**TEHNIČKO VELEUČILIŠTE U ZAGREBU**

**POLITEHNIČKI SPECIJALISTIČKI DIPLOMSKI STRUČNI STUDIJ**

SPECIJALIZACIJA INFORMATIKA

A blue circle with black text

AI-generated content may be incorrect.

Grgo Lovrić

**Bayesova mreža naučena**

**na water\_datasetu**

Kolegij: NOSQL i napredna big data analitika

Zagreb, Veljača 2025.

**Karakteristike modela Bayesove mreže**

Bayesove mreže su grafički probabilistički model koji koristi uvjetne vjerojatnosti ulaznih varijabli u svrhu predviđanja izlazne varijable. Uvjetne vjerojatnosti pohranjene su u usmjerenom acikličkom grafu u kojem čvorovi predstavljaju ulazne varijablu i veze koje predstavljaju uvjetne vjerojatnosti među varijablama. U slučaju da je neka ulazna varijabla uočen događaj (evidence), mreža koristeći Bayesov teorem ažurira vjerojatnosti drugih varijabli u grafu, a tako i ciljne varijable. Model podržava rad sa diskretnim i kontinuiranim podacima te različite vrste zadataka kao što su klasifikacija, regresija, pronalazak ekstremnih vrijednosti i sl.

**Analiza osjetljivosti**

Ažurirajući vjerojatnosti ulaznih varijabli s najvećim IV dobivenim tokom pretprocesiranja podatka *Dissolved oxygen (Maximum)*, *Specific conductance (Maximum)*, *pH, standard units (Minimum)* i *Specific conductance (Minimum)* može se uočiti najveći pomak u izlaznim vjerojatnostima varijable *Target*, što je očekivano. Rezultati *Target* varijable u navedenim slučajevima je najčešće 90-10. Mreža je također naučila dvije nove direktne veze s Target varijablom preko ulaznih varijabli *Dissolved oxygen (Mean)* i *Temperature (Minimum)*. Ažurirajući vrijednosti ovih varijabli vjerojatnosti izlaza nisu usmjerene previše u korist željene *Target* varijable, s prosječnim rezultatom 50-50. Sveukupno *Target* varijablu direktno uvjetuju 6 ulaznih varijabli, od kojih su 4 forsirane zbog velikog IV-a, a 2 su naučene algoritmom prepoznaje veze.

A diagram of a network

AI-generated content may be incorrect.

Slika Grafički prikaz naučenog modela

**Značajne karakteristike naučenog modela**

Mreža naučena na ovom datasetu stvorila je graf takav da se skoro sve ulazne varijable koriste u izračunu uvjetnih vjerojatnosti osim varijable *pH, standard units (Maximum)* koja je izolirana. Također varijabla *Temperature (Maximum)* nije korištena u izračunu Target varijable, no direktno je uvjetno povezana varijablom *Temperature (Mean)*. Varijabla *Specific conductance (Minimum)* jedina je koja posjeduje samo izlazne veze, što označava da nju niti jedna druga varijable ne uvjetuje. Ostale varijable su međusobno isprepletene, imaju više ulaznih i izlaznih veza.

**Skoriranje cjelokupnog seta podataka**

Korištenjem cijelog podatkovnog seta za validiranje naučenog dobiju se slijedeći rezultati za *Target* varijablu:

A close up of numbers

AI-generated content may be incorrect.

Slika Točnost modela

Model puno bolje predviđa klasu 0 nego klasu 1, što sveukupno rezultira točnošću od cca. 62%.

A close-up of a number

AI-generated content may be incorrect.

Slika Konfuzijska matrica

Također se vidi kroz konfuzijsku matricu da je model puno skloniji predviđanju klase 0.

A graph with green line

AI-generated content may be incorrect.

Slika ROC krivulja

ROC krivulja pokazuje omjer True positive i False positive rezultata za klasu 0. U idealnom slučaju bi krivulja bila točka u gornjem lijevom uglu gdje je True positive rate 1, a False positive rate 0, no u ovom slučaju je nešto iznad linije što označava da model daje bolju predikciju od nasumičnog pogađanja, a goru od idealnog slučaja.