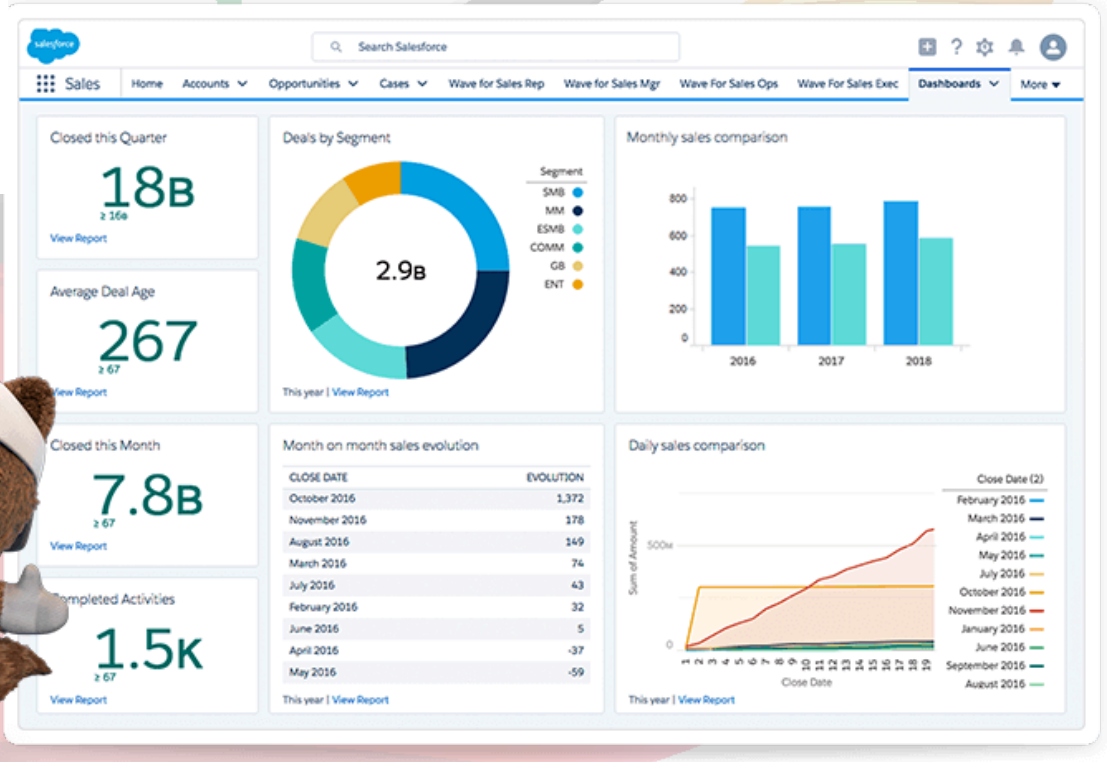
Application: CRM

HubSpot

salesforce

- What parameters have Increased the number of leads?
- What kinds of leads are more likely to be converted to a contract?
- …….





Definitions:

Consider a set of items $I=\{I1,I2,....., In\}$:

- A Transaction Ti is a items Ij where $1 \leq j \leq n$ and $Ti \subseteq I$
- $T=\{T1,T2,.....\}$ a set of Transactions
- An "Association Rule (AR)" written in the form of $X \Rightarrow Y$ where $X \cap Y = \emptyset$ and $X, Y \subseteq I$.

Example:

I={milk, sugar, salt, yogurt, cheese, butter, vegetable, bottled water, rice, fruit}

T1={milk, cheese, butter}, T2={milk, rice, salt}, T3={vegetable, fruit, rice}, T4={yogurt, salt, rice}, T5={milk, sugar, salt, butter, rice}, T6={vegetables, bottled water}, T7={vegetables, fruit}

AR1:{rice}⇒{salt}, AR2:{cheese, butter}⇒{milk}, AR2: {fruit}⇒{bottled water,

Metrics: SUPPORT

DEF

SUPP(X⇒Y)=
(# Transaction that contain X and Y)/#Total Transactions

EX

SUPP({whole milk}⇒{yogurt})=3/9
SUPP({tropical fruit}
⇒{rice})=0/9=0
SUPP({yogurt}⇒{whole
{}})=3/9

NOTE

SUPP(X⇒Y) =
SUPP(Y⇒X)
Range:[0,1]

TRANSACTION	ITEMS
T1	Tropical fruit, yogurt, coffee
T2	Whole milk
T3	Whole milk, butter, yogurt, rice
T4	Whole milk, cereals
T5	Citrus fruit, tropical fruit, whole milk, butter, curd, yogurt, flour, bottled water, dishes
T6	chicken, tropical fruit
T7	Root vegetables, other vegetables, whole milk, beverages, sugar
T8	Berries, yogurt
T9	Whole milk, curd, yogurt, pastry

Metrics: CONFIDENCE

DEF

CONF(X⇒Y)= SUPP(X⇒Y)/ SUPP(X)

EX

CONF({whole milk} ⇒{yogurt})=3/6=50%
CONF({tropical fruit}⇒{rice})=0/9=0
CONF({yogurt}⇒{whole milk}))= 3/5 60%

NOTE

CONF(X⇒Y) ≠ CONF(Y⇒X)
Range:[0,1]

