

**Izrada bayesove mreže i prediktivnog modela  
za potrošnju energije uređaja u stanu**

# Sadržaj

1.	Uvod.....	1
2.	Bayesova mreža .....	2
3.	Analiza modela .....	4
3.1	Točnost modela.....	4
3.2	Konfuzijska matrica .....	5
4.	ROC krivulje .....	6
4.1	ROC krivulja - s1_below_50_00 (AUC = 0,829976).....	6
4.2	ROC krivulja - s2_50_00_60_00 (AUC = 0,726859).....	7
4.3	ROC krivulja - s3_60_00_70_00 (AUC = 0,732672).....	7
4.4	ROC krivulja - s4_70_00_110_00 (AUC = 0,808998).....	8
4.5	ROC krivulja - s5_110_00_up (AUC = 0,817166).....	9
4.6	ROC krivulje – zaključak .....	9
5.	Scenariji osjetljivosti.....	10
5.1	Scenarij 1 – Noć, niska temperatura dnevnog boravka, slabo korištenje svjetla ..	10
5.2	Scenarij 2 – Večer, upaljena svjetla, umjerna temperatura u dnevnom boravku..	11
5.3	Scenarij 3 – Poslijepodne, visoka vanjska vlažnost, niska temperatura dnevnog boravka	12
6.	Zaključak.....	14

## Popis slika

Slika 1 Prikaz bayesove mreže.....	2
Slika 2 Izgled dijela novo generiranog dokumenta "Validate_Appliances" .....	3
Slika 3 Konfuzijska matrica .....	5
Slika 4 ROC krivulja - s1_below_50_00 (AUC = 0,829976) .....	6
Slika 5 ROC krivulja - s2_50_00_60_00 (AUC = 0,726859).....	7
Slika 6 ROC krivulja - s3_60_00_70_00 (AUC = 0,732672).....	8
Slika 7 ROC krivulja - s4_70_00_110_00 (AUC = 0,808998).....	8
Slika 8 ROC krivulja - s5_110_00_up (AUC = 0,817166).....	9
Slika 9 Početno stanje mreže s target varijablom "Appliances" .....	10
Slika 10 Scenarij 1 – Noć, niska temperatura dnevnog boravka, slabo korištenje svjetla .....	11
Slika 11 Večer, upaljena svjetla, umjerna temperatura u dnevnom boravku .....	12
Slika 12 Poslijepodne, visoka vanjska vlažnost, niska temperatura dnevnog boravka .....	13

# 1. Uvod

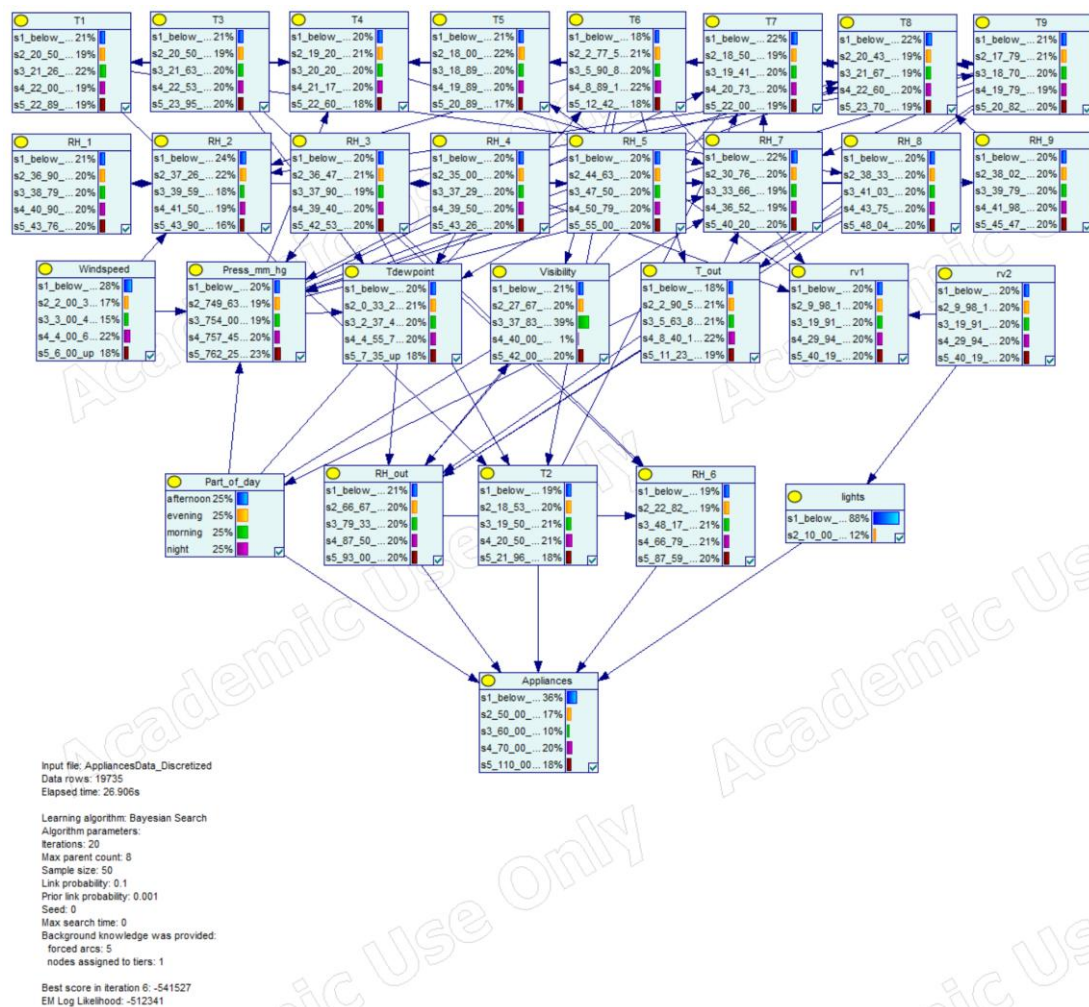
Bayesove mreže su probabilistički grafički modeli koji služe za prikazivanje i analizu uvjetnih ovisnosti između skupa varijabli. Sastoje se od čvorova, koji predstavljaju varijable, i usmjerenih veza, koje označavaju uzročno-posljedične odnose ili uvjetne zavisnosti između njih. Svaki čvor povezan je s tablicom uvjetnih vjerojatnosti (CPT) koja kvantificira vjerojatnost njegovih stanja s obzirom na stanja povezanih varijabli. U ovom projektu koristimo bayesovu mrežu kako bi shvatili kako su varijable povezane te kako utječu jedne na druge. Time je moguće napraviti kvalitetniju analizu podataka tj. donijeti neke korisne zaključke o promjenama koje pojedine varijable mogu uzrokovati. Pomoću tih zaključaka možemo odlučiti kako promijeniti ili prilagoditi situaciju na koju se varijable u pravom svijetu odnose.

