

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

ФАКУЛЬТЕТ ЕКОНОМІЧНОЇ ІНФОРМАТИКИ

КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ

Пояснювальна записка

до дипломної роботи

МАГІСТРА

на тему: «Інформаційні технології в управлінні реакцією клієнта на
маркетингову компанію підприємства»

Виконала: студентка 2 року навчання,
за освітнім ступенем «магістр»
зі спеціальності 126 «Інформаційні системи та
технології»

Пастушенко Ю.Е.

Керівник: к.е.н., доц. Ушакова І.О.

Харків – 2019 рік

РЕФЕРАТ

Магістерська дипломна робота: 97 сторінок, 48 рисунків, 6 таблиць, 51 джерело, 2 додатки.

Об'єкт дослідження – методи прогнозування відгуку клієнтів на маркетингову компанію.

Мета дослідження – адаптація методів прогнозування до конкретної теми та розробка моделі для прогнозування поведінки клієнтів з ціллю підвищення рівня відгуку на нові товари та пропозиції.

Методи дослідження – методи системного аналізу, методи порівняльного аналізу, методи статистичного моделювання та прогнозування, методи опитування та анкетування.

Пояснювальна записка магістерської дипломної роботи містить результати дослідження методів клієнтського аналізу. Створено модель для управління відгуками клієнтів на маркетингову компанію.

АЛГОРИТМ СПИСКУ РІШЕНЬ, ДЕРЕВО РІШЕНЬ, КЛІЄНТСЬКА АНАЛІТИКА, ПРОГНОЗНА АНАЛІТИКА, УПРАВЛІННЯ ВЗАЄМВІДНОСИНАМИ З КЛІЄНТАМИ, BIG DATA, DATA ANALYTICS, DATA MINING, IBM SPSS MODELER.

ABSTRACT

Master's degree thesis: 97 pages, 48 figures, 6 tables, 51 sources, 2 applications.

Object of research - methods of predicting customer response to a marketing company.

The purpose of the research is to adapt forecasting methods to the topic of work and develop a model for predicting customer behavior in order to increase the level of response to new products and offers.

Method of research – methods of system analysis, methods of comparative analysis, methods of statistical modeling and forecasting, methods of interviewing and questionnaires.

The explanatory note of the master's thesis contains the results of research of methods of client analysis. Created a model for managing customer response of an advertising company.

DECISION LIST ALGORITHM, DECISION TREES, CUSTOMER ANALYTICS, CUSTOMER RELATIONSHIP MANAGEMENT, ПРОГНОЗНА АНАЛІТИКА, BIG DATA, DATA ANALYTICS, DATA MINING, IBM SPSS MODELER.

ЗМІСТ

Вступ.....	7
1. Інформаційні технології клієнтської аналітики в дослідженні поведінки клієнтів	9
1.1. Цілі і завдання клієнтської аналітики.....	9
1.2. Етапи розвитку та технологічні тренди в обробці даних та аналітиці	16
1.3. Інформаційні технології клієнтської аналітики	24
1.4. Інструментальні засоби клієнтської аналітики.....	28
2. Аналіз методів прогнозування відгуку клієнтів на маркетингову компанію	36
2.1. Дерево рішень	37
2.2. Алгоритм списку рішень.....	41
2.3. Висновки.....	46
3. Побудова моделі для управління відгуками клієнтів на маркетингову компанію	49
3.1. Постановка завдання	49
3.2. Збір , попередній аналіз та підготовка даних для моделювання	55
3.3. Побудова прогнозної моделі на основі ibm spss modeler	66
3.4. Аналіз отриманих результатів.....	76
Висновки	84
Список використаної літератури	87
Додаток А Відповіді на опитування ТОВ «Тубний Завод»	93
Додаток Б Оброблені дані опитування ТОВ «Тубний Завод».....	95

ВСТУП

Маркетинг на сьогоднішній день орієнтований на клієнта і в значній мірі на довгострокову стратегічну перспективу, завдяки якій організації здатні реагувати на постійні зміни зовнішнього середовища. Переваги і звички змінюються в результаті змін в ринковому середовищі, в тому числі на них впливають активність конкурентів і їх пропозиції. Згідно з концепцією управління маркетингом, основне завдання, яке стоїть перед будь-якою комерційною організацією, полягає в підвищенні задоволеності клієнтів.

