

IMPLEMENTATION AND RESEARCH OF THE RANDOM FOREST ALGORITHM TO SOLVE CLASSIFICATION PROBLEMS

Summary. Introduction. *Solution trees are one of the most commonly used ways to build recognition algorithms in modern information processing systems due to their many advantages, in particular, the possibility of their application in expert systems. One of the problems solved by expert systems is the classification problem, which is to determine the class of an object from a set of classes known to the expert system. This article is devoted to the possibility of using one type of decision tree ensemble - a random forest - to solve the classification problem.*

The purpose of this paper is to implement and study the Random Forest algorithm and its application to solving classification problems.

Results. *The problem of classification by means of a random forest is considered in the work. This is a formalized task that specifies a set of objects that are divided into classes in some way. A finite set of objects is specified for which classes they are known. The Random Forest algorithm is implemented in the Octave environment. The decision tree is implemented according to the CART algorithm. Implemented programs allow us to solve both the problem of binary (two-class classification) and multi-class classification.*

To study the Random Forest algorithm, the following data sets were considered: passenger classification, seed classification, mobile operator customer outflow classification, credit issuance classification, iris classification, wine classification. The quality of the algorithm was evaluated by the share of correct answers in the deferred sample: 89.04% received correct answers for the Titanic data set, 94.11% for Seeds, 91.66% for Telecom_churn, 90.47% for CreditScoring, 95.94% for Iris, and 93.65% for Wine Quality.

Conclusion. *As a result of this work, the effectiveness of the application of ensemble methods based on decision trees, which are built using the CART algorithm, was confirmed. Testing showed the accuracy of the results, more than 90%, which indicates the high quality of the algorithm. It is clear that this quality is obtained through the use of several different decision trees, the results of which are summarized in the derivation of the final assessment.*

Keywords: *machine learning, classification problem, decision tree, Random Forest.*

*Одержано редакцією 20.12.2019 р.
Прийнято до публікації 24.02.2020 р.*

УДК 378:517

DOI 10.31651/2076-5886-2020-1-77-86

PACS 02.60.-x

ТАРАСЕНКОВА Ніна Анатоліївна,
доктор педагогічних наук, професор,
завідувач кафедри математики та
методики навчання математики,
Черкаський національний університет
імені Богдана Хмельницького
e-mail: ntaras7@ukr.net
ORCID: 0000-0002-6418-6380

СЕРДЮК Зоя Олексіївна,
кандидат педагогічних наук, доцент,
доцент кафедри математики та методики
навчання математики,
Черкаський національний університет
імені Богдана Хмельницького
e-mail: serdyuk_z@ukr.net
ORCID: 0000-0002-9376-4346

**ОСОБЛИВОСТІ ВИКЛАДАННЯ КУРСУ МАТЕМАТИЧНОГО АНАЛІЗУ
ДЛЯ ФАХІВЦІВ З АНАЛІЗУ ДАНИХ**

У статті розглянуто специфіку організації вивчення математичного аналізу студентами – майбутніми фахівцями з аналізу даних з урахуванням особливостей освітнього контенту спеціальності 126 «Інформаційні системи та технології»; зазначено загальні компетентності з даної навчальної дисципліни для зазначеної спеціальності та вказано шляхи їх досягнення, зокрема схарактеризовано особливості застосування схематизації під час вивчення та закріплення теоретичного матеріалу в курсі математичного аналізу студентами спеціальності 126 «Інформаційні системи та технології»; наведено конкретні приклади завдань із застосуванням різноманітних схем.

Ключові слова: математичний аналіз, аналіз даних, схематизація, схеми, студенти-програмісти.

Вступ

Курс математичного аналізу є базовим курсом циклу математичних дисциплін, які вивчають студенти більшості нематематичних спеціальностей класичних, педагогічних, технічних ЗВО. Наприклад, у класичних університетах – це спеціальності «Інформаційні системи та технології», «Прикладна математика», «Програмна інженерія», «Системний аналіз», «Фізика та астрономія», «Прикладна фізика та наноматеріали» тощо. Даний курс спрямований не лише на засвоєння студентами основних математичних понять і фактів диференціального та інтегрального числення функцій однієї чи кількох змінних та їх застосування до дослідження функцій, обчислення довжин кривих, площ поверхонь, моментів інерції та статичних моментів, знаходження кратних і контурних інтегралів тощо, а й на формування в них спроможності доказово і несуперечливо міркувати, аналізувати, порівнювати, узагальнювати тощо, загалом, уміння робити правильні висновки та будувати реалістичні прогнози, застосовувати отримані знання, навички й уміння до розв'язування різноманітних практичних задач. Деякі аспекти вивчення даного курсу висвітлено у працях [1; 2; 3]. Проте специфіка організації вивчення математичного аналізу з урахуванням особливостей освітнього контенту спеціальності 126 «Інформаційні системи та технології» ще не була предметом розгляду і залишається у фокусі наших наукових розвідок.

