**ЗАТВЕРДЖЕНО**

Науково-методичною радою Державного університету «Житомирська політехніка»

протокол від \_\_ \_\_\_\_\_\_\_ 20\_\_ р. №\_\_

**МЕТОДИЧНІ РЕКОМЕНДАЦІЇ**

**для проведення практичних занять**

**з навчальної дисципліни**

**«Цифрове підприємництво в дії»**

для здобувачів вищої освіти освітнього ступеня «магістр»

спеціальності 051 «Економіка»

освітньо-професійна програма «Економіка»

факультет бізнесу та сфери обслуговування

(назва факультету)

кафедра цифрової економіки та міжнародних економічних відносин

(назва кафедри)

Рекомендовано на засіданні кафедри цифрової економіки та міжнародних економічних відносин

28 серпня 2021 р.,

протокол № 13

The Data Entrepreneurship in Action syllabus is developed in the framework of ERASMUS+ CBHE project “Digitalization of economic as an element of sustainable development of Ukraine and  Tajikistan” / DigEco 618270-EPP-1-2020-1-LT-EPPKA2-CBHE-JP

*This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained there in.*

*Цей проєкт фінансується за підтримки Європейської Комісії. Цей документ відображає лише погляди автора, і Комісія не несе відповідальності за будь-яке використання інформації, що міститься в документі.*

**Житомир  
2021-2022 н.р.**

Інформаційне підприємництво в дії  
[Електронний ресурс] : методичні вказівки до практичної роботи з дисципліни «Цифрове підприємництво в дії» для здобувачів вищої освіти «Магістр» за спеціальністю: 051 «Економіка». – Житомир : Державний університет «Житомирська політехніка», 2021. – Режим доступу:

Методичні вказівки до спрактичної роботи розроблено в рамках проєкту Erasmus+ “Діджиталізація економіки як елемент сталого розвитку України та Таджикистану (DigEco) 618270-EPP-1-2020-1-LT-EPPKA2-CBHE-JP”/ The proramm is developed in the framework of ERASMUS+ CBHE project “Digitalization of economic as an element of sustainable development of Ukraine and Tajikistan” / DigEco 618270-EPP-1-2020-1-LT-EPPKA2-CBHE-JP

Цей проект фінансується за підтримки Європейської Комісії. Цей документ відображає лише погляди автора, і Комісія не несе відповідальності за будь-яке використання інформації, що міститься в документі/This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained there in.

Розробник: аспірант кафедри інформаційних систем в управлінні та обліку, Дмитро Миколайович Захаров; доцент кафедри інформаційних систем в управлінні та обліку, кандидат економічних наук, доцент, Городиський Микола Петрович.

© Державний університет «Житомирська політехніка», 2021

© М. П. Городиський, 2021

***ПРИМІТКА. До початку виконання практичних робіт обдумайте вибір товару (послуги) та сформулюйте досліджувану підприємницьку ідею. Всі практичні роботи виконуються з огляду на зроблений вами вибір.***

**Змістовий модуль 1. Методи аналізу даних**

**Тема 1. Статистичні методи аналізу даних**

Практична робота 1.1

РОЗРАХУНОК СТАТИСТИЧНИХ ХАРАКТЕРИСТИК І ПОБУДОВА ГІСТОГРАМ

ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

ДИСПЕРСІЙНИЙ АНАЛІЗ

КОРЕЛЯЦІЙНИЙ АНАЛІЗ

Завдання:

**І. Розрахунок статистичних характеристик і побудова гістограм**

1. Підготувати вихідні дані для моделювання вибірки випадкових величин.

2. Виконати наступний алгоритм розрахунків:

2.1. Здійснити імітацію значень вибірки за допомогою відповідної методики.

2.2. Знайти кількість інтервалів, на який необхідно розбити масив вихідних даних, визначити ширину і границі інтервалів.

2.3. Підрахувати емпіричні частоти для кожного інтервалу.

2.4. Обчислити відносні частоти для кожного інтервалу.

2.5. Побудувати експериментальну функцію розподілу у виді гістограми.

2.6. Для обчислення теоретичних частот нормального розподілу знайти середини часткових інтервалів.

2.7. Обчислити середнє значення й оцінку середньоквадратичного відхилення, медіану і розмах.

2.8. Пронормувати границі інтервалів.

2.9. Використовуючи функцію Лапласа, знайти теоретичні ймовірності влучення в інтервали.

2.10. Визначити теоретичні частоти нормального розподілу.

2.11. Побудувати теоретичну криву розподілу.

3. Визначити значення статистичних характеристик і побудувати гістограми за допомогою пакету Statistica і додатка MS Excel.

4. Порівняти і проаналізувати результати, отримані по складеній програмі і за допомогою пакету Statistica і додатка MS Excel.

**ІІ. Перевірка статистичних гіпотез про однорідність вибірок**

1. Підготувати вихідні дані для вихідні дані для моделювання на ЕОМ послідовностей псевдовипадкових значень порівнюваних вибірок.

2. Виконати наступний алгоритм розрахунків:

2.1. Здійснити імітацію значень порівнюваних вибірок за допомогою методики, викладеної в Додатку А.

2.2. Розрахувати значення Т-критерію, F-критерію, G-критерію і U-критерію для порівнюваних вибірок.

3. Визначити значення критеріїв для перевірки однорідності порівнюваних вибірок за допомогою пакету Statistica і додатку MS Excel.

4. Порівняти і проаналізувати результати, отримані за складеною програмою і за допомогою пакету Statistica і додатку MS Excel.

**ІІІ. Дисперсійний аналіз**

1. Підготувати вихідні дані для дисперсійного аналізу.

2. Обробити вихідні дані методом дисперсійного аналізу.

3. Виконати розрахунок за допомогою модуля ANOVA/MANOVA з пакету Statistica.

4. Виконати розрахунок за допомогою засобу статистичного аналізу даних «Дисперсійний аналіз» з додатку MS Excel.

**ІІІ. Кореляційний аналіз**

1. Підготувати вихідні дані для моделювання на ЕОМ значень для вхідних факторів і вихідних параметрів досліджуваного об’єкта (процесу).

2. Виконати наступний алгоритм розрахунків:

2.1. Здійснити імітацію значень для вхідних факторів і вихідних параметрів за допомогою методики, викладеної в Додатку А.

2.2. Розрахувати значення парних коефіцієнтів кореляції між вхідними факторами, між вихідними параметрами і між вхідними факторами і вихідними параметрами.

2.3. Розрахувати коефіцієнти лінійних рівнянь регресії, що зв’язують вихідні параметри з вхідними факторами.

3. Провести кореляційний і регресійний аналізи для встановлення наявності зв’язку між вхідними і вихідними змінними досліджуваного об’єкта (процесу) і побудувати математичну модель за допомогою пакету Statistica і додатку MS Excel.

4. Порівняти і проаналізувати результати, отримані за складеною програмою і за допомогою пакету Statistica і додатку MS Excel.

МЕТОДИЧНІ ВКАЗІВКИ

***РОЗРАХУНОК СТАТИСТИЧНИХ ХАРАКТЕРИСТИК І ПОБУДОВА ГІСТОГРАМ***

*Використання вмонтованих статистичних функцій з пакету аналізу MS Excel*

В програмі MS Excel представлено багато вмонтованих статистичних функцій, за допомогою яких можна розрахувати всі статистичні характеристики, що зазвичай використовуються на практиці. Для виклику потрібної вмонтованої статистичної функції використовується команда Вставка Функція ... чи кнопка Вставка функції, і у вікні Майстер функцій вибрати категорію «статистичні» і вказати потрібну функцію. Після вибору і клацання по кнопці OK з’являється вікно Аргументи функції, за допомогою якого здійснюється введення адрес комірок таблиці із значеннями оброблюваного масиву і, при необхідності, інших необхідних величин (наприклад, ступенів вільності).

Для побудови гістограми використовується команда Сервіс Аналіз даних, після чого вибирається інструмент аналізу – Гістограма.

У вікні Гістограма задається вхідний інтервал, де вказуються адреси комірок з вихідними даними і інтервал кишень, де вказуються адреси комірок із значеннями границь інтервалів. Потім задаються бажані параметри виводу після чого потрібно клацнути кнопку OK.

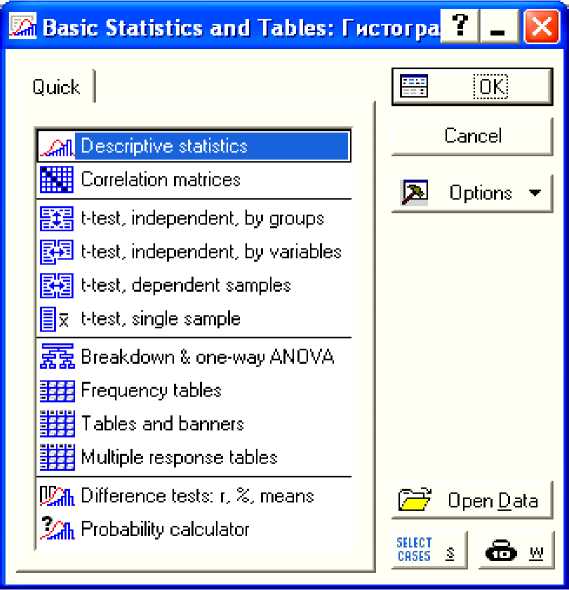
*Розрахунок статистичних характеристик і побудова гістограми в пакеті Statistica*

У версії пакета 6.0 розрахунок статистичних характеристик і побудова гістограми визначаються процедурою, що складається з наступних кроків:

Крок 1: Ввести чи імпортувати (наприклад, з Excel) вихідні дані в робочу книгу (Workbook) системи Statistica.

Крок 2: Виділити стовпець з уведеними чи імпортованими даними, для якого необхідно розрахувати статистичні характеристики і побудувати гістограму. При необхідності можна змінити назву стовпця.

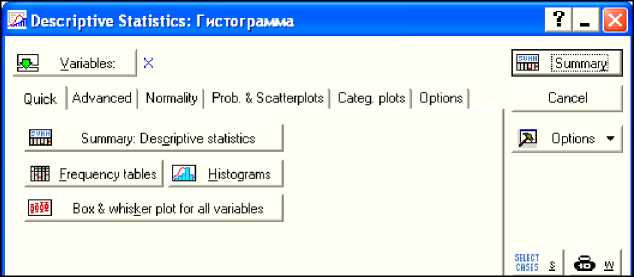
Крок 3: Клацнути по кнопці Start menu ..., що розташована в лівому нижньому куті вікна додатка і у меню вибрати Statistics Basic Statistics and Tables Descriptive Statistics.



Крок 4: Необхідні статистичні характеристики (описові статистики) задаються у вікні Descriptive Statistics на вкладці Advanced.



Крок 5: Для побудови таблиці з абсолютними і відносними частотами попадання даних в автоматично обраний інтервал і гістограми використовуються кнопки Frequency table, Histograms на вкладці Quick вікна Descriptive Statistics.



***Імітаційне моделювання псевдовипадкових чисел***

За допомогою функції RAND () програми Excel можна здійснити моделювання вибірок будь-якого обсягу з некорельованими або корельованими нормально-розподіленими псевдовипадковими числами.

Для моделювання незалежних нормально-розподіленого випадкового числа використовується формула:

=$A$1 + $B$1\*КОРІНЬ(-LN(СЛЧІС()))\*COS(6,28\*(RAND ()))

де в комірки $A$1 заноситься мат. ожіданіе $B$1 заноситься ср. кв. откл.

Для моделювання корельованих випадкових чисел:

1. Моделюється 1-е випадкове число з будь-яким математичним очікуванням і великим середньоквадратичним відхиленням (наприклад, в комірці D4)

2. Моделюється 2-е випадкове число з будь-яким математичним очікуванням (В$A$2) і невеликим середньоквадратичним відхиленням ($B$2) і до нього додається перше число (занесене, наприклад, в комірку D4).

= $ A $ 2 + $ B $ 2 \* КОРІНЬ (-LN (СЛЧІС ()))\* COS (6,28 \* (СЛЧІС ()))+ D4

3. Необхідна кількість чисел (вибірку обсягу n) отримують автозбереженням.

***ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ ПРО ОДНОРІДНІСТЬ ВИБІРОК***

*Використання вмонтованих статистичних функцій з пакету аналізу MS Excel*

Для перевірки статистичних гіпотез можуть бути використані наступні вмонтовані функції:

* ТТЕСТ – для визначення імовірності того, що дві вибірки узяті з генеральних сукупностей з однаковим математичним очікуванням;
* ZTECT – для визначення імовірності того, що вибірка узята з визначеної нормально розподіленої генеральної сукупності. Можна використовувати цю функцію щоб оцінити імовірність того, що конкретне спостереження взяте з конкретної генеральної сукупності;
* СТЬЮДРАСП – для розрахунку критичних значень розподілу Ст’юдента (*t*-розподілу);
* НОРМРАСП – повертає нормальну функцію розподілу для зазначеного середнього і стандартного відхилення;
* ДОВІРИТЬ – повертає довірчий інтервал для середньо визначеної нормально розподіленої генеральної сукупності;
* ФТЕСТ – повертає однобічну імовірність того, що дисперсії вибірок розрізняються;
* ХИ2ТЕСТ – повертає значення для розподілу хі-квадрат (використовується як критичне значення X2 -критерій Пірсона).

У засобах статистичного аналізу, що викликаються командою Аналіз даних меню Сервіс, для перевірки статистичних гіпотез пропонуються наступні інструменти:

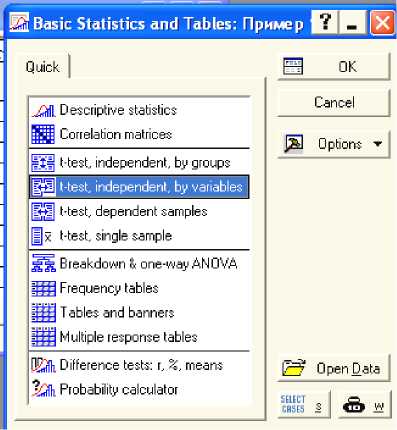
* Двовибірковий F-тест для дисперсії дозволяє перевірити гіпотезу про рівність дисперсій для двох вибірок;
* Парний двовибірковий t-тест для середніх: для перевірки гіпотези про рівність середніх для двох вибірок за допомогою *t*-критерію за умови рівності обсягу вибірок;
* Двовибірковий t-тест з однаковими дисперсіями: двовибірковий t-тест Стьюдента служить для перевірки гіпотези про рівність середніх для двох вибірок за умови рівності дисперсій;
* Двовибірковий t-тест із різними дисперсіями: двовибірковий *t*-тест Ст’юдента використовується для перевірки гіпотези про рівність середніх для двох вибірок за умови неоднорідності дисперсій. Вигляд вікон для *t*-тестів аналогічний як і для *F*-тесту;
* Z-тест: двовибірковий z-тест для середніх з відомими дисперсіями використовується для перевірки гіпотези про розходження між середніми двох генеральних сукупностей за умови, що дисперсії відомі, але гіпотези про їхню однорідність не перевірялися.

*Перевірка статистичних гіпотез у пакеті Statistica*

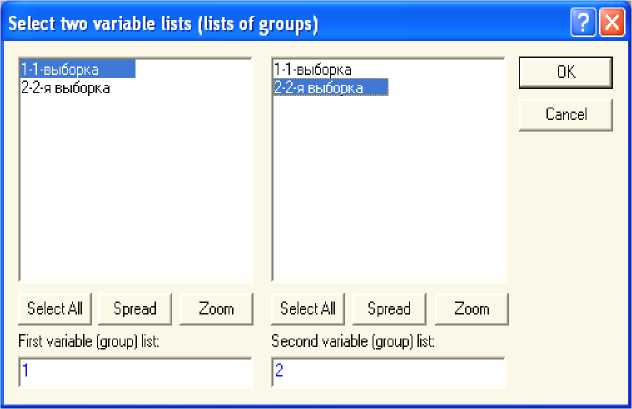
Розглянемо на прикладі перевірки гіпотез про однорідність середніх і дисперсій двох вибірок.