Знання тренду суспільних настроїв, викликаних різними соціально-економічними процесами, дозволяє за рахунок компетентних рішень керівників нівелювати негативні явища, підтримувати і розвивати позитивне, що сприяє розвитку держави, зміцненню його економічної і військово-політичної могутності, забезпечує зростання народного добробуту, в першу чергу особливо потребуючих.

Наявність інформації, яка збирається в ході досліджень громадської думки, недостатньо для вироблення позитивних управлінських впливів. Необхідна їй науково-обґрунтована статистична обробка з метою отримання узагальнених результатів, вилучення з масивів вихідних даних нетривіальних, уявлення підсумкових показників в наочному графічному вигляді.

Зростаючі можливості збору і зберігання даних надали бізнесу розширені можливості в ретроспективному і real-time аналізі. Тепер можна простежити закономірності, зробити певні висновки про минулі невдачі. Клієнтська аналітика залучає безліч методів статистики, інтелектуального аналізу даних, аналізує як поточних даних, так і даних за минулі періоди, на основі яких складаються прогнози про майбутні події. У бізнесі моделі прогнозування використовують патерни, складені на основі даних за певний період, щоб оцінити потенційні ризики та можливості.

Метою даної роботи є адаптація методів прогнозування до конкретної предметної області та застосування сучасних інформаційних технологій для розроблення моделі прогнозування поведінки клієнтів для підвищення рівня відгуку на нові товари та пропозиції.

Об'єктом дослідження виступають процеси класифікації та прогнозування відгуку клієнтів на маркетингову компанію.

Предметом дослідження є методи класифікації та прогнозування відгуку клієнтів на маркетингову компанію, такі як метод дерева рішень та метод списку рішень. На основі особливостей, недоліків та переваг кожного методу буде обрано той, який найбільше відповідає початковим цілям роботи.

В якості інструментарію моделювання буде використаний засіб моделювання бізнес-процесів Ramus, інструментальний засіб моделювання IBM SPSS Modeler.

В ході дослідження для досягнення поставленої мети повинно бути вирішено наступні завдання:

- Обґрунтувати роль маркетингового аналізу клієнтської бази в діяльності підприємств;
- Провести аналіз основних інструментальних засобів клієнтської аналітики;
- Здійснити порівняння методів класифікації та прогнозування клієнтської поведінки;
- Створити модель прогнозування відгуку клієнтів підприємства на нові пропозиції та проведено аналіз результатів цієї моделі.

Для розв'язання визначених завдань використовувався комплекс взаємодоповнюючих методів дослідження: методи системного аналізу, методи порівняльного аналізу, методи опитування та анкетування (на етапі збору інформації), методи класифікації для побудови прогнозної моделі.

1. ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ КЛІЄНТСЬКОЇ АНАЛІТИКИ В ДОСЛІДЖЕННІ ПОВЕДІНКИ КЛІЄНТІВ

1.1. Цілі і завдання клієнтської аналітики

Використання клієнтської аналітики стає все більш важливим в сучасному світі. Це тому, що клієнти стають все більш впливовими і пов'язаними, ніж будь-коли. Клієнти тепер озброєні великою кількістю інформації, і вони мають доступ до інформації в будь-якому місці і в будь-який час. Це змушує організацію використовувати аналітику клієнтів. Поведінка клієнтів швидко змінюється в сучасному світі. Якщо ви розумієте купівельні звички і їх поведінку, то ви можете передбачити поведінку клієнтів в майбутньому, що допоможе вам своєчасно запускати відповідні продукти для клієнтів. Це також збільшить прибуток вашого бізнесу [1].