У Стандарті Вищої Освіти України для галузі знань 12 «Інформаційні технології» спеціальності 126 «Інформаційні системи та технології» першим серед програмних результатів навчання зазначено наступний: «ПР 1. Знати лінійну та векторну алгебру, диференціальне та інтегральне числення, теорію функцій багатьох змінних, теорію рядів, диференціальні рівняння для функції однієї та багатьох змінних, операційне числення, теорію ймовірностей та математичну статистику в обсязі, необхідному для розробки та використання інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації» [4, с. 8].

Цей програмний результат навчання сприяє формуванню у студентів таких загальних компетентностей:

- КЗ 1. Здатність до абстрактного мислення, аналізу та синтезу;
- КЗ 2. Здатність застосовувати знання у практичних ситуаціях;
- КЗ 3. Здатність до розуміння предметної області та професійної діяльності;
- КЗ 6. Здатність до пошуку, оброблення та узагальнення інформації з різних джерел [4].

Під час вивчення математичного аналізу якраз і створюються комфортні умови для ефективного формування у студента зазначених компетентностей. Вивчення курсу математичного аналізу сприяє не лише оволодінню студентами основними теоретичними знаннями, практичними навичками й уміннями, а й забезпечує досягнення такого рівня засвоєння матеріалу, який був би достатнім для їхньої професійної діяльності.

Забезпечення належного рівня математичної освіти набуває на сучасному етапі розвитку суспільства особливого значення. Математична компетентність для фахівців різних напрямів, зокрема інженерів-програмістів – основне підґрунтя для засвоєння фахових дисциплін і майбутньої професійної діяльності. У ході вивчення математичного аналізу закладають не тільки методологічний, а й психофізіологічний фундамент системного, логічного та критичного мислення, що є життєво необхідним.

Мета статті – розглянути особливості організації вивчення курсу математичного аналізу студентами спеціальності 126 «Інформаційні системи та технології» з опорою на схематизацію як специфічну знаково-символічну діяльність.

Виклад основного матеріалу

Аналіз даних – розділ математики, що займається розробкою методів обробки даних незалежно від їх природи. Аналіз даних включає виконання послідовних, логічних дій з інтерпретації зібраних даних (наприклад, відповідей респондентів) та їх перетворення на статистичні форми, потрібні для ухвалення маркетингових та керівницьких рішень.

Можна виділити такі етапи аналізу даних: отримання даних, обробка, аналіз та інтерпретація результатів обробки.

Останнім часом значного поширення набувають нові технології і методи аналізу даних, зокрема методи інтелектуального аналізу даних (Data Mining), які використовують для виявлення прихованих закономірностей у великих масивах даних, та нейроінформатики, а також методи й засоби статистичного контролю за якістю на виробництві та в управлінні організаціями. Основні процедури аналізу даних найчастіше реалізують за допомогою сучасних комп'ютерних технологій [5].

Аналіз даних по суті є прикладним розділом математичної статистики, а як відомо, то більшість фактів, формул, теорем математичної статистики базуються на теоретичних знаннях та практичних навичках з математичного аналізу. Тому під час підготовки майбутніх інженерів-програмістів – студентів спеціальності 126 «Інформаційні системи та технології», – важливо, щоб вони отримали міцну теоретичну базу та набули практичних навичок з тих чи тих тем математичного аналізу, які будуть необхідні їм у подальшому професійному становленні, зокрема: «Числові ряди», «Степеневі ряди», «Інтегральне числення функції однієї змінної», «Інтегральне числення функцій кількох змінних» тощо.

На жаль, за останні роки кількість годин, що відводиться на вивчення даного курсу значно зменшилося, проте вимоги до змістового наповнення суттєво не змінилися. Тому доцільно змінити підхід до структурування змісту курсу та відпрацювання навичок й умінь студентів у нових «складніших» умовах. Наразі структура навчальної дисципліни «Математичний аналіз» для студентів спеціальності 126 «Інформаційні системи та технології» є такою, як показано у таблиці 1.

