

კომპანია, რომელშიც მუშაობთ როგორც მონაცემთა მეცნიერი, თანამშრომლობს ერთ-ერთ საავადმყოფოსთან, სადაც იკვლევენ დიაბეტს. ამ საავადმყოფომ შეაგროვა ქალი პაციენტების მონაცემები. მათი მთავარი მიზანია დიაბეტის აღრეული დაგენოსტირება პაციენტებში. თქვენი ამოცანაა დაგალებასთან თანდართული მონაცემთა ფაილის ანალიზი, მოდელის შექმნა და დატრენინგებული მოდელის შეფასება.

ფაილი შედგება 9 სვეტისგან:

1. Pregnancies - რამდენჯერ იყო პაციენტი ორსულად
2. Glucose - პლაზმური გლუკოზის კონცენტრაცია
3. BloodPressure - დიასტოლური არტერიული წნევა
4. SkinThickness - ტრიცეფსის კანის ნაკეცის სისქე
5. Insulin - ინსულინის დოზე
6. BMI - სხეულის მასის ინდექსი
7. DiabetesPedigreeFunction - დიაბეტის ალბათობა ოჯახის ისტორიაზე დაყრდნობით
8. Age - პაციენტის ასაკი
9. Outcome - ტესტის შედეგები (1 - პაციენტს აქვს დიაბეტი, 0 - არ აქვს)

შესაბამისად,

- გადაწყვიტეთ რომელი ML-ის ტიპს და ამოცანას მიეკუთვნება ეს პრობლემა.
- განახორციელეთ მონაცემთა ანალიზი.
- გამოსახეთ მონაცემები ორ განმომილებიან გრაფიკზე განზომილების შემცირების ტექნიკის საშუალებით.
- გაამზადეთ მონაცემები სატრენინგოდ.
- ამოარჩიეთ 3 ალგორითმი პრობლემის შესაბამისად მოცემული სიიდან:
 - Linear Regression
 - Ridge
 - PCA
 - Logistic Regression
 - Decision Tree
 - Random Forest
 - XGBoost
 - LightGBM
 - KNN
 - K-Means
 - SVM
- ამორჩეული ალგორითმებისთვის განახორციელეთ ჰიპერპარამეტრების ოპტიმიზაცია.
- საბოლოოდ, ამოარჩიეთ საუკეთესო ალგორითმი ოპტიმალური ჰიპერპარამეტრებით.
- შეაფასეთ საბოლოო მოდელი, როგორც სატრენინგო, ასევე სატესტო მონაცემებზე. ამისთვის გამოთვალეთ:

- აკურატულობა
- დაბნეულობის მაგრიცა
- სიზუსტე
- გახსენება
- F1 ჯელა
- AUROC (ასევე გამოსახეთ გრაფიკზე ROC მრუდი)

შესრულებული სამუშაო წარმოადგინეთ კოდისა და წერილობითი რეპორტის სახით (კომენტარები/markdown აღწერები). ეს იქნება თქვენ მიერ შესრულებული სამუშაოსა და მიღებული შედეგების აღწერა კომპანიისთვის.

გაითვალისწინეთ, თქვენ მიერ დაწერილი კოდი უნდა იყოს განმეორებადი (*reproducible*).