# Detecția anomaliilor în rezultatele studenților folosind învățarea automată

Solschi Eliza

Scopul lucrării este de a determina valori unice, diferite în rezultatele studenților în decursul unei materii și de a le analiza. În realizarea acestui proiect s-au utilizat două fișiere excel reprezentând o anumită materie, pe o perioadă de doi ani: logs\_pl2\_2018.xlsx și logs\_pl2\_2019.xlsx. Log-urile conțin toate acțiunile realizate de studenți și profesor în decursul unui semestru.

Pentru a putea ajunge la partea de detecție asupra informațiilor a fost necesară o operație de agregare din care a rezultat un singur excel bazat pe datele celor două log-uri din 2018 și 2019: 10\_log\_extended2.xlsx. Noul excel conține 212 linii reprezentând fiecare student. Coloane reprezintă proprietăți, subliniind diferite acțiuni realizate de studentul în cauză.

'User full name' – codul de identificare al studentului

'Year' – anul din care face parte

'NoIP' – numărul de adrese IP utilizate de student pentru conectare

'Created' – numărul de creații realizate de student de-alungul semestrului

'Deleted'- numărul de ștergeri realizate de student de-alungul semestrului

'Downloaded - numărul de descărcări realizate de student de-alungul semestrului

'Ended' - numărul de activități încheiate realizate de student de-alungul semestrului

'Reset' - numărul de reset-uri realizate de student de-alungul semestrului

'Shown'- numărul de shown arătate de student de-alungul semestrului

'Started'- numărul de activități începute de student de-alungul semestrului

'Submitted'- numărul de activități trimise de student de-alungul semestrului

'Updated'- numărul de activități actualizate de student de-alungul semestrului

'Uploaded' - numărul de activități încărcate de student de-alungul semestrului

'Viewed' - numărul de vizualizări realizate de student de-alungul semestrului

'Holiday'- numărul de accesări in perioada de vacanta a studentului

'Week 01' - numărul de accesări in săptămâna 1 a studentului

'Week 02' - numărul de accesări in săptămâna 2 a studentului

'Week 03' - numărul de accesări in săptămâna 3 a studentului

'Week 04' - numărul de accesări in săptămâna 4 a studentului

'Week 05' - numărul de accesări in săptămâna 5 a studentului

'Week 06' - numărul de accesări in săptămâna 6 a studentului

'Week 07' - numărul de accesări in săptămâna 7 a studentului

'Week 08' - numărul de accesări in săptămâna 8 a studentului

'Week 09' - numărul de accesări in săptămâna 9 a studentului

'Week 10' - numărul de accesări in săptămâna 10 a studentului

'Week 11' - numărul de accesări in săptămâna 11 a studentului

'Week 12' - numărul de accesări in săptămâna 12 a studentului

'Week 13' - numărul de accesări in săptămâna 13 a studentului

'Week 14' - numărul de accesări in săptămâna 14 a studentului

'Mean events' – media evenimentelor realizate de-alungul semestrului de student

'Max events' – numărul maxim de evenimente realizate de student

'Min events' – numărul minim de evenimente realizate de student

'Total events' – numărul total de evenimente de-alungul semestrului, săptămânile 1-14 ale studentului

'1st half sem'- împărțirea semestrului in 2 jumătăți, prima jumătate

'2nd half sem'- împărțirea semestrului in 2 jumătăți, a doua jumătate

'Visit days' – numărul total de zile in care utilizatorul a realizat cel puțin o accesare

'Events' - numărul total de evenimente de-alungul semestrului dar si in perioada de vacanta a studentului

'Day events min' – numărul minim de accesări ale studentului într-o zi

'Day events max'- numărul maxim de accesări ale studentului într-o singura zi

'Weekend days' – numărul de accesări ale studentului in weekend-uri

'Class days'- numărul de accesări in zilele de curs ale studentului

'Out days' – numărul de accesări in zilele fără curs ale studentului

'Afternoon' – numărul de accesări in perioada de după masa ale studentului

'Evening' - numărul de accesări seara ale studentului

'Morning' - numărul de accesări dimineața ale studentului

'Night' - numărul de accesări in perioada de noapte ale studentului

'Class' - numărul de accesări in timpul cursului ale studentului

'Out' - numărul de accesări in afara cursului ale studentului

'Frequency'

'Recency'

'T'

Prin activitate si evenimente se înțelege orice acțiune ce are efect in acea categorie.

Prima etapă a reprezentat investigarea datelor. Data Frame-ul conține 212 înregistrări și 51 de coloane. Am înlocuit celulele NaN cu valoarea 0 pentru simplitate.

În continuare m-am focusat pe determinarea outliners-urilor, a valorilor anormale prin mai multe metode:

1. QUANTILE

2. STANDARD DEVIATION METHOD

3. INTERQUARTILE RANGE METHOD

Am observat că există anomalii pentru următoarele caracteristici și le-am filtrat pe acestea pentru fiecare metodă de mai sus: NoIP, Viewed, Class days.