Крок 1: Ввести або імпортувати (наприклад, з Excel) вихідні дані в робочу книгу (Workbook) системи Statistica. При перевірці гіпотез у відношенні двох вибірок вводяться два масиви (групи) даних. При необхідності можна ввести назву для таблиці і змінити назву стовпця.

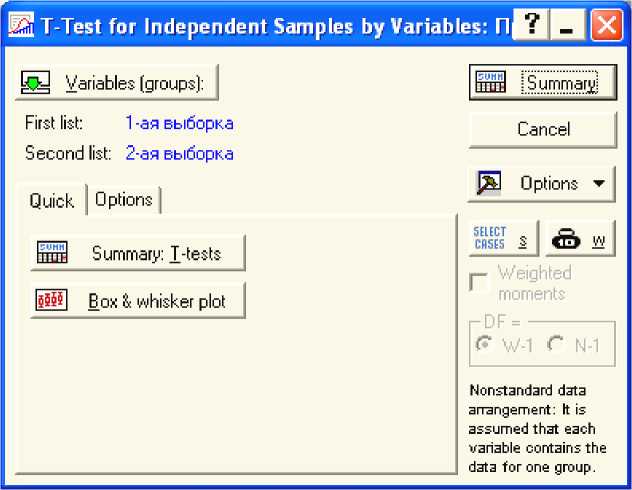
Крок 2: Клацнути по кнопці Start menu ..., що розташована в лівому нижньому куті вікна додатка і у меню вибрати Statistics Basic Statistics and Tables t-test, independent, by variables.



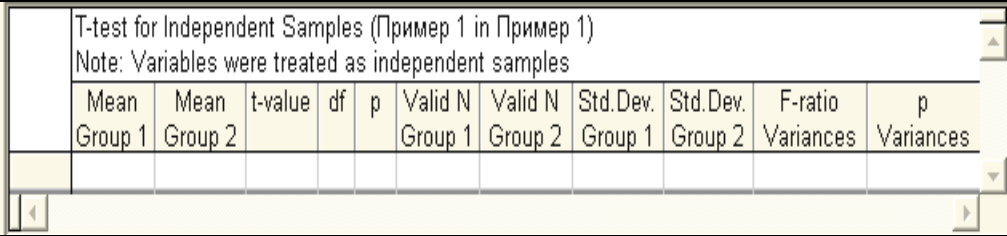
Крок 3: У вікні T-Test for Independent Samples by Variables після клацання по кнопці Variables (Groups) - вказати на стовпці з першою і другою вибіркою.



Крок 4: Для перевірки гіпотез про однорідність дисперсій і середніх вибирається кнопка Summary:T-tests.



Крок 5: Після клацання по кнопці Summary:T-tests з’являється вікно зі значеннями середніх, середньоквадратичних, обсягами вибірок і результатами перевірки гіпотез - розрахунковими значеннями *t*-критерію і *F* -критерію.



***ДИСПЕРСІЙНИЙ АНАЛІЗ***

***Використання пакету аналізу MS Excel***

Інструменти дисперсійного аналізу доступні через команду Аналіз даних меню Сервіс. Існує кілька видів дисперсійного аналізу. Необхідний варіант вибирається з урахуванням числа факторів і наявних вибірок з генеральної сукупності.

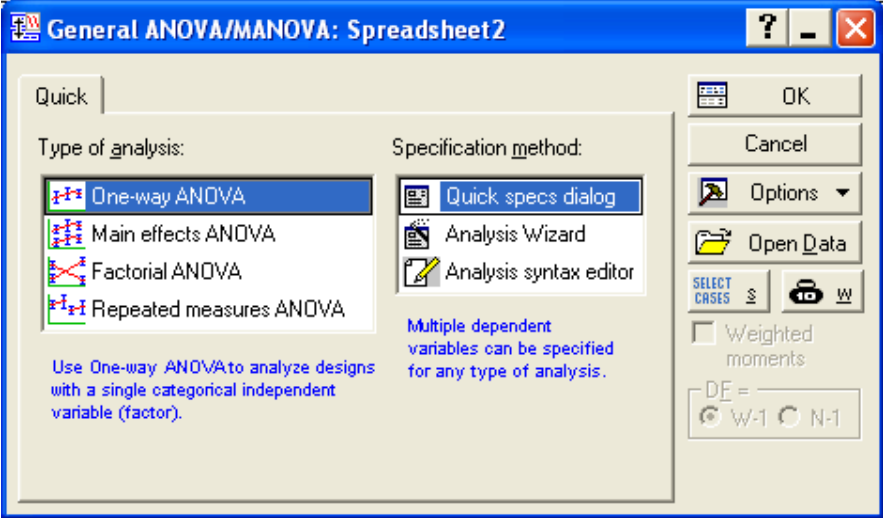
Однофакторний дисперсійний аналіз використовується для перевірки гіпотези про подібність середніх значень двох чи більше вибірок, що належать одній генеральної сукупності.

Двофакторний дисперсійний аналіз з повтореннями представляє собою більш складний варіант дисперсійного аналізу з декількома вибірками для кожної групи даних, його називають також дисперсійним аналізом при класифікації з групуванням. Двофакторний дисперсійний аналіз без повторення представляє собою двофакторний аналіз дисперсії, що не включає більше однієї вибірки на групу, його називають також дисперсійним аналізом при класифікація з пересічними факторами.

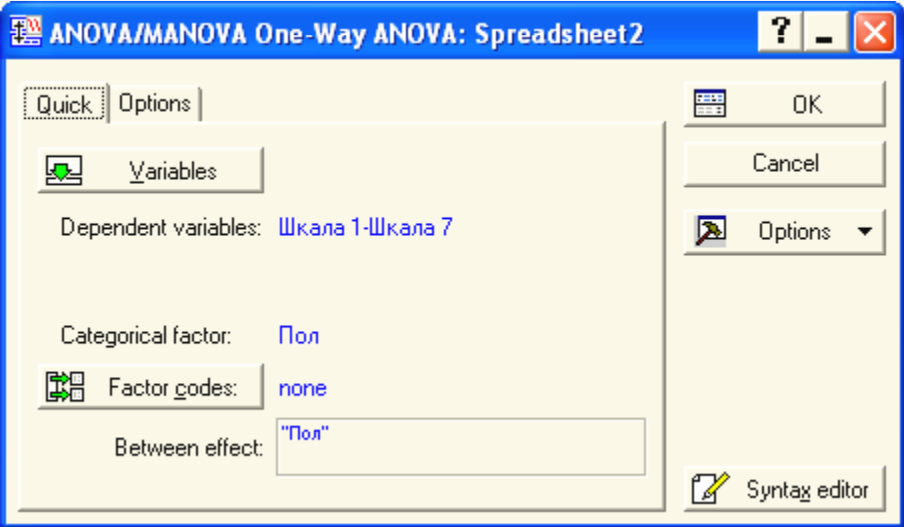
***Дисперсійний аналіз у пакеті Statistica***

Крок 1: Ввести або імпортувати (наприклад, з Excel) вихідні дані в робочу книгу (Workbook) системи STATISTICA, виділити їх, увести назву таблиці вихідних даних і назви змінних.

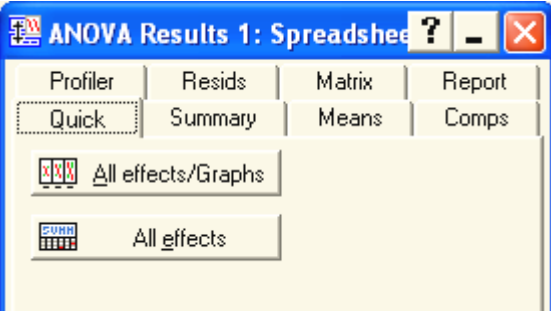
Крок 2: Клацнути по кнопці Start menu ..., розташованій в лівому нижньому куті вікна додатка і у меню вибрати Statistics → ANOVA. Після чого з’являється вікно General ANOVA/MANOVA → Quick.



Крок 3: Після вибору рядка One-way ANOVA з’являється відповідне вікно, за допомогою якого вибираються залежні змінні (dependent variable(s)) і змінні, що групують (categorical predictor variables).



Крок 4: Потім клацнути по кнопці OK і з’являється вікно ANOVA Results → Quick, де вибрати кнопку All effects, після чого і з’являється вікно Multivariate Tests of Significance з результатами аналізу



***КОРЕЛЯЦІЙНИЙ АНАЛІЗ***

***Використання вмонтованих статистичних функцій з пакету аналізу MS Excel***

Для проведення кореляційного аналізу можуть бути використані функції:

− *КОВАР* − повертає значення коваріації, тобто середнє добутків відхилень для кожної пари точок даних. Коваріація використовується для визначення зв’язку між двома множинами даних і дає можливість установити, чи асоційовані набори даних по величині, тобто великі значення з одного набору даних зв’язані з великими значеннями іншого набору (позитивна коваріація), або, навпаки, малі значення одного набору зв’язані з великими значеннями іншого (негативна коваріація), чи дані двох діапазонів ніяк не зв’язані (коваріація близька до нуля).

− *КОРРЕЛ* − повертає коефіцієнт кореляції. Коефіцієнт кореляції використовується для визначення наявності взаємозв’язку між двома вибірками.

У вікні Аналіз даних для кореляційного аналізу використовується інструмент − *Кореляція* та *Коваріація*.

Для проведення регресійного аналізу можуть бути використані функції:

− *ЛИНЕЙН* − розраховує за допомогою методу найменших квадратів параметри для лінійного рівняння регресії;

− *ПРЕДСКАЗ* − передбачає значення на основі рівняння лінійної регресії. Цю функцію можна використовувати для передбачення майбутніх продаж, потреб в устаткуванні чи тенденцій споживання;

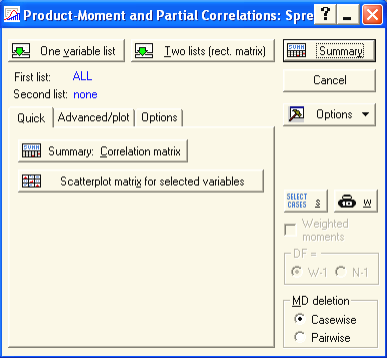
– *ТЕНДЕНЦІЯ* − повертає значення відповідно до лінійного тренду. У вікні Аналіз даних для регресійного аналізу використовується інструмент − Регресія, користуючись яким можна отримати таблиці з регресійною статистикою, коефіцієнтами регресії, стандартними помилками, *t*-статистикою та ін.

***Кореляційний аналіз у пакеті Statistica***

Крок 1: Увести або імпортувати (наприклад, з Excel) вихідні дані в робочу книгу (Workbook) системи Statistica, виділити їх, увести назву таблиці вихідних даних і назви змінних.

Крок 2: Клацнути по кнопці *Start menu* …, розташованій в лівому нижньому куті вікна додатка і у меню вибрати *Statistics → Basic Statistics and Tables → Correlation matrices*.

Крок 3: У вікні *Product-Moment and Partial Correlations → Quick* вибрати кнопку *Summary: Correlation matrix*, після чого з’являється вікно з кореляційною матрицею по обраним змінним.

****

Практична робота 1.2.

ФАКТОРНИЙ АНАЛІЗ

КЛАСИФІКАЦІЇ ДАНИХ

ПОБУДОВА Й ДОСЛІДЖЕННЯ РЕГРЕСІЙНИХ МОДЕЛЕЙ

Завдання:

**І. Факторний аналіз**

1. Підготувати вихідні дані для моделювання експерименту.

2. Виконати наступний алгоритм розрахунків:

2.1. Скласти матрицю планування ПФЕ для заданого числа факторів.

2.2. Здійснити імітацію експерименту на ЕОМ скориставшись методикою, викладеної в Додатку А.

2.3. Заповнити експериментальними даними, отриманими в результаті імітаційного моделювання, матрицю планування.

2.4. Обчислити коефіцієнти рівняння регресії;

2.5. Зробити статистичну оцінку відтворюваності досвідів і значимості коефіцієнтів рівняння регресії.

2.6. Перевірити адекватність рівняння регресії.

2.7. Перейти від рівняння регресії в нормованому виді до рівняння в звичайному масштабі.

3. Використовуючи отримане рівняння регресії, передбачити значення цільової функції для заданого викладачем набору значень вхідних факторів.

4. За допомогою пакету Statistica скласти матрицю планування експериментів, заповнити її експериментальними даними, отриманими в результаті імітаційного моделювання, обробити і проаналізувати результати експерименту.

5. Порівняти і проаналізувати результати, отримані за складеною програмою і за допомогою пакету Statistica.

**ІІ. Класифікація даних**

**ІІІ. Побудова й дослідження регресійних моделей**

1. Підготувати вихідні дані для моделювання на ЕОМ значень для вхідних факторів і вихідних параметрів досліджуваного об’єкта (процесу).

2. Виконати наступний алгоритм розрахунків:

2.1. Здійснити імітацію значень для вхідних факторів і вихідних параметрів за допомогою методики, викладеної в Додатку А.

2.2. Розрахувати значення парних коефіцієнтів кореляції між вхідними факторами, між вихідними параметрами і між вхідними факторами і вихідними параметрами.

2.3. Розрахувати коефіцієнти лінійних рівнянь регресії, що зв’язують вихідні параметри з вхідними факторами.

3. Провести кореляційний і регресійний аналізи для встановлення наявності зв’язку між вхідними і вихідними змінними досліджуваного об’єкта (процесу) і побудувати математичну модель за допомогою пакету Statistica і додатку MS Excel.

4. Порівняти і проаналізувати результати, отримані за складеною програмою і за допомогою пакету Statistica і додатку MS Excel.

МЕТОДИЧНІ ВКАЗІВКИ

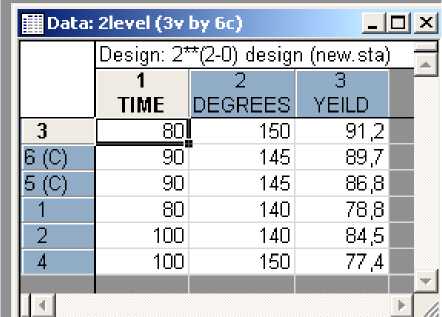
***ПЛАНУВАННЯ ПОВНИХ ФАКТОРНИХ ЕКСПЕРИМЕНТІВ***

*Планування експерименту в пакеті Statistica*

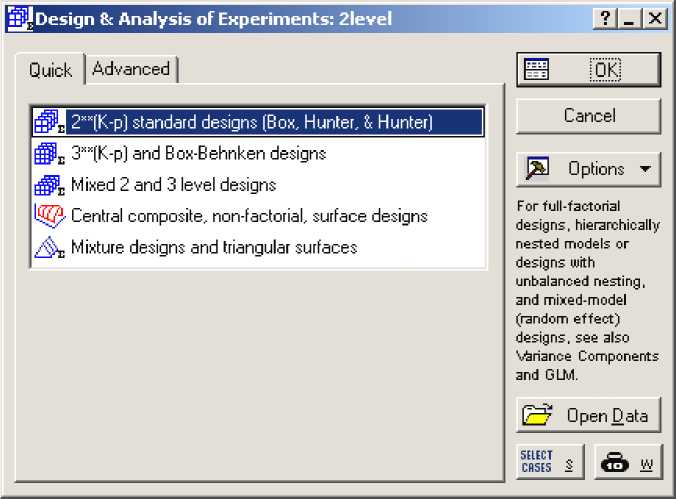
В пакеті Statistica є могутній модуль планування експериментів, що дозволяє здійснити планування повних і дробових факторних експериментів з варіюванням факторів на 2-х і/чи 3-х рівнях, побудувати центральні композиційні плани, латинські і греко- латинські квадрати, робастні (стійкі) плани і ряд інших планів.

Обмежимося розглядом тільки технології побудови планів ПФЕ та їхнім аналізом.