Розуміння клієнтів - це ключ до довгострокової взаємодії, лояльності та прибутковості. Зростаюча чисельність каналів, з якими клієнти можуть взаємодіяти, робить доступними величезний обсяг даних для отримання інформації про клієнтів та ефективного маркетингу. Інтеграція цих багатоканальних даних стає все більш складною, через що багато маркетологів перевантажені і не можуть отримати осмислену інформацію [2].

Аналіз клієнтів - це процес, який дає організаціям більш глибокі знання про поведінку клієнтів, які можна використовувати для прийняття ділових рішень. Аналітика клієнтів допомагає перетворити ваші дані про клієнтів в щось, що представляє цінність для організації. Вона перетворює ваші дані в маркетинговий інструмент, який допоможе перетворити всі ваші перешкоди в можливості. Клієнтська аналітика служить основою всієї маркетингової діяльності, такої як прогнозне моделювання, візуалізація даних, управління інформацією та сегментація.

Аналітика клієнтів стає критично важливою в будь-якому типі бізнесу, від малого до великого. Клієнти стали більш впливовими, ніж раніше, і вони також стали більш пов'язаними, ніж будь-коли [3]. Аналіз клієнтів дає

компаніям повне уявлення про те, як клієнти використовують їхні продукти. Аналітика особливо корисна для компаній, що пропонують технологічні пропозиції, оскільки вони можуть збирати покрокові дані про те, як клієнти, користувачі або передплатники проходять через свої сайти або додатки. На макрорівні це розкриває основні тенденції, такі як те, як клієнти виявляють свій продукт, які товари їм подобаються найбільше, де вони знаходять цінність і що змушує їх йти. На мікрорівні аналітика клієнтів дозволяє компаніям зрозуміти, хто їх користувачі як окремі особи. Вони можуть сегментувати користувачів по демографії, інтересам, поведінці і переглядати їх унікальні переміщення від компанії до компанії. Ці знання допомагають підприємствам краще обслуговувати кожного клієнта. Elavon, наприклад, це мобільне платіжний додаток, яке виявило, що його користувачі скаржаться, що не можуть завантажити додаток. Використовуючи програмне забезпечення Mixpanel для аналізу клієнтів, вони змогли миттєво ідентифікувати всіх користувачів, які намагаються завантажити додаток в несумісних системах ОС, зв'язатися з ними і запропонувати виправлення. Інший клієнт Mixpanel, STARZ PLAY, зміг сегментувати своїх клієнтів з поведінки, щоб виявити шахрайство і скоротити його в 1000 разів [4].

Коло питань, на які може відповісти клієнтська аналітика дуже велике.

Платформи аналітики клієнтів можуть відповідати на питання з усього, що може бути виміряно або відстежено. Проте, відповіді, які вони дають, так само хороші, як і питання, що задаються їм. На відкриті, суб'єктивні або відносні запитання наприклад «Чи є наш продукт кращим?» Важко відповісти однозначно. Хороші питання є конкретними, легко перевіряються або спростовувати і пов'язані з основними бізнес-цілями, такими як придбання, отримання доходу, утримання і взаємодія.

Основні напрямки клієнтської аналітики та загальні питання з аналізу клієнтів наведені далі:

1. Залучення клієнтів:
 - 1.1. Які канали залучають самих нових клієнтів?

1.2. Який найбільш поширений шлях клієнта від обізнаності до пропаганди?

2. Прибутковість клієнтів:

2.1. Які наші найприбутковіші канали доходу?

2.2. Які клієнти найприбутковіші?

3. Утримання клієнтів:

3.1. Де ми втрачаємо клієнтів і чому?

3.2. Яка види поведінки пов'язані з високим рівнем утримання?

4. Взаємодія з клієнтами:

4.1. Які особливості резонують з якими клієнтами?

4.2. Який оптимальний досвід для клієнтів?

Щоб отримувати точні відповіді після аналізу клієнтів, підприємства збирають і аналізують свої дані. Для цього правильна реалізація є ключовою [4].

Щоб реалізувати аналітику клієнтів і отримати корисні відомості, компанії повинні чітко слідувати правилам.

