

Izveštaj o Analizi Podataka i Modeliranju za Klasifikaciju Pečuraka

1. Pregled Skupa Podataka

Skup podataka sadrži informacije o 8124 uzorka pečuraka, sa 23 kolone koje predstavljaju različite karakteristike, uključujući klasu (jestiva ili otrovna).

- Sve kolone su nominalne vrednosti, što zahteva enkodiranje pre modeliranja.
- Kolona *veil-type* je uklonjena, jer sadrži samo jednu vrednost za sve uzorke.

Ključne informacije:

- Podaci su kompletni, bez nedostajućih vrednosti.
- Ciljna varijabla (*class*) označava jestive (0) i otrovne (1) pečurke.

2. Pretprocesiranje Podataka

- **Enkodiranje:** Sve kategorijske vrednosti su konvertovane u numeričke pomoću **LabelEncoder** za obradu od strane modela.
- **Vizualizacija:** Distribucije vrednosti za svaku kolonu su prikazane pomoću pie grafika.
 - Neke karakteristike poput *odor* pokazuju značajne razlike između vrednosti, što može ukazivati na značaj za klasifikaciju.

3. Podela na Trening i Test Skupove

- Podaci su podeljeni u trening (80%) i test (20%) skupove kako bi se procenila performansa modela.

4. Modeliranje: Decision Tree Classifier

- **Hiperparametarska Optimizacija:**

Pomoću **GridSearchCV**, ispitani su sledeći parametri:

- Kriterijumi: *gini* i *entropy*
- Maksimalna dubina stabla: [3, 5, 7, 10, None]
- Minimalni broj uzoraka za podelu: [2, 10, 20]
- Minimalni broj uzoraka u listu: [1, 5, 10]

Najbolji Parametri:

- Kriterijum: *gini*
- Maksimalna dubina: Nema ograničenja

- Minimalni broj uzoraka za podelu: 2
- Minimalni broj uzoraka u listu: 1

5. Evaluacija Modela

- Na test podacima, model je postigao savršenu tačnost: **100%**.
- **Matrica konfuzije:**

	Predviđeno 0	Predviđeno 1
Stvarno 0	843	0
Stvarno 1	0	782

- **Izveštaj klasifikacije:**
 - Preciznost, odziv i F1-skore za obe klase su 1.00.

6. Kros Validacija

- Model je testiran putem **5-fold Cross-Validation**, gde je prosečna tačnost iznosila **91.43%**.
 - Individualni rezultati pokazuju visoku tačnost u većini podela, ali su varijacije moguće zbog neravnomerne distribucije podataka.

7. Vizualizacija Odlučujućeg Stabla

- Vizualizacija stabla pokazuje ključne karakteristike koje model koristi za donošenje odluka.
 - Na primer, *gill-color* se pojavljuje kao jedna od prvih odluka u stablu, što ukazuje na njegov značaj za klasifikaciju.

8. Ključni Uvidi i Zapažanja

- Model je izuzetno efikasan u klasifikaciji jestivih i otrovnih pečuraka, što potvrđuje 100% tačnost na test podacima.
- Razlike u kros validaciji sugerišu mogućnost daljeg istraživanja, poput balansiranja podataka ili primene složenijih modela.

9. Zaključak

Analiza je uspešno pokazala kako Decision Tree Classifier može tačno razlikovati jestive od otrovnih pečuraka. Model pruža visoku tačnost i jasan uvid u značaj karakteristika.