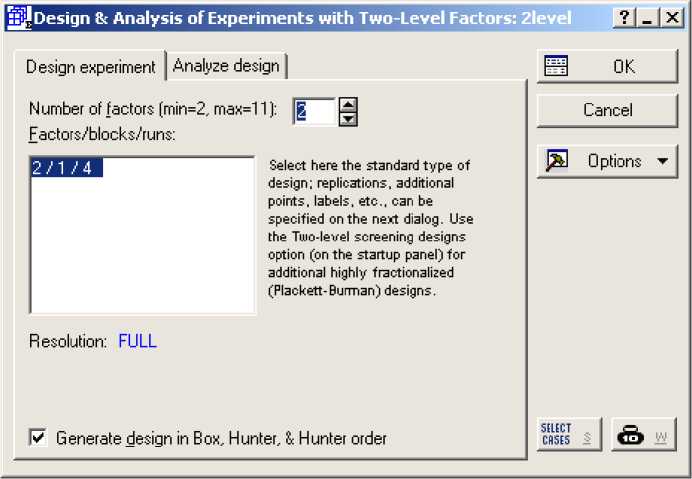
Крок 1: Ввести або імпортувати (наприклад, з Excel) вихідні дані в робочу книгу (Workbook) системи Statistica, виділити їх, увести назву таблиці вихідних даних і назви змінних. Або можна вибрати файл із уже існуючих, для прикладу візьмемо файл 2level.sta, що знаходиться в папці Examples і поставляється разом із системою Statistica.



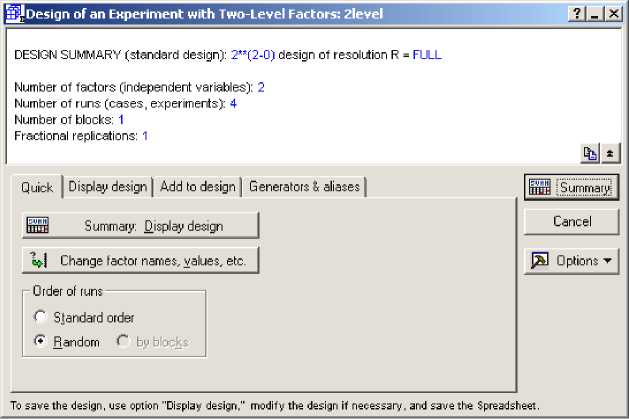
Крок 2: Клацнути по кнопці Start menu ..., розташованої в лівому нижньому куті вікна додатка і у меню вибрати Statistics (Industrial Statistics & Six Sigma (Experimental Design (DOE). Після чого з’являється вікно Design and Analysis of Experiments, у якому після вибору потрібного плану (у даному прикладі вибирається план 2\*\*(K-p) standard designs (Box, Hunter & Hunter) клацнути по кнопці OK.

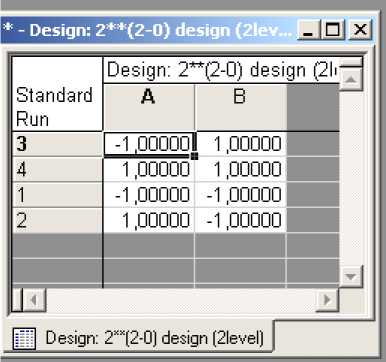


*Крок 3: У вікні* Design & Analysis of Experiments with Two-Level Factors ^ Design Experiment *задати на лічильнику* Number of factors (min=2, max=11) *(у даному прикладі число факторів - 2) і клацнути по кнопці OK.*

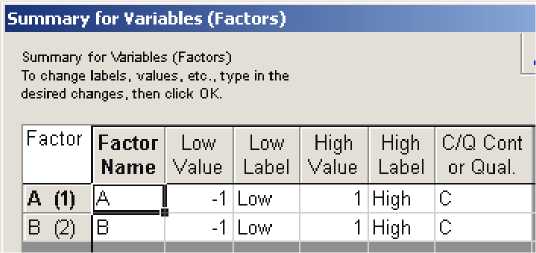


*Крок 4: У вікні* Design of an Experiment with Two-Level Factors → Quick вибір кнопки Summary: Display design дозволяє вивести вікно з матрицею планування ПФЕ (у даному прикладі це матриця виду 22).

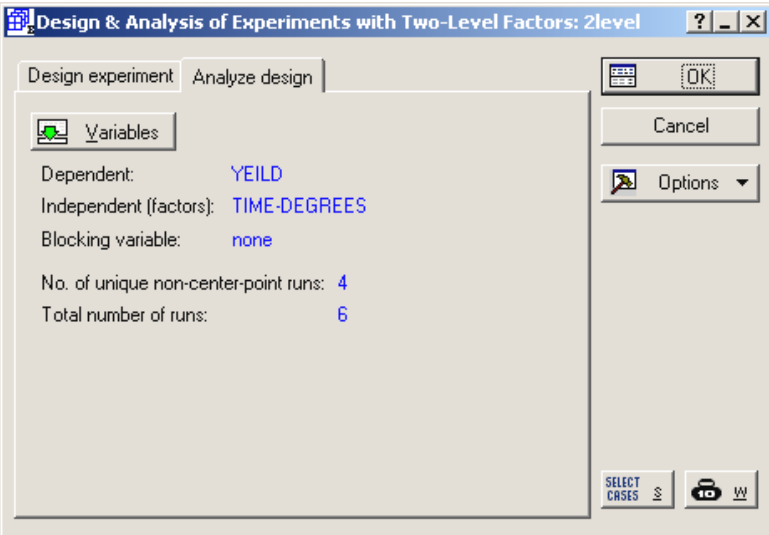


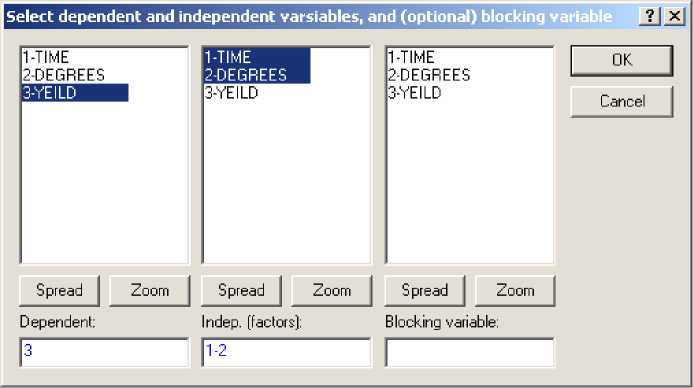


Вибір кнопки Change factor names, values, etc показує кодування факторів у матриці планування.



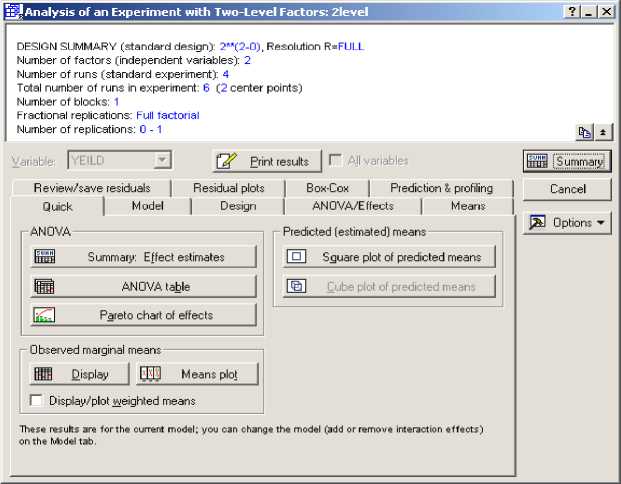
*Крок 5: Повернутися до вікна* Design & Analysis of Experiments with Two-Level Factors, *в якому обрати вкладку -* Analyze Design Tab *і клацнути по кнопці OK. У вікні* Design & Analysis of Experiments with Two-Level Factors → Analyze Design клацнути по кнопці Variables і визначити залежну змінну (dependent), відгук і незалежні змінні (indeptndent (factors)) фактори. У даному прикладі визначення змінних показане в наступному вікні.





Клацнути по кнопці OK.

Крок 6: За допомогою вікна, що з’явилося, можна зробити всебічний аналіз результатів експерименту за досліджуваним планом. Для даного прикладу це вікно буде мати наступний вид:



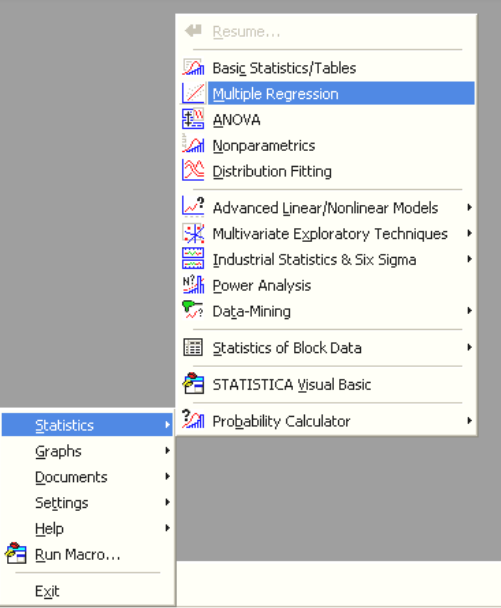
Вибір кнопки Summary: Effect estimates виводить вікно з ефективними коефіцієнтами моделі.



***Регресійний аналіз у пакеті Statistica***

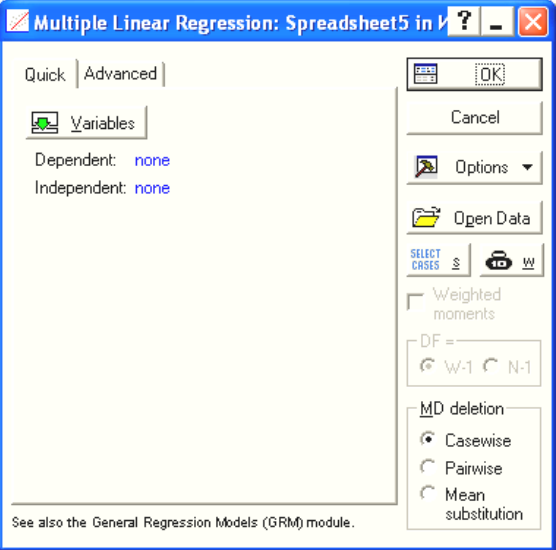
Крок 1: Ввести чи імпортувати (наприклад, з Excel) вихідні дані в робочу книгу (Workbook) системи Statistica, виділити їх, увести назву таблиці вихідних даних і назви змінних.

Крок 2: Клацнути по кнопці *Start menu* ..., розташованій в лівому нижньому куті вікна додатка і у меню вибрати *Statistics → Multiple Regression*.

****

Крок 3: У вікні *Multiple Linear Regression*: ... клацнути по кнопці *Variables*, після чого з’являється вікно *Select dependent and independent variable list*, в якому необхідно задати вихідну (залежну) змінну і вхідні (незалежні) змінні.

Крок 4: Клацнути по кнопці OK і з’являється вікно *Multiple Regression Results*, у якому відбиті результати аналізу і знаходяться вкладки, що дозволяють одержати більш детальну інформацію. За допомогою кнопки *Summary: Regression results* на вкладці *Quick* викликається вікно *Regression Summary for Dependent Variable*, у якому виведені коефіцієнти рівняння регресії і їхніх оцінок.

****

***Тема 2. Методи інтелектуального аналізу даних***

Практична робота 2.1.

ПРОГРАМА WEKA. НАБІР «ІРИС ФІШЕРА»

КРОС-ВАЛІДАЦІЯ

АЛГОРИТМИ ZeroR, OneR

СПРОЩЕНИЙ АЛГОРИТМ БАЙЄСА

WEKA EXPERIMENTER

Завдання:

**І. Програма Weka. Набір «ірис Фішера»**

1. Розпочніть роботу в Explorer з вкладки Preprocess. Відкрийте підготовлений файл з набором даних для аналізу кнопкою Open file ...

2. З наборів даних відкрийте iris.arff.

3. У вікні Preprocess знайдіть інформацію про набір – скільки в ньому елементів, скільки атрибутів в описі кожного елемента. Переходячи за списком атрибутів, опишіть, якого типу їх значення, в яких діапазонах змінюються. Для атрибута class напишіть, які класи визначені для елементів набору, скільки елементів в кожному класі. За кнопкою Edit можна переглянути або відредагувати вміст набору.

4. Перейдіть на вкладку Classify. За умовчанням вибраний класифікатор ZeroR. Натисніть кнопку Choose, в списку алгоритмів виберіть trees і J48. У розділі Test Options виберіть Percentage split і поставте значення 80. Натисніть Start, щоб провести навчання.

5. Ознайомтеся зі звітом. Опишіть, скільки елементів з тестової вибірки було правильно розцінена. Ознайомтеся з побудованим деревом рішень. Опишіть отриманий класифікатор.

5. У вікні Classify знайдіть кнопку More options, змініть параметр Random seed for XVal /% Split з 1 на 2. Запустіть процес навчання класифікатора. Знайдіть в звіті, елемент якого класу був неправильно класифікований (вихідний клас і клас, призначений класифікатором).

6. Встановіть параметр Random seed for XVal /% Split назад в 1. Повторіть експеримент з розбиття на навчальний і тестовий набори в пропорції 60/40. Дайте відповідь на питання: Як змінилася точність класифікатора? З чим ви пов’язуєте зміну точності?

7. Порівняйте три отриманих дерева рішень. Дайте відповідь на питання: Чи є в них відмінності? Чи можна говорити про те, що зниження точності прогнозу в останньому експерименті пов’язане з недоліком даних для навчання?

**ІІ. Крос-валідація. алгоритми ZEROR, ONER. Спрощений алгоритм Байєса. Weka experimenter**

1. Для набору iris.arff і алгоритму J48 з параметрами за замовчуванням проведіть навчання класифікатора з крос-валідація. Порівняйте модель і оцінки точності, з моделлю і оцінками, отриманими в попередній роботі.

2. Для набору даних iris.arff виконайте класифікацію з використанням алгоритмів ZeroR і OneR (обидва алгоритму засновані на правилах і в WEKA вони розташовані в WEKA → Classifiers → Rules). Виконайте навчання при використанні налаштувань за замовчуванням і крос-валідації, опишіть в звіті отримані моделі і їх точність.

3. На вкладці Classify при обраному алгоритмі OneR клацніть по рядку з назвою і параметрами алгоритму. Натисніть кнопку More, ознайомтеся з описом використовуваного алгоритму і його параметрів. Коротко опишіть в звіті.

4. У WEKA Explorer завантажте набір vote.arff. Він містить дані про результати голосування в Конгресі США в 1980-х рр. Опишіть в звіті цей набір даних – скільки в ньому елементів, скільки атрибутів в описі кожного елемента, які значення можуть приймати атрибути, які значення може приймати атрибут, що містить клас елемента.

5. Для набору даних vote.arff виконайте класифікацію за допомогою спрощеного алгоритму Байеса (classifiers → bayes → NaiveBayes) c параметрами за замовчуванням і крос-валідація. Опишіть отриману модель, охарактеризуйте її точність. Аналогічно завдання 3 познайомтеся з довідковою інформацією по параметрам алгоритму. Дайте відповідь на питання: Яким параметром можна скористатися, якщо ми захочемо використовувати спрощений алгоритм Байеса на наборі даних з речовими числовими значеннями незалежних атрибутів (такими як в іриси Фішера)?

6. У вікні Weka GUI натисніть кнопку Experimenter. У вікні програми Experimenter на вкладці Setup натисніть кнопку New (якби у нас був файл з описом існуючого експерименту, можна було б його відкрити по кнопці Open). Вирішимо задачу порівняння точності алгоритмів NaiveBaeys, J48, OneR, ZeroR з параметрами за замовчуванням при класифікації набору vote.arff.

У вікні Datasets додайте набір vote.arff, у вікні Algorithms – перераховані вище алгоритми в зазначеному порядку. Інші налаштування залишаємо без змін. Для оцінки точності буде використовуватися крос-валідація на 10 підмножинах. Експеримент буде проводитися 10 разів з різними значеннями, не започатковано генератор випадкових чисел.

Перейдіть на вкладку Run і натисніть кнопку Start, щоб запустити експеримент. Коли він буде закінчений, перейдіть на вкладку Analyse, де натисніть кнопку Experiment (аналізується поточний експеримент), і далі – кнопку Perform test.

Зверніть увагу на налаштування significance (рівень статистичної значущості), яка за замовчуванням дорівнює 0,05.

