

ავტომშიფრავი და ლოჯისტიკური რეგრესია: საკრედიტო ბარათების თაღლითობის გამოვლენის ალგორითმები და მათი შედეგები

ერეკლე შიშნიაშვილი, ლიზი მამისაშვილი, ანანო თურქიაშვილი

ელ-ფოსტა: ersho.ge@gmail.com

კომპიუტერული მეცნიერებები: ტექნიკური ინფორმატიკა, კომპიუტერული მეცნიერება, თბილისის სახელმწიფო უნივერსიტეტი, უნივერსიტეტის ქ. N13

ელ-ფოსტა: mamisashvili.lizi@gmail.com

კომპიუტერული მეცნიერებები: ტექნიკური ინფორმატიკა, კომპიუტერული მეცნიერება, თბილისის სახელმწიფო უნივერსიტეტი, უნივერსიტეტის ქ. N13

ელ-ფოსტა: turkiashvilianano@gmail.com

კომპიუტერული მეცნიერებები: ტექნიკური ინფორმატიკა, კომპიუტერული მეცნიერება, თბილისის სახელმწიფო უნივერსიტეტი, უნივერსიტეტის ქ. N13

ონლაინ ტრანზაქციები 21-ე საუკუნიდან გახდა პოპულარული, რაც ნიშნავს იმას, რომ მომხმარებელს თანხის გადახდა შეუძლია ონლაინ, საკრედიტო ბარათის გამოყენების გარეშე. მიუხედავად იმისა, რომ ამგვარი სისტემა მოსახერხებელია, ბარათის მფლობელისთვის უდიდეს რისკსაც შეიცავს. სავაჭრო კომპანია „Shift Credit Card Processing“-ის მიხედვით, 2018 წელს საკრედიტო ბარათებიდან თაღლითურად 24.26 მილიარდი დოლარი იქნა გადარიცხული.

საკრედიტო ბარათების თაღლითობასთან გამკლავების ყველაზე თანამედროვე გზა ჭკვიანი სისტემებია, რომელთაც ვალიდური და თაღლითური ტრანზაქციების განსხვავება შეუძლიათ, თუმცა გამოწვევები მაინც არსებობს. მაგალითად, მონაცემები, რომლებიც ასახავს ტრანზაქციების ისტორიას, ძალიან მცირე რაოდენობით შეიცავს თაღლითური გადარიცხვების ნიმუშებს. მეორეც, მონაცემთა კონფიდენციალურობის გამო, პოტენციურად საჭირო ინფორმაცია ხშირად მიუწვდომელია (Jurgovskya et al, 2018). ბარიერს ქმნის ის გარემოებაც, რომ სხვადასხვა სახის მისკლასიფიკაციას სხვადასხვა წონა აქვს, რაც ნიშნავს, რომ მცდარი პოზიტივებისა და მცდარი ნეგატივების არსებობა განსხვავებულად მოქმედებს მოდელის უტყუარობის შეფასებაზე (Gomez et al, 2018).

Autoencoder- ნეიონული ქსელის ერთ-ერთი განსაკუთრებული მოდელია, რომელსაც ხშირად იყენებენ თაღლითობის გამოსავლენად. თუმცა, არასდროს ჩატარებულა კვლევა იმის დასადგენად, რამდენად კარგად მუშაობს ეს მოდელი სხვადასხვა ჰიპერპარამეტრებთან ერთად. ჩვენ საიტ Kaggle-დან ავიღეთ დაუბალანსებელი მონაცემები, რომელიც შეიცავს სამი დღის განმავლობაში განხორციელებულ ტრანზაქციებს. აქედან, თაღლითური გადარიცხვები შეადგენს მონაცემების მხოლოდ 0.172%-ს. ჩვენ მიერ შექმნილი ავტოდამშიფრავის გამოყენებით, 30 მახასიათებელი 10-მდე შევამცირეთ. ასევე, გამოვიყენეთ ავტომშიფრავის მიერ მიღებული შედეგები, რათა გაგვეწვრთნა ლოჯისტიკური რეგრესიის მოდელი. რადგანაც ჩვენი მიზანი იყო ნეირონული ქსელის მოდელისა და კლასიკური მანქანური სწავლების ალგორითმის მუშაობის ხარისხის შედარება, ლოჯისტიკური რეგრესია ცალკე გამოვიყენეთ თავდაპირველ მონაცემებზე, ხოლო მოდელის გაწვრთნის შემდეგ მოდელის წარმატებულობის შესაფასებლად გამოვიყენეთ PR (სიზუსტე vs. სისრულე) მრუდი. საბოლოოდ, მიმოფანტვის მატრიცისა და AUCPR ქულით შევადარეთ ავტოდამშიფრავისა და ლოჯისტიკური რეგრესიის და უბრალოდ ლოჯისტიკური რეგრესიის ეფექტურობა.

ლიტერატურა

"Credit Card Fraud Statistics". 2020. Shiftprocessing.Com.

Gómez, Jon Ander, Juan Arévalo, Roberto Paredes, and Jordi Nin. 2018. "End-To-End Neural Network Architecture For Fraud Scoring In Card Payments". *Pattern Recognition Letters* 105: 175-181.

Jurgovsky, Johannes, Michael Granitzer, Konstantin Ziegler, Sylvie Calabretto, Pierre-Edouard Portier, Liyun He-Guelton, and Olivier Caelen. 2018. "Sequence Classification For Credit-Card Fraud Detection". *Expert Systems With Applications* 100: 234-245.