Оскільки курс математичного аналізу зазвичай насичений великою кількістю досить складних математичних фактів (теорем, формул тощо), які студенти спеціальності 126 «Інформаційні системи та технології» насамперед повинні вміти застосовувати у своїй безпосередній професійній діяльності, то ми маємо за мету не лише їх «бездумне» запам'ятовування, адже тренувати пам'ять можна й іншими способами, а саме вміння їх застосовувати в тій чи тій практичній ситуації та для розв'язання різних прикладних задач. Згідно з нашими спостереженнями, засвоєння та, найголовніше, застосування тих чи тих фактів у процесі розв'язування задач часто зумовлюють появу труднощів у студентів-програмістів. Тому методика їх вивчення має будуватися як специфічна для такої категорії студентів.

Таблиця 1

Структура навчальної дисципліни

Форма навчання	Семестр у межах дисципліни	Кількість кредитів	Загальна кількість годин	Аудиторна робота		Самостійна робота		Форма підсумкового контролю
				лекції	практичні	розрахункові роботи	підготовка до занять	
денна	перший	5	150 год	50 год /33 %		100 год /67 %		іспит
				26 год	24 год	30 год	70 год	
денна	другий	5	150 год	50 год /33 %		100 год /67 %		іспит
				26 год	24 год	30 год	70 год	

Ураховуючи останні тенденції освітньої політики та специфіку підготовки фахівців зі спеціальності 126 «Інформаційні системи та технології», а саме – схильність до алгоритмізації у розв’язанні тих чи тих практичних завдань, то доцільним, на нашу думку, буде використання схематизації під час вивчення математичного аналізу вищезазначеними студентами – і під час опанування теоретичного матеріалу, і в ході розв’язування задач.

Схематизація як знаково-символічна діяльність у навчанні математичного аналізу студентів-програмістів посідає значне місце. Як зазначає Н. А. Тарасенкова, її сутність полягає в тому, що навчальне пізнання здійснюється з опорою на певну схему, що відображає структуру реальності мовними чи немовними засобами [6].

У ході ознайомлення студентів з новим матеріалом та під час застосування знань можна виділити кілька різновидів навчальних ситуацій, в яких вони стикаються з діяльністю схематизації. Розглянемо їх детальніше.

Перший випадок організації схематизації ми пов’язуємо з використанням відомих схем діяльності для отримання суб’єктивно нових знань загального характеру. Такі ситуації виникають, наприклад, під час застосування певного методу доведення теорем з курсу математичного аналізу чи окремого прийому встановлення математичного факту тощо. При цьому важливо, щоб студенти не просто слідували за викладом лектора, а просувалися по зарані розставлених орієнтирах [7]. Сукупність таких орієнтирів і виступатиме схемою діяльності студентів.

Не менш важливу роль схематизація може відігравати у розвитку візуального мислення студентів. Для цього потрібно, щоб схема діяльності фіксувалася графічними засобами, істотні зв’язки між елементами схеми були виведені назовні й так, щоб утворювалися можливості декодувати інформацію поза її вербалізації.

До наступного виду схематизації ми відносимо використання плану викладу змісту певної навчальної теми. При цьому схеми діяльності можуть подаватися учням як у розгорнутій словесній формі, так і за допомогою немовних знаково-символічних засобів [6].

Наприклад, план викладу може бути пред’явлений студентам у формі опорного конспекту. Так, під час введення означення визначеного інтеграла доцільно скористатися наступним покроковим планом-конспектом.

План-конспект

Нехай функція $y = f(x)$ визначена на відрізку $[a; b], a < b$ (рис. 1).

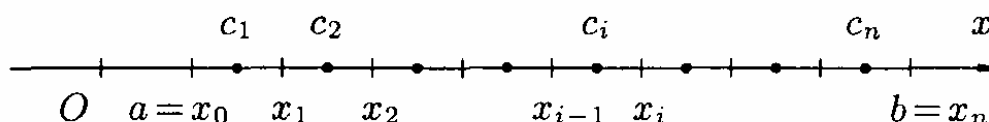


Рис. 1

Виконаємо наступні дії.