6. Повторіть описаний експеримент, перш ніж можна буде його результатів на вкладці Analyse відзначте галочкою пункт Show std. deviations і після цього натисніть кнопку Perform test. У звіті Experimenter додатково буде відображатися значення стандартного відхилення. Опишіть отримані результати в звіті про лабораторну роботу. Дайте відповідь на питання: Які рекомендації можна дати на їх основі?

7. На вкладці Analyse натисніть кнопку Test base і в якості основи для порівняння виберіть алгоритм OneR. Не повторюючи весь експеримент, натисніть кнопку Perform test. За новим звітом визначте, чи є алгоритми, точність яких перевищує OneR. Дайте відповідь на питання: Чи є алгоритми, перевага яких статистично значимо для заданого за умовчанням рівня 5% (0,05)? Чи є алгоритми, які OneR перевершує за точністю, і це перевага статистично значимо?

МЕТОДИЧНІ ВКАЗІВКИ

***ПРОГРАМА WEKA. НАБІР «ІРИС ФІШЕРА»***

WEKA (Waikato Environment for Knowledge Analysis) – це вільне програмне забезпечення, що розробляється в Новій Зеландії в Університеті Waikato і включає написані на мові Java бібліотеки для аналізу даних і програмні засоби, їх використовують. Назва програми збігається з назвою птаха, що живе в Новій Зеландії. Звідси і зображення птаха, що є символом цього проекту.

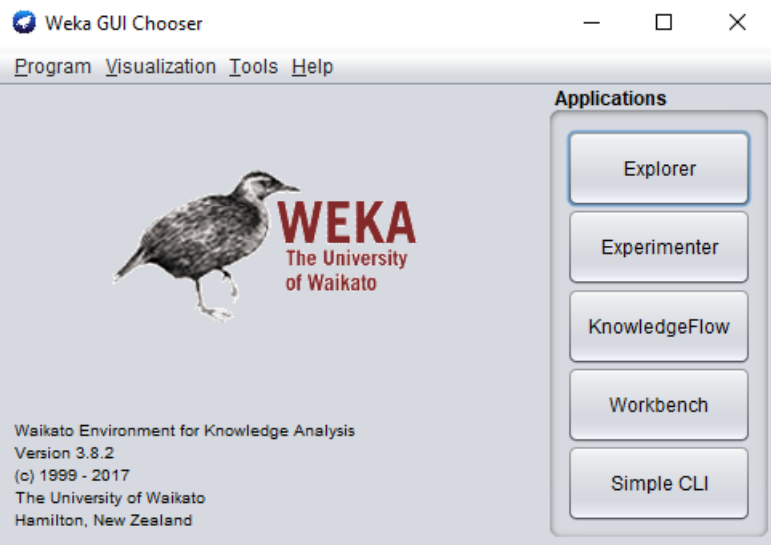
Головна сторінка проекту: <https://www.cs.waikato.ac.nz/ml/weka/index.html>.

У Windows після установки йдемо в основне меню, знаходимо папку Weka 3.8.2 (номер залежить від встановленої версії) і запускаємо Weka 3.8 (номер залежить від версії). Якщо програма встановилася коректно і у вас встановлена ​​відповідна версія Java, то з’явиться вікно Weka GUI (рис. 2.1.1).

З набору кнопок з назвами додатків вибираємо Explorer – це одна з програм, що забезпечують графічний інтерфейс для роботи з бібліотеками машинного навчання WEKA.

Робота в Explorer починається з вкладки Preprocess (рис. 2.1.2). Нам потрібно буде відкрити підготовлений файл з набором даних для аналізу. Натисніть кнопку Open file ...

WEKA Explorer в Windows за замовчуванням відкриває папку «Мої документи» (Documents). Щоб було швидше працювати, має сенс скопіювати в неї з C: \ Program Files \ Weka-3-8 (або іншого каталогу, куди встановилася програма) папку data з демонстраційними наборами даних. В «Моїх документах» цю папку краще назвати WEKAdata (або якось в цьому стилі), щоб відразу було зрозуміло, про що мова.



Pіc. 2.1.1. Weka GUI

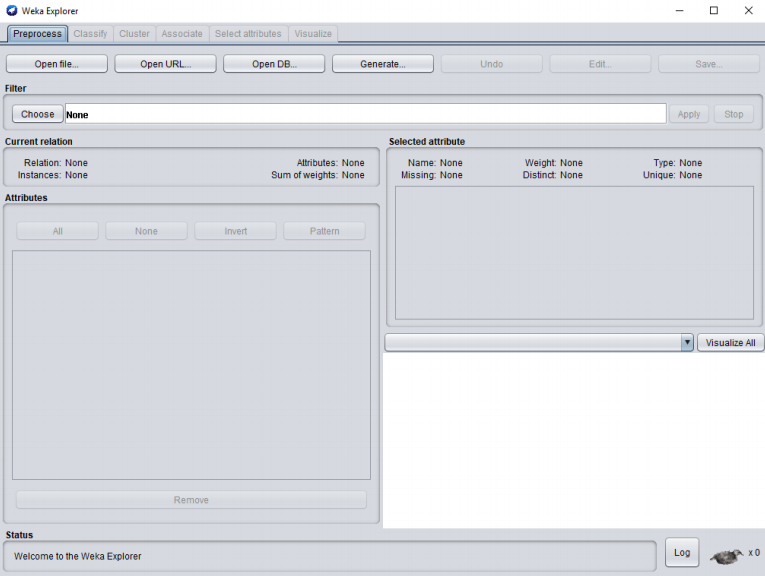


Рис. 2.1.2. Додаток WEKA Explorer

З наборів даних потрібно відкрити iris.arff. Цей файл містить відомий набір, званий «Іриси Фішера». Опис набору можна легко знайти в Інтернеті, набравши в пошуковій системі назву. Щоб було зрозуміло, про що мова, прочитайте один з описів, краще з фотографіями ірисів.

Завдання 3. У вікні Preprocess знайдіть інформацію про набір – скільки в ньому елементів, скільки атрибутів в описі кожного елемента. Переходячи за списком атрибутів, опишіть, якого типу їх значення, в яких діапазонах змінюються. Для атрибута class напишіть, які класи визначені для елементів набору, скільки елементів в кожному класі. За кнопкою Edit можна переглянути або відредагувати вміст набору.

Тепер, коли ми трохи познайомилися з набором даних, створимо на його основі модель для класифікації ірисів по зазначеним в наборі ознаками. Будемо використовувати дерева рішень. Детальніше про це класі алгоритмів можна дізнатися з посібників.

Завдання 4. Перейдіть на вкладку Classify. За умовчанням вибраний класифікатор ZeroR. Натисніть кнопку Choose, в списку алгоритмів виберіть trees і J48. У розділі Test Options виберіть Percentage split і поставте значення 80 (рис. 2.1.3). Обрана опція говорить про те, що набір даних буде розбитий на дві частини – 80% для навчання моделі, 20% – для тестування. Натисніть Start, щоб провести навчання.

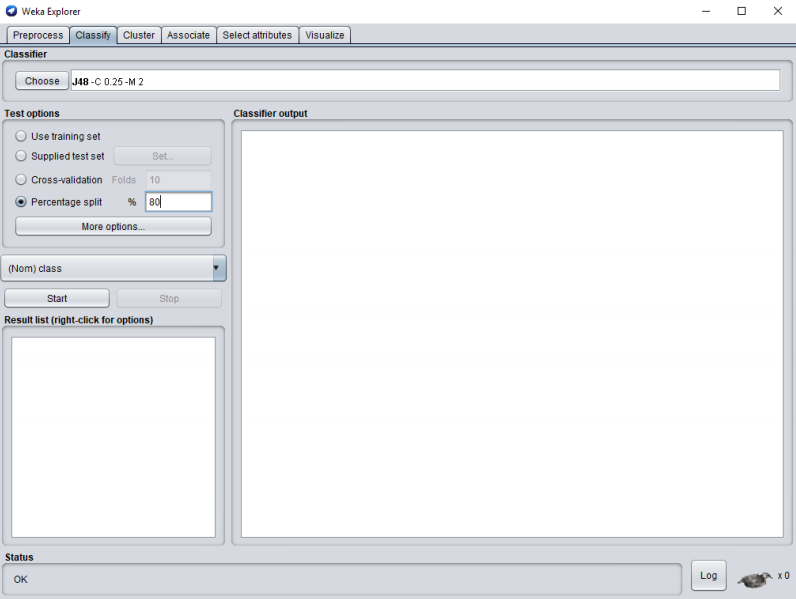


Рис. 2.1.3. Налаштування для використання J48 і розбиття 80/20

Завдання 5. Ознайомтеся зі звітом. Опишіть, скільки елементів з тестової вибірки було правильно розцінена. Ознайомтеся з побудованим деревом рішень. В текстовому вигляді воно представлено в звіті, крім того, в контекстному меню можна вибрати пункт Visualize tree (рис. 2.1.4) і отримати зображення в окремому вікні. Опишіть отриманий класифікатор.

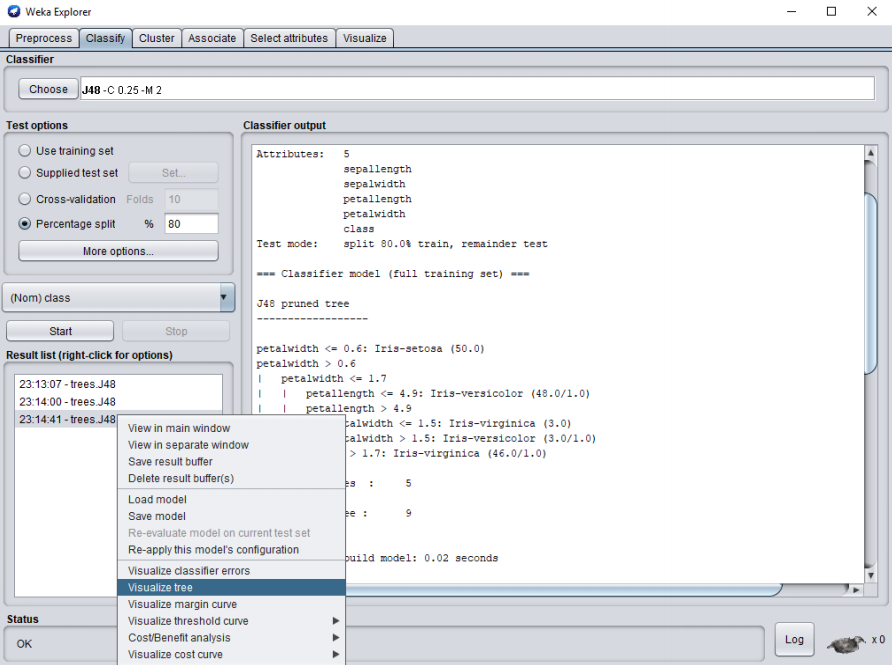


Рис. 2.1.4. Вибір пункта меню, який показує дерево рішень

Завдання 5. Розбиття на навчальний і тестовий набори проводиться на підставі послідовності, яка видається генератором випадкових чисел. Генератор ініціалізується числом, званим seed. У вікні Classify (рис. 2.1.3) знайдіть кнопку More options, Змініть параметр Random seed for XVal /% Split з 1 на 2. Знову запустіть процес навчання класифікатора. Тепер замість 30 з 30 повинно бути правильно розцінено 29 з 30 варіантів. Знайдіть в звіті, елемент якого класу був неправильно класифікований (вихідний клас і клас, призначений класифікатором).

Завдання 6. Встановіть параметр Random seed for XVal /% Split назад в 1. Повторіть експеримент з розбиття на навчальний і тестовий набори в пропорції 60/40. Як змінилася точність класифікатора? З чим ви повязуєте зміну точності? В ході цієї роботи ми познайомилися з побудовою моделі для класифікації. Класифікація відноситься до завдань навчання з учителем, коли ми знаємо правильну відповідь і використовуємо це знання для навчання моделі. Для перевірки точності класифікації важливо використовувати нові дані, які не були задіяні в процесі навчання. Саме тому ми разбивали набір на навчальну і тестову вибірки і першу використовував в процесі навчання, а другу – для перевірки точності.

При класифікації атрибут класу повинен мати дискретне значення. У нашому випадку це назви класів ірисів.

Якщо даних для навчання Бракує, точність класифікатора знижується (модель недостатньо точно описує предметну область). У той же час необхідно надати і достатню кількість нових даних для тестування, щоб воно адекватно оцінювало точність класифікатора. При використанні невеликих наборів ці вимоги вступають в протиріччя, і доводиться шукати компроміс.

Завдання 7. Порівняйте три отриманих дерева рішень.

Практична робота 2.2.

ВІЗУАЛІЗАЦІЯ ДАНИХ

АЛГОРИТМ К-НАЙБЛИЖЧИХ СУСІДІВ

ВИКЛЮЧЕННЯ АТРИБУТІВ

ВИКОРИСТАННЯ ФІЛЬТРІВ

1. Відкриємо набір даних з ірисами Фішера (iris.arff) і перемкнемося на вкладку Visualize.

Виберіть зображення, де по осі X розташовані значення атрибута petallength, по осі Y – petalwidth. Існує інший спосіб будь-який варіант зображення набору і в випадних списках у верхній частині вікна виберіть відповідні атрибути (petallength і petal-width) для осей X і Y.

Для набору iris.arff виконайте дії. Подивіться на уявлення набору даних при виборі інших пар атрибутів. Дайте відповідь на питання: Якщо використовувати для класифікації тільки два атрибути, яке поєднання, на ваш погляд, буде більш інформативно?

2. Алгоритм k-найближчих сусідів (k Nearest Neighbor) визначає приналежність об’єкта класифікації до класу виходячи з приналежності найближчих до нього об’єктів навчальної вибірки. У WEKA реалізація алгоритму називається IBk і знаходиться в групі lazy. На вкладці Classify виберіть цей алгоритм, залиште налаштування за замовчуванням і кроссвалідацію для 10 підмножин. Запустіть навчання класифікатора. У сформованому програмою звіті знайдіть оцінку точності алгоритму, проаналізуйте матрицю помилок (Confusion Matrix). Дайте відповідь на питання: Виявилося чи ваше припущення вірним?

3. Скористайтеся довідковою системою WEKA, щоб познайомитися з описом параметрів, які можна задати алгоритму IBk. Дайте відповідь на питання: Який сенс і значення за замовчуванням у параметра windowSize? Коли його варто задавати явно? Яка кількість «сусідів» класифікатор використовує за замовчуванням?

Проведіть повторні експерименти, використовуючи крос-валідацію і значення k-рівні 2, 3, 4. Дайте відповідь на питання: Чи змінилися точність і матриця помилок? Виходячи з опису алгоритму, наведеного за посиланням в завданні 2, що можна сказати щодо вибору значення k? Чому, на ваш погляд, для набору «іриси Фішера» зміна значення k несуттєво впливало на точність класифікації?

4. На вкладці Classify в списку Result list в лівій нижній частині вікна програми виберіть результати першого експерименту (з настройками IBk за замовчуванням). Клацніть на ньому правою клавішею миші і з контекстного меню виберіть Visualize classifier errors. У вікні, по осі X розташовані фактичні значення класів, по осі Y – призначені класифікатором. Квадратиком обведені неправильно класифіковані елементи. На графіку у вас буде 5 точок: 3 відповідні випадків правильної класифікації (в наборі всього 3 класу), ще 2 у квадратиках – випадків неправильної класифікації. Якщо посунути вправо «повзунок» з підписом Jitter (від англ. – тремтіти), то відповідні окремих елементах точки випадковим чином зміщуються щодо своєї точної координати. Випадки як правильною, так і помилковою класифікації перестають зливатися. Клацанням миші по окремій точці можна отримати інформацію про відповідному їй елементі.

5. Виконайте описаний експеримент. Отримайте інформацію про всі семи випадках неправильної класифікації, додайте її в звіт. У вікні візуалізації помилок виберіть по осі X атрибут petaUength, по Y – petalwidth. Дайте відповідь на питання: В якій області графіка розташовані помилково класифіковані елементи?

6. У WEKA на вкладці Preprocess можна прибрати непотрібні атрибути. Для цього виділіть їх в списку і натисніть кнопку Remove. Погодження про надання послуг атрибутів можна по кнопці Undo у верхній частині вікна програми. Зміни відбуваються в оперативній пам’яті, але, якшо по кнопці Save, змінений набір можна зберегти в файл.

