



თბილისის თავისუფალი უნივერსიტეტი
მათემატიკის, კომპიუტერული მეცნიერების და ინჟინერიის
სკოლა (MACS[E])

კომპიუტერული მეცნიერების და მათემატიკის პროგრამა

მარიამ ჯავახიშვილი

სამუშაო პრაქტიკის ანგარიში

თბილისი
2024

ანოტაცია

სამუშაო პრაქტიკა დავიწყო 2024 წლის აპრილში საქართველოს ერთ-ერთ წამყვან ტექნოლოგიურ კომპანია - "თიბისი ბანკში" მონაცემთა მეცნიერის პოზიციაზე. "თიბისი ბანკი" საქართველოს და მსოფლიო მასშტაბით ემსახურება თითქმის 700,000+ მომხმარებელს, არის მრავალპროფილური და ცდილობს თავისი წვლილი შეიტანოს როგორც საბანკო, ასევე ტექნოლოგიურ სფეროში. რაც შეეხება მონაცემთა მეცნიერებას - ეს არის მუდმივად მზარდი სფერო, განსაკუთრებით ბოლო ათწლეულში მოთხოვნა საგრძნობლად გაიზარდა ამ მიმართულებით. ტექნოლოგიურ განვითარებასთან ერთად ამ სფეროს თითქმის ყველა განხრამი მნიშვნელოვანი წინსვლაა. გაადვილდა მონაცემთა შეგროვება და შენახვა ინტერნეტიდან, მოხერხდა აპლიკაციებისა და სხვა საბანკო ბაზების საშუალებით. აქედან გამომდინარე, იმისათვის რომ მოხდეს მომგებიანი გადაწყვეტილებების მიღება მარკეტინგული თუ სხვა საბანკო მიზნით, აუცილებელი გახდა იმ ადამიანების ჩართულობა რომლებიც სპეციალიზდებიან სწორედ შემდეგ უნარებში: კომპლექსურ მონაცემებში სწრაფად და დეტალურად გარკვევა, ანალიტიკური კვლევა და დასკვნების გამოტანა. შედეგად, "თიბისი ბანკს" ჰყავს თავისი მონაცემთა მეცნიერების გუნდი, რომელიც იყოფა 4 სხვადასხვა ჯგუფად და შედგება 20-მდე სპეციალისტისგან. ოთხივე გუნდი მუშაობს ხელოვნური ინტელექტისა თუ მანქანური სწავლების პროექტებზე.

არსებული ოთხი გუნდიდან მე უშუალოდ ვექვემდებარები ბიზნეს მონაცემთა მეცნიერების ჯგუფს. ჩემი პოზიცია მოიაზრებს ბიზნეს ანალიტიკოსებთან, მარკეტინგის ან/და ბანკის სხვა გუნდებთან კოლაბორაციას. მათთვის სასურველი მონაცემების დამუშავებას, შეფასებას, ინტერპრეტაციას და შესაბამისი დასკვნების გაკეთებას, იმისათვის, რომ დაინტერესებულმა მხარეებმა გაიგონ არსებული მონაცემების ზოგადი ფონი და მიიღონ მათთვის სასურველი და მომგებიანი გადაწყვეტილებები.

როგორც მონაცემთა მეცნიერის, ჩემს პასუხისმგებლობაშია მონაცემთა პროცესინგი, ანალიზი პატერნების, მიმდინარე ტრენდების მოძებნა და ანომალიების

იდენტიფიცირება. ასევე სხვადასხვა სტატისტიკური ტექნიკების და შესაბამისი პროგრამების გამოყენება მონაცემთა შერჩევისა და ვიზუალიზაციისათვის. ამ ეტაპების შემდეგ მიღებული შედეგების მიხედვით საუკეთესო მოდელის შერჩევა, აგება, გატესტვა, ვალიდაცია და პრეზენტირება შესაბამისი მხარეებისათვის. მნიშვნელოვანია ისიც, რომ ეს პოზიცია მოითხოვს კარგ კომუნიკაციის უნარს, რადგან გიწევს იმუშაო სხვადასხვა პროფესიის ადამიანებთან, შესაბამისად შენი შედეგები მაქსიმალურად გასაგებად და ნათლად უნდა მიაწოდო მათ. ასევე გარემოცული ხარ ერთ-ერთი საუკეთესო პროფესიონალებით, რომლებიც გამოცდილები არიან ამ სფეროში და აქტიურად გიზიარებენ საკუთარ შეხედულებებს და შეფასებებს.