Ciljna varijabla ovog projekta je varijabla „Appliances“ koja predstavlja potrošnju energije uređaja, a na ostatak varijabli (temperatura, vlažnost, vanjski vremenski uvjeti, svjetla...) gledamo kao na faktore koji mogu ili nemoraju imati utjecaj na ciljnu varijablu. Zadatak ovog projekta je saznati koje varijable jako utječu na potrošnju energije uređaja. Te varijable treba u potpunosti analizirati te opisati njihov utjecaj uz pomoć analize njihovog IV i WoE. Naposljetku, treba napraviti završnu analizu i prediktivni model izradom bayesove mreže s fokusom na varijable za koje je izračunato da imaju najveći IV prema ciljnoj varijabli.

Prilikom analize IV i WoE, varijable koje su se istaknule su: Part\_of\_day, RH\_out, T2, RH\_6 i lights. Varijabla Part\_of\_day nam govori kojem vremenskom periodu u danu podaci pripadaju, RH\_out nam daje vrijednosti vanjske vlažnosti, RH\_6 vrijednosti vanjske vlažnost sa sjeverne strane zgrade (u kojoj se mjerenja odvijaju), T2 podatke o unutrašnjoj temperaturi dnevnog boravka, a lights nam govori kolika se koriste unutrašnja svjetla.

## 2. Bayesova mreža

Nakon diskretizacije podataka u prethodnom projektnom zadatku i spremanjem tih diskretiziranih podataka u novu csv datoteku, učitani su u GeNIe program. Iz tih podataka naučena je bayesova mreža s fokusom na varijable s najvećim IV vrijednostima i njihovim utjecajem na ciljnu varijablu (background knowledge je opisan tako da je 5 najjačih varijabli po IV-u direktno povezano s ciljnom varijablom „Appliances“). Nakon što je završila generacija bayesove mreže naučeni su i parametri (također, varijable su u prikazu grupirane po kategorijama „temperatura“, „vlažnost“, „ostalo“ te najjače varijable po IV-u i naposljetku ciljna varijabla) te je mreža izvedena u xdsl obliku („AppliancesData\_Discretized\_Network.xdsl“).



Slika 1 Prikaz bayesove mreže

Nakon toga napravljena je validacija po ciljnoj varijabli „Appliances“ i generiran novi dokument Validate\_Appliances.txt s originalnim podacima (može se koristiti i na nekim

drugim podacima) i prediktivnim podacima o ciljnoj varijabli (stupac „Appliances\_predicted“).

Windspeed	Visiblity	Tidepoint	nr1	nr2	Part_of_day	Appliances_a1_below_50_00	Appliances_a2_50_00_60_00	Appliances_a3_60_00_70_00	Appliances_a4_70_00_110_00	Appliances_a5_110_00_up	Appliances_predicted
s5_6.00_up	s5_42.00_up	s4_4.55_7.35 s2_9.98_19.91	s2_9.98_19.91	afternoon	0.457143	0.457143	0.0285714	0.0285714	0.0285714	s1_below_50_00	
s5_6.00_up	s5_42.00_up	s4_4.55_7.35 s2_9.98_19.91	s2_9.98_19.91	afternoon	0.457143	0.457143	0.0285714	0.0285714	0.0285714	s1_below_50_00	
s5_6.00_up	s5_42.00_up	s4_4.55_7.35 s3_19.91_29.94	s3_19.91_29.94	afternoon	0.457143	0.457143	0.0285714	0.0285714	0.0285714	s1_below_50_00	
s4_4.00_6.00	s5_42.00_up	s4_4.55_7.35 s5_40.19_up	s5_40.19_up	afternoon	0.457143	0.457143	0.0285714	0.0285714	0.0285714	s1_below_50_00	
s4_4.00_6.00	s5_42.00_up	s4_4.55_7.35 s2_9.98_19.91	s2_9.98_19.91	afternoon	0.457143	0.457143	0.0285714	0.0285714	0.0285714	s1_below_50_00	
s4_4.00_6.00	s5_42.00_up	s4_4.55_7.35 s5_40.19_up	s5_40.19_up	afternoon	0.457143	0.457143	0.0285714	0.0285714	0.0285714	s1_below_50_00	
s4_4.00_6.00	s3_37.83_40.00	s4_4.55_7.35 s5_40.19_up	s5_40.19_up	evening	0.0571429	0.114286	0.0285714	0.314286	0.485714	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s4_4.55_7.35 s4_29.94_40.19	s4_29.94_40.19	evening	0.0571429	0.114286	0.0285714	0.314286	0.485714	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s4_4.55_7.35 s4_29.94_40.19	s4_29.94_40.19	evening	0.0571429	0.114286	0.0285714	0.314286	0.485714	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s4_4.55_7.35 s1_below_9.98	s1_below_9.98	evening	0.122222	0.0111111	0.122222	0.122222	0.622222	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s4_4.55_7.35 s2_9.98_19.91	s2_9.98_19.91	evening	0.122222	0.0111111	0.122222	0.122222	0.622222	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s4_4.55_7.35 s1_below_9.98	s1_below_9.98	evening	0.122222	0.0111111	0.122222	0.122222	0.622222	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s4_4.55_7.35 s4_29.94_40.19	s4_29.94_40.19	evening	0.0181818	0.109091	0.109091	0.2	0.563636	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s2_9.98_19.91	s2_9.98_19.91	evening	0.0181818	0.109091	0.109091	0.2	0.563636	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s4_29.94_40.19	s4_29.94_40.19	evening	0.575	0.075	0.0125	0.2	0.1375	s1_below_50_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s3_19.91_29.94	s3_19.91_29.94	evening	0.575	0.075	0.0125	0.2	0.1375	s1_below_50_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s4_29.94_40.19	s4_29.94_40.19	evening	0.575	0.075	0.0125	0.2	0.1375	s1_below_50_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s5_40.19_up	s5_40.19_up	evening	0.23913	0.26087	0.130435	0.217391	0.152174	s2_50_00_60_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s2_9.98_19.91	s2_9.98_19.91	evening	0.23913	0.26087	0.130435	0.217391	0.152174	s2_50_00_60_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s4_29.94_40.19	s4_29.94_40.19	evening	0.23913	0.26087	0.130435	0.217391	0.152174	s2_50_00_60_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s5_40.19_up	s5_40.19_up	evening	0.23913	0.26087	0.130435	0.217391	0.152174	s2_50_00_60_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s2_9.98_19.91	s2_9.98_19.91	evening	0.0175439	0.0701754	0.0701754	0.526316	0.315789	s4_70_00_110_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s4_29.94_40.19	s4_29.94_40.19	evening	0.0175439	0.0701754	0.0701754	0.526316	0.315789	s4_70_00_110_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s1_below_9.98	s1_below_9.98	evening	0.0175439	0.0701754	0.0701754	0.526316	0.315789	s4_70_00_110_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s2_9.98_19.91	s2_9.98_19.91	evening	0.0175439	0.0701754	0.0701754	0.526316	0.315789	s4_70_00_110_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s2_9.98_19.91	s2_9.98_19.91	evening	0.0175439	0.0701754	0.0701754	0.526316	0.315789	s4_70_00_110_00	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s1_below_9.98	s1_below_9.98	evening	0.0594595	0.167568	0.00540541	0.248649	0.518919	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s2_9.98_19.91	s2_9.98_19.91	evening	0.0594595	0.167568	0.00540541	0.248649	0.518919	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s5_40.19_up	s5_40.19_up	evening	0.0333333	0.166667	0.2	0.266667	0.333333	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s2_9.98_19.91	s2_9.98_19.91	evening	0.0333333	0.166667	0.2	0.266667	0.333333	s5_110_00_up	
s4_4.00_6.00	s3_37.83_40.00	s3_2.37_4.55 s3_19.91_29.94	s3_19.91_29.94	evening	0.0333333	0.166667	0.2	0.266667	0.333333	s5_110_00_up	
s5_6.00_up	s3_37.83_40.00	s3_2.37_4.55 s4_29.94_40.19	s4_29.94_40.19	evening	0.0333333	0.166667	0.2	0.266667	0.333333	s5_110_00_up	
s5_6.00_up	s3_37.83_40.00	s3_2.37_4.55 s3_19.91_29.94	s3_19.91_29.94	evening	0.0333333	0.166667	0.2	0.266667	0.333333	s5_110_00_up	
s5_6.00_up	s3_37.83_40.00	s3_2.37_4.55 s1_below_9.98	s1_below_9.98	evening	0.0594595	0.167568	0.00540541	0.248649	0.518919	s5_110_00_up	
s5_6.00_up	s3_37.83_40.00	s3_2.37_4.55 s5_40.19_up	s5_40.19_up	evening	0.0594595	0.167568	0.00540541	0.248649	0.518919	s5_110_00_up	
s5_6.00_up	s3_37.83_40.00	s3_2.37_4.55 s4_29.94_40.19	s4_29.94_40.19	evening	0.0594595	0.167568	0.00540541	0.248649	0.518919	s5_110_00_up	
s5_6.00_up	s3_37.83_40.00	s3_2.37_4.55 s4_29.94_40.19	s4_29.94_40.19	evening	0.0594595	0.167568	0.00540541	0.248649	0.518919	s5_110_00_up	
s5_6.00_up	s2_27.67_37.83	s3_2.37_4.55 s1_below_9.98	s1_below_9.98	evening	0.0333333	0.166667	0.2	0.266667	0.333333	s5_110_00_up	
s5_6.00_up	s2_27.67_37.83	s3_2.37_4.55 s4_29.94_40.19	s4_29.94_40.19	evening	0.46875	0.125	0.125	0.21875	0.0625	s1_below_50_00	