У наборі «іриси Фішера» приберіть атрибути sepallength і sepalwidth. Натиснувши кнопку Edit, перегляньте отриманий новий набір. Закрийте вікно редагування і збережіть набір в новий файл. У WEKA виконайте класифікацію з використанням алгоритму IBk з параметрами за замовчуванням, а також зі значеннями k, рівними 2, 3 і 4. Дайте відповідь на питання: Що можна сказати про оцінку точності класифікації? В якому з восьми проведених експериментів (з різними значеннями k на повному наборі і на наборі з віддаленими атрибутами) вона була набольшей і найменшою? Чи можна сказати, що ця різниця істотна? Чи можна сказати, що виключення половини незалежних атрибутів (двох з чотирьох) знизило точність класифікації?

7. У наборі можуть виявитися повторювані записи. У якихось випадках це нормально, в інших – від дублікатів треба попередньо позбутися. Використовуйте набір з трьома атрибутами, отриманий після експериментів із завдання 5 (при необхідності завантажте його з файлу, куди він був збережений).

Для відкидання дублікатів застосуєте до нього фільтр Re- moveDuplicates, який знаходиться в розділі filters → unsupervised → instance (т.т. це фільтр, робота якого не залежить від атрибута класу, застосовується до примірника даних цілком). Дайте відповідь на питання: Скільки елементів даних залишилося після фільтрації? Як це вплинуло на точність алгоритму IBk ck = 2? З чим це може бути пов’язано? Чи варто було в даному випадку відкидати повторення?

8. Скасуйте всі зміни в наборі або заново завантажте набір iris.arff. Побудуємо на його основі набір, в якому не буде iris-versicolor. Для цього використовуємо фільтр RemoveWith Values, який також знаходиться в розділі filters → unsupervised → instance.

9. Поверніться до початкового набору даних. Дайте відповідь на питання: Що станеться, якщо застосувати до нього фільтр RemoveWithValues ​​з настройками за замовчуванням (їх можна подивитися в довідці або змінити фільтр і потім знову повернутися до RemoveWithValues)? Як ви поясните результат?

10. Знову поверніться до вихідного набору. Дайте відповідь на питання: Чи дозволяє WEKA один за іншим застосувати кілька фільтрів, наприклад RemoveDuplicates і RemoveWithValues? Чи були співпадаючі записи в вихідному наборі (чи змінилося число варіантів при застосуванні першого фільтра)? Якщо так, до яких класів вони ставилися (можна подивитися кількість примірників кожного класу)?

МЕТОДИЧНІ ВКАЗІВКИ

***ВІЗУАЛІЗАЦІЯ***

При аналізі даних важливу роль відіграє наявність у аналітика гіпотез щодо наявних в наборі взаємозв’язків. У якихось випадках ці припущення будуються на основі знань про предметну область, в інших допомогу може надати візуальний аналіз наявних в наборі даних. У WEKA Explorer засоби візуалізації набору даних зібрані на вкладці Visualize.

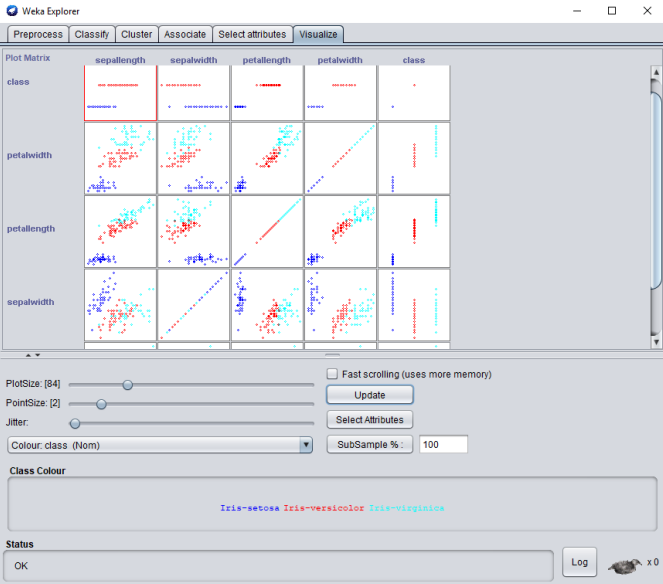


Рис. 2.2.1. Графічне представлення набору даних в різних проекціях

Відкриємо набір даних з ірисами Фішера (iris.arff) і перемкнемося на вкладку Visualize. Там будуть зображені проекції набору даних на різні пари атрибутів (рис.2.2.1).

Виберіть зображення, де по осі X розташовані значення атрибута petallength, по осі Y – petalwidth (рис. 2.2.2). Існує інший спосіб будь-який варіант зображення набору і в випадних списках у верхній частині вікна виберіть відповідні атрибути (petallength і petal-width) для осей X і Y.



Рис. 2.2.2. Візуальне представлення набору даних для атрибутів petaІength (по X) і petalwidth (по У)

Завдання 1. Для набору iris.arff виконайте дії. Подивіться на уявлення набору даних при виборі інших пар атрибутів.

***АЛГОРИТМ К-НАЙБЛИЖЧИХ СУСІДІВ***

Завдання 2. Алгоритм k-найближчих сусідів (k Nearest Neighbor) визначає приналежність об’єкта класифікації до класу виходячи з приналежності найближчих до нього об’єктів навчальної вибірки. Детальніше про алгоритм k-найближчих сусідів можна прочитати, наприклад, тут <http://www.machinelearning.ru/wiki/index.php?title=KNN>.

У WEKA реалізація алгоритму називається IBk і знаходиться в групі lazy (від англ. Лінивий, так як алгоритм, по суті, не будує модель, а зберігає навчальну вибірку і використовує її, коли потрібно провести класифікацію). На вкладці Classify виберіть цей алгоритм, залиште налаштування за замовчуванням і кросс-валідацію для 10 підмножин. Запустіть навчання класифікатора. У сформованому програмою звіті знайдіть оцінку точності алгоритму, проаналізуйте матрицю помилок (Confusion Matrix).

Завдання 3. Скористайтеся довідковою системою WEKA (у властивостях алгоритму - кнопка More), щоб познайомитися з описом параметрів, які можна задати алгоритму IBk. Проведіть повторні експерименти, використовуючи крос-валідацію і значення k-рівні 2, 3, 4.

***ВІЗУАЛІЗАЦІЯ ПОМИЛОК***

Завдання 4. На вкладці Classify в списку Result list в лівій нижній частині вікна програми виберіть результати першого експерименту (з настройками IBk за замовчуванням). Клацніть на ньому правою клавішею миші і з контекстного меню виберіть Visualize classifier errors. У вікні, по осі X розташовані фактичні значення класів, по осі Y – призначені класифікатором. Квадратиком обведені неправильно класифіковані елементи. На графіку у вас буде 5 точок: 3 відповідні випадків правильної класифікації (в наборі всього 3 класу), ще 2 у квадратиках – випадків неправильної класифікації. Якщо посунути вправо «повзунок» з підписом Jitter (від англ. – тремтіти), то відповідні окремих елементах точки випадковим чином зміщуються щодо своєї точної координати. Випадки як правильною, так і помилковою класифікації перестають зливатися (рис. 2.2.3). Клацанням миші по окремій точці можна отримати інформацію про відповідному їй елементі. На малюнку 2.2.4 представлений приклад такого опису.

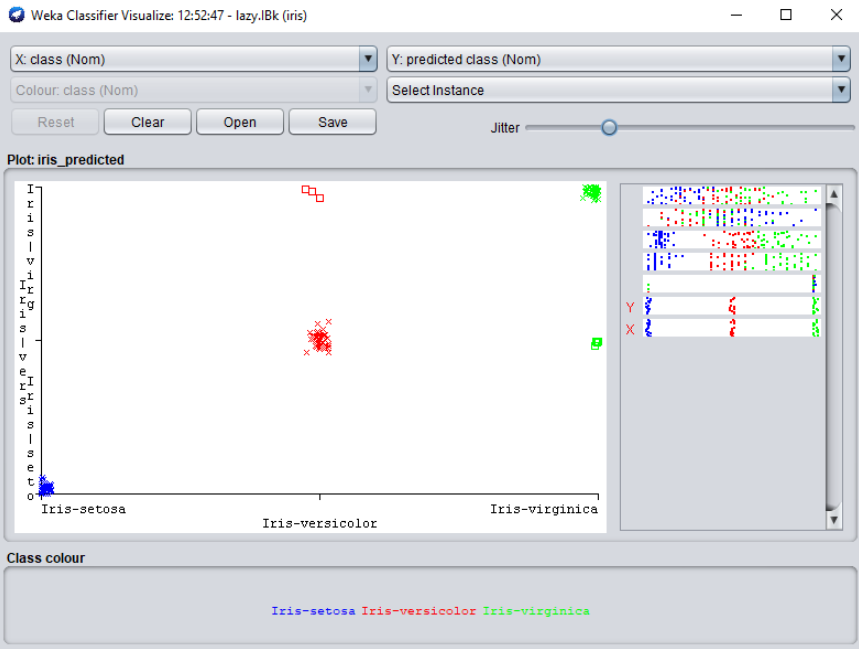


Рис. 2.2.3. Візуалізація помилок класифікатора з «тремтінням» (Jitter)

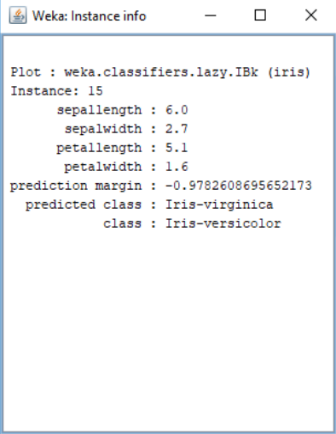


Рис. 2.2.4. Опис окремої помилки класифікатора

Завдання 5. Виконайте описаний експеримент. Отримайте інформацію про всі семи випадках неправильної класифікації, додайте її в звіт. У вікні візуалізації помилок виберіть по осі X атрибут petaUength, по Y – petalwidth.

***ВИКЛЮЧЕННЯ АТРИБУТІВ***

Точність класифікатора істотно залежить від набору аналізованих атрибутів. Очевидно, що адекватна модель не може бути побудована, якщо ми не враховуємо в ній якісь важливі параметри, істотно що характеризують об’єкти або явища в досліджуваній предметній області. З іншого боку, зайві атрибути, аналізовані алгоритмом, теж можуть спотворити картину. Це справедливо не тільки для класифікаторів, але і для інших типів моделей інтелектуального аналізу даних.

У WEKA на вкладці Preprocess можна прибрати непотрібні атрибути. Для цього виділіть їх в списку і натисніть кнопку Remove. Погодження про надання послуг атрибутів можна по кнопці Undo у верхній частині вікна програми. Зміни відбуваються в оперативній пам’яті, але, якшо по кнопці Save, змінений набір можна зберегти в файл.

Завдання 6. У наборі «іриси Фішера» приберіть атрибути sepallength і sepalwidth. Натиснувши кнопку Edit, перегляньте отриманий новий набір. Закрийте вікно редагування і збережіть набір в новий файл. У WEKA виконайте класифікацію з використанням алгоритму IBk з параметрами за замовчуванням, а також зі значеннями k, рівними 2, 3 і 4.

Якщо перевірку точності проводити не за допомогою крос-валідації, а на навчальній множині (опція use training set), то на початковому наборі iris.arff алгоритм IBk з настройками за замовчуванням покаже 100% -ву точність. У інших алгоритмів (J48, OneR) точність буде вище, ніж при «чесної» перевірці, але не досягне 100%.

Даний приклад ще раз показує, що тестування треба проводити на незалежних даних - окремому наборі, шляхом поділу на навчальну і тестову вибірку або за допомогою крос-валідації.

***ФІЛЬТРИ***

Можливість прибрати окремі атрибути – корисна, але далеко не єдина опція, що надається WEKA Explorer на вкладці Preprocess. До набору можна застосувати різні фільтри. Розглянемо деякі приклади.

Завдання 7. У наборі можуть виявитися повторювані записи. У якихось випадках це нормально, в інших – від дублікатів треба попередньо позбутися. Використовуйте набір з трьома атрибутами, отриманий після експериментів із завдання 5 (при необхідності завантажте його з файлу, куди він був збережений).

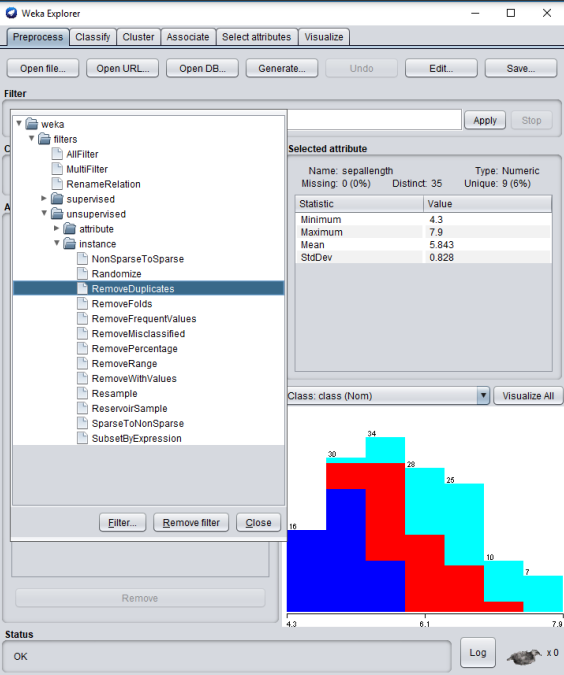


Рис. 2.2.5. Фільтр RemoveDuplicates

Для відкидання дублікатів застосуєте до нього фільтр Re- moveDuplicates (рис. 2.2.5), який знаходиться в розділі filters → unsupervised → instance (т. Е. Це фільтр, робота якого не залежить від атрибута класу, застосовується до примірника даних цілком). Скільки елементів даних залишилося після фільтрації?

Завдання 8. Скасуйте всі зміни в наборі або заново завантажте набір iris.arff. Побудуємо на його основі набір, в якому не буде iris-versicolor. Для цього використовуємо фільтр RemoveWith Values, який також знаходиться в розділі filters → unsupervised → instance. Налаштування фільтру не зовсім очевидна, тому розберемося з нею докладніше (рис. 2.2.6).

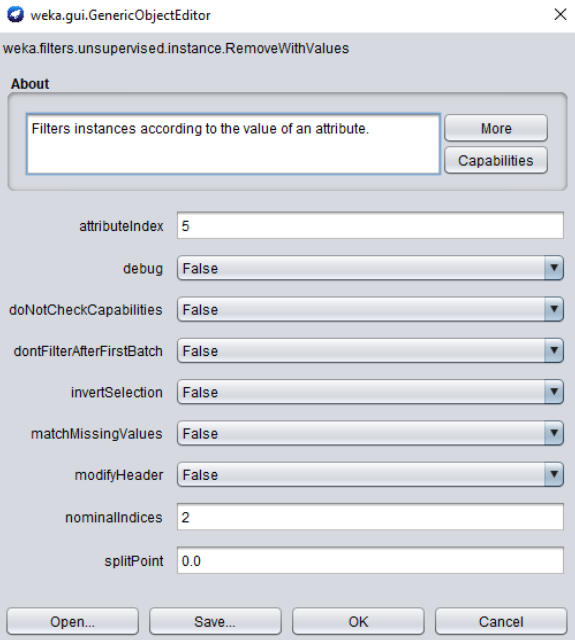


Рис. 2.2.6. Налаштування фільтру RemoveWithValues

Нас цікавить значення атрибута class, який в переліку атрибутів, який відображається на вкладці Preprocess, варто п’ятим, до того ж на останньому місці. Тому в параметрі фільтра attri-butelndex можна поставити 5 або залишити last (за замовчуванням). Значення атрибута номінальне (одне з трьох назв класів ірисів), тому далі в фільтрі налаштовуємо атрибут nominalIndices. Відкинути нам треба екземпляри з одним значенням цього атрибуту – iris-versicolor – в списку значень атрибута class на вкладці Preprocess воно йде другим. Тому в фільтрі в nominalIndices ставимо 2. Виконайте настройку фільтра і застосуєте його. За кнопці Edit на вкладці Preprocess переконайтеся, що в списку залишилося тільки два класи. Збережіть змінений набір під новим ім’ям.

