

მწველობის სტატუსის პროგნოზირება Kaggle-ის მონაცემებზე დაყრდნობით

მომხსენებლები: ლუკა ნამორაძე, ლუკა კაკრიაშვილი, გიორგი აფციაური, ნიკა ინიაშვილი

ელ-ფოსტა: nika.iniaashvili767@ens.tsu.edu.ge

კომპიუტერული მეცნიერება, ზუსტი და საბუნებისმეტყველო მეცნიერებების ფაკულტეტი, თბილისის სახელმწიფო უნივერსიტეტი, თბილისი, საქართველო.

მოცემული პროექტი მიზნად ისახავს ინდივიდის მწველობის სტატუსის პროგნოზირებას Kaggle-ის მონაცემთა ნაკრების ანალიზით. აღნიშნული ამოცანა მნიშვნელოვანია საზოგადოებრივი ჯანდაცვისთვის, რადგან დროული და ზუსტი იდენტიფიკაცია ხელს უწყობს პრევენციული ღონისძიებების დაგეგმვას.

პროექტი ეფუძნება მონაცემთა მეცნიერების თანამედროვე მეთოდოლოგიებს და მოიცავს მონაცემთა წინასწარ დამუშავებას, მახასიათებლების შერჩევას, კლასიფიკაციის ალგორითმების გამოყენებასა და შედეგების ანალიზს. გამოყენებულია ისეთი ალგორითმები, როგორიცაა ლოგისტიკური რეგრესია, Random Forest, Gradient Boosting (XGBoost, LightGBM), რომლებიც განხორციელდა Python-ის შესაბამისი ბიბლიოთეკებით: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn.

მოდელის შეფასება განხორციელდა მეტრიკებით: Accuracy, ROC-AUC, Precision, Recall და F1 Score. საუკეთესო შედეგი აჩვენა Gradient Boosting ალგორითმმა (ROC-AUC $\approx$ 92%). გამოვლენილ იქნა მწველობასთან ასოცირებული ძირითადი ფაქტორები: ასაკი, სქესი და ცხოვრების წესი.

პროექტი წარმოადგენს მონაცემთა მეცნიერების პრაქტიკულ გამოყენებას მედიცინაში და ხელს უწყობს ჯანდაცვის გადაწყვეტილებების მიღებას მონაცემებზე დაყრდნობით.

ლიტერატურა

[1] Kaggle – Smoking Status Prediction Dataset. <https://www.kaggle.com>