Slika 2 Izgled dijela novo generiranog dokumenta "Validate\_Appliances"

### 3. Analiza modela

Model je validiran na temelju ciljne varijable „Appliances“. Rezultati su dobiveni na temelju diskretizacije svih varijabli u 5 binova (kod onih kod kojih je to moguće i tako da su što ravnomjernije raspoređeni) i varijable „date“ u 4 kategorije (dijelovi dana „morning“, „afternoon“, „evening“ i „night“). Pokušano je razviti model diskretizacijom varijabli na više binova zbog moguće bolje preciznosti, ali program GeNIe se u tim okolnostima zamrznuo i ugasio pa je za broj binova odabran broj 5 kako bi GeNIe mogao uspješno napraviti mrežu.

#### 3.1 Točnost modela

Ukupnu točnost modela možemo definirati kao postotak primjera koje je model točno klasificirao, a u slučaju ovog modela ona iznosi 0.520851 tj. oko 52%. To je napredak u odnosu na nasumično pogađanje, ali jako minimalan jer se nasumičnim pogađanjem definira točnost od 50%. Ovaj podatak nam govori da, što se tiče točnosti, model donosi napredak od samo 2%, ali ukupna točnost nije jedina karakteristika koja je bitna pri analizi performansi modela (detaljna analiza kategorizacija te njima prouzročenih slučajeva i ROC krivulja nam mogu donijeti jako korisne funkcije, možda i korisnije od točnosti). Pogledamo li točnost po segmentima varijable „Appliances“ možemo uočiti sljedeće:

- $s1\_below\_50\_00 = 0.815063$  (81,5%, 6.082/7.462)
- $s2\_50\_00\_60\_00 = 0.143205$  (14,3%, 470/3.282)
- $s3\_60\_00\_70\_00 = 0.0461538$  (4,6%, 72/1.560)
- $s4\_70\_00\_110\_00 = 0.590747$  (59,1%, 2.324/3.934)
- $s5\_110\_00\_up = 0.380612$  (38,1%, 1.331/3.497)

U slučajevima gdje je potrošnja manja od 50, model ima iznimno dobru sposobnost predikcije od 81.5%. Nadalje, kod potrošnje između 70 i 110 ta sposobnost se smanjuje, ali je također respektabilna (59.1%). Točnost se nastavlja smanjivati kada potrošnja raste iznad brojke od 110 te u tim slučajevima iznosi 38.1%. Možemo uočiti kako u ovim slučajevima model već ima velike poteškoće pri predviđanju ishoda. Kategorije od 50 do 60 i 60 do 70 dodatno otežavaju situaciju modelu, a to se može i lagano vidjeti iz podataka od točnosti (14.3% i 4.6%). Zaključak je da nešto u diskretizaciji ili prirodi varijabli (podataka) značajno olakšava predviđanje modelu u određenim slučajevima, a u drugim slučajevima mu taj zadatak značajno otežava te nije u stanju određivati varijablu s velikom točnošću. Ovo bih se moglo detaljnije proučiti (možda čak i unaprijediti) drugačijom diskretizacijom varijabli, ali zbog već prije navedenih limitacija programa GeNIe, to nije bilo moguće.

