

ოპერაციების კვლევა Data mining ტექნოლოგიის საფუძველზე

ხათუნა აბულაშვილი, ეკატერინე მაღრაძე
საქართველოს ტექნიკური უნივერსიტეტი, ეკონომიკური ინფორმატიკის კათედრა

ანოტაცია:

ნაშრომში წარმოდგენილია *Data Mining* ტექნოლოგიით შემუშავებული პროგნოზირება საგადასახადო სტრუქტურაში, კერძოდ გადამხდელის პირადი ბარათის საფუძველზე. *Data Mining* ტექნოლოგიის კლასიფიკაციის მეთოდის და ალგორითმ *Naivy Bayes* კრიტერიუმების შერჩევის საფუძველზე ვლებულობთ პროგნოზს, რომლის შედეგადაც შეგვიძლია მივიღოთ სასარგებლო ინფორმაცია საგადასახადო სისტემის მუშაობისათვის.

საჯანძო სიტყვები: *Data Mining*, ინტელექტუალური ანალიზი, ცოდნის მოძიება, *Oracle Data Miner*, *Naivy Bayes* ალგორითმი, პროგნოზირება.

თანამედროვე ბიზნეს სამყაროს საქმიანობის ნებისმიერ სფეროში ინფორმაციის მოცულობა ორგანიზაციებში საკმაოდ დიდია და ორაგნიზაციის წარმატებული მუშაობა დამოკიდებულია მასში არსებული, კუთვნილი ინფორმაციიდან მაქსიმუმის გამოყენებისაგან. საკმაოდ რთულია მონაცემთა მნიშვნელოვნების შეფასება, რომლებიც უწყვეტად გროვდება ადამიანების საქმიანობის პროცესში, რომლებიც მონაწილეობენ ბიზნესის მართვასა და წარმოებაში, საბანკო საქმიანობაში, სამეცნიერო, ინჟინერული თუ სამედიცინო ამოცანების გადაწყვეტაში. მძლავრი კომპიუტერული სისტემები, რომლებიც მოიცავენ ინახავენ და მართავენ უზარმაზარ მონაცემთა ბაზებს წარმოადგენენ განუყორელ ატრიბუტს მსხვილი კორპორაციების საქმიანობისათვის და არა მარტო მსხვილი კორპორაციებისათვის. მაგრამ მონაცემების მხოლოდ არსებობობა არ არის საკმარისი მუშაობის მაჩვენებლების გასაუმჯობესებლად. საჭიროა არსებული მონაცემების ტრანსფორმირების ცოდნა, მათგან სარგებლიანობის მისაღწევად და გადაწყვეტილებების მისაღებად.[1] ამისათვის არის წარმოდგენილი ტექნოლოგია *Data Mining*.

მომხმარებლის სათანადო მონაცემების ანალიზის მეშვეობით შესაძლოა ვიწინასწარმეტყველოთ კლიენტის მოქმედება და აგრეთვე მოვახდინოთ მასზე ზეგავლენაც კი.

Data Mining-ის გამოყენების სფერო არ არის განსაღვრული - ის ყველგან არის სადაც კი არსებობს მონაცემები. მაგრამ პირველ რიგში *Data Mining* -ის მეთოდებმა დააინტრიგა კომერციული საწარმოები, რომლებიც თავიანთ პროექტებს ახორციელებენ ინფ. მონაცემთა საცავის (*Data Warehousing*) საფუძველზე ასეთი საწარმოების მაგალითი გვიჩვენებს, რომ *Data Mining*-ის გამოყენების ეფექტიანობა აღწევს 100%.

Data Mining-ი წარმოადგენს დიდ ფასეულობას ანალიტიკოსებისათვის მათ ყოველდღიურ საქმიანობაში. საქმიანმა ხალხმა კარგად გაითავისა, რომ *Data Mining*-ის მეთოდების მეშვეობით მათ შეუძლიათ მიიღონ საგრძნობი უპირატესობა კონკურენტულ ბრძოლაში.