Завдання 9. Поверніться до початкового набору даних.

Завдання 10. Знову поверніться до вихідного набору.

Практична робота 2.3.

КЛАСТЕРИЗАЦІЇ

Завдання:

1. Використаємо алгоритм кластеризації X-means, якого немає в складі дистрибутива. Додамо його. У WEKA GUI – виберіть меню Tools і в ньому Package Manager.

В поле для пошуку Package Search наберіть XMeans. З’явиться запис про пакет, виділіть її та встановіть.

2. У WEKA Explorer відкрийте набір даних з ірисами Фішера. Перейдіть на вкладку Cluster. Натисніть кнопку Ignore attributes, у вікні, виберіть атрибут class (його не враховуємо при кластеризації) і натисніть Select.

Запустіть кластеризацію з використанням алгоритму X-Means c настройками за замовчуванням. Дайте відповідь на питання: Скільки вийшло кластерів? Відкрийте панель настройки алгоритму. Що означають параметри maxNumClusters, minNumClusters, distanceF і які у них налаштування за замовчуванням? Чому вийшов такий результат кластеризації?

3. Перевірте, що у вас стоїть позначка Store clusters for visualization (розділ Cluster Mode). У списку Result list клацніть правою клавішею на результатах кластеризації і виберіть пункт Visualize cluster assignments. Виберіть по осі X атрибут petallength, а по осі Y – petalwidth. Дайте відповідь на питання: Чи можете ви пояснити результати кластеризації?

4. Виберіть алгоритм EM, ознайомтеся з його описом. Дайте відповідь на питання: Що означає numClusters і яке у цього параметра значення за замовчуванням? Перевірте настройки Ignore Attribute. Виконайте кластеризацию, проаналізуйте отримані результати.

5. Виберіть для кластеризації алгоритм k-середніх (в WEKA називається SimpleKMeans). Дайте відповідь на питання: Який сенс мають параметри numClusters, maxIterations, initializationMethod, distanceFunction і які у них значення за замовчуванням? Перевірте настройки Ignore Attribute. Повторіть кластеризацию набору з параметрами за замовчуванням. Як отримані результати співвідносяться з раніше виконаними завданнями?

6 Виконайте новий експеримент, використовуючи алгоритм k-середніх і явно задавши число кластерів = 3. Перевірте настройки Ignore Attribute. Дайте відповідь на питання: Який результат ви отримали? Як його можна пояснити?

7. Перейдіть на вкладку Preprocess. У фільтрах виберіть unsupervised → attribute → AddCluster. Цей фільтр додає до набору стовпець з номером кластера. В налаштуваннях фільтра в якості алгоритму кластеризації виберіть SimpleKMeans з числом кластерів, рівним 3. Задайте параметру ignored- AttributeIndices таке значення, щоб при кластеризації не враховувався атрибут класу. Застосуйте фільтр. За кнопці Edit відрийте отриманий набір даних (або його можна зберегти у форматі текст з роздільником – \* .csv – і відрити в текстовому редакторі або в програмі для обробки електронних таблиць). Дайте відповідь на питання: В яких випадках поділ на кластери збіглося з початковим розподілом на класи, а коли збігу не було?

МЕТОДИЧНІ ВКАЗІВКИ

***ВСТАНОВЛЕННЯ ДОДАТКОВИХ МОДУЛІВ***

WEKA має багатою бібліотекою алгоритмів і фільтрів. Не всі з них відразу встановлюються разом з програмою. У цій лабораторній ми будемо використовувати алгоритм кластеризації X-means, якого немає в складі дистрибутива. Почнемо з того, що додамо його. У WEKA GUI (рис. 2.3.1) виберіть меню Tools і в ньому Package Manager.

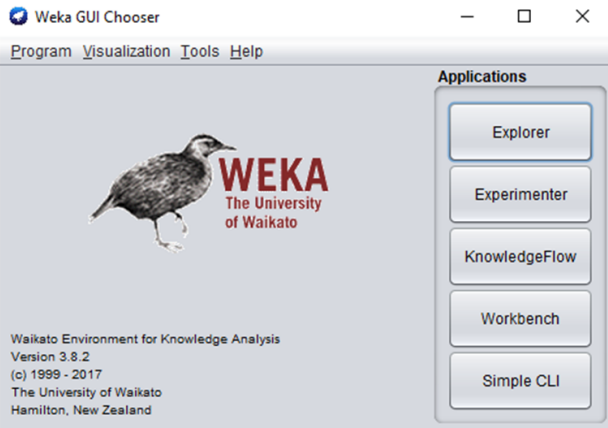


Рис. 2.3.1. Weka GUI

В поле для пошуку Package Search наберіть XMeans. З’явиться запис про пакет, виділіть її та встановіть. На малюнку 2.3.2 показаний вже встановлений пакет. Алгоритм X-Means розширює можливості алгоритму K-means (K-середніх, додаючи можливість визначення числа кластерів). Але X-Means не може працювати з номінальними атрибутами.



Рис. 2.3.2. Встановлений пакет XMeans

Детальніше про k-середніх можна прочитати, в статті X-means: Extending K-means with Efficient Estimation of the Number of Clusters доступна за посиланням: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=0E5148D5EBE3>IBB3C3340B1F7E4E2672? Doi = 10.1.U9.3377 & rep = rep1 & type = pdf.

***КЛАСТЕРИЗАЦІЇ***

Є різні алгоритми кластеризації. Зокрема, k-середніх відноситься до методів жорсткої кластеризації (кожен варіант відноситься тільки до одного кластеру), інший алгоритм, який розглядається в лабораторній, – максимізація очікування EM – алгоритм м’якою кластеризації (варіант може належати кільком кластерам з різними можливостями).

У WEKA Explorer відкрийте набір даних з ірисами Фішера. Перейдіть на вкладку Cluster. Ви пам’ятаєте, що атрибут class містить найменування класу, до якого належить певний ірис.

Натисніть кнопку Ignore attributes, у вікні, виберіть атрибут class (його не враховуємо при кластеризації) і натисніть Select.

Завдання 1. Запустіть кластеризацію з використанням алгоритму X-Means c настройками за замовчуванням.

Завдання 2. Перевірте, що у вас стоїть позначка Store clusters for visualization (розділ Cluster Mode). У списку Result list клацніть правою клавішею на результатах кластеризації і виберіть пункт Visualize cluster assignments. За аналогією з попередньою лабораторної, виберіть по осі X атрибут petallength, а по осі Y – petalwidth. Чи можете ви пояснити результати кластеризації?

Завдання 3. Виберіть алгоритм EM, ознайомтеся з його описом. Перевірте настройки Ignore Attribute. Виконайте кластеризацию, проаналізуйте отримані результати.

Завдання 4. Виберіть для кластеризації алгоритм k-середніх (в WEKA називається SimpleKMeans).

Завдання 5. Виконайте новий експеримент, використовуючи алгоритм k-середніх і явно задавши число кластерів = 3. Перевірте настройки Ignore Attribute.

Завдання 6. Перейдіть на вкладку Preprocess. У фільтрах виберіть unsupervised → attribute → AddCluster. Цей фільтр додає до набору стовпець з номером кластера. В налаштуваннях фільтра в якості алгоритму кластеризації виберіть SimpleKMeans з числом кластерів, рівним 3. Задайте параметру ignored- AttributeIndices таке значення, щоб при кластеризації не враховувався атрибут класу. Застосуйте фільтр. За кнопці Edit відрийте отриманий набір даних (або його можна зберегти у форматі текст з роздільником – \* .csv – і відрити в текстовому редакторі або в програмі для обробки електронних таблиць).

**Змістовий модуль 2. Інформаційне підґрунтя розробки бізнес-ідеї та започаткування підприємницької діяльності / Розробка бізнес-моделі**

**Тема 3. Оцінка ринку сфер економічної діяльності**

Практична робота 3.1

АНАЛІЗ РИНКУ

Завдання:

1. Проаналізувати ринок для обраного товару. Зробити висновки про основні ринкові проблеми та можливості.

2. Визначити причини проблем та розробити гіпотези для їх вирішення або використання можливостей.

3. Розробити план збору інформації для перевірки гіпотез.

4. Здійснити збір та аналіз первинної інформації.

5. На підставі отриманої первинної інформації зробити висновок про пріоритетні варіанти вирішення ринкових проблем та використання можливостей.

МЕТОДИЧНІ ВКАЗІВКИ

Аналіз ринку – це оцінка розмірів ринку як за обсягом, так і за вартістю, аналіз сегментів клієнтів і моделі покупок, конкуренції й економічного середовища з точки зору бар’єрів для входу та регулювання.

Для цього рекомендується аналіз ключових компонентів:

* Галузь
* Цільовий ринок
* Конкуренція

**ГАЛУЗЕВИЙ АНАЛІЗ**

Галузевий аналіз оцінює загальне галузеве середовище, в якому бізнес конкурує. Необхідно представити статистичні дані про розмір галузі та темпи її зростання за останні кілька років і відповісти на питання: Чи розширюється галузь, скорочується чи тримається стабільно? Чому?

Також потрібно вказати важливі тенденції, які можуть вплинути на галузь. Наприклад, значні зміни на цільовому ринку, в технології або в інших суміжних галузях можуть вплинути на сприйняття ринку пропонованого товару або на прибутковість підприємства: скажімо, тренд на екоспоживання та підвищення екосвідомості може підвищити рівень залученості споживачів бренду, що акцентує увагу на переробці відходів власного виробництва. Сучасні виробництва не можуть ігнорувати цей факт, й, можливо, деяким доведеться змінювати підходи, аби відповідати запитам цільової аудиторії.

Така інформація часто доступна безкоштовно з наступних джерел:

* Дані торгових асоціацій
* Галузеві публікації та бази даних
* Публічна інформація у формі відкритих даних (наприклад, Єдиний державний веб-портал відкритих даних, Загальнодоступна інформаційна база даних НКЦПФР)
* Дані та думки аналітиків про найбільших гравців галузі (наприклад, звіти Standard & Poor’s, цитати з поважних джерел новин)

***ЦІЛЬОВИЙ РИНОК***

Визначити, хто є ідеальним клієнтом/замовником бізнесу? Ці дані повинні включати демографічні показники щодо групи, на яку орієнтується бізнес, включаючи вік, стать, рівень доходу та стиль життя. Цей розділ також повинен містити дані про розмір цільового ринку, потенціал придбання та мотивацію аудиторії та те, як бізнес має намір вийти на ринок.

Ресурси для визначення цільової аудиторії в аналізі ринку:

* Пошук за ключовими словами може забезпечити загальне уявлення потенційного попиту на товар чи послугу залежно від кількості пошукових запитів.
* Аналіз Google Trends може повідомити, як кількість пошукових запитів змінювалася з часом.
* Кампанії в соціальних мережах можуть вказувати на потенційний інтерес клієнта до вашої бізнес-ідеї.
* Pew Research Center надає дуже конкретну інформацію про світогляд та соціальні й демографічні тенденції цільового ринку

***КОНКУРЕНЦІЯ***

Потрібно дати чітке уявлення про те, з ким змагається бізнес. Необхідно пояснити *позиціювання конкурентів* і описати їх сильні та слабкі сторони. Ідея тут полягає в тому, щоб знайти слабкість конкурента, яку компанія зможе використовувати у своєму позиціюванні на ринку. Варто дати відповідь на наступні питання: «Хто є конкурентом підприємства? Які його сильні та слабкі сторони? Які потенційні перешкоди заважають компанії вийти на ринок?».

Один зі способів проведення аналізу – *порівняти конкурента* з кожним із ключових факторів попиту на ринку (ціна, якість, додаткові послуги тощо). Потрібно знати якомога більше про особи конкурентів і подробиці їхньої діяльності. Варто дослідити їх рекламу, зробити у них кілька покупок, вивчити їх ціни та поспілкуватися з клієнтами своїх конкурентів.

Базова інформація, яку кожна компанія повинна знати про своїх конкурентів, включає:

* Цінову політику компаній
* Сильні та слабкі сторони, позиціювання конкурентів на ринку
* Фінансові показники та об’єми продажів
* Канали дистриб’юції
* Оцінку та сприйняття цільовими покупцями товарів або послуг конкурентів

Цю інформацію можливо дізнатись:

1. Аналізуючи сайти та соціальні мережі конкурентів. Слід вивчати ключових гравців та прямих конкурентів

2. Через сервіси аналізу конкурентів: SEMRush, Advodka, Spywords, AdVse, Serpstat, SimilarWeb та інші.

3. Фінансові звіти. Метод не є універсальним, але великі компанії почасти публікують такі звіти у відкритому доступі.

4. Дослідження, що містять фізичні розвідки цін, асортименту, роботи персоналу конкурентів.

5. Сервіси моніторингу цін: Priceva, Competera, Price2Spy та інші.

***БАР’ЄРИ ДЛЯ ВХОДУ***

Бар’єри для входу на ринок – це ті фактори, які заважають новому бренду легко вступити у галузь чи сферу бізнесу. Серед таких бар’єрів можна виділити:

* Інвестиції. Однією з перешкод є пошук ресурсу для започаткування справи. Для того щоб уникнути ризиків, почати слід з мінімально життєздатного продукту (продукт з мінімальними функціями, що створюється для отримання зворотного зв’язку від перших споживачів для поліпшення продукту в подальшому) або придбати вже налагоджений на ринку бізнес (наприклад, у рамках франшизи).
* Технології. Складністю можуть стати технології розробки веб-сайту. На початкових етапах радимо використовувати ринок Etsy або You Tube як платформу для продажів.
* Створення бренду. Для досягнення певного рівня впізнаваності можуть знадобитися великі маркетингові витрати. Вірусний маркетинг може стати у пригоді для підприємців початківців, адже дозволяє скоротити маркетингові витрати на залучення нових продажів. Також для економії обсягу можна використовувати вже наявний бренд, налагоджуючи зв’язки та вступаючи у партнерства.
* Обладнання. Для зменшення постійних витрат бізнес може надавати перевагу лізингу, а не придбанню обладнання/торгових майданчиків.

Практична робота 3.2

SWOT-АНАЛІЗ РИНКОВОГО СПРЯМУВАННЯ

Завдання:

Провести SWOT-аналіз бізнесу, сформувати поле проблем розвитку бізнесу та комплекс заходів щодо їх усунення чи запобігання проявів.

Методичні вказівки

SWOT-аналіз – один із найпоширеніших аналітичних методів, який дозволяє в комплексі оцінити сильні й слабкі сторони компанії, а також можливості й загрози, що впливають на неї.

Матриця SWOT містить:

S (strengths) – сильні сторони. Характеристики бізнесу, які вирізняють його на фоні конкурентів. Наприклад: кращий клієнтський сервіс на ринку, більш доступні ціни.

W (weaknesses) – слабкі сторони. Ознаки, які роблять компанію вразливою на ринку. Наприклад: неефективна реклама, недостатня кількість співробітників.

O (opportunities) – можливості. Їх компанія може використовувати для розвитку свого бізнесу. Наприклад: правильне розміщення виробництва.

T (threats) – загрози. Вони можуть завдати компанії збитків. Наприклад: висока конкуренція на ринку.

Уперше термін «SWOT-аналіз» був застосований американським академіком Кеннетом Ендрюсом у 1963 році в Гарварді на конференції з проблем бізнес-політики.

Для виходу стартапу на ринок потрібно розуміти всі внутрішні та зовнішні можливості й загрози. Наприклад, створюючи нову соціальну мережу, необхідно оцінювати інтенсивність конкурентної боротьби, оскільки соціальних мереж існує велика кількість, загрози від альтернатив продукту або послуги, ринкову владу споживачів; вирішити, що саме є вашою УТП (унікальною торговельною пропозицією) та ін.

Великі гравці ринку мають проводити SWOT-аналіз організації щонайменше раз на рік. Це надає можливість підтримувати свою актуальність, бо зміни відбуваються щодня.