### 3.2 Konfuzijska matrica

Konfuzijska matrica nam otkriva ono što smo već uočili pri analizi segmenata varijable „Appliances“. Redovi označavaju stvarne vrijednosti, a stupci predviđene vrijednosti tj. kategorije koje je model predvidio. Svaki element u matrici predstavlja broj primjera koji su stvarno pripadali određenom segmentu (red) i koji su od strane modela klasificirani u određeni segment (stupac). Dijagonala nam otkriva točne predikcije, primjere koji su stvarno pripadali određenom segmentu i koje je model točno klasificirao u taj segment. Ovi brojevi su ključni za izračun točnosti po segmentima. Vrijednost s dijagonale podijeljena s ukupnim brojem stvarnih vrijednosti za taj stupac, govori nam kolika je točnost modela za tu kategoriju. Ukupna točnost izračunata je zbrojem svih vrijednosti s dijagonale i dijeljenjem te vrijednosti s ukupnim zbrojem svih vrijednosti iz konfuzijske matrice.

		Predicted				
		s1_below_5...	s2_50_00_...	s3_60_00_...	s4_70_00_...	s5_110_00...
Actual	s1_below_50_00	<b>6082</b>	260	13	800	307
	s2_50_00_60_00	1911	<b>470</b>	24	631	246
	s3_60_00_70_00	629	145	<b>72</b>	511	203
	s4_70_00_110_00	805	142	34	<b>2324</b>	629
	s5_110_00_up	805	98	14	1249	<b>1331</b>

Slika 3 Konfuzijska matrica

Po konfuzijskoj matrici možemo ponovo donijeti iste zaključke. Model pokazuje dobru točnost u segmentu s1 (81,5%) i umjereno dobru u s4 (59,1%). Ovo ukazuje na to da model dobro radi u nekim područjima podataka. Model ima nisku točnost u s5 (38,1%) te vrlo nisku točnost u s2 (14,3%) i s3 (4,6%). Dodatnom analizom možemo uočiti da je velik broj pogrešnih klasifikacija, npr. 1.911 primjera iz s2 klasificirano u s1, sugerira da model često miješa ove segmente. Ovo može biti uzrokovano sličnostima u podacima ili nedostatkom dovoljno različitih značajki između segmenata. Sljedeći korak je analiza ROC krivulja kako bi vidjeli je li model u tom smislu koristan.

## 4. ROC krivulje

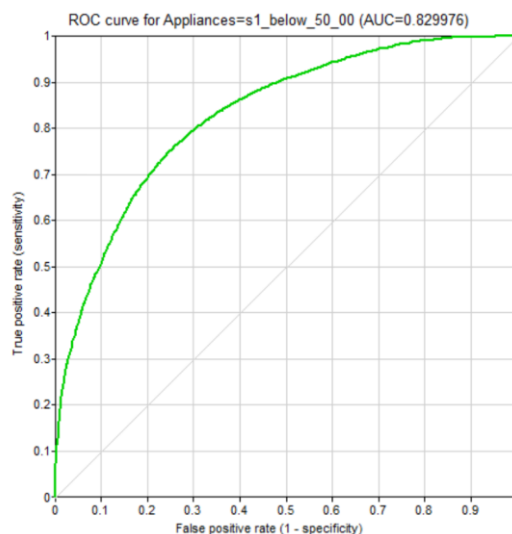
ROC (Receiver Operating Characteristic) krivulje prikazuju performanse modela za svaki segment (s1\_below\_50\_00, s2\_50\_00\_60\_00, s3\_60\_00\_70\_00, s4\_70\_00\_110\_00, s5\_110\_00\_up) u kontekstu ciljne varijable "Appliances". Svaka krivulja dolazi s pridruženim AUC (Area Under the Curve) vrijednostima, koje mjere diskriminativnu moć modela za razlikovanje između klasa (npr. pozitivnih i negativnih ishoda). Krivulja bliža lijevom gornjem kutu (1,1) označava bolji model, dok krivulja blizu dijagonalne linije ( $y = x$ ) ukazuje na performanse poput nasumičnog pogađanja ( $AUC \approx 0,5$ ). AUC (Area Under the Curve) predstavlja broj između 0 i 1 koji kvantificira ukupnu sposobnost modela za razlikovanje klasa:

- $AUC \approx 0,5$ : Nasumično pogađanje (slab model).
- $AUC > 0,7$ : Pristojna diskriminativna moć.
- $AUC > 0,8$ : Dobra diskriminativna moć.
- $AUC > 0,9$ : Izvrsna diskriminativna moć.

Slijedi analiza ROC krivulje svake kategorije ciljne varijable.

### 4.1 ROC krivulja - s1\_below\_50\_00 ( $AUC = 0,829976$ )

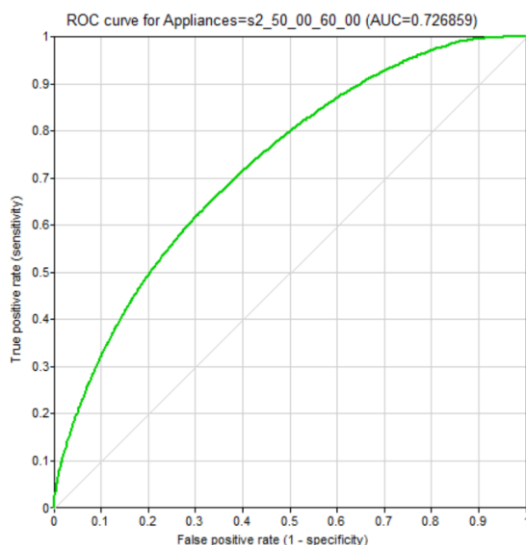
Krivulja je izražena i nalazi se znatno iznad dijagonalne linije, što ukazuje na dobru diskriminativnu moć. AUC od 0,83 sugerira da model odlično razlikuje klase u ovom segmentu, bolje od svih drugih i puno bolje od nasumičnog pogađanja. Model je vrlo učinkovit u razdvajanju pozitivnih i negativnih ishoda u ovoj kategoriji, što može značiti da podaci u s1 imaju jasne i razlikovne značajke.



Slika 4 ROC krivulja - s1\_below\_50\_00 ( $AUC = 0,829976$ )

## 4.2 ROC krivulja - s2\_50\_00\_60\_00 (AUC = 0,726859)

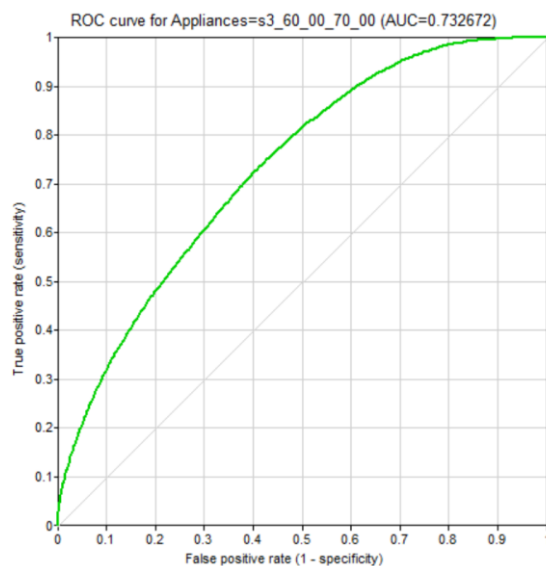
Krivulja je izražena i nalazi se znatno iznad dijagonalne linije, što ukazuje na pristojnu diskriminativnu moć. AUC od 0,73 sugerira da model solidno razlikuje klase u ovom segmentu, bolje od nasumičnog pogađanja, ali ne tako dobro kao u nekim drugim segmentima. Model je umjereno učinkovit u razdvajanju pozitivnih i negativnih ishoda u ovoj kategoriji, što može značiti da podaci u s2 imaju neke razlikovne značajke, ali možda i određena preklapanja s drugim segmentima.