TRANSACTION	ITEMS
T1	Tropical fruit, yogurt, coffee
T2	Whole milk
T3	Whole milk, butter, yogurt, rice
T4	Whole milk, cereals
T5	Citrus fruit, tropical fruit, whole milk, butter, curd, yogurt, flour, bottled water, dishes
T6	chicken, tropical fruit
T7	Root vegetables, other vegetables, whole milk, beverages, sugar
T8	Berries, yogurt
T9	Whole milk, curd, yogurt, pastry

Metrics: LIFT

DEF

LIFT(X⇒Y)= CONF(X⇒Y)/ SUPP(Y)

EX

LIFT({whole milk}⇒{yogurt})=(3/6)/ (5/9)=0.9
LIFT({tropical fruit}⇒{rice})=0
LIFT({yogurt}⇒{whole milk}})= (3/5)/ 6/9)=0.9

NOTE

LIFT(X⇒Y) = LIFT(Y⇒X)
Range:[0,∞)

TRANSACTION	ITEMS
T1	Tropical fruit, yogurt, coffee
T2	Whole milk
T3	Whole milk, butter, yogurt, rice
T4	Whole milk, cereals
T5	Citrus fruit, tropical fruit, whole milk, butter, curd, yogurt, flour, bottled water, dishes
T6	chicken, tropical fruit
T7	Root vegetables, other vegetables, whole milk, beverages, sugar
T8	Berries, yogurt
T9	Whole milk, curd, yogurt, pastry

Metrics: LEVERAGE

DEF

$$LEV(X \Rightarrow Y) = SUPP(X \Rightarrow Y) - SUPP(X).SUPP(Y)$$

EX

$LEV(\{whole\ milk\} \Rightarrow \{yogurt\}) = -1/27$

$LEV(\{tropical\ fruit\} \Rightarrow \{rice\}) = 0-(3/9)(1/9)=-1/27$

$LEV(\{yogurt\} \Rightarrow \{whole\ milk\}) = -1/27$

$LEV(\{curd\} \Rightarrow \{yogurt\}) = 2/9-(2/9)(5/9)=8/81$

NOTE

$LEV(X \Rightarrow Y) = LEV(Y \Rightarrow X)$

$Range:[-1,1]$

TRANSACTION	ITEMS
T1	Tropical fruit, yogurt, coffee
T2	Whole milk
T3	Whole milk, butter, yogurt, rice
T4	Whole milk, cereals
T5	Citrus fruit, tropical fruit, whole milk, butter, curd, yogurt, flour, bottled water, dishes
T6	chicken, tropical fruit
T7	Root vegetables, other vegetables, whole milk, beverages, sugar
T8	Berries, yogurt
T9	Whole milk, curd, yogurt, pastry

Metrics: CONVICTION

DEF

$$CONV(X \Rightarrow Y) = \frac{(1-SUPP(Y))}{(1-CONF(X \Rightarrow Y))}$$

EX

$CONV(\{whole\ milk\} \Rightarrow \{yogurt\}) = 8/9 \simeq 0.89$

$CONV(\{tropical\ fruit\} \Rightarrow \{rice\}) = 8/9/(1-0) \simeq 0.89$

$CONV(\{yogurt\} \Rightarrow \{whole\ milk\}) = 5/6 \simeq 0.83$

$CONV(\{curd\} \Rightarrow \{yogurt\}) = 4/9/(1-1) = \infty$

NOTE

$CONV(X \Rightarrow Y) \neq CONV(Y \Rightarrow X)$

$Range:[0,\infty)$

TRANSACTION	ITEMS
T1	Tropical fruit, yogurt, coffee
T2	Whole milk
T3	Whole milk, butter, yogurt, rice
T4	Whole milk, cereals
T5	Citrus fruit, tropical fruit, whole milk, butter, curd, yogurt, flour, bottled water, dishes
T6	chicken, tropical fruit
T7	Root vegetables, other vegetables, whole milk, beverages, sugar
T8	Berries, yogurt
T9	Whole milk, curd, yogurt, pastry

FREQUENT ITEMSET

DEF

An item set X where SUPP(X)≥ minsup is called a frequent itemset

EX

minsup=⅓

frequent itemsets:

SUPP({whole milk})=2/3

SUPP({yogurt})=5/9

SUPP({yogurt,whole milk})=1/3

SUPP({tropical fruit)=1/3

TRANSACTION	ITEMS
T1	Tropical fruit, yogurt, coffee
T2	Whole milk
T3	Whole milk, butter, yogurt, rice
T4	Whole milk, cereals
T5	Citrus fruit, tropical fruit, whole milk, butter, curd, yogurt, flour, bottled water, dishes
T6	chicken, tropical fruit
T7	Root vegetables, other vegetables, whole milk, beverages, sugar
T8	Berries, yogurt
T9	Whole milk, curd, yogurt, pastry

FREQUENT ITEMSET

Q

How to find frequent item sets?

EX

Algorithms

- ***Apriori***
- ***FP-growth***
- ***ECLAT***

Different methods, same results

TRANSACTION	ITEMS
T1	Tropical fruit, yogurt, coffee
T2	Whole milk
T3	Whole milk, butter, yogurt, rice
T4	Whole milk, cereals
T5	Citrus fruit, tropical fruit, whole milk, butter, curd, yogurt, flour, bottled water, dishes
T6	chicken, tropical fruit
T7	Root vegetables, other vegetables, whole milk, beverages, sugar
T8	Berries, yogurt
T9	Whole milk, curd, yogurt, pastry