**Переваги й недоліки**

**Переваги:**

1. Це універсальний метод, який допомагає бізнесу дізнатися і використовувати:

* внутрішні можливості (у чому проект/колектив/команда випереджає інших на ринку). Наприклад: сильна команда технічної підтримки, інноваційна бізнес-модель;
* внутрішні загрози (що всередині проекту/команди/колективу послаблює потенційні можливості). Наприклад: недостатньо укомплектована команда, застаріла техніка продажів;
* зовнішні можливості (ті фактори, які дають додаткові можливості для досягнення кінцевої мети і результату). Наприклад: зростаючий інтерес до омніканального маркетингу, розширення ринку;
* зовнішні загрози (можливі фактори, які, навпаки, уповільнюють процес досягнення мети команди). Наприклад: велика конкуренція.

1. Допомагає виявити потенційні загрози і побудувати стратегію захисту.
2. SWOT досить нескладний у проведенні і не потребує спеціальної вузькопрофільної освіти.

**Недоліки:**

1. SWOT-аналіз не покаже чітку аналітику. За його допомогою можна отримати структуровані дані, а все інше – робота вашого аналітика чи будь-якого іншого співробітника, який допоможе із формуванням маркетингової стратегії на базі отриманих даних.
2. SWOT показує стан справ на поточний момент і залишається статичним. Саме тому його потрібно проводити не менш ніж один раз на рік.
3. Є досить суб’єктивним аналізом і залежить від того, хто його здійснює.
4. Для якісного аналізу потрібний великий обсяг інформації з усіх сфер (логістична, комерційна, маркетингова, фінансова й ін.), що може бути витратним.

**Методика SWOT-аналізу**

Перш за все треба створити матрицю й розмістити SWOT-фактори за такою схемою:



У готовій матриці потрібно проставити бали від 1 до 5. У списку сильних сторін проставляються тільки позитивні оцінки (якщо «5» – відповідний сильний фактор компанії має найвищий рівень, а якщо «1» – найнижчий). Слабкі сторони оцінюються за тим самим принципом, тільки максимум – для негативних значень (якщо «5» – за цим фактором компанія вкрай слабка, якщо «1» – цей фактор слабкості практично невластивий компанії).

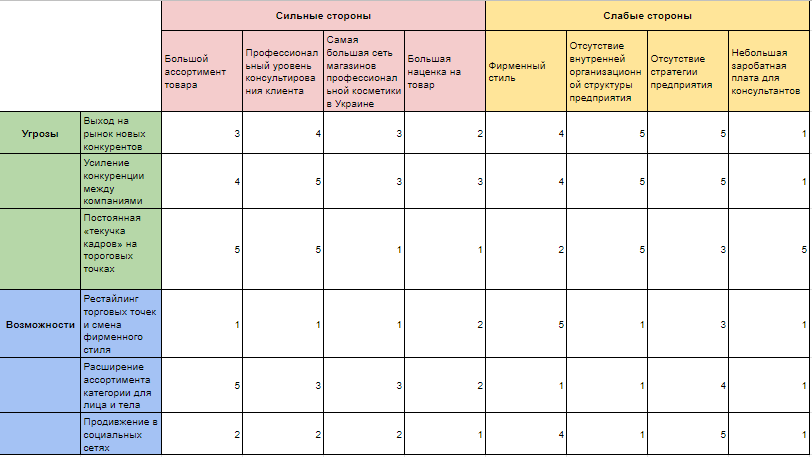
Наприклад:

* Сильні сторони: відоме ім’я на ринку – 5, широкий асортимент товару – 2.
* Можливості: зростання доходів населення – 1, розвиток інформаційних технологій – 4.
* Слабкі сторони: високі ціни на товар – 3, відсутність бюджету для маркетингу – 5.
* Загрози: зростання кількості конкурентів на ринку – 4, проблеми з пошуком нових постачальників – 1.

*Перший етап* – створюємо матрицю SWOT. Розглянемо найпростішу матрицю на прикладі роздрібної мережі магазинів косметики:



*Другий етап* (поле оцінок) – перехресне оцінювання сильних і слабких сторін, можливостей і загроз. На перетині рядків і стовпчиків СВОТ-матриці слід поставити оцінку значущості конкретної пари факторів. Найвищий рівень значущості отримує високу оцінку – 5 балів, найнижчий відповідно 1 бал (див. приклад вище):

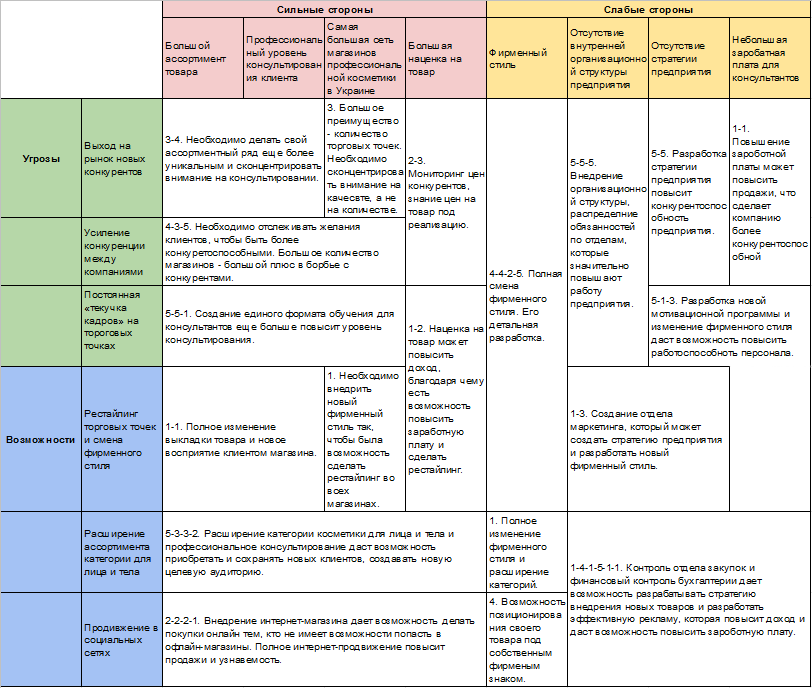


Такий аналіз дозволяє зрозуміти, наскільки суттєвими перевагами й недоліками є сильні та слабкі сторони, а також оцінити важливість загроз і можливостей зовнішнього середовища. Розраховувати оцінку треба таким чином: Великий асортимент товару – сума всіх вертикальних оцінок у стовпчику, так само як Поява на ринку нових конкурентів – сума по горизонталі та ін.



*Третій етап* (перехресне оціночне поле): після визначення кількісних оцінок необхідно сформулювати проблеми для кожної комбінації сильних і слабких сторін із загрозами та можливостями. Таке формулювання називають проблемним полем.

Під пунктами проблемного поля маються на увазі заходи, які необхідно практично реалізувати за даних умов роботи підприємства. Кожна цифра – це кількісна оцінка з таблиці вище. Наприклад: перетин Велика націнка на товар «2» та Посилення конкуренції між компаніями «3». На перетині ми отримаємо 2-3.



*Четвертий етап*: проблеми, сформульовані в такий спосіб, можуть оцінити експерти (провідні менеджери компанії). Оцінка однієї проблеми складається з суми експертних комбінацій сильних і слабких сторін, можливостей і загроз.



На завершення проведення SWOT-аналізу формується комплекс заходів (проблем), які мають бути закладені в основу стратегії, та визначення порядку їхньої практичної реалізації. Тобто при цьому варіанті проведення SWOT-аналізу не тільки оцінюються фактори середовища, але також практично вказується, для чого робиться така оцінка.

Інформаційний ресурс для SWOT-аналізу: <https://www.canva.com/uk_ua/grafiky/swot-analiz/>

Практична робота 3.3

АНАЛІЗ КОНКУРЕНТНОГО СЕРЕДОВИЩА ЗА МОДЕЛЛЮ «5 СИЛ ПОРТЕРА»

Завдання:

Провести SWOT-аналіз бізнесу, сформувати поле проблем розвитку бізнесу та комплекс заходів щодо їх усунення чи запобігання проявів.

МЕТОДИЧНІ ВКАЗІВКИ

Ця модель дозволяє проаналізувати рівень конкуренції в конкретній галузі та тиск різних факторів на бізнес. Використовується у випадку відкриття нового бізнесу, розширення діяльності у новому секторі, випуску нового товару чи послуги.

[Модель 5 сил Портера](https://www.business-to-you.com/porters-five-forces/) описує фактори, що формують конкурентне середовище. Додаткову цінність вона має тому, що в неї легко можна додати фактор корупції і врахувати бізнес-реалії конкретної держави. Знайти баланс у цій моделі – означає утримати прибутковість, втратити його – отримати збитки, вийти з ринку або провалитися під час входу на нього.

***Перша сила. Нові конкуренти***

Якщо відкрити бізнес у галузі складно, вірогідність того, що з’являться нові гравці, невелика. Складність виходу на ринок може проявлятись у браку технологій, відсутності спеціалістів, великих початкових витратах, відсутності каналів дистрибуції, необхідності отримати ліцензію тощо. Складність потрібно оцінювати не в загальному, а для конкретного бізнес-плану.

Наприклад, щоб відкрити нову авіакомпанію, необхідно придбати або взяти в лізинг літаки, знайти відповідний персонал, отримати ліцензію, дозволи та маршрути, налагодити продажі. Зробити все це досить важко і здається, що галузь авіаперевезень надійно захищена від новачків. Насправді лібералізація ринку і доступний лізинг дозволяють компаніям швидко виходити на ринок і захоплювати значну його частку – це легко побачити на прикладі [діяльності лоукостерів в Україні](https://www.ukrinform.ua/rubric-tourism/2583582-castka-loukostiv-u-miznarodnih-rejsah-sagnula-majze-30-omelan.html).

***Друга сила. Ринкова сила постачальників***

Ця сила показує, наскільки постачальники контролюють ринки: якою мірою вони спроможні впливати на ціну і якість продукції або послуг. Чим більше на ринку продавців і чим менший їхній потенційний вплив, тим легше вибрати контрагента, отримати знижку або пільгову ціну та утримувати видатки компанії на певному рівні.

Але не завжди мала кількість продавців означає високі ціни, шантаж і тиск. [У цивільному авіабудуванні лише дві компанії Airbus і Boeing утримують величезну частку ринку](https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D0%BD%D0%BA%D1%83%D1%80%D0%B5%D0%BD%D1%86%D0%B8%D1%8F_%D0%BC%D0%B5%D0%B6%D0%B4%D1%83_Airbus_%D0%B8_Boeing), але це не заважає їм виготовляти якісну продукцію і надавати гідний рівень сервісу. Цьому сприяє  дуже напружена конкуренція між компаніями, що вимагає постійного розвитку та створення нових конкурентних переваг.

***Третя сила. Ринкова сила покупців***

На одному ринку покупці можуть легко змінити постачальника, знайти товару заміну або бойкотувати компанію через будь-яку причину. На іншому – знайти альтернативу або відмовитись від продукції дуже важко. Спробуйте придбати автомобіль – знайдете в інтернеті тисячі пропозицій різних марок, моделей машин, нових і вживаних. Спробуйте змінити операційну систему на вашому комп’ютері – доведеться обирати серед версій Windows, можливо, вдасться впоратися з Linux, останній варіант – операційна система від Apple в комплекті з їхньою продукцією.

Якщо повернутися до авіакомпаній: пасажири мають безліч опцій при виборі авіарейсу – є прямі та з пересадкою; сайти-агрегатори пропозицій допомагають зекономити, а для того, щоб змінити авіакомпанію, не треба витрачати додаткові кошти – доведеться лише перейти з одного сайту на інший.

***Четверта сила. Доступність альтернативних продуктів***

Часто споживач може не тільки змінити постачальника, а навіть відмовитись від товару чи послуги на користь альтернативного продукту, продукту-замісника. Можна обирати напій серед десятків брендів кави, а придбати чай. В підсумку споживачу потрібно вгамувати спрагу та отримати заряд бадьорості, як це зробити – його вибір.

Користувачу авіаліній насамперед необхідне переміщення в просторі або враження в пункті призначення. Перше з успіхом забезпечують: залізниця, автобусні перевезення, власний автомобіль. Друге – журнали National Geographic і такі програми, як “Орел і решка”, якщо ви турист, або месенджери та соціальні мережі, якщо вам потрібно їхати у відрядження.

***П’ята сила. Суперництво наявних конкурентів***

В цій частині моделі Портера вивчається можливість конкурентів, які вже існують, впливати на те, що відбувається в галузі. На них теж діють всі вищевказані сили, і вони теж мають боротися з таким самим тиском.

Якщо склалась ситуація, що у конкурентів висока [маржинальність](http://national-fc.com.ua/shho-take-marzhinalnist-analiz-i-rozrahunok-pokaznika/), то є велика ймовірність, що одна чи декілька компаній почнуть цінову або рекламну війну з метою збільшити власну частку ринку. Так само може вчинити фірма, яка вже домінує в галузі і бажає ще більше послабити позиції інших організацій. Зараз це можна [спостерігати на прикладі демпінгу цін на нафту](https://www.epravda.com.ua/rus/columns/2020/03/13/657966/) з боку держав, які є найбільшими виробниками.

В глобальній авіаіндустрії не існує монополії, і навіть найбільші ринки – висококонкурентні. (до речі, найбільша авіакомпанія світу [American Airlines](https://www.aa.com/homePage.do)) утримує 17,6% національних продажів, а [United Airlines](https://www.united.com/), яка знаходиться на 4 місці, – 14,9%. Компаній на ринку багато, можливість підвищувати ціни на квитки обмежують лоукостери, тож суперництво в галузі високе.

***Методика аналізу***

Потрібно проаналізувати всі сили окремо та в сукупності. Для цього необхідно скласти таблицю. Визначте, за якими параметрами ви будете оцінювати кожну силу тиску. Встановіть бали за різні рівні тиску. Наприклад, 3 бали за високий тиск, 2 – за середній, 1 – за низький. Далі проведіть оцінку, вибравши відповідні варіанти. Якщо в вашій галузі компаній багато, то за кількість ви ставите найвищий бал. Якщо при цьому кожна з компаній пропонує продукти, які значно відрізняються один від одного, то за рівень розмаїття продукту бал буде найнижчим, тому що це зменшує конкуренцію. І так далі.

Приклад заповнення матриці «Сила тиску конкурентів»

|  |  |  |  |
| --- | --- | --- | --- |
| Параметр оцінки | Оцінка | | |
| 3 | 2 | 1 |
| Кількість гравців | Високий рівень насичення рику | Середній рівень насичення рику | Невелика кількість гравців |
|  |  | + |
| Темп зростання ринку | Стагнація або зменшення об’єму ринку | Повільне зростання | Швидке зростання |
|  | + |  |
| Рівень диференціації | Компанії продають стандартизований товар | Товар стандартизований за ключовими ознаками, але відрізняється додатковими перевагами | Продукти компаній значно відрізняються між собою |
| + |  |  |
| Обмеження в підвищенні цін | Жорстка цінова конкуренція на ринку | Існує можливість підвищення ціни лише в межах покриття зростання витрат | Завжди є можливість підвищення цін для покриття витрат та підвищення доходу |
|  | + |  |
| Підсумковий бал | 8 | | |
| 4 бали | Низький рівень середньогалузевої конкуренції | | |
| 5-8 балів | Середній рівень середньогалузевої конкуренції | | |
| 9-12 балів | Високий рівень середньогалузевої конкуренції | | |

У прикладі показаний підсумковий бал 8, що потрапляє в категорію: середній рівень конкуренції всередині галузі.

