

LØSNINGSSKISSE TIL EKSAMENSOPPGAVE I FAG TDT4300 – MAI 2020

NB! Dette er ikkje fullstendige løysingar på oppgåvene, kun skisse med viktige element, hovudsakleg laga for at vi skal ha oversikt over arbeidsmengda på eksamen, og som huskeliste under sensurering. Det er også viktig å være klar over at det også kan vere andre svar enn dei som er gjeve i skissa som vert rekna som korrekt om ein har god grunngjeving eller dei gjev meining utifrå kva det vert spurd om.

Oppgave 1 – Modellering og OLAP – 20 %

a) Her er det flere mulige svar utifra forutsetningene man gjør, her er en variant:

Faktatabell:

Delivery(time_k, cust_k, rest_k, pris)

Dimensjonstabeller:

Time(time_k, minute, hour, day, month, year)

Customer(cust_k, cust_name, address, city, email)

Restaurant(rest_k, rest_name, address, city)

Legg merke til at det ikke er sagt noe i oppgaven om (for eksempel) pris per rett etc. Det er ikke galt å ha med dette, men ikke nødvendig. Et viktig moment er forståelse for hva som skal være i faktatabell, og hva som skal være i dimensjonstabeller.

b)

1) Nei, lokasjon mangler

2) Ja, rollup av street til city, og seleksjon på 2006

3) Nei, kan ikke gjøre drill-down til city på materialisert kube

Oppgave 2 – Klynging og klyngingsvalidering – 30 %

a) Svar: 2 klynger (P1,P2,P6,P9) og (P3,P4,P7,P8), en outlier som er P5

b) $a=1$, $b=\min((3+4)/2, (2+3)/2)=\min(3.5, 2.5) = 2.5$
 $s = (b-a)/\max(a,b)=(2.5-1)/\max(2.5,1)=1.5/2.5=0.6$

Oppgave 3 – Klassifisering – 20 %

Gini i rotnode:

$p(J|Parent) = 6/15 = 0.4$, $p(N|Parent) = 9/15 = 0.6$

$GI(J, N) = 1 - 0.4 * 0.4 - 0.6 * 0.6 = 1 - 0.16 - 0.36 = 0.48$

1) **Splitting på A:**

$S1 = "L"$

$J1=2$, $N1=8$, $GI(J1, N1)=GI(2,8)=1 - 2/10 * 2/10 - 8/10 * 8/10 = 0.32$

$S2 = "H"$

$J2=3$, $N2=1$, $GI(J2, N2)=GI(4,1)=1 - 4/5 * 4/5 - 1/5 * 1/5 = 0.32$

$GAIN(A1) = 0.48 - 10/15 * 0.32 - 5/15 * 0.3200 = 0.16$

2) Splitting på C

S1="G"

J1=1, N1=3, GI()=1-1/4*1/4-3/4*3/4=0.375

S2="H"

J2=2, N2=3, GI()=1-2/5*2/5-3/5*3/5=0.48

S3="S"

J3=3, N3=3, GI()=1-3/6*3/6-3/6*3/6=0.5

$$\text{GAIN (A2)} = 0.48 - 4/15 * 0.375 - 5/15 * 0.48 - 6/15 * 0.5 = 0.02$$

Vi vel attributtet med størst GAIN, dvs. **A** blir foretrukket for første splitting av treet.

Oppgave 4 – Assosiasjonsregler – 30 %

a)

C1: A:7, B:6, C:1, D:3, E:2, F:7, G:1, H:5

F1: A:7, B:6, F:7, H:5

C2: AB:5, AF:6, AH:5, BF:6, BH:4, FH:4

F2: AB:5, AF:6, AH:5, BF:6, BH:4, FH:4

C3: ABF:5, ABH:4, AFH:4, BFH:4

F3: ABF:5, ABH:4, AFH:4, BFH:4

C4: ABFH: 4

F4: ABFH: 4

b)

A->BH	4/7	0.57	
AB->H	4/5	0.8	*
B->AH	4/6	0.67	
BH->A	4/4	1.0	*
H->AB	4/5	0.8	*
AH->B	4/5	0.8	*