Slika 5 ROC krivulja - s2\_50\_00\_60\_00 (AUC = 0,726859)

## 4.3 ROC krivulja - s3\_60\_00\_70\_00 (AUC = 0,732672)

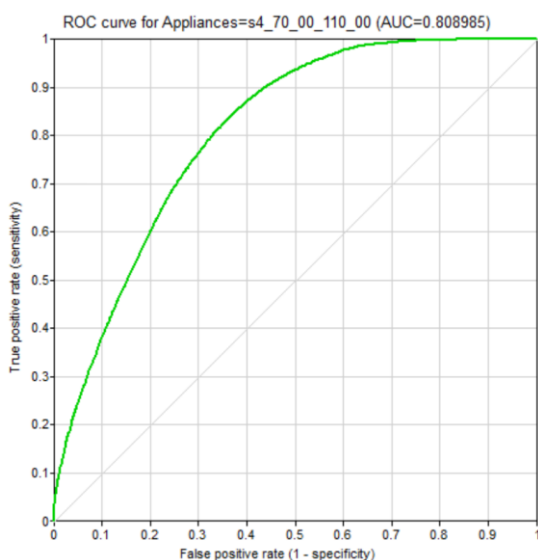
Slično kao i krivulja kategorije s2\_50\_00\_60\_00, ova krivulja je izražena i nalazi se znatno iznad dijagonalne linije, ali ne toliko visoko kao kod kategorija s najjačom diskriminativnom moći. Ovo nam govori da je u ovoj kategoriji diskriminativna moć pristojna. AUC od 0,73 sugerira da model solidno razlikuje klase u ovom segmentu, dosta bolje od nasumičnog pogađanja. Model je umjereno učinkovit u razdvajanju pozitivnih i negativnih ishoda u ovoj kategoriji, što može značiti da podaci u s3 imaju određene razlikovne značajke, ali možda i neuravnotežene ili prirodno nezgodne podatke.



Slika 6 ROC krivulja - s3\_60\_00\_70\_00 (AUC = 0,732672)

#### 4.4 ROC krivulja - s4\_70\_00\_110\_00 (AUC = 0,808998)

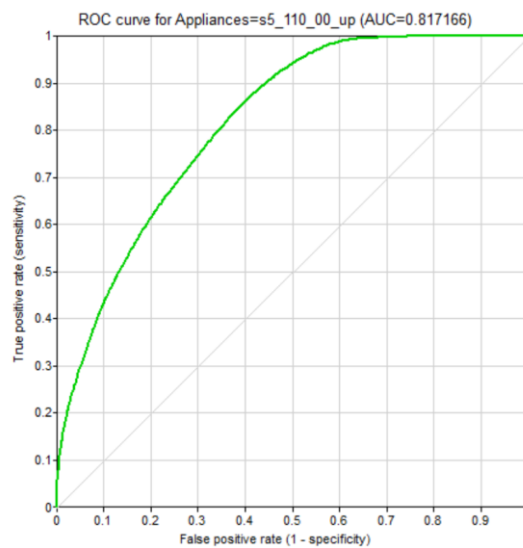
Krivulja je izražena i nalazi se znatno iznad dijagonalne linije, što ukazuje na dobru diskriminativnu moć. AUC od 0,81 sugerira da model odlično razlikuje klase u ovom segmentu, bolje od većine drugih kategorija i puno bolje od nasumičnog pogađanja. Model je vrlo učinkovit u razdvajanju pozitivnih i negativnih ishoda u ovoj kategoriji, što može značiti da podaci u s4 imaju jasne i razlikovne značajke.



Slika 7 ROC krivulja - s4\_70\_00\_110\_00 (AUC = 0,808998)

#### 4.5 ROC krivulja - s5\_110\_00\_up (AUC = 0,817166)

Krivulja je izražena i nalazi se znatno iznad dijagonalne linije, što ukazuje na dobru diskriminativnu moć. AUC od 0,82 sugerira da model slično kao i kod kategorije s4, odlično razlikuje klase u ovom segment. Ovo je kategorija s drugom najvećom AUC vrijednosti te u njoj model ima puno manje problema što se tiče razlikovanja od ostalih kategorija. Vrlo je učinkovit u razdvajanju pozitivnih i negativnih ishoda u ovoj kategoriji, što može značiti da podaci u s5 imaju vrlo jasne i pogodne podatke.



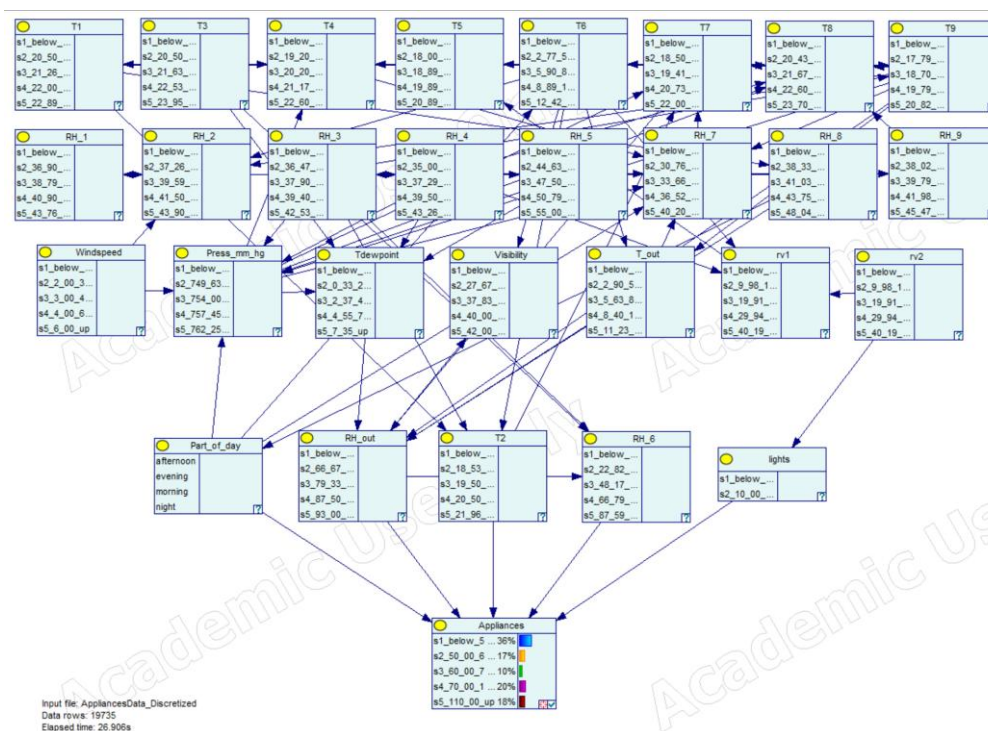
Slika 8 ROC krivulja - s5\_110\_00\_up (AUC = 0,817166)