1. За допомогою точок $x_0 = a, x_1, x_2, \dots, x_n = b$ ($x_0 < x_1 < \dots < x_n$) **розіб'ємо** відрізок $[a, b]$ на n **частинних** відрізків $[x_0; x_1], [x_1; x_2], \dots, [x_{n-1}; x_n]$.
2. У кожному частинному відрізку $[x_{i-1}; x_i], i = 1, 2, \dots, n$ **виберемо довільну точку** $c_i \in [x_{i-1}; x_i]$ й обчислимо значення функції в ній, тобто знайдемо $f(c_i)$.
3. **Помножимо** знайдене значення функції $f(c_i)$ на довжину відповідного частинного відрізка $\Delta x_i = x_i - x_{i-1}$, тобто $f(c_i) \cdot \Delta x_i$.
4. **Складемо суму** S_n усіх можливих добутків:

$$S_n = f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \dots + f(c_n)\Delta x_n = \sum_{i=1}^n f(c_i)\Delta x_i. \quad (1)$$

Сума S_n виду (1) називається **інтегральною сумою** функції $y = f(x)$ на відрізку $[a; b]$.

5. **Знайдемо границю інтегральної суми** S_n (1) за умови, що $n \rightarrow \infty$ так, щоб $\lambda \rightarrow 0$, де λ – довжина найбільшого частинного відрізка, тобто $\lambda = \max \Delta x_i$ ($i = 1, 2, \dots, n$).
6. Якщо при цьому інтегральна сума S_n має границю, яка не залежить ні від способу розбиття відрізка $[a; b]$ на частинні відрізки, ні від вибору точок в них, то вона називається **визначеним інтегралом** від функції $y = f(x)$ на відрізку $[a; b]$ і позначається

$$\int_a^b f(x)dx = \lim_{\substack{n \rightarrow \infty \\ (\lambda \rightarrow 0)}} \sum_{i=1}^n f(c_i)\Delta x_i \quad (2)$$

Якщо опорний конспект формувати разом зі студентами протягом вивчення нової теми, зокрема визначеного інтеграла, й відновити у процесі актуалізації знань, тоді цей опорний конспект цілком може використовуватися в якості плану вивчення матеріалу і в подальшому, а саме – під час введення й засвоєння понять інших видів інтегралів: подвійного, потрійного, криволінійного 1 роду, криволінійного 2 роду, поверхневого 1 роду тощо. Поняття кратних, криволінійних та поверхневих інтегралів вводяться вже в другому семестрі вивчення курсу математичного аналізу, коли студенти є більш підготовленими і можуть брати активну участь у складанні плану-конспекту введення означень даних понять, а деякі з них можуть робити це й самостійно. Доцільно в плані-конспекті виділити *слова (фрази)-індикатори*, щоб студентам було легше засвоїти вищезазначені математичні поняття. У наведеному нами плані-конспекті ці індикатори виділені жирним курсивом. Для інших понять зазначені індикатори будуть такими ж і допоможуть студентам у кращому засвоєнні даних означень.

Більш того, результати наших спостережень показали, що така форма введення нових понять студентам спеціальності 126 «Інформаційні системи та технології» у

вигляді плану-конспекту виявилася найбільш ефективною, тобто відтворити означення даного поняття студентам саме у запропонованому вигляді найлегше.

Крім планів-конспектів під час вивчення нового матеріалу доцільно пропонувати студентам також короткі схеми, у яких зазначено основні формули чи випадки того чи того математичного факту.

Наприклад, під час пояснення теми «Частинні похідні вищих порядків функції кількох змінних» варто запропонувати студентам наступну схему (рис. 2).

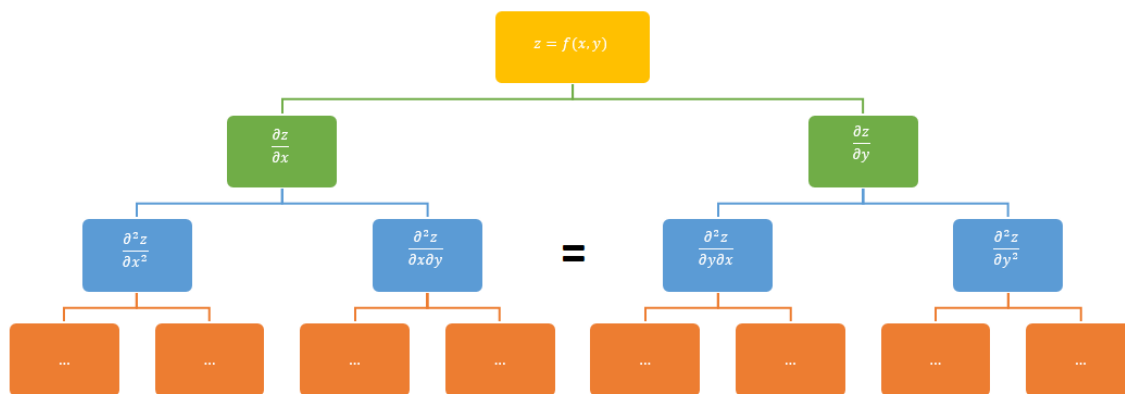


Рис. 2

Як показали наші спостереження, за допомогою саме такої схеми студентам набагато легше запам'ятати послідовність (алгоритм) відшукування похідних вищих порядків на прикладі функції двох змінних, а також, що є особливо важким для запам'ятовування – форма запису частинних похідних та спосіб їх відшукування відповідно до візуального образу.