Data Mining-ის აპლიკაციები ვრცლად არის გავრცელებული: საბითუმო ვაჭრობაში, მარკეტინგში, ფინანსებში, მედიცინაში, წარმოებაში და სხვა სფეროებში. [2]

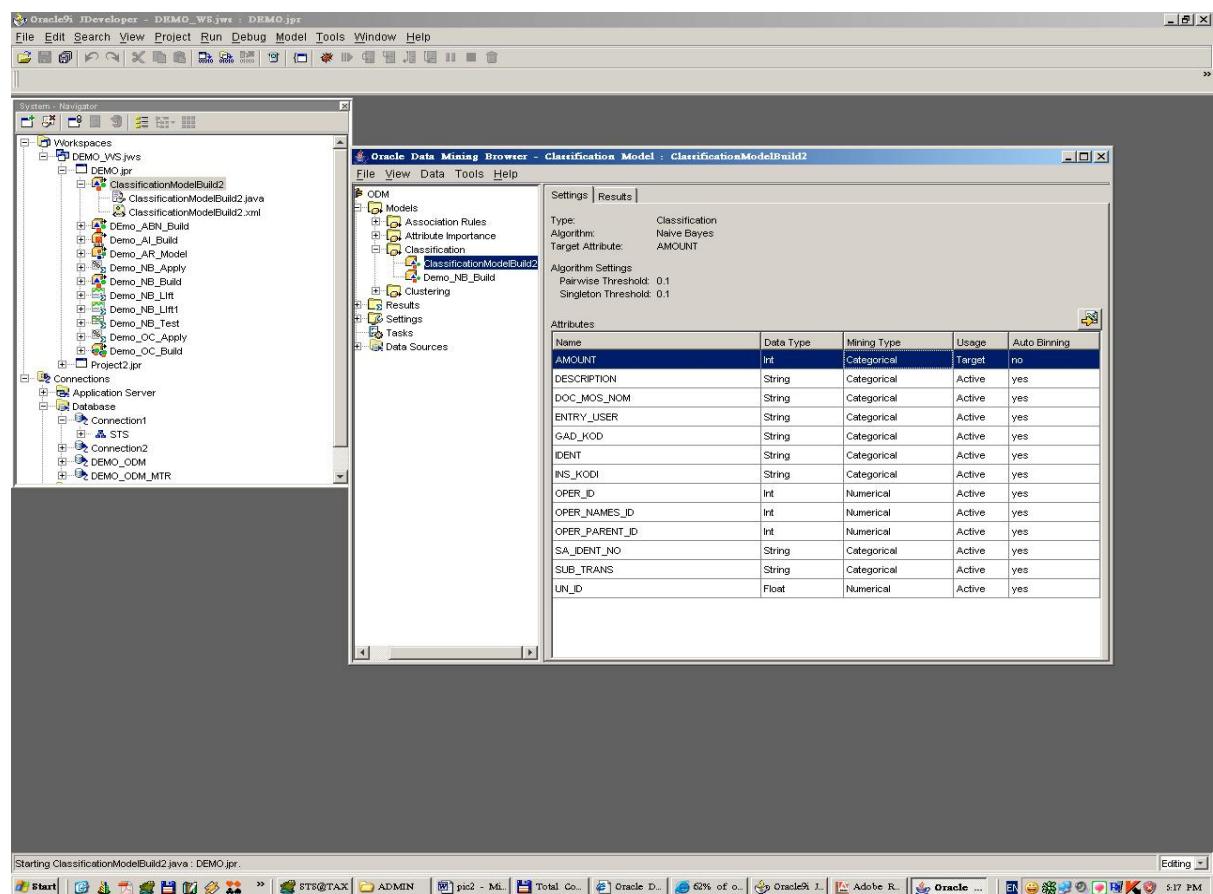
საქართველოში საბაზრო ეკონომიკის დამყარებასა და განვითარებასთან ერთად რა თქმა უნდა იზრდება მფლობელი სუბიექტების (მცირე თუ მსხვილი საწარმოების და ა.შ.), ანუ გადამხდელების რაოდენობა. საგადასახადო სტრუქტურაში დასამუშავებელი ინფორმაციის მოცულობა ყოველდღიურად იზრდება, რაც მოითხოვს ახალი თანამედროვე და ეფექტური ინსტრუმენტებისა და გადაწყვეტილებების დანერგვას. ერთ-ერთი ყველაზე ძირითადი სახელმწიფო ორგანოს ფუნქციონირების ასპექტი, რომელიც არეგულირებს გადასახადების მოგროვების პროცესს, წარმოადგენს ინფორმაციული ტექნოლოგიების განვითარება, რომელიც იძლევა საშუალებას უფრო ოპერატიულად და ზუსტად დავამუშავოდ ინფორმაციის დიდი ნაკადები. გადამხდელისაგან მიღებული ინფორმაციის დამუშავების პრობლემის გადაწყვეტისა და მისგან ცოდნის მოპოვების საშუალებას იძლევა Data Mining ტექნოლოგიის გამოყენება. ცოდნის მოპოვება ამ ტექნოლოგიის და კერძოდ პროდუქტ Oracle Data Miner გამოყენება ჩვენი თვალსაზრისით მოქვედის საშუალებას შევამციროდ დროითი და შრომითი დანახარჯები გადამხდელთა პირადი აღრიცხვის ბარათის ინფორმაციის დამუშავებისას, რაც გამოიწვევს სამუშაოს საკმაო ოპტიმიზაციას და საგადასახადო სტრუქტურის მუშაობის ეფექტიანობას. უნდა აღინიშნოს, რომ არსებული ტექნოლოგია როგორც წესი წარმოდგენილია და აღწერილია შემდეგი სფეროებისათვის: საბანკო საქმე, საბიუმო ვაჭრობა, ტელეკომუნიკაცია, დაზღვევა, მედიცინა, მაგრამ ამ დროისათვის ამ ტექნოლოგიის გამოყენება საგადასახადო სისტემისათვის არ არის გავრცელებული.[3]