#### 4.6 ROC krivulje – zaključak

Model pokazuje pristojnu do dobru diskriminativnu moć preko svih segmenata (s AUC-ovima između 0,73 i 0,83). Diskriminativna moć modela odnosi se na njegovu sposobnost da razlikuje ili "diskriminira" između različitih kategorija u klasifikacijskom problemu, čak i ako točnost varira. Visoka diskriminativna moć znači da model može jako dobro identificirati koja instanca pripada kojoj kategoriji, bez obzira na to jesu li konačne predikcije uvijek točne. Ova karakteristika ga čini korisnim alatom za procjenu vjerojatnosti i rangiranje. Model je dobar za primjene gdje je ključna diskriminacija, ali zahtijeva daljnju optimizaciju kako bi se poboljšale točne predikcije u slabijim segmentima. Ova analiza potvrđuje ranije zaključke o modelu i pruža jasnu putanju za buduća poboljšanja.

## 5. Scenariji osjetljivosti

U ovom poglavlju radit će se i objasniti tri scenarija osjetljivosti. Oni su napravljeni kako bi saznali kako se ciljna varijabla mijenja kada se mijenjaju vjerojatnosti kategorija varijabla koje su ključni prediktori za ciljnu varijablu, analizom tih scenarija možemo napraviti korisne zaključke. Svaki scenarij imat će sliku koja prikazuje rezultate i biti će objašnjeno kako su ključni prediktori tj. varijable postavljeni/prilagođeni. Da bi mogli usporediti stanje napravljenih scenarija, moramo prvo znati kako izgleda početna situacija te je iz toga razloga priložena slika početnog stanja mreže (Slika 9).

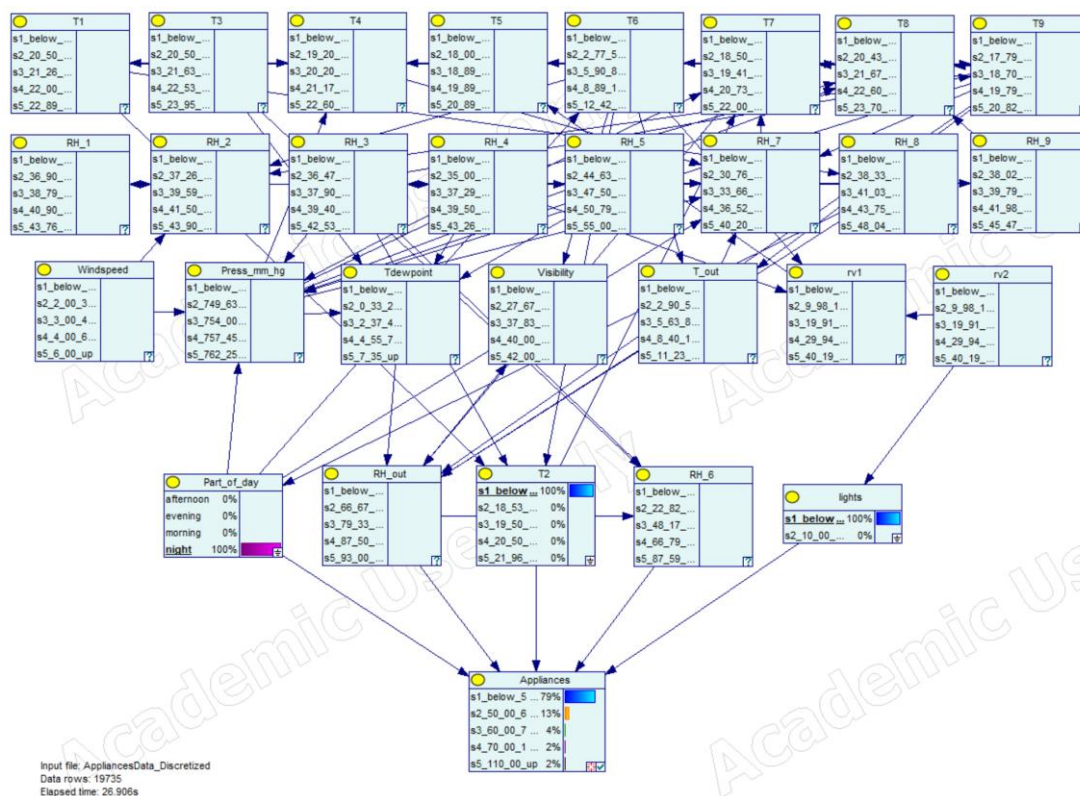


Slika 9 Početno stanje mreže s target varijablom "Appliances"

### 5.1 Scenarij 1 – Noć, niska temperatura dnevnog boravka, slabo korištenje svjetla

Prvi scenarij proučava potrošnju energije dok je temperatura dnevnog boravka niska, svjetla se slabo koriste, a vremenski period dana je noć. Ove promjene na ključnim prediktorima značajno utječu na ciljnu varijablu „Appliances“. Sa slike (Slika 10) možemo isčitati da su sve kategorije ciljne varijable, osim one koja označuje slabu potrošnju energije, značajno pale što se tiče njihove vjerojatnosti. Uzrok ovih promjena je postavljanje varijabli: Part\_of\_day na 100% noć, T2 na ispod 18.5 celzijevih stupnjeva u 100% slučajeva i lights na ispod 10 Wh u 100% slučajeva. U takvom scenariju u čak 79% slučajeva, potrošnja energije je mala. Možemo zaključiti da se u promatranom prostoru po noći uređaji u većini slučajeva manje koriste. Niska

temperatura dnevnog boravka iz dodatno pomaže slučaju manje potrošnje (možemo pretpostaviti da grijanje/hlađenje po noći ne radi te je potrošnja automatski manja). Nadalje, slabo korištenje svjetla ili njihovo potpuno nekorištenje dodatno osnažuju slučaj manje potrošnje. Možemo pretpostaviti da ukućani spavaju ili nisu kod kuće te se energija iz tog razlog ne troši kao inače. Sve u svemu, zaključak i korisna informacija je da se u ovakvim situacijama energija puno slabije koristi te to možemo koristiti za buduću optimizaciju potrošnje ili analizu navika ukućana.