საყურადღებოა, რომ ამ და სხვა დავალებების პარალელურად ჩართული ვიყავი აქტიური თვით-განვითარების პროცესში, მოვახერხე უნივერსიტეტიდან მიღებული ცოდნის გამყარება მანქანური სწავლების განხრით, გავიარე რამდენიმე საჯარო კურსი და დავეუფლე იმ პრაქტიკებს რომელიც აუცილებელია ჩემი, როგორც მონაცემთა მეცნიერის განვითარებისთვის.

პრაქტიკის პერიოდში მომიწია მემუშავა სხვადასხვა კომპლექსურ კლასიფიკაციის ამოცანებზე. ერთ-ერთი მათგანი ემსახურება მოსახლეობის იმ პოპულაციის, რომელთაც ხელფასის სახით არ ერიცხება ანაზღაურება და შესაბამისად სესხის აღება უჭირს ან საერთოდ ვერ ახერხებს, სავარაუდო შემოსავლის დადგენა და შესაბამისად სესხის გაცემათა რიცხვის გაზრდა მომხმარებელთა ამ ნაწილზე. კომპანიაში მუშაობის დაწყებიდან ერთ-ერთი პირველი დავალების ფარგლებში მომიწია ამ პრობლემის ჩემი გუნდის მიერ უკვე არსებული მოდელის შეფასება, ახალ მონაცემებზე გატესტვა, კლასიფიკაცია და ვალიდაცია. ისეთ დიდ კომპანიაში მუშაობა როგორც თიბისი ბანკია განსაკუთრებით საინტერესოა, რადგან გუნდებს უწევთ არამარტო სამეცნიერო კუთხით მუშაობა, არამედ საბანკო სისტემაში რეალურ, ყოველდღიურობაში დასმული ამოცანების გადაწყვეტა.

I started working in April 2024 as a data scientist in one of the leading technological companies of Georgia - "TBC Bank". "TBS Bank" serves almost 700,000+ customers in Georgia and worldwide. The bank is multi-profile and strives to contribute both in the banking and technological fields. As for data science - this is a constantly growing field, especially in the last decade, the demand in this direction has increased significantly. Along with technological development, there are significant advances in almost every aspect of this field. It has become easier to collect and store data from the Internet, mobile applications and other banking databases. Therefore, in order to make profitable decisions for marketing or other banking purposes, It became necessary to involve people who specialize in the following skills: quickly and thoroughly analyzing complex data, conducting analytical research, and drawing conclusions. As a result, "TBS Bank" has its data science team, which is divided into 4 different groups and consists of about 20 specialists. All four teams work on artificial intelligence or machine learning projects.

Of the four existing teams, I report directly to the Business Data Science group. My position involves collaboration with business analysts, marketing and/or other teams in the bank. To process, evaluate, interpret and draw appropriate conclusions from the data they want, in order for the interested parties to understand the general background of the available data and make their desired and beneficial decisions.

As a data scientist, my responsibility is to process and analyze data to identify patterns, trends and anomalies, As well as the use of various statistical techniques and relevant programs for data selection and visualization. After these stages, according to the results obtained, the best model is selected, built, tested, validated and presented to the relevant parties. It is also important that this position requires good communication skills, because you have to work with people from different teams, so you have to deliver your results as clearly as possible. You are also surrounded by some of the best professionals in the field. They are experienced in this field and actively share their views and assessments with you.

It is noteworthy that in parallel with these and other tasks, I was involved in an active self-development process, I managed to consolidate the knowledge I received from the university in the field of machine learning, I took several public courses and mastered the practices necessary for my development as a data scientist.