ამ შემთხვევაში კლევის ობიექტს წარმოადგენს ტექნოლოგია Data Mining. ხოლო კლევის საგანს წარმოადგენს გადამხდელთა პირადი აღრიცხვის ბარათი საქართველოს რესპუბლიკის საგადასახადო დეპარტამენტის ინფორმაციული ტექნოლოგიების სამართველოში არსებული მონაცემთა საცავების საფუძველზე.

გამოიყოფა ხუთი კანონზომიერების სტანდარტული ტიპი, რომელიც იძლევა საშუალებას გამოვააშკარაოდ Data Mining- ის მეთოდები.

- 1 ასოციაცია
- 2 თანმიმდევრობა
- 3 კლასიფიკაცია
- 4 კლასტერიზაცია
- 5 პროგნოზირება

ჩვენ ვემთხვევაში ვიყენებთ Data Mining- კლასიფიკაციის მეთოდს, რომელიც იყენებს Naive Bayes ალგორითმს (ნახ.1).



ნახ.1 წარმოდგენილია კლასიფიკაციის მოდელის შექმნის პროცესის ფრაგმენტი NB ალგორითმის საფუძველზე

იმისათვის, რომ წარმოდგენა შეგვექმნას თუ რა პროგნოზირების ჩატარება არის შესაძლო Data Mining ტექნოლოგიის მეშვეობით თავდაპირველად ვაყალიბებთ კვლევის ობიექტის აღწერას ანუ გადამხდელის პირადი აღრიცხვის ბარათის აღწერას.[2]

გადამხდელთა პირადი აღრიცხვის ბარათი (გპაბ) წარმოადგენს მკაცრი აღრიცხვის დოკუმენტს, რომელშიც აისახება გადამხდელების/საგადასახადო აგენტების ან სხვა ვალდებული პირების მიერ დეკლარირებული და გადახდილი თანხები, აგრეთვე საგადასახადო და სხვა უფლებამოსილი ორგანოების მიერ საქართველოს კანონმდებლობით გათვალისწინებული გადასახდელად დარიცხული ან შემცირებული თანხები, ვადაგადაცილებული დავალიანებები, ზედმეტად გადახდილი თანხები, სანქციები, ჯარიმები და საურავები.

გადასახადის გადამხდელის/საგადასახადო აგენტის ან სხვა ვალდებული პირის მიერ საგადასახადო ვალდებულების შესრულების აღრიცხვის წარმოების მიზნით საგადასახადო ორგანოებში იხსნება გპაბ, სადაც შესაძლოა კონკრეტული პერიოდისათვის ბარათზე ჩატარებული ოპერაციების დათვალიერება და შესაბამისად შედეგების ნახვა. ნახ2.