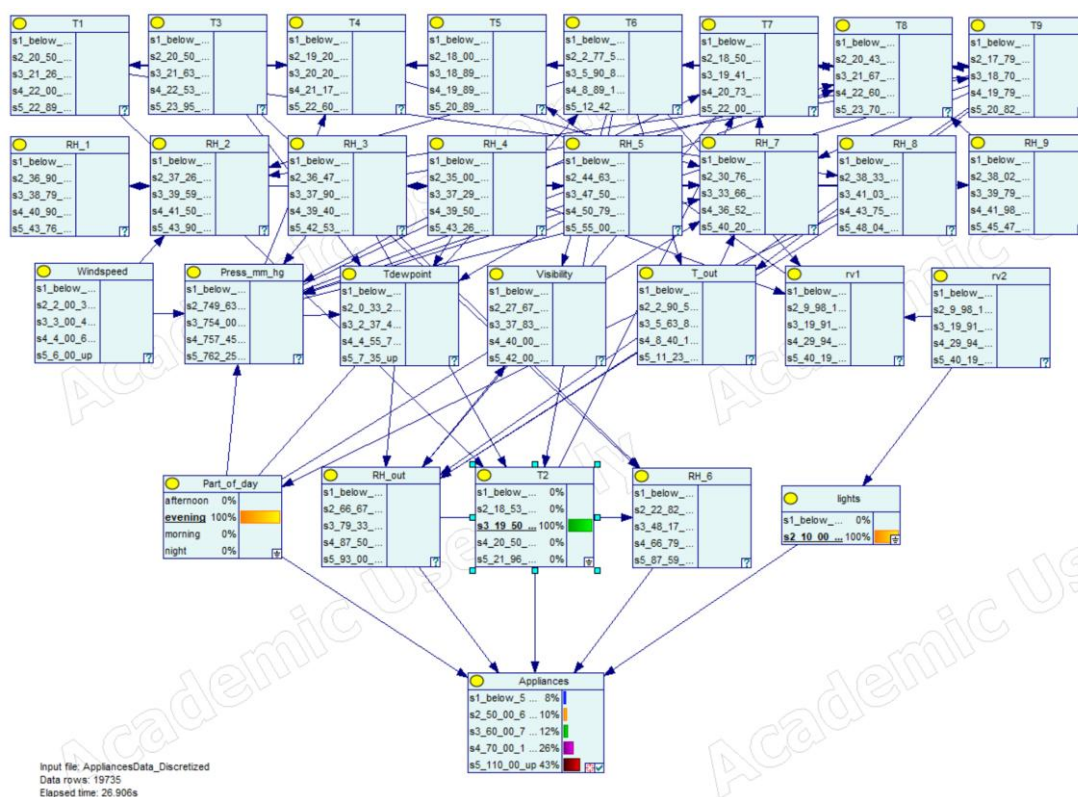


Slika 10 Scenarij 1 – Noć, niska temperatura dnevnog boravka, slabo korištenje svjetla

## 5.2 Scenarij 2 – Večer, upaljena svjetla, umjerna temperatura u dnevnom boravku

Drugi scenarij proučava potrošnju energije u uvjetima gdje je vremenski period dana večer, svjetla su upaljena, a temperatura u dnevnom boravku je umjerena. Ove promjene ključnih prediktora imaju značajan utjecaj na ciljnu varijablu „Appliances“. Na osnovu slike (Slika 11), možemo isčitati da su vjerojatnosti različitih kategorija ciljne varijable distribuirane na sljedeći način: velika potrošnja energije (kategorije s iznad 70 Wh) je predviđena u čak 69% slučajeva, dok ostale kategorije (manja potrošnja) postižu značajno niži postotak vjerojatnosti (odnosno

31% sve zajedno). Uzrok ovih promjena leži u postavljanju varijabli: Part\_of\_day na 100% večer, lights na iznad 10 Wh u 100% slučajeva (svjetla su upaljena), te T2 na između 19.5 i 20.5 celzijevih stupnjeva u 100% slučajeva (umjerena temperatura u dnevnom boravku). U ovom scenariju, najvjerojatniji slučaj potrošnje energije je da je ona visoka, što ukazuje na to da se u tom periodu koristi velika količina uređaja. Možemo zaključiti da se u večernjim satima, kada su svjetla upaljena i temperatura u dnevnom boravku umjerena, ukućani vjerojatno nalaze u prostoru i koriste puno kućanskih uređaja (npr. televiziju, grijanje/hlađenje, punjače...). Sve u svemu, ovaj scenarij sugerira tipično večernje ponašanje ukućana, gdje se energija malo intenzivnije koristi, što može poslužiti kao korisna informacija za analizu navika potrošnje energije i potencijalnu optimizaciju te potrošnje u budućnosti.