During my time in TBC, I worked on various complex classification tasks. One of them involved estimating the income of individuals who do not receive a regular salary, making it difficult for them to obtain loans. My task was to evaluate the existing model developed by my team, test it on new data, classify the results, and validate the model's performance. Working at TBC Bank has been particularly interesting because the teams must address not only scientific challenges but also solve practical, everyday issues within the banking system.

სარჩევი

შესავალი	1
ძირითადი ნაწილი	3
დასკვნა.....	5

შესავალი

სამუშაო პრაქტიკა 2024 წლის 22 აპრილს დავიწყე "თიბისი ბანკში" როგორც ბიზნეს მონაცემთა მეცნიერების გუნდის მეცნიერმა. კარიერის სწორად შერჩევა საკმაოდ საპასუხისმგებლოა. უნივერსიტეტში უამრავი კურსი გავიარე რომელიც მოიცავდა სხვადასხვა ტექნოლოგიურ მიმართულებას, მაგალითად: Java/C++/python Developement, Android/iOS Developement და სხვა. ბევრი ფიქრისა და კვლევის შედეგად გადაწყვიტე რომ ცოდნა ისეთი კუთხით გამეღრმავებინა რომელიც ყველაზე მეტად მაინტერესებდა და თან მექნებოდა როგორც კარიერული, ასევე თვითგანვითარების ფართო არჩევანი. სამწუხაროდ ამ პროფესიაში გამოცდილებას დიდი მნიშვნელობა აქვს და რთული აღმოჩნდა პირველი-მეორე კურსიდანვე კონკრეტული სამსახურის მოძიება. ამ პერიოდში სამუშაო გამოცდილება დავაგროვე როგორც Quality Assurance(QA) Engineer (ვამოწმებდი კოდის ხარისხს და დეველოპერის მიერ შექმნილი პროდუქტის მოთხოვნასთან შესაბამისობას) ერთ-ერთ მსხვილ კომპანიაში, რომელიც მომხმარებლებს სთავაზობდა სხვადასხვა Software სერვისს. იმავდროულად ვცდილობდი, რომ ფეხი ამეწყო მანქანური სწავლების და ხელოვნური ინტელექტის სფეროს სიახლეებისათვის.

საბოლოოდ ამ წლის დასაწყისში გავგზავნე ჩემი რეზიუმე მონაცემთა მეცნიერის პოზიციაზე თიბისი ბანკში - ეს გადაწყვეტილებაც ლოგიკური იყო, რადგან ბანკები საქართველოში გამოირჩევიან მონაცემთა სიმრავლით, "თიბისი ბანკი" კი წარმოადგენს მონაცემთა მეცნიერების ცოდნის ცენტრს, რაც დამატებით მიზეზს მაძლევდა. გავიარე ორი გასაუბრება. აქედან ერთი - ტექნიკური გასაუბრება მოიცავდა ჩემი ალბათობა/სტატისტიკის უნარებისა და Python-ის ცოდნის შემოწმებას, რომლის წარმატებით გავლის შემდეგ შევუერთდი ბიზნეს მონაცემთა მეცნიერების გუნდს. ჩემთვის მნიშვნელოვანია პრაქტიკულ ამოცანებზე მუშაობა, რაც ამ გუნდის მთავარი მიზანია. შედეგად, დღესაც ვარ თიბისის გუნდის წევრი.

მოკლედ თიბისიზეც რომ ვისაუბრო ის წარმოადგენს ტექნოლოგიურ კომპანიას. ბანკი იცავს კორპორაციული მართვის უმაღლეს სტანდარტებს, რომლებიც საერთაშორისო

პრაქტიკას შეესაბამება. თიბისის ხედვა დაფუძნებულია გუნდზე, რომელიც არის "გამარჯვებული, ბედნიერი, ცნობისმოყვარე, შედეგზე ორიენტირებული, ღია ახალი შესაძლებლობებისთვის, კეთილსინდისიერი, და ყოველთვის ასრულებს დანაპირებს" (TBC Bank, n.d.), ასევე "არ ეშინია შეცდომების და ინოვაციურად აზროვნებს" (TBC Bank, n.d.). წარმატებული გუნდები სწორედ ამ პრინციპებით იქმნება და შესაბამისად ჩემთვისაც მოტივაციაა ვიყო ამ ხედვის თანაზიარი.

