

УДК 004.82:004:85

*Резнік Р.Ю., студент 1 курсу спеціальності  
«Комп'ютерні технології обробки даних  
(DataScience)»*

*Нескородєва Т.В., к.т.н., доцент, доцент  
кафедри комп'ютерних наук та  
інформаційних технологій*

## **КЛАСИФІКАЦІЯ ЗРАЗКІВ СКЛА НА ОСНОВІ ХІМІКО-ФІЗИЧНИХ ВЛАСТИВОСТЯХ МЕТОДАМИ СТАТИСТИЧНОГО НАВЧАННЯ**

*Донецький національний університет імені Василя Стуса, м. Вінниця*

Наш світ не ідеальний, щосекунди по всьому світі відбуваються злочини. Як правило на місці злочину залишаються докази, і нерідко це є зразки скла. Вірна класифікація зразків скла може допомогти криміналістам зрозуміти обставини злочину та в подальшому розкрити його.

Аналіз проведемо за допомогою пакету R. Розглянемо набір даних «Glass Identification»[1] який був спеціально сформований для таких цілей. В даному наборі даних міститься 9 числових предикторів:

- RI – коефіцієнт заломлення світла;
- Na: ваговий відсоток натрію у відповідному оксиді скла;
- Mg: ваговий відсоток магнію у відповідному оксиді скла;
- Al: ваговий відсоток алюмінію у відповідному оксиді скла;
- Si: ваговий відсоток силіцію у відповідному оксиді скла;
- K: ваговий відсоток калію у відповідному оксиді скла;
- Ca: ваговий відсоток кальцію у відповідному оксиді скла;
- Ba: ваговий відсоток барію у відповідному оксиді скла;
- Fe: ваговий відсоток феруму у відповідному оксиді скла,

та власне вихідна змінна яку потрібно буде класифікувати Type of glass (тип скла) яка може приймати значення: 1 – термополіроване будівельне скло; 2 – будівельне скло; 3 – термополіроване автомобільне скло; 5 – контейнер; 6 – посуд; 7 – скло з ліхтарів.

В даному випадку будемо досліджувати які з предикторів дозволяють як найточніше класифікувати зразки скла, для цього сформуємо матрицю діаграм розсіювання та кореляційну матрицю[2]:

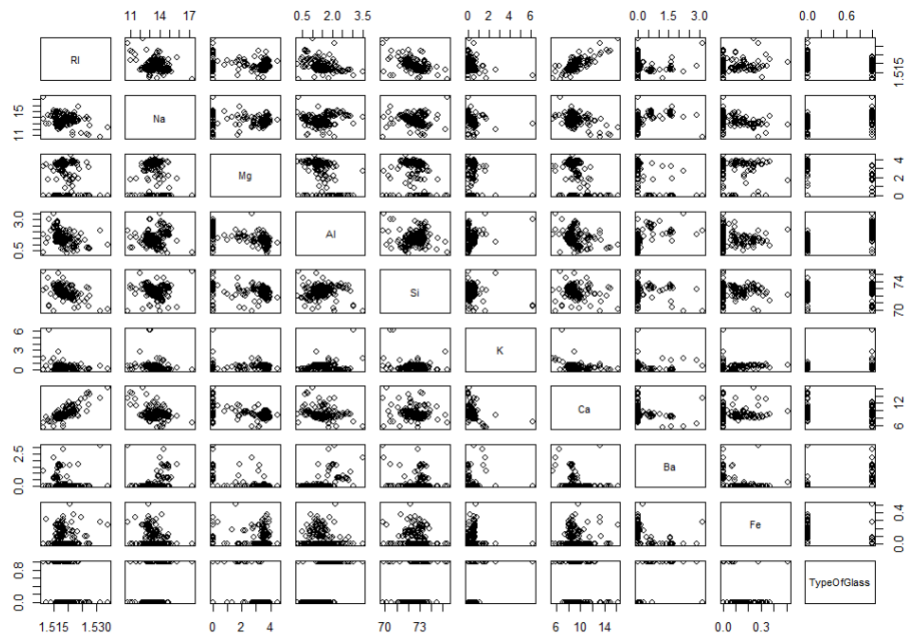


Рисунок 1 – Матриця діаграм розсіювання

	RI	Na	Mg	Al	Si
RI	1.000000000	-0.19188538	-0.122274039	-0.40732603	-0.54205220
Na	-0.1918853790	1.000000000	-0.273731961	0.15679367	-0.06980881
Mg	-0.1222740393	-0.27373196	1.000000000	-0.48179851	-0.16592672
Al	-0.4073260341	0.15679367	-0.481798509	1.000000000	-0.00552372
Si	-0.5420521997	-0.06980881	-0.165926723	-0.00552372	1.000000000
K	-0.2898327111	-0.26608650	0.005395667	0.32595845	-0.19333085
Ca	0.8104026963	-0.27544249	-0.443750026	-0.25959201	-0.20873215
Ba	-0.0003860189	0.32660288	-0.492262118	0.47940390	-0.10215131
Fe	0.1430096093	-0.24134641	0.083059529	-0.07440215	-0.09420073
TypeOfGlass	-0.1642372146	0.50289804	-0.744992888	0.59882921	0.15156526
	K	Ca	Ba	Fe	TypeOfGlass
RI	-0.289832711	0.8104026963	-0.0003860189	0.143009609	-0.1642372146
Na	-0.266086504	-0.2754424856	0.3266028795	-0.241346411	0.5028980423
Mg	0.005395667	-0.4437500264	-0.4922621178	0.083059529	-0.7449928875
Al	0.325958446	-0.2595920102	0.4794039017	-0.074402151	0.5988292084
Si	-0.193330854	-0.2087321537	-0.1021513105	-0.094200731	0.1515652579
K	1.000000000	-0.3178361547	-0.0426180594	-0.007719049	-0.0100544638
Ca	-0.317836155	1.0000000000	-0.1128409671	0.124968219	0.0009522246
Ba	-0.042618059	-0.1128409671	1.0000000000	-0.058691755	0.5751614590
Fe	-0.007719049	0.1249682190	-0.0586917554	1.000000000	-0.1882775640
TypeOfGlass	-0.010054464	0.0009522246	0.5751614590	-0.188277564	1.0000000000

Рисунок 2 – Кореляційна матриця

З (рис.1) та (рис.2) видно, що вихідна змінна Type of glass сильно корелює з предикторами Na, Mg, Al, Ba, тому виведемо двовимірні залежності відгука від цих предикторів[3].

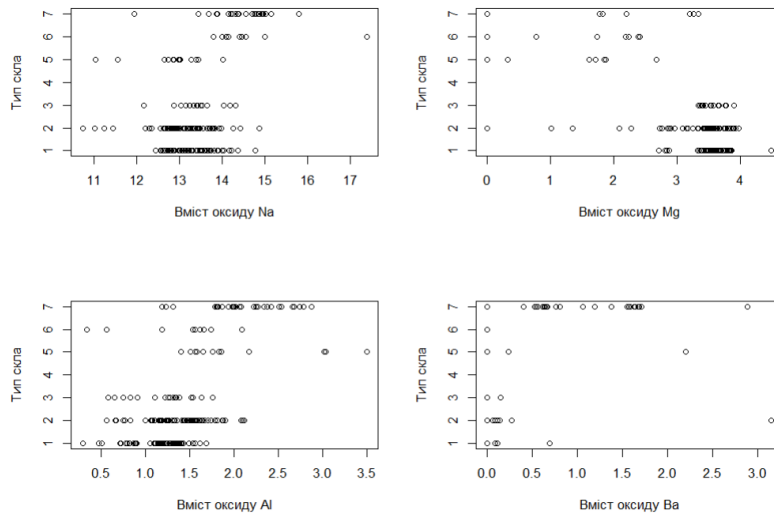


Рисунок 3 – Двовимірні залежності відгука від найзначущих предикторів

З (рис.3) видно, що зразки скла 1, 2, 3 та 5 класів мають менший вміст Na ніж 6 і 7. Також видно, що зразки скла 1, 2, та 3 класів містять великий вміст Mg а 5, 6 та 7 класи менший. По вмісту Al йде розділення на 2 «табори» 1, 2, 3 класи мають низький його вміст, а 5, 6, 7 вище середнього. А по вмісту Ba яскраво виділяється 7 клас – скло для ліхтарів.

