

NALOGA 1

Zaporniki se na otoku lahko prosto gibljejo. Do kopnega je 8 km in pogosto lahko vidimo v morju morske pse. Nekateri zaporniki so poskušali pobegniti s plavanjem ponoči, a samo nekaterim uspe doseči kopno. V tabeli so opisi nekaterih poskusov pobega.

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

Uporabi naivni Bayes z m-oceno ($m=5$) in oceni verjetnost, da je bil zapornik dober plavalec, če ni dosegel kopna med polno luno ter so ga med pobegom opazili pazniki.

NALOGA 2

Ko je novopečeni gobar prišel iz gozda, je nesel polno košaro gob, čeprav ni vedel, ali so užitne ali ne. Prosil je izkušenega gobarja, naj jih razdeli v tri košare: v prvo vse užitne, v drugo vse strupene (ki so seveda neužitne) in v tretjo nestrupene a neužitne gobe. Novopečeni gobar je sestavil učno množico tako, da je za vsako gobo zapisal barvo klobuka in barvo beta, nato pa je preštel gobe iz vsake podskupine. Za vsako košaro je dobil po eno tabelo:

UŽITNE	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

NEUŽITNE STRUPENE	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

NEUŽITNE NESTRUPENE	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

Kakšna je verjetnost, da je goba z rjavim betom in rdečim klobukom užitna, če naivni Bayes uporablja m-oceno in je $m = 10$?

NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, od bronastih ni bil nobeden ponarejen in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja. Naj bo to naša učna množica (2 atributa: vrsta cekina, simbol gusarja; 2 razreda: pravi/ponarejen; 100 učnih primerov).

- a) Nariši odločitveno drevo, ki vsebuje celotno zgoraj opisano informacijo, torej v vsakem listu mora biti napisano število cekinov, ki ustreza temu listu in verjetnost pravilnega odgovora. Pri gradnji drevesa za izbiro atributa uporabi **informacijski prispevek**,
- b) Z odločitvenim drevesom klasificiraj 101. cekin, ki ga je potapljač kasneje našel na obali, kjer ga je bil izgubil. Cekin je bil srebrn z vtisnjenim simbolom gusarja.

NALOGA 4

Da bi zagovili likvidnost banke, so začeli zbirati podatke o posojilojemalcih in o tem, ali je bilo posojilo vrnjeno. Zbrali so podatke v spodnji tabeli.

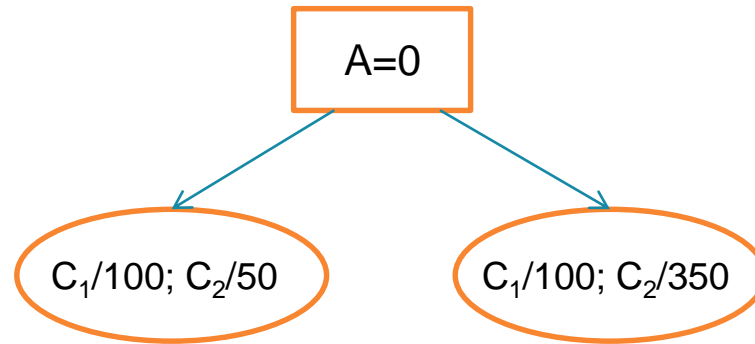
Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

Direktor banke se je odločil, da bo napovedoval, ali bo posojilo vrnjeno, kar s pomočjo odločitvenega drevesa, ki ga je zgradil z uporabo **gini-indeksa** iz zgornje podatkovne baze o posojilojemalcih.

- Nariši direktorjevo odločitveno drevo.
- Kakšna je napoved za direktorjevo ženo, ki ni v kazenskem postopku, nikoli ni imela kreditov in bi si sposodila 20 000 evrov?

NALOGA 5

Spodaj je odločitveno drevo s frekvencami primerov v listih za razreda C1 in C2:



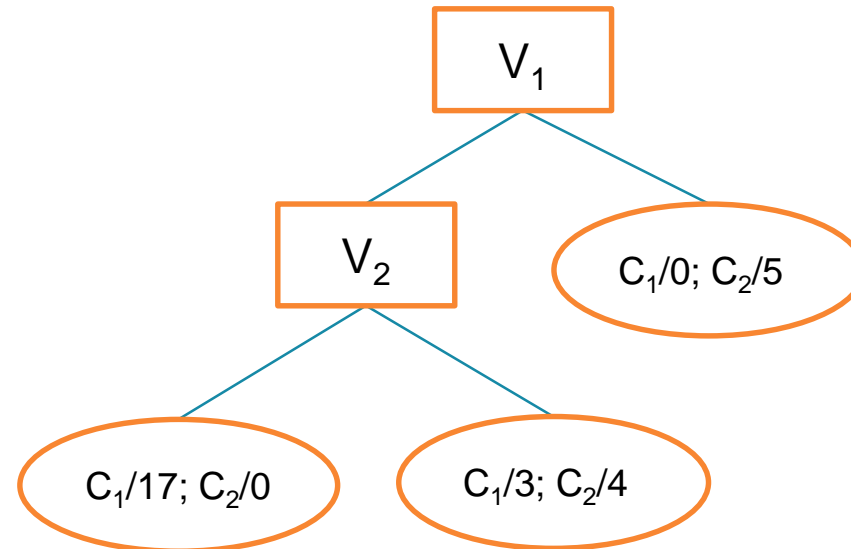
Izračunaj pričakovano klasifikacijsko točnost drevesa z m-oceno, če je:

- a) $m = 0$,
- b) $m = 1000$

NALOGA 6

Podano je odločitveno drevo za klasifikacijo v razrede z naslednjimi apriornimi verjetnostmi razredov: $p_a(C_1)=0.6$ in $p_a(C_2)=0.4$.

Obreži podano odločitveno drevo s postopkom minimizacije napake in vrednostjo $m = 10$.



NALOGA 7

Klasifikator je na 4-razrednem problemu dosegel naslednje rezultate, predstavljene z matriko zmot. V vsaki celici je vneseno ustrezno število_testnih_primerov/cena_napačne_klasifikacije:

Pravi razred	Napovedani razred			
	C_1	C_2	C_3	C_4
C_1	12/0	0/1	4/2	4/2
C_2	5/1	12/0	2/1	5/3
C_3	5/10	0/3	20/0	15/2
C_4	8/2	0/1	2/4	46/0

Izračunaj:

- klasifikacijsko točnost klasifikatorja,
- pričakovano točnost večinskega klasifikatorja (predpostavi, da je verjetnostna distribucija po razredih v testni množici enaka distribuciji v učni množici),
- povprečno ceno napačne klasifikacije.