თიბისი მომხმარებელთა ყოველდღიურობის ნაწილია და მათ ციფრულად ემსახურება. ძირითად პროდუქციას წარმოადგენს სწრაფი სამომხმარებლო სესხები, ანაზრები და ა.შ. ასევე ხშირია კონკრეტული ამოცანები ბანკის სხვა გუნდების მხრიდან, რომელსაც საბოლოოდ სწორედ ბიზნეს გუნდი ასრულებს. Computer Vision, NLP, AI- ეს ამოცანების მცირე ჩამონათვალია რომელთა გამოყენებითაც მონაცემთა მეცნიერების გუნდი ცდილობს "გაუმარტივოს ადამიანებს ცხოვრება" (TBC Bank, n.d.) - რაც სწორედ "თიბისი ბანკის" მისიას წარმოადგენს.

ძირითადი ნაწილი

ჩემი პირველი ნაბიჯები თიბისიში როგორც მონაცემთა მეცნიერი იყო უკვე არსებული პროექტების დეტალურად გაცნობა და ცოდნის გაღრმავება. ამ პროცესში მქონდა ხშირი კომუნიკაცია გუნდის ლიდერთან, რომელიც მაწვდიდა მასალებს და უკუკავშირს სხვადასხვა მანქანური სწავლების ალგორითმებზე. ასევე ვეცნობოდი მონაცემთა მეცნიერების სხვა გუნდების საქმიანობას და მათ პროექტებს. პარალელურად მქონდა რამდენიმე დავალება შესწავლილი ცოდნის პრაქტიკაში გამოყენებისათვის. ერთ-ერთი იყო შემოსავლის შეფასების კლასიფიკაცია, და მიღებულ კლასებზე რეგრესიის დაწერა. მიზანი იყო ინდივიდების კლასიფიკაცია ორ შემოსავლის ჯგუფად: 1500 ლარზე ნაკლები და 1500 ლარზე მეტი. დავიწყე არსებული მონაცემების ანალიზით. შემდეგ ეტაპი იყო არარსებული მონაცემების გადაყრა ან სავარაუდო შედეგებით შევსება. რის შემდეგაც გადავარჩიე ცვლადები. აქაც რამდენიმე სავარაუდო ალგორითმი ვცადე და საბოლოოდ ავიღე ის ცვლადები, რომელთა საშუალებითაც რეალურ და შეფასებულ მონაცემებს შორის სხვაობა Gini მეტრიკით საუკეთესო იყო. უშუალოდ მოდელის შესადგენად ლოჯისტიკური რეგრესიის, გადაწყვეტილების ხეების, Random Forest-ის ალგორითმების განხილვის მიუხედავად აზრი შევაჩერე XGBClassifier-ზე რომელიც უფრო მეტად ოპტიმიზებული და მოქნილია. გარდა Gini მეტრიკისა, გამოვიყენე F1-score კონკრეტული წერტილების სიზუსტის საჩვენებლად. დამაკმაყოფილებელი შედეგის შემდეგ სატრენინგო მონაცემებზე, ჩავატარე ჰიპერპარამეტრების ტუნინგი რომ გამეუმჯობესებინა მოდელის სიზუსტე. საბოლოოდ, შევაფასე მიღებული მოდელის პერფორმანსი სატესტო მონაცემებზე და გავაკეთე შედეგების ვიზუალიზაცია სხვადასხვა სახელფასო ჯგუფებისთვის.