Наприклад, для закріплення даного матеріалу доцільно запропонувати студентам наступне завдання.

Завдання 1. Довести, що $\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}$, якщо $z = e^x \cos y + e^y \cos x$.

Розв'язання будемо проводити за наступною схемою:

1. Знаходимо $\frac{\partial z}{\partial x} = e^x \cos y - e^y \sin x$.
2. Знаходимо $\frac{\partial^2 z}{\partial x \partial y} = -e^x \sin y - e^y \sin x$.
3. Знаходимо $\frac{\partial z}{\partial y} = -e^x \sin y + e^y \cos x$.
4. Знаходимо $\frac{\partial^2 z}{\partial y \partial x} = -e^x \sin y - e^y \sin x$.
5. Звідси маємо, що $\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}$, що й треба було довести.

Наведемо ще один приклад використання схем. Під час вивчення методів інтегрування, а саме методу інтегрування частинами, у студентів особливі утруднення виникають під час застосування цього методу в розв'язуванні завдань, наприклад, коли потрібно правильно вибрати функцію для заміни через змінну u . Студенти зазвичай роблять це несвідомо, або ж методом підбору, що не є правильним. І через певний час, або ж під час застосування в інших завданнях, студенти не можуть пригадати, яку саме

заміну треба зробити, тобто не можуть відтворити отримані знання в іншій ситуації. Тому для кращого засвоєння пропонуємо наступну схему (рис. 3).

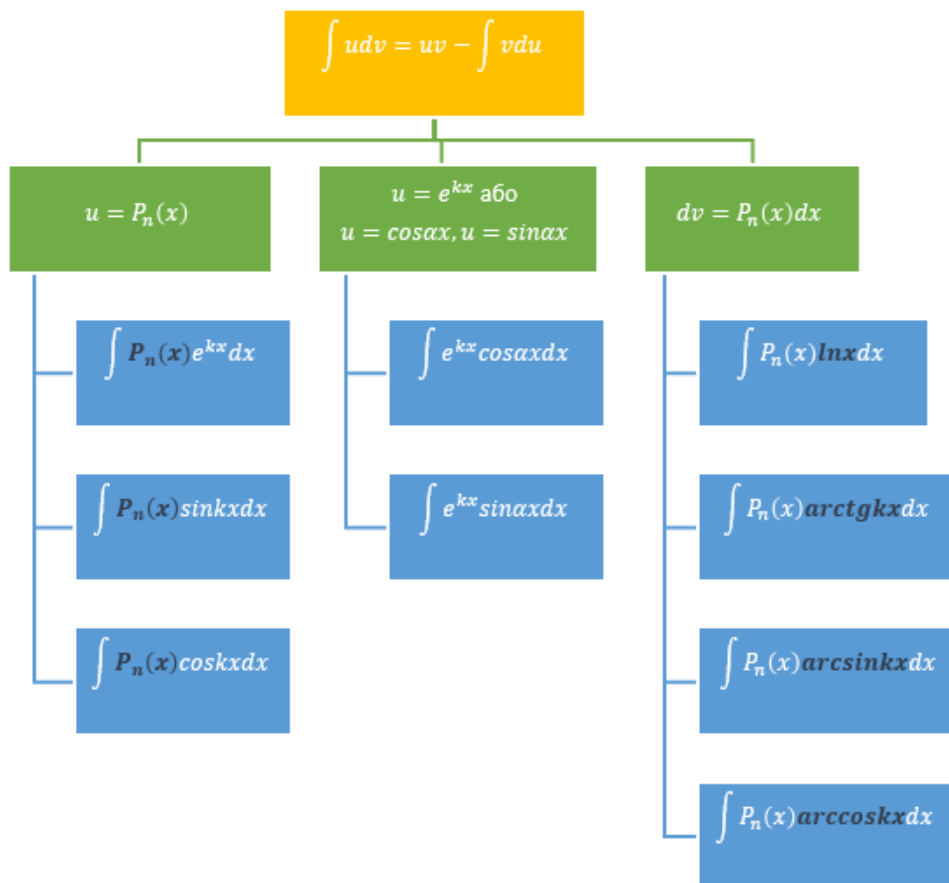


Рис. 3

Розв’язування задач алгоритмічного чи евристичного характеру є другим випадком навчальних ситуацій, коли діяльність студентів є схематизацією. Тут, використовуючи відомі схеми, студенти здебільшого отримують математичні результати окремого характеру.