თარიღი	გადაუხდებული მოცულობა	კუთხეობის მოცულობა	კუთხეობის მოცულობა	კუთხეობის მოცულობა	კუთხეობის მოცულობა	კუთხეობის მოცულობა	კუთხეობის მოცულობა
20.01.2005							
16.02.2005	გადაუხდებული მოცულობა 01.1.9.3	16.02.2005	0	0	0	0	0
16.02.2005	გადაუხდებული მოცულობა 01.1.9.3	15.02.2005	632000.00	0	0	0	0
28.02.2005	გადაუხდებული მოცულობა 01.1.9.3	27.02.2005	0	0	150000	0	0
28.02.2005	გადაუხდებული მოცულობა 01.1.9.3	27.02.2005	0	0	50000	0	0
21.03.2005	გადაუხდებული მოცულობა 01.1.9.3	15.03.2005	0	0	644412.7	0	0
19.04.2005	გადაუხდებული მოცულობა 01.1.9.3	17.04.2005	1400363.98	0	0	0	0
01.05.2005	გადაუხდებული მოცულობა 01.1.9.3	28.04.2005	0	0	50000	0	0
22.05.2005	გადაუხდებული მოცულობა 01.1.9.3	22.05.2005	0	0	0	0	0
22.05.2005	* კა	22.05.2005	0	0	0	0	0
22.06.2005	გადაუხდებული მოცულობა	22.06.2005	2032452.04	2251224.3	250000	0	0.00

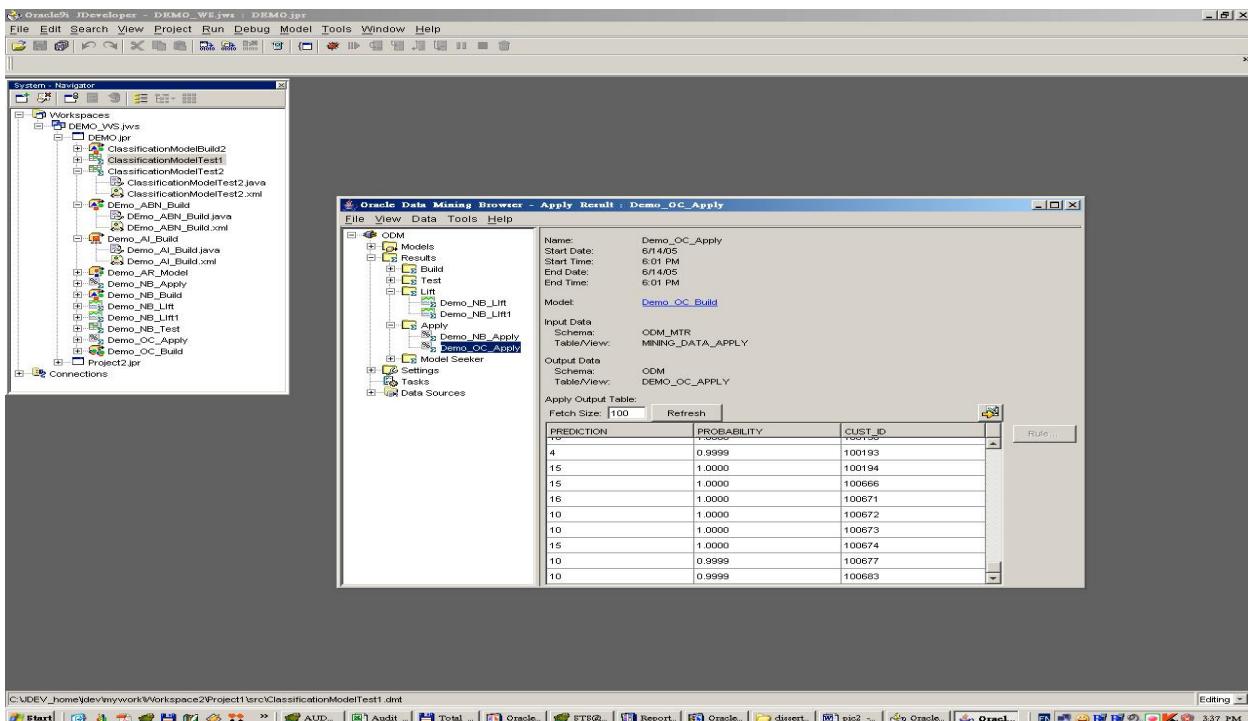
ნახ.2. გადამხდელის პირადი ბარათი

განვიხილოთ კონკრეტული ბარათის წარმოების პროცესი: 2005 წლის იანვრის თვის პირველ რიცხვში კონკრეტულ გადამხდელს ათვლის სათავეში ჰქონდა შემდეგი ნაშთი: ძირითადი ნარჩენი 100ლარი (წარმოქმნილი კუთვნილი გადასახადის (დარიცხული) თანხის გადაუხდელობით), სანქციის ნარჩენი 50 ლარი (წარმოიქმნება დარიცხული სანქციის თანხის გადაუხდელობით), საურავი 20 ლარი (ყოველ ვადაგადაცილებულ დღეზე ძირითადი თანხის 0.07%, აისახება ნაზარდი ჯამით), ამის შედეგად ვლინდება, რომ საერთო დავალიანება ამ თარიღისთვის წარმოადგენს 170 ლარს.