Аналітика клієнта вимагає поєднання збору даних, обґрунтованих управлінських рішень на основі цих даних і готовності експериментувати з альтернативними стратегіями для створення нових даних. Ваші послуги контролера можуть допомогти вам створювати звіти з аналізу клієнтів. Основні заходи, які необхідно відстежувати і намагатися поліпшити, включають наступне:

1. Залучення клієнтів

Заходи по залученню клієнтів включають витрати, коефіцієнти конверсії і розбивку по маркетинговим каналам і кампаніям. Згідно з опитуванням DataMatics, компанії, які відстежують показники ефективності залучення клієнтів, в 23 рази частіше говорять, що вони випереджають своїх конкурентів в цій галузі.

2. Міграція в прибуткові сегменти

Багато підприємств пропонують рівні обслуговування або продукту, де більш низькі рівні можуть мати більш низьку маржу або навіть бути збитковими. Мета полягає в тому, щоб вивести клієнтів на більш високий рівень.

Щоб дізнатися, чи працює ця стратегія, вам потрібно знати кількість нових і перенесених клієнтів на кожному рівні, скільки часу знадобилося перенесеному клієнту для переміщення, що змушує клієнтів мігрувати і якщо будь-які ваші продажі або маркетингові зусилля роблять міграцію більш-менш вірогідною.

Компанії, які відстежують ці заходи, в 21 разів частіше повідомляють про те, що вони перевищили показники з міграції в прибуткові сегменти.

3. Рентабельність клієнта

Відстеження показників прибутковості клієнтів призводить до того, що в 18,8 разів більше шансів повідомити про те, що вони перевершують конкурентів за прибутковістю на одного клієнта. Показники доходу повинні включати загальний обсяг продажів клієнту, середній розмір продажів, час між покупками і те, які конкретні товари або послуги купує клієнт.

Показники витрат повинні включати як прямі, так і непрямі витрати. Прямі витрати включають такі речі, як вартість проданих товарів, вартість доставки і збитки, пов'язані з поверненням. Непрямі витрати включають в себе приховані витрати, такі як час, який торгові представники повинні витратити з кожним клієнтом або необхідність відволікати ресурси для виконання незвичайних запитів.

4. Лояльність та утримання клієнтів

Лояльність і утримання клієнтів ділиться на дві частини - ідентифікацію та відстеження.

По-перше, ви повинні вміти ідентифікувати окремих клієнтів. Це легко зробити, якщо у вас є облікові записи служб або ви збираєте інформацію про доставку для онлайн-замовлень, але це стає складніше в роздрібній торгівлі.

Одним з найпростіших і розповсюджених способів відстеження роздрібних клієнтів є використання карт лояльності або дисконтних карт.

Як тільки ви впізнаєте свого клієнта, відстеження терміну зберігання - це питання визначення середнього проміжку часу між першим і останнім замовленням клієнта і тим, скільки клієнтів не розміщують друге замовлення. Щоб поліпшити показники і терміни зберігання, ви можете отримувати дані для пошуку тенденцій, які окремі клієнти зберігали протягом тривалого і короткого періоду.

Лояльність можна відстежувати через середній час між замовленнями і які конкретні товари купує клієнт. Це може допомогти вам визначити, чи є клієнт лояльним до вашого бренду, чи обирає ціну між конкурентами або вважає, що незручно мати справу з вами через відстань або тривалий час доставки.

Чим більше ви розумієте своїх клієнтів, тим краще ви зможете передбачити і задовольнити їх потреби, і збільшення продажів природно піде вгору[5].

На рисунку 1.1 представлено зростання показників ефективності організацій, за рахунок більш якісної обробки даних в системах аналітики і подальшого прийняття управлінських рішень на основі аналізу [6].



Рисунок – 1.1 Вплив застосування систем аналітики на показники ефективності клієнтського сервісу

Відповідно до представленої графіки організації, які використовують системи аналітики клієнтського сервісу отримують куди більш відчутний приріст ключових показників ефективності. Ці показники відносяться до різних аспектів діяльності: рівень задоволеності клієнтів, операційна ефективність і фінансові результати.