კომპანიაში მნიშვნელოვანი ადგილი ეთმობა თვითგანვითარებას. კვირიდან ერთი დღე შემიძლია გუნდის წევრებთან ერთად თუ დამოუკიდებლად ვიმუშაო იმაზე, რაც მგონია რომ როგორც ჩემი კარიერული, ასევე ზოგადი ცოდნის და პირადი ტექნოლოგიური ინტერესების გაღრმავებისათვის გამოსადეგი იქნება. ამის ფარგლებში

წამოვყენე ინიციატივა, რომელიც მოიაზრებდა ალგორითმების და სტრუქტურების მიმართულებით არსებული ცოდნის გაზიარებას. შევიკარით რამდენიმე თანამოაზრე და რადგან ჩვენი საქმიანობა Python-ის აქტიურად გამოყენებას მოიცავს, და ეფექტურობა როგორც დროის ასევე მეხსიერების მხრივ ძალიან მნიშვნელოვანია. გადავწყვიტეთ რომ გავცნობოდით უკვე არსებულ ალგორითმებს და გამოყოფილ დროში ჩვენი ამოგვეხსნა სხვადასხვა ამოცანები შერჩეული თემატიკის ფარგლებში. ამ შეკრებებში საკმაოდ კრიტიკული იყო ჩემი როგორც კომპიუტერული მეცნიერებების სტუდენტის ჩართულობა, რადგან ეს კურსი მე უკვე გავლილი მქონდა და შესაბამისად უფრო გამიადვილდა არსებული ცოდნით ამოცანების სწორი მიდგომით ეფექტურად ამოხსნა და ასევე იმ ალგორითმების გახსენება რომელიც სხვებისთვის უცნობი იყო. თავისუფალ უნივერსიტეტში გავლილი სხვადასხვა კურსიდან ყველაზე მეტად ალბათობა/სტატისტიკის და მანქანური სწავლების კურსი დამეხმარა. ასევე მივიღე ისეთი უნარები, როგორიცაა კრიტიკული აზროვნება, სწრაფად და დეტალურად ახალი ინფორმაციის მიღება და დამუშავება, შედეგად ადვილია ერთი დავალებიდან მეორეზე ეფექტურ დროში გადასვლა. ასევე მნიშვნელოვანია, რომ უნივერსიტეტში დაგროვილი ზოგადი ცოდნა მაგალითად ფილოსოფია, ისტორია, პოლიტიკური იდეოლოგიები, ლიტერატურა და ხელოვნება დამეხმარა რომ ჩემს თანაგუნდელებთან მეპოვა საერთო ინტერესები და მათთან კომუნიკაცია ადვილად დამემყარებინა არასამუშაო თემატიკაზეც. ჩემი მოლოდინები გუნდის მიმართ გამართლდა, შექმნილი გარემო მუშაობას კიდევ უფრო საინტერესოს ხდის. ხშირად ვიღებ დახმარებას გამოცდილი თანაგუნდელებისაგან და ვცდილობ ჩემი იდეებიც გავაცნო მათ. რადგანაც ჩემი დაინტერესება მანქანური სწავლების და მონაცემთა მეცნიერების მიმართ დიდია, თიბისიში მუშაობა ხელს მიწყოფს განვითარებაში, თანამშრომლების მოტივაცია და გამოცდილება კი მეხმარება არ შევუშინდე გამოწვევებს და თვალყური ვადევნო ამ სფეროში მიმდინარე სიახლეებს.

დასკვნა

ჩემი პრაქტიკის განმავლობაში, მივიღე გამოცდილება მონაცემთა ანალიზში, მანქანურ სწავლებაში, მოდელის განვითარებაში. ყოველდღიურობაში წამოჭრილ პრობლემებზე ფიქრი და მუშაობა საშუალებას მაძლევს პრაქტიკაში გამოვიყენო კომპლექსური მონაცემთა დამუშავებისა და ცვლადების გადარჩევის ალგორითმები, მოდელის ვალიდაციისა და ოპტიმიზაციის მიდგომები. ეს გამოცდილება კიდევ უფრო ამყარებს ჩემს ინტერესს მონაცემთა მეცნიერების მიმართ. შესაბამისად ჩემი მიზანია გავიღრმავო ცოდნა ამ მიმართულებით და გამოვიმუშაო შესაბამისი უნარები, რომ შევიტანო ჩემი წვლილი ისეთ პროექტებში, რომელიც გადაჭრის უფრო დიდი მასშტაბის რეალურ პრობლემებს ხარისხიანად და ეფექტურად.

References:

TBC Bank. (n. d.). *მისია და ხედვა*. Retrieved June 16, 2024, from
<https://www.tbcbank.ge/web/ka/web/guest/mission-vision>