

კომპანია, რომელშიც მუშაობთ როგორც მონაცემთა მეცნიერი, თანამშრომლობს ერთ-ერთ საავადმყოფოსთან, სადაც იკვლევენ დიაბეტს. ამ საავადმყოფომ შეაგროვა ქალი პაციენტების მონაცემები. მათი მთავარი მიზანია დიაბეტის აღრეული დიაგნოსტიკა პაციენტებში. თქვენი ამოცანაა დავალებასთან თანდართული მონაცემთა ფაილის ანალიზი, მოდელის შექმნა და დაგრენინგებული 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 დონე
6. BMI - სხეულის მასის ინდექსი
7. DiabetesPedigreeFunction - დიაბეტის ალბათობა ოჯახის ისტორიაზე დაყრდნობით
8. Age - პაციენტის ასაკი
9. Outcome - ტესტის შედეგები (1 - პაციენტი აქვს დიაბეტი, 0 - არ აქვს)

შესაბამისად,

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

- აკურატულობა
- დაბნეულობის მატრიცა
- სიზუსტე
- გახსენება
- F1 ქულა
- 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).*