Під час вивчення теми «Невласні інтеграли», наприклад, невластні інтеграли першого роду розв’язуються на основі засвоєних раніше операцій інтегрування та граничного переходу. Відповідні правила для студентів є об’єктами засвоєння. Отже, розв’язуючи такі завдання з опорою на відповідні правила, студенти виконують діяльність схематизації. Розглянемо етапи розв’язування, наступного завдання.

Завдання 2. Обчислити невластний інтеграл або встановити його розбіжність: $\int_1^{+\infty} \frac{dx}{x^2}$

Під час введення поняття невластного інтеграла першого роду ми запропонували студентам наступну схему (рис. 4) для того, щоб вони могли краще розібратися, як саме застосувати дане означення до розв’язування конкретних завдань.

Розв’язування завдання 2 доцільно супроводжувати такими роз’ясненнями:

1) перепишемо даний невластний інтеграл у вигляді границі, згідно з означенням:

$$\int_1^{+\infty} \frac{dx}{x^2} = \lim_{t \rightarrow +\infty} \int_1^t \frac{dx}{x^2};$$

2) обчислимо визначений інтеграл, користуючись методом безпосереднього інтегрування (за допомогою таблиці інтегралів): $\int_1^t \frac{dx}{x^2} = \int_1^t x^{-2} dx = -\frac{1}{x} \Big|_1^t = -\left(\frac{1}{t} - 1\right) = -\frac{1}{t} + 1$;

3) наступним кроком обчислимо границю отриманого виразу, враховуючи, що $t \rightarrow \infty$: $\lim_{t \rightarrow \infty} \left(-\frac{1}{t} + 1\right) = 0 + 1 = 1$;

4) на основі отриманого значення границі (у нас це 1) робимо висновок про те, що даний невластний інтеграл є збіжним і дорівнює 1.

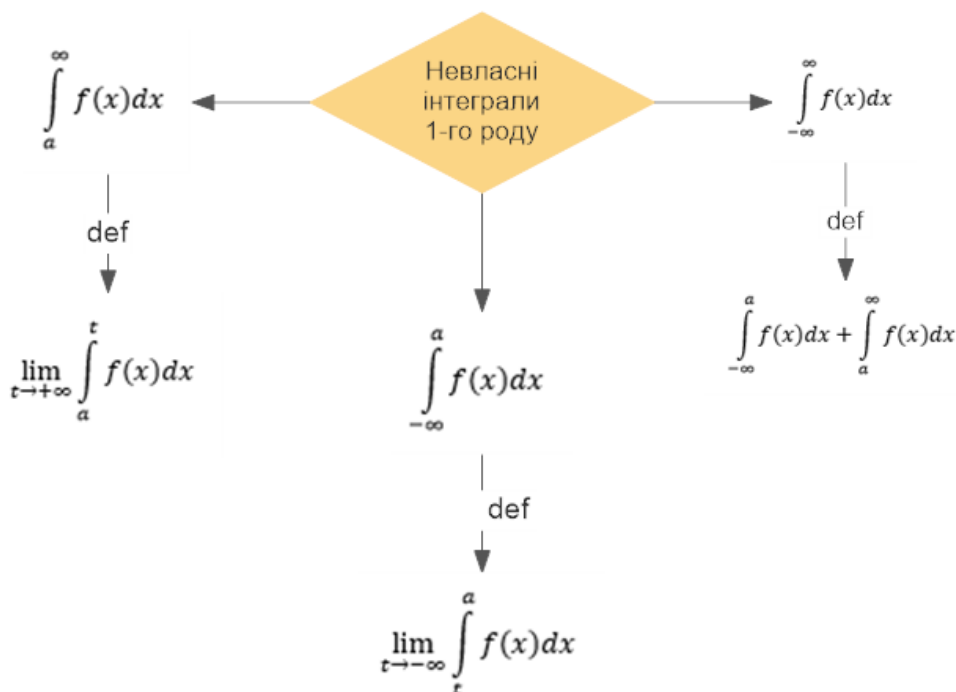


Рис. 4

У наведеному прикладі навчальної ситуації відома студентам схема діяльності представлена у словесній формі – через пояснення і обґрунтування.