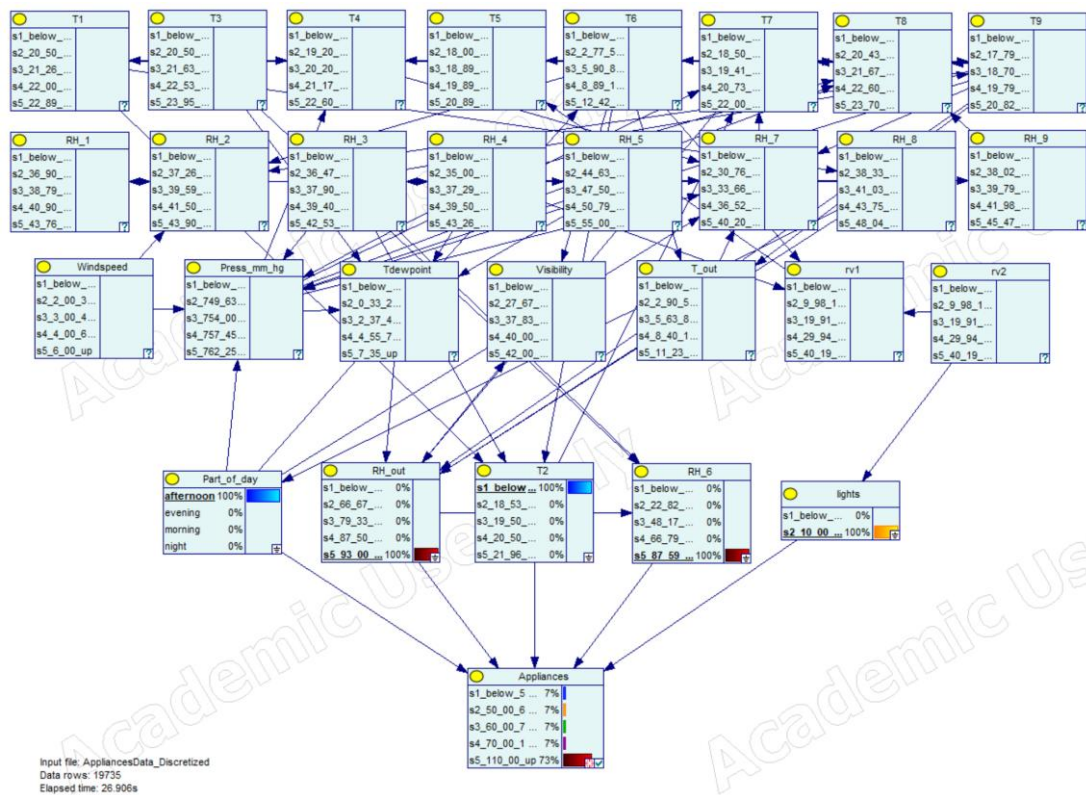


Slika 11 Večer, upaljena svjetla, umjerena temperatura u dnevnom boravku

### 5.3 Scenarij 3 – Poslijepodne, visoka vanjska vlažnost, niska temperatura dnevnog boravka

Treći scenarij analizira potrošnju energije u uvjetima gdje je vremenski period dana poslijepodne, vanjska vlažnost je visoka, a temperatura u dnevnom boravku je niska. Ove promjene ključnih prediktora značajno utječu na ciljnu varijablu „Appliances“. Na osnovu slike (Slika 12), možemo isčitati da su vjerojatnosti različitih kategorija ciljne varijable

distribuirane na sljedeći način: velika potrošnja energije (iznad 110 Wh) ima daleko najveću vjerojatnost od 73%, a ostatak kategorija ima vjerojatnost od po samo 7% (ukupno 28%). Uzrok ovih promjena leži u postavljanju varijabli: Part\_of\_day na 100% poslijepodne, RH\_out na iznad 93% u 100% slučajeva (visoka vanjska vlažnost), RH\_6 na iznad 87.6% u 100% slučajeva (visoka vlažnost s vanjske sjeverne strane zgrade) te T2 na ispod 18.5 celzijevih stupnjeva u 100% slučajeva (niska temperatura u dnevnom boravku). U ovom scenariju, najvjerojatnija je visoka potrošnja energije, što ukazuje na intenzivnu upotrebu uređaja u tom periodu. Možemo zaključiti da u poslijepodnevni satima, kada je vanjska vlažnost visoka i temperatura u dnevnom boravku niska, ukućani više koriste kućanske uređaje. Niska temperatura u dnevnom boravku može sugerirati da je grijanje/hlađenje tek uključeno (trenutak kada ukućani dođu s posla i pale grijanje ili hlađenje), što automatski povećava potrošnju energije. Visoka vanjska vlažnost može dodatno utjecati na ponašanje ukućana, možda ih potičući na češću upotrebu sustava za grijanje ili hlađenje kako bi se osigurali ugodni uvjeti u unutrašnjem prostoru, posebno ako je zgrada osjetljiva na vanjske klimatske uvjete poput visoke vlage s sjeverne strane. Sve u svemu, ovaj scenarij ukazuje na visoku potrošnju energije, što može poslužiti kao važna informacija za razumijevanje navika ukućana i planiranje mjera za smanjenje energetske potrošnje (ako je to cilj) u sličnim uvjetima.



Slika 12 Poslijepodne, visoka vanjska vlažnost, niska temperatura dnevnog boravka

## 6. Zaključak

Ovaj projekt demonstrirao je primjenu bayesovih mreža za analizu potrošnje energije uređaja (ciljna varijabla „Appliances“) u odnosu na ključne prediktorske varijable poput vremenskog perioda dana (Part\_of\_day), vanjske vlažnosti (RH\_out, RH\_6), temperature u dnevnom boravku (T2) te korištenja svjetla (lights). Bayesova mreža, razvijena u programu GeNIe uz diskretizaciju podataka u pet binova, omogućila je vizualizaciju i promatranje uvjetnih ovisnosti između varijabli, pružajući korisne uvide u faktore koji utječu na potrošnju energije. Iako je ukupna točnost modela relativno skromna (52%), što predstavlja minimalan napredak u odnosu na nasumično pogađanje, detaljna analiza segmenta ciljne varijable, konfuzijske matrice i ROC krivulja otkrila je značajne koristi i razlike u performansama modela preko različitih kategorija potrošnje energije. Model pokazuje dobru diskriminativnu moć (AUC iznad 0.8) u segmentima s niskom (s1\_below\_50\_00) i visokom (s4\_70\_00\_110\_00, s5\_110\_00\_up) potrošnjom, a malo slabiju (AUC iznad 0.7), ali pristojnu diskriminativnu moć u srednjim kategorijama potrošnje (s2\_50\_00\_60\_00, s3\_60\_00\_70\_00). Ovi nalazi potvrđuju da model ima potencijal za praktičnu primjenu, posebno u predviđanju ekstremnih vrijednosti potrošnje, ali zahtijeva daljnje unapređenje, možda kroz drugačiju diskretizaciju ili dodatnu obuku na većem i raznolikijem skupu podataka. Scenariji osjetljivosti dodatno su istaknuli praktičnu korisnost modela, pokazujući kako se potrošnja energije mijenja u različitim uvjetima. Na primjer, noćni period s niskom temperaturom i slabim korištenjem svjetla u većini slučajeva rezultira malom potrošnjom energije (79%), sugerirajući neprisutnost ili spavanje ukućana. Nasuprot tome, večernji period s upaljenim svjetlima i umjerenom temperaturom te poslijepodne s visokom vanjskom vlažnošću i niskom temperaturom u dnevnom boravku pokazuju visoku (iznad 70Wh) potrošnju energije (preciznije 69% i 80%). Ovi uvidi omogućuju bolje razumijevanje navika ukućana i pružaju temelj za optimizaciju energetske učinkovitosti, poput prilagodbe sustava grijanja, hlađenja ili osvjetljenja u različitim dijelovima dana. Zaključno, ovaj projekt pokazuje kako bayesove mreže mogu biti moćan alat za analizu kompleksnih međuovisnosti u podacima. Buduća istraživanja mogu se usmjeriti na alternativnu diskretizaciju, proširenje podataka i unapređenje modela kako bi se povećala točnost predviđanja (posebno u izazovnijim segmentima) te omogućila još preciznija optimizacija potrošnje energije u stvarnom svijetu.