Побудуємо класифікатор на основі тих змінних які виглядають найбільш тісно пов'язані з відгуком. Використаємо KNN-модель. Спершу я розділив дані на навчальні – для побудови моделі та тестові – для оцінки точності моделі.

```
[1] 0.271028
knn.pred  1  2  3  5  6  7
1  30  9  2  1  0  0
2   5 25  2  2  0  1
3   0  2  4  0  0  0
5   0  0  0  4  0  0
6   0  2  0  0  3  2
7   0  0  0  0  1 12
```

Рисунок 4 – Матриця сплутувань KNN моделі.

Дана модель помиляється в 27.1% випадків. З матриці сплутувань видно, що модель сильно путає між собою 1(термополіроване будівельне скло) та 2(будівельне скло) класи, також модель часто відносить автомобільні вікна до будівельних.

Для покращення даної моделі в подальшому можна використати декілька класифікаторів. Тобто, об'єднати в один клас всі будівельні вікна, і після того як перший класифікатор віднесе зразок скла до будівельного, подати цей зразок на інший класифікатор який буде розрізняти між собою лише будівельні вікна.

Список використаних джерел:

1. Glass Identification Data Set [Електронний ресурс] UC Irvine Machine Learning Repository. – Режим доступу до ресурсу: <https://archive.ics.uci.edu/ml/machine-learning-databases/glass/glass.data>
2. Gareth James, Trevor Hastie, Robert Tibshirani, Daniela Witten – «An Introduction to Statistical Learning: with Applications in R» – Springer, 2013. 456р.
3. Крохмалюк В.В. Нескородева Т.В. Аналіз даних про серцевні захворювання методами статистичного навчання. Матеріали всеукраїнської науково-практичної конференції для студентів, аспірантів та молодих вчених "Прикладні інформаційні технології" (29 квітня 2020 року) - Вінниця: ДонНУ імені Василя Стуса., с.29-33.

**УДК 519.2:004.8:004.62**

*Солодун Т.Р., студентка 3 курсу  
спеціальності 122 «Комп'ютерні науки»  
Потапова Н. А., к.е.н., доцент, доцент  
кафедри інформаційних технологій*

## **МАТЕМАТИЧНА ФОРМАЛІЗАЦІЯ ПРОЦЕСІВ В КОМП'ЮТЕРНОМУ МОДЕЛЮВАННІ**

*Донецький національний університет імені В. Стуса, м. Вінниця*

Моделювання є одним із способів пізнання та дослідження будь-якої предметної області. В основу моделювання покладено визначення моделі. Модель – це спосіб відображення найсуттєвіших характеристик об'єктів, що вивчають системи, явища або процеси; створюється для отримання інформації про об'єкт, необхідну для вирішення поставленої задачі.

Моделювання вважається мистецтвом та наукою. На початкових етапах моделювання виділяються ознаки досліджуваного об'єкта і дається розгорнутий змістовний опис зв'язку між ними – здійснюється так звана постановка задачі. Моделювання будь-якої системи неможливе без попередньої формалізації, адже це перший та дуже важливий етап створення інформаційної моделі. Точніше формалізація – етап переходу від текстового опису зв'язків між визначеними ознаками об'єкта до опису, що створюється за допомогою кодування. Її суть – описати і визначити певні характеристик видуманого об'єкта або процесу і спрогнозувати його поведінку, аби уникнути зупинок під час виконання певного етапу роботи у реальному часі та ситуації.

Одним із напрямів сучасних досліджень є комп'ютерне моделювання. В основу комп'ютерного моделювання покладається комп'ютерна імітація процесів. Імітаційне моделювання дозволяє здійснити множину прогнозів за різними сценаріями залежно від динамічного формування різноманітних ситуацій практично необмеженої складності. Необхідними умовами створення мінімалістичних моделей, адекватних реальним системам та поставленим