Приклад заповнення матриці «Ринкова сила постачальників»

|  |  |  |
| --- | --- | --- |
| Параметр оцінки | Оцінка | |
| 2 | 1 |
| Кількість постачальників | Незначна кількість постачальників або монополія | Значна кількість постачальників |
|  |  |
| Обмеженість ресурсів постачальників | Обмеженість в обсягах ресурсів | Необмеженість в обсягах ресурсів |
|  |  |
| Видатки переходу | Високі видатки переходу на інших постачальників | Невисокі видатки переходу на інших постачальників |
|  |  |
| Пріоритетність галузі для постачальника | Низька | Висока |
|  |  |
| Підсумковий бал |  | |
| 4 бали | Низький рівень впливу постачальників | |
| 5-8 балів | Середній рівень впливу постачальників | |
| 9-12 балів | Високий рівень впливу постачальників | |

Приклад заповнення матриці «Ринкова сила нових конкурентів»

|  |  |  |  |
| --- | --- | --- | --- |
| Параметр оцінки | Оцінка | | |
| 3 | 2 | 1 |
| Сильні гравці з високим рівнем знань та лояльності | Відсутні крупні гравці | 2-3 крупних гравця тримають близько 50% ринку | 2-3 крупних гравця тримають більше 80% ринку |
|  |  |  |
| Диференціація продукту | Низький рівень різноманітності товару | Існують мікро-ніші | Всі ніші зайняті гравцями |
|  |  |  |
| Рівень інвестицій і витрат для входу в галузь | Низький (окупність 1-3 місяця) | Середній (окупність до 1 року) | Високий (окупність більше 1 року) |
|  |  |  |
| Доступ до каналів розподілу | Повністю відкритий | Вимагає помірних інвестицій | Обмежений |
|  |  |  |
| Політика уряду | Немає обмежень з боку держави | Є певні державні обмеження | Держава повністю регламентує галузь |
|  |  |  |
| Готовність гравців до зниження цін | Гравці не підуть на зниження цін | Крупні гравці не підуть на зниження цін | За будь-якої дешевшої пропозиції всі гравці знижують ціни |
|  |  |  |
| Темпи росту галузі | Високий і зростаючий | Уповільнення | Стагнація або скорочення |
|  |  |  |
| Підсумковий бал |  | | |
| 8 бали | Низький рівень загрози входу нових конкурентів | | |
| 9-16 балів | Середній рівень загрози входу нових конкурентів | | |
| 17-24 балів | Високий рівень загрози входу нових конкурентів | | |

Таким чином проводиться аналіз для всіх п’яти ринкових сил, які визначають конкурентне середовище бізнесу. Наприкінці підрахуйте бали і на базі отриманих результатів зробіть висновки, чи варто заходити на новий ринок, чи є сенс залишатися на старому, що змінити для того, щоб покращити власні позиції та обійти конкурентів. Наприклад, якщо на ринку багато компаній, які пропонують майже однаковий продукт, варто подумати, як зробити власну пропозицію унікальною. Можливо, варто об’єднатися з деякими конкурентами в мережу, щоб збільшити частку ринку. Або більше уваги приділити не товарам, а сервісу.

З наявних загроз обираються ті, які становлять найбільшу загрозу вашому бізнесу, і за ними розробляються заходи для мінімізації або ліквідації наслідків. Наприклад, якщо немає загрози появи товарів-замінників, раціонально підтримувати та удосконалювати унікальність власного товару. При наявності високого рівня загрози появи нових гравців оптимальним варіантом буде проведення безперервного моніторингу пропозицій конкурентів, щоб своєчасно відреагувати.

Таким чином, після проведення аналізу буде сформована чітка стратегія розвитку бізнесу на кілька років вперед по кожній силі Портера. Це дозволить мінімізувати ризики появи збитків і банкрутства.

Практична робота 3.4

ФОРМУВАННЯ ПОРТРЕТУ СПОЖИВАЧА

Завдання:

Для обраного товару (бізнесу) провести статистичне дослідження покупців і скласти портрет споживача.

МЕТОДИЧНІ ВКАЗІВКИ

Розуміння своєї цільової аудиторії дозволяє бізнесу спростити пошук потенційних споживачів, які з великою ймовірністю зацікавлені в товарах і послугах, які пропонуються. Під портретом покупця розуміють визначення поведінки, потреб та інших характеристик, які характерні для більшості клієнтів. Наявність такої інформації спрощує співробітникам компанії створення маркетингових матеріалів, розробку стратегії просування інтернет-магазину та збільшення продажів.

Сучасні споживачі стикаються із сотнями маркетингових повідомлень щодня, і більшість з них «фільтрує» ті звернення, які не мають відношення до них особисто. Для фахівця це означає, що потрібно мати хороше уявлення про свою цільову аудиторію – хто ці люди і чого вони хочуть. Така інформація дозволяє виділитися серед загального інформаційного шуму і підвищити шанси донести свою пропозицію до тих людей, які є потенційними покупцями.

Наявність профілю клієнта дозволяє структурувати інформацію і перелік робіт, необхідних для досягнення поставлених перед бізнесом цілей. Завдяки демографічним, поведінковим та іншим характеристикам своєї цільової аудиторії ви отримуєте чітке уявлення про те, на кого орієнтовані товари або послуги.

Наприклад, можна застосувати ці знання при розробці стратегії контент-маркетингу, адже незалежно від того, який тип контенту ви створюєте, потрібно запитати себе: «Чи буде це цікаво моїй цільової аудиторії? Чи буде він стимулювати читачів до здійснення потрібної бізнесу цільової дії? «. Якщо відповідь «Так», отже робота рухається у правильному напрямку.

Одне з найважливіших завдань, яке стоїть перед бізнесом, – розуміння свого клієнта. Це допомагає визначити, наприклад, чим ваш бізнес відрізняється від інших, які пропонують такі ж товари. Навколо цих відмінностей і вибудовується поняття бренду.

Чим інформативнішим буде портрет покупця, тим легше знайти можливості для розширення клієнтської бази та росту продажів. Крім того, додаткове сегментування на підгрупи спрощує персоналізацію e-mail-розсилок і повідомлень за іншими каналами комунікації. Звернутися до потенційного покупця за допомогою таких засобів, як оголошення і банери, в контекстній рекламі набагато простіше і ефективніше, ніж роздавати листівки оффлайн.

На основі портрета покупця розробляються і відповідні маркетингові кампанії. Наприклад, якщо ви випускаєте шкіряні аксесуари преміум-сегмента, основними покупцями яких є топ-менеджери і підприємці, фахівець вочевидь розуміє, які налаштування таргетингу використовувати, замість того щоб відкручувати покази реклами на всю аудиторію.

***ЯК СКЛАСТИ ПОРТРЕТ ПОКУПЦЯ?***

Інформація, яку необхідно зібрати, залежить від спрямованості бізнесу. Якщо бізнес орієнтований на індивідуальних споживачів (B2C), потрібно буде дізнатися їхні демографічні характеристики, купівельні звички, доходи і канали реклами, які допоможуть залучити їх. Якщо ж робота здійснюється за B2B-моделлю, тобто орієнтована на продаж товарів/послуг підприємствам, необхідно з’ясувати, в якому секторі економіки вони працюють, скільки зазвичай витрачають і з якими постачальниками працюють.

Щоб скласти портрет покупця для B2C-сегменту, знадобляться такі критерії:

* демографічні: вік, стать, середній дохід;
* поведінкові: пріоритети, хобі, рід занять;
* географічні: регіони і окремі міста;
* мотиви покупки: проблеми чи потреби;
* сімейний статус (за необхідності).

Важливим кроком є збір інформації щодо купівельних звичок цільової аудиторії. Наприклад, потрібно визначити, чому той чи інший сегмент клієнтів завжди купує найдешевший продукт або ж воліє купувати нові моделі певних товарів.

Не потрібно надто заглиблюватися в збір інформації, деталізація повинна бути помірною. Навряд чи інформація про те, що клієнт надає перевагу шоколаду замість мармеладу і обожнює свіжий морквяний сік, стане в нагоді, якщо ви торгуєте технікою для геймерів. Потрібно визначити пріоритети покупців і фактори, які впливають на прийняття ними рішення, а не дізнаватися, який одяг вони воліють одягати вранці щопонеділка.

Якщо бізнес уже працює протягом хоча б декількох місяців, обов’язково проаналізуйте особливості трафіку на його сайт за допомогою інструментів веб-аналітики. Або ж доручіть це завдання фахівцям, які займаються його просуванням. Ви можете бути приємно здивовані тим, як багато інформації надасть аналіз трафіку та поведінки відвідувачів на сайті. Google Analytics є надзвичайно потужним інструментом для збору і сегментації даних, але для їхньої правильної інтерпретації краще скористатися послугами фахівців.

Найкращий спосіб зрозуміти, які потреби дійсно є у клієнтів і що вони хотіли б отримати від бізнесу – запитати у них самих. Для повноти отриманої інформації краще надати аудиторії максимальну кількість інструментів зворотного зв’язку. Це може бути і електронна пошта, і сторінки в соціальних мережах, телефонні інтерв’ю та опитування на сайті інтернет-магазину.

Іноді налагодити комунікацію з клієнтом і вибудувати з ним довірливі відносини не так уже й легко. У таких випадках незайвим буде обдумати способи заохочення, які можна запропонувати тим людям, які захочуть висловити свою думку. Використовуйте, наприклад, невеликий подарунок, купон або промокод на знижку. Переконайтеся у тому, що незалежно від того, з якою метою ви намагаєтеся налагодити зворотний зв’язок з аудиторією, цей процес буде для обох сторін приємним.

Розуміння своїх клієнтів має для бізнесу вирішальне значення при створенні ефективної стратегії, спрямованої на досягнення поставлених цілей. Дуже ризиковано будувати маркетингову стратегію, ґрунтуючись на одних тільки припущеннях. У цьому випадку ризик витратити час і ресурси для реалізації свідомо неправильної маркетингової стратегії буде занадто великим. Саме тому вам потрібно перестати відштовхуватися від здогадок, а приступити до формування портрета клієнта та опису цільової аудиторії, на задоволення потреб яких і орієнтований бізнес.

Чим більше є інформації про клієнтів, тим легше буде побудувати з ними довірливі відносини та домогтися лояльності з їхнього боку. Важливо відзначити, що час не стоїть на місці, і звички, інтереси та переваги потенційних клієнтів будуть змінюватися. А значить інформацію про цільову аудиторію теж потрібно буде постійно актуалізувати. У будь-якому випадку бізнес перебуває у процесі постійного розвитку, і це абсолютно нормально в умовах сучасної економіки. Адже в багатьох нішах конкуренція перебуває на такому рівні, що навіть для того, щоб перебувати на місці, доводиться постійно бігти.

*Основні методи складання портрета цільової аудиторії:*

1. Метод п’яти питань (або 5W)

В основі методики 5W авторства Марка Шеррингтона лежать п’ять питань. Відповіді на них заносяться в спеціальну таблицю для наочності.

« What? » – що за пропозицію (різновид товару) надходить покупцеві?

« Who? » – хто є клієнтами бізнесу?

« Why? » – чому необхідно купувати саме у Вас?

« When? » – коли (в який час) продукт потрібний клієнтам?

« Where? » – де саме покупцеві зручно купувати продукт?

Отримавши найбільш детальні відповіді на запропоновані питання, Ви створите аватар клієнта і визначите найкращий спосіб продати йому Вашу продукцію.

1. Метод від результату

Головним орієнтиром цього методу є кінцевий результат. Перш за все необхідно зрозуміти, яку вигоду отримає покупець, уклавши угоду. Далі можна описати шлях, пройшовши по якому, людина досягне необхідного результату.

Потім слід з’ясувати потреби, що змушують його йти по цьому маршруту. І останнім кроком є аналіз ЦА для визначення сегмента, у якого ці потреби виникнуть найімовірніше.

В результаті повинні бути сформовані кілька груп ЦА. Використовуючи ці дані, запустіть тестову кампанію, спрямовану на окремі сегменти, і оцініть правильність гіпотези.

1. Метод від особливостей продукту

Якщо товар від подібних відрізняють будь-які нестандартні характеристики, слід при пошуку цільової аудиторії відштовхуватися саме від них. Щоб визначитися з цільовою аудиторією виконайте наступні дії:

* Виберіть кілька продуктів конкурентів, проведіть їх оцінку і порівняння зі своїм;
* Організуйте опитування для виявлення особливостей і потреб існуючих покупців;
* Проведіть короткий аналіз свого продукту: плюси та мінуси, можливості для реклами і потенційні загрози;
* Розділіть всіх клієнтів на 3 підгрупи: вже купують товар, потенційні покупці та ті, хто точно не зацікавиться продуктом;
* Створіть план маркетингових активностей для залучення уваги потенційної аудиторії.

Залишається тільки тестувати свої гіпотези і спостерігати за реакцією на продукт.

1. Метод визначення ЦА від ринку

Метод заснований на обліку поточного стану ринку.

План дій такий:

* Аналіз ринку, пошук відкритих статистичних даних або замовлення комплексного звіту у компанії, що спеціалізується на цій діяльності;
* Провести опитування серед поточних і потенційних покупців;
* Вивчення конкурентів: їх УТП і особливостей продукту;
* Складання попереднього портрета клієнта;
* Тестування гіпотези.

Важливо! Незалежно від обраного методу створення портрета ЦА, не забувайте про тестування. Ніякі висновки не можна назвати справжніми, поки вони не перевірені на практиці.

*Приклад портрета ЦА для дитячого магазину*

Загальний рівень:

Ядром ЦА є жінки 35-45 років (ширший сегмент – 25-55) із середнім і низьким рівнем доходу, що живуть в невеликих населених пунктах і виховують хоча б одну дитину не доросліше семи років. Вони цінують сталість, практичність і доцільність. Їх вибір завжди раціональний і заснований на відгуках і рекомендаціях, оскільки у них відсутня матеріальна можливість здійснювати імпульсивні покупки. Думка близького кола має вирішальне значення. Схильні до впливу реклами.

Рівень товарної категорії:

Відвідують магазин не частіше ніж один раз на місяць і тільки в разі потреби: купівля одягу для дитини, заміна меблів в дитячій або подарунок при наявності вагомого приводу. Не люблять витрачати час на тривалі поїздки, тому вибирають магазини «біля будинку» і вважають за краще купити все в одному місці. Завжди відслідковують акції та знижки, порівнюють ціни в декількох торгових точках і читають відгуки перш ніж купити конкретний товар. Потенційний ризик і невизначеність не дозволяють їм робити замовлення через інтернет – товар обов’язково потрібно побачити, помацати, приміряти. Невідповідну з яких-небудь причин продукцію повертають або міняють негайно. Середній чек на придбання іграшок – не більше 800 гривень, на покупку одягу – 2 000 гривень, а на замовлення великогабаритних товарів – 3 000 гривень. Обирають магазини з доступними цінами, хорошим вибором товарів відомих марок і душевним ставленням продавців.

Рівень бренду:

Здійснюють покупки в конкретному магазині тільки тому, що так зручніше, і не є справжніми прихильниками торгової точки. Якщо ціни або асортимент перестануть їх задовольняти, відразу підуть до конкурентів. Не обговорюють магазин зі знайомими. Вважають важливим, щоб продавці їх впізнавали, пам’ятали їх потреби та рекомендували потрібні товари. Це економить час покупки.

Інформаційні ресурси для вивчення покупців:

<https://chrome.google.com/webstore/detail/website-seo-checker/cokmobldpihbibodplbdfhkbiibganfh?hl=ru&>

<https://chrome.google.com/webstore/detail/full-page-screen-capture/gfalfgfjjngkcfhadnehacimlfhenjbo>

**Тема 4. Оцінка й обґрунтування вибору організаційно-правової форми бізнесу**

Практична робота 4

ОБҐРУНТУВАННЯ ВИБОРУ ОРГАНІЗАЦІЙНО-ПРАВОВОЇ ФОРМИ БІЗНЕСУ

Завдання:

1. Проаналізуйте модельний статут ТОВ. Дайте відповідь на питання:

як формується статутний капітал товариства?

які права має учасник товариства щодо його статутного капіталу?

який порядок вступу до товариства?

який порядок виходу з нього?

що таке дивіденди і як врегульовувала виплата дивідендів модельним статутом?

які органи управління має ТОВ і яка їх компетенція?

який порядок прийняття рішення органами управління ТОВ?

що таке значні правочини та правочини щодо яких є заінтересованість?

2. Охарактеризуйте вид підприємницької діяльності, яким ви б хотіли займатися. Яку організаційно-правову форму краще вибрати для цього виду діяльності? Обґрунтуйте свою позицію керуючись вимогами чинного законодавства та за допомогою аналізу установчих документів. Порівняйте обраний вами установчий документ із статутом ТОВ.

Інформаційний ресурс: <https://zakon.rada.gov.ua/laws/show/367-2019-%D0%BF#Text>