Під час вивчення курсу математичного аналізу доцільно акцентувати увагу студентів на застосуванні отриманих ними теоретичних основ дисципліни, а саме – розв’язування задач. Достатньо посутню частку всіх розглядуваних задач складають задачі з геометричним та фізичним (чи то механічним) змістом. Їх розв’язування викликає у студентів значні утруднення, оскільки тут перетинаються дві складові – якісна та кількісна (правильний вибір і вміння застосовувати ті чи ті математичні формули). Тому спочатку доцільно добре відпрацювати саме вміння студентів працювати з математичними формулами.

Висновки. Подальші дослідження ми вбачаємо у розробці компетентісно орієнтованої потемної системи завдань з математичного аналізу з урахуванням специфіки підготовки майбутніх фахівців з аналізу даних.

Список використаної літератури:

- Сердюк З. О. Математичний тезаурус як інтелектуальний засіб навчання студентів фізичних спеціальностей ВНЗ / З. О. Сердюк, Т. М. Христенко // Засоби і технології сучасного навчального середовища // Матеріали всеукраїнської науково-практичної конференції (20-21 травня 2011 року, м. Кіровоград). – С. 78-79.
- Сердюк З. О. Особливості вивчення навчальної дисципліни «Математичний аналіз» для студентів

- фізичних спеціальностей ВНЗ / З. О. Сердюк // Актуальні проблеми і перспективи дидактики фізики // Матеріали Всеукраїнської науково-практичної конференції (26-28 квітня 2012 року, м. Черкаси). – Черкаси, ЧНУ ім. Б. Хмельницького. – С. 51-52.
3. Сердюк З. О. Реалізація компетентнісного підходу під час вивчення курсу математичного аналізу в ВНЗ. – Вісник Черкаського університету, Випуск № 8 (341): серія «Педагогічні науки». – Черкаси: Вид. від. ЧНУ ім. Б. Хмельницького, 2015. – С. 101-106.
 4. Стандарт вищої освіти України: перший (бакалаврський) рівень, галузь знань 12 – Інформаційні технології, спеціальність 126 – Інформаційні системи та технології / [Електронний ресурс]. – Режим доступу: <https://mon.gov.ua/ua/npa/pro-zatverdzhennya-standartu-vishoyi-osviti-za-specialnistyu-126-informacijni-sistemi-i-tehnologiyi-dlya-pershogo-bakalavrskogo-rivnya-vishoyi-osviti>.
 5. Бахрушин В. Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
 6. Тарасенкова Н. А. Використання знаково-символічних засобів у навчанні математики : [монографія] / Н. А. Тарасенкова. – Черкаси : Відлуння-плюс, 2002. – 400 с.
 7. Талызина Н. Ф. Педагогическая психология / Н. Ф. Талызина. – М. : Академия, 1999. – 288 с.

Bibliography:

1. Serdiuk, Z. (2011). Mathematical thesaurus as an intellectual means of teaching students of physical specialties of higher education. Means and technologies of modern educational environment. Proceedings of the All-Ukrainian scientific-practical conference, 78-79. [in Ukrainian]
2. Serdiuk, Z. (2012). Peculiarities of studying the discipline "Mathematical analysis" for students of physical specialties of higher education. Actual problems and prospects of didactics of physics. Proceedings of the All-Ukrainian scientific-practical conference, 51-52. [in Ukrainian]
3. Serdiuk, Z. (2015). Implementation of the competence approach during the study of the course of mathematical analysis in higher education. Bulletin of Cherkasy University, 8 (341): series "Pedagogical Sciences", 101-106. [in Ukrainian]
4. Standart of higher education of Ukraine: first (bachelor's) level, branch of knowledge 12 Information technologies, specialty 126 Information systems and technologies, (2020). Retrieved from <https://mon.gov.ua/ua/npa/pro-zatverdzhennya-standartu-vishoyi-osviti-za-specialnistyu-126-informacijni-sistemi-i-tehnologiyi-dlya-pershogo-bakalavrskogo-rivnya-vishoyi-osviti> [in Ukrainian]
5. Bakhrushin, V. (2011). Methods of data analysis: a textbook for students. [in Ukrainian]
6. Tarasenkova, N. (2002). The use of sign-symbolic means in teaching mathematics: [monograph] [in Ukrainian]
7. Talyzina, N. (1999). Pedagogical psychology. M.: Akademiya. [in Russian]

