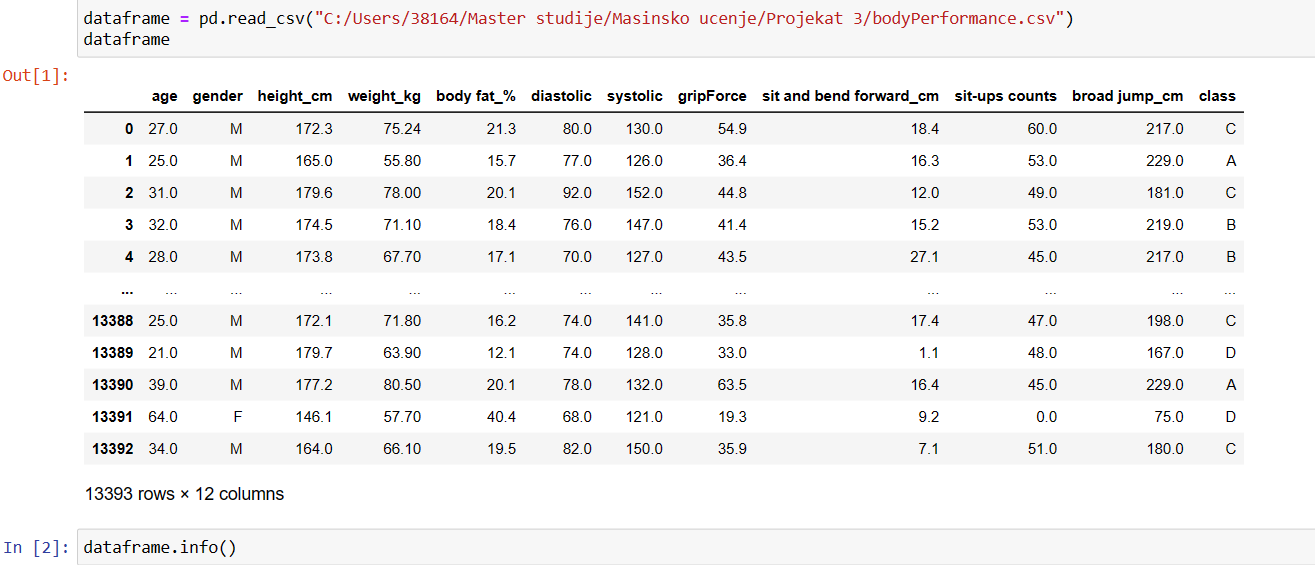
O datasetu:

Dataset sadrzi atribute na osnovu kojih treba predvideti nivo fitnesa opservacija koji moze biti A, B, C ili D (A je najbolji, a D je najgori nivo fitnesa). Atributi su:

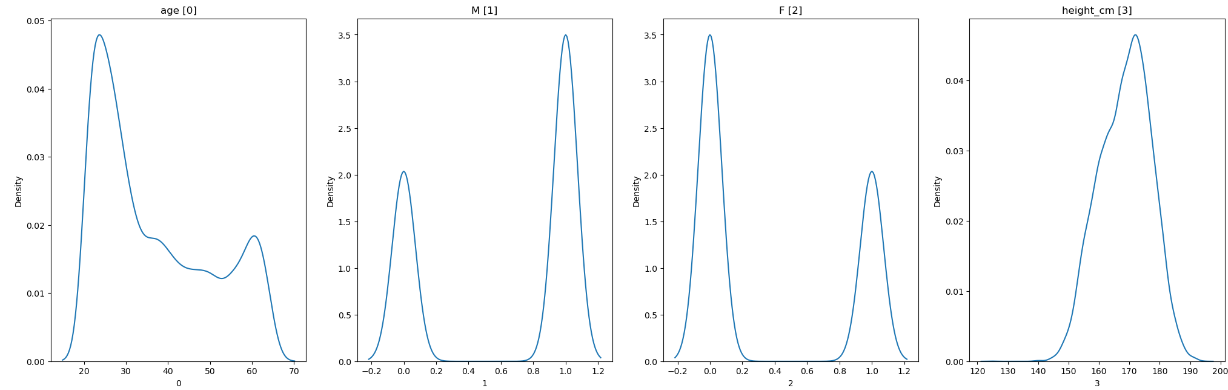
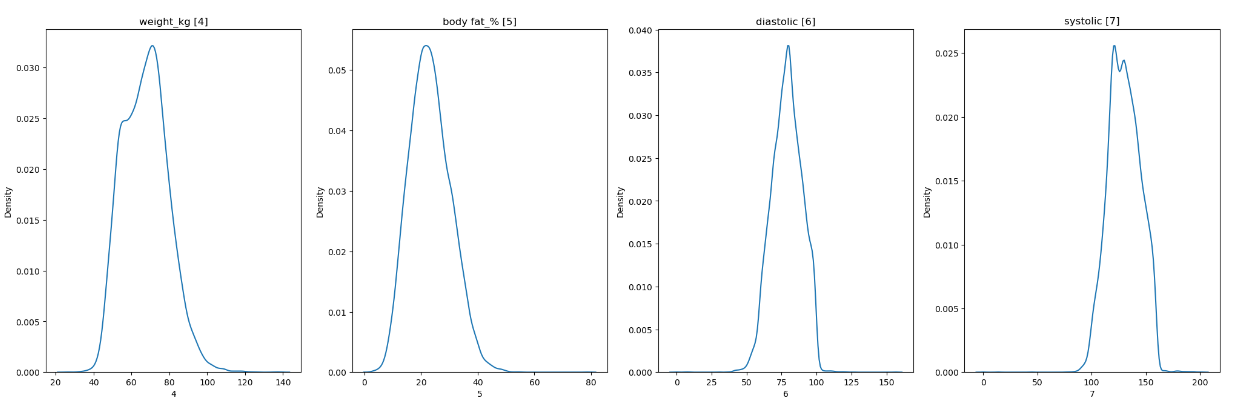
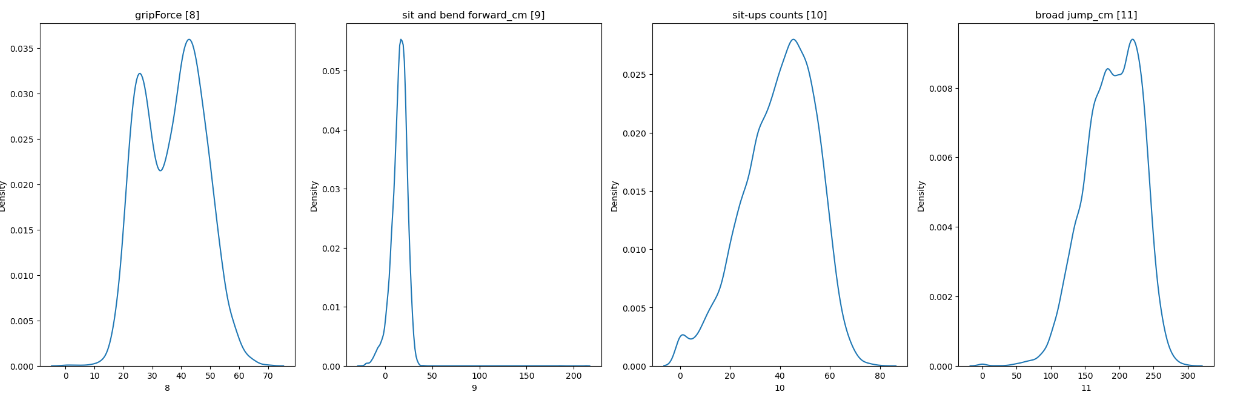
* Age
* Sex
* Height\_cm
* Weight\_kg
* Body\_fat\_%
* Diastolic
* Systolic
* gripForce
* sit\_and\_bend\_forward\_cm
* sit\_ups\_counts
* broad\_jump\_cm
* class (atribut koji se predvidja)

Osim pola, svi atributi su kontinualni numericki.

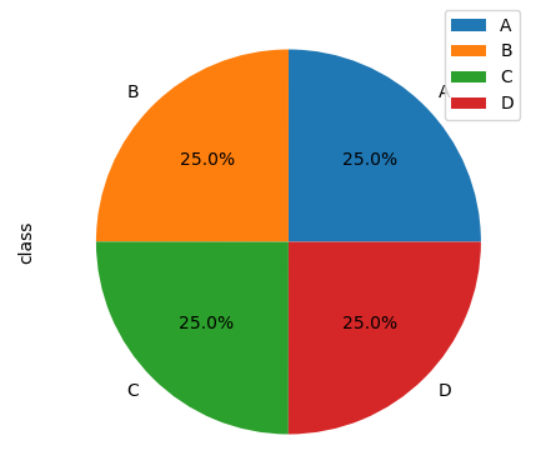


O radu:

Buduci da su podaci kontiualni, algoritam koji se implementira je Gausov model naivne Bajesove metode. Klasni atribut, kao I atribut “sex” su One-Hot enkodovani. Vecina atributa je normalno rasporedjena. Nema korelisanosti izmedju atributa. Nema nedostajucih vrednosti i nadjen je jedan duplikat koji je uklonjen. Iz grafika raspodele vidi se da nema znacajnih outlier-a.



Klasa je skoro savrseno balansirana:



**Implementacija Gausovog modela naivne Bayes-ove metode:**

Metoda je implementirana u klasi NaiveBayes. Njeni atributi su *features\_probs* u kojoj se cuvaju verovatnoce feature-a za svaku od klasa, *num\_of\_features* i *num\_of\_classes* koje predstavljaju broj feature-a i klasa. Metode koje koristi ova klasa su *fit(features, target)* u kojoj se vrsi treniranje modela, tj. u kojoj se postavljaju vrednosti atributa, kao i *predict(features)* koja za svaku opsevaciju vraca najvecu od verovatnoca pripadanja klasama. Pomocna funkcija za racunanje verovatnoce za jednu opservaciju je funkcija *f(x, mi, teta)*.

Rucna implementacija vraca accuracy 0.5505, dok Scikit-Learn verzija vraca 0.5504. Rezultat je pogorsan normalizacijom i standardizacijom podataka.