2005 წლის 25 იანვრისათვის გადამხდელს ჰქონდა წარმოდგენილი დეკლარაცია თანხაზე 350 ლარი, რის შედეგადაც ძირითადი ნარჩენის თანხა გახდა 450 ლარი, სანქციის ნარჩენი 50 ლარი - რადგან სანქციის ნარჩენი წარმოებს ნაზარდი ჯამით, რადგან 1 იანვარს გადაუხდელი საურავი ჰქონდა 20 ლარი და 25 იანვრამდე დავალიანება იყო 100 ლარი, ამიტომ დარიცხული საურავი ანგარიშდება შემდეგი ფორმულით $100 * 0.07% * 25$.

ამ ოპერაციების მიხედვით შეგვიძლია ვივარაუდოთ, რომ თუ გამხდელს გააჩნია დავალიანება და არის აქტიური გადამხდელი მოსალოდნელია, რომ ბიუჯეტი მიღებს მისგან ნავარაუდებ თანხას. ყოველი გადამხდელისაგან მიღებული ინფორმაციით ვაჯგუფებთ საქმიანობის სახეებით, რის შედეგადაც ვღებულობთ მონაცემებს სხვადასხვა საქმიანობაზე.[4]

ალგორითმ NB-ს(ნახ.3) მიღებული რეზულტატიდან შეგვიძლია გავაანალიზოთ შემდეგი სიტუაცია. PREDICTION სვეტში მიღებული 1-ანი ნიშნავს, რომ პროგნოზი არსებობს, ხოლო მიღებული 0 გულისხმობს, რომ გადამხდელის მონაცემები პროგნოზში არ ღებულობს მონაწილეობას. სვეტი PROBABILITY გვიჩვენებს პროგნოსის შესრულების ალბათობას. შესაბამისად ჩვენ ვირჩევთ იმ გადამხდელებს, რომლეთა PREDICTION არის 1 და PROBABILITY-ის მნიშვნელობა მაქსიმალური.



ნახ.3 ალგორითმ NB-ს კრიტერიუმების შერჩევის საფუძველზე მიღებული პროგნოზი

კრიტერიუმების შერჩევის საფუძველზე ვღებულობთ პროგნოზს, რომლის შედეგადაც შეგვიძლია გამოვავლინოთ ის დარგები რომლებიც განვითარების სტატიაზეა, განვითარებულ საფეხურზეა და ის დარგები, რომლებიც არიან ცუდად განვითარებულნი. ამ ინფორმაციის საფუძველზე შეგვიძლია დავადგინოთ, თუ რომელ დარგს შეუძლია მოუტანოს ბიუჯეტს მეტი შემოსავალი და რომელ დარგებს უნდა მიექციოს ყურადღება.

ლიტერატურა:

- ნ. კრეჩეტოვი. პროდუქტები მონაცემთა ინტელექტუალური ანალიზისათვის. პროგრამული საშუალებების ბაზარი, №14-15-97, გვ. 32-39
- Knowledge Discovery Through Data Mining: What is Knowledge discovery? – Tandem Computers Inc., 1996
- მ. კისელიოვი, ე. სოლომატინი. ცოდნის მოპოვების საშუალებები ბიზნესსა და ფინანსებში. - ღია სისტემები, №14, 1997, გვ. 41-44.
- საქართველოს ფინანსთა სამინისტროს ბრძანება №228 “გადამხდელთა პირადი აღრიცხვის ბარათების წარმოების წესის შესახებ”//2005 წ. 7 აპრილი, ქ. თბილისი. მუხლი 1-5.
- Oracle Documentation: “Solving Business Problems with Oracle Data Mining” Oracle Inc., 2004.

ნახატების რაოდენობა: 3