TARASENKOVA Nina,

Doctor of Science (Pedagogical Sciences), Professor, Head of the Department of Mathematics and Methods of Learning of Mathematics, Cherkasy Bohdan Khmelnytsky National University, Ukraine

SERDIUK Zoia,

PhD (Pedagogical Sciences), Associate Professor of the Department of Mathematics and Methods of Learning of Mathematics, Cherkasy Bohdan Khmelnytsky National University, Ukraine

PECULIARITIES OF TEACHING THE COURSE OF MATHEMATICAL ANALYSIS FOR SPECIALISTS IN DATA MINING

Summary. Introduction. *The course is aimed not only at mastering basic mathematical concepts and facts of differential and integral calculus of functions of one or more variables and their application to the study of functions, calculation of curve lengths, surface areas, moments of inertia and static moments, finding multiple and contour integrals, etc., but also to form in them the ability to reasonably and consistently reason, analyze, compare, summarize, etc., in general, the ability to draw correct conclusions and make realistic predictions, apply the acquired knowledge, skills and abilities to solve various practical problems. However, the specifics of the organization of the study of mathematical analysis, taking into account the peculiarities of the educational content of the specialty 126 «Information Systems and Technologies» has not yet been considered and remains the focus of our scientific research.*

Purpose – *to consider features of the organization of studying of a course of the mathematical analysis by students of a specialty 126 «Information systems and technologies» by means of schematization.*

Originality. *In this paper, given the latest trends in educational policy and the specifics of training specialists in the specialty 126 «Information Systems and Technologies», namely – the*

tendency to algorithmization in solving certain practical problems, it is advisable to use schematization when studying mathematical analysis by the above students – and during mastering the theoretical material, and in solving problems. Schematization as a sign-symbolic activity in the teaching of mathematical analysis to students-programmers occupies a significant place. During the acquaintance of students with the new material and during the application of knowledge, several types of learning situations are identified in which they encounter the activities of schematization.

Conclusion. We see further research in the development of a competency-based dark system of tasks in mathematical analysis, taking into account the specifics of training future professionals in data analysis.

Keywords: mathematical analysis, data mining, schematization, schemes, students-programmers.

Одержано редакцією 03.12.2019 р.
Прийнято до публікації 24.02.2020 р.

УДК 004.85:519.6

DOI 10.31651/2076-5886-2020-1-86-100

PACS 02.70.Wz, 07.05.Kf, 07.05.Mh,
07.05.Tr

КОВАЛЕНКО Олена Сергіївна,
магістрантка спеціальності «Прикладна
математика» Черкаського національного
університету імені Богдана
Хмельницького

СЕРДЮК Олександр Анатолійович,
кандидат економічних наук, старший
викладач кафедри прикладної математики
та інформатики Черкаського
національного університету імені Богдана
Хмельницького
e-mail: serdyuk@ukr.net
ORCID 0000-0002-3919-4661

ВИКОРИСТАННЯ КЛАСИЧНИХ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ КЛАСИФІКАЦІЇ ТЕКСТІВ У ПРОГРАМАХ ГЕНЕРАЦІЇ АВТОМАТИЧНИХ ВІДПОВІДЕЙ

У статті подано опис методів машинного навчання для розв'язання задач класифікації текстів з метою їх подальшого використання у програмах автоматичної генерації відповідей на основі аналізу контексту питань користувача. Розглянуто кроки підготовки та проведення класифікації текстів разом з фрагментами програмного коду мовою Python з використанням бібліотеки обробки природної мови. Проведено попередній огляд автоматичних генераторів відповідей та вибрано метод автоматичної генерації відповідей. Розроблено структуру програми автоматичної генерації відповідей та визначено напрямки подальшої роботи.

Ключові слова: машинне навчання, класифікація текстів, чат-боти, програмування, Python.

Вступ

Методи машинного навчання наразі широко використовуються для задач обробки природної, чи людської, мови (Natural Language Processing) з метою її розпізнавання, розуміння, інтерпретації та генерації комп'ютерними алгоритмами. У сучасному світі інформаційних технологій постійно збільшується спектр задач, що можуть використовувати результати інтерпретації людської мови, зокрема це автоматична класифікація текстів для класифікаторів електронних бібліотек, рубрикація новин,