Щоб отримати максимальну віддачу від своїх даних, виконайте наступні рекомендації:

1. Об'єднання сховищ даних - використовуйте інструменти, які об'єднують дані з багатьох джерел для кращого виявлення тенденцій і закономірностей, які виявлятимуть клієнтські настрої, больові точки клієнтів і потреби на різних етапах процесу прийняття рішень.

2. Використання даних в режимі реального часу - дані і аналітика в режимі реального часу дозволяють надавати цільові, персоналізовані повідомлення, які відповідають потребам клієнтів в потрібний час і в потрібних точках доступу протягом усього шляху клієнта.

3. Зміна необроблених даних на практичну інформацію - необроблені дані корисні тільки в тому випадку, якщо ви можете отримати з них інформацію. Використовуйте рішення, які дозволяють легко витягувати корисні відомості з ваших даних для прийняття рішень.

4. Забезпечте постійну роботу всіх каналів. Нинішні клієнти взаємодіють з брендами з безлічі каналів і точок взаємодії. Зосередьтеся на створенні узгодженого досвіду у всіх каналах, таких як електронний маркетинг, соціальні мережі, відео і всі точки взаємодії, через які клієнти взаємодіють з вашим брендом.

5. Зосередитися на особистості. Сегментація колись була цінною стратегією, і хоча вона все ще може виявитися корисною, сьогоднішні клієнти очікують, що до нього будуть ставитися як до окремих осіб. Орієнтовані на клієнта організації можуть досягти цього рівня персоналізації тільки тоді, коли у них є атомарне уявлення про кожного клієнта.

Поставивши клієнта в центр своєї маркетингової стратегії і використовуючи передові, комплексні аналітичні інструменти, забезпечуючи досвід клієнтського попиту і випереджаючи конкуренцію, можна досягти успіху для організацій всіх розмірів [7].

Незалежно від розміру компанії, аналіз даних про клієнтів є одним з пріоритетів, оскільки він допомагає досягти наступних 10 цілей:

1. Короткострокові:
 - 1.1. Сегментування клієнтів на основі схожих моделей поведінки для кращої орієнтації на їх потреби;
 - 1.2. Розуміння пересувань клієнтів (які продукти і коли вони купують, які канали вони вважають кращими).
2. Середньострокова перспектива:
 - 2.1. Прогнозування продажів;
 - 2.2. Прогнозування реакції клієнтів на маркетингові заходи і пропозиції їм відповідних продуктів і акцій;
 - 2.3. Зниження витрат на проведення маркетингових кампаній;
 - 2.4. Встановлення адресної / персоналізованої комунікації.
3. Довгострокові:
 - 3.1. Підвищення задоволеності клієнтів;
 - 3.2. Підвищення лояльності й утримання клієнтів;
 - 3.3. Оптимізація портфеля продуктів для кращого задоволення потреб клієнтів.

На щастя, аналітика даних про клієнтів, підтримувана технологіями big data і business intelligence, не відстає від потреб бізнесу і може забезпечити продажі, маркетинг, розробку продуктів і обслуговування клієнтів з будь-якої галузі з необхідною інформацією [8].

У минулому технологія ефективного і точного стиснення великих обсягів клієнтських даних для досягнення дієвої аналітики була просто недоступна. Потім, коли технологія стала доступною, компанії, які

ефективно використовували її, отримали значну конкурентну перевагу від використання аналітики даних про клієнтів.

Реальність сьогодні така, що більшість компаній використовують ту чи іншу форму клієнтської аналітики, тим самим нейтралізуючи конкурентну перевагу, яке вона колись забезпечувала. Тому компанії прагнуть йти далі, ніж дозволяє "стандартна" аналітика даних про клієнтів, щоб підвищити ефективність маркетингу клієнтів і відновити значну конкурентну перевагу, яке клієнтська аналітика запропонувала своїм раннім послідовникам.

Але існують деякі проблеми в аналізі клієнтських даних.

Перший рівень складності в отриманні точної і дієвої інформації про клієнтів з даних компанії полягає в тому, що відповідні дані розкидані між різними базами даних, в різних відділах, використовуючи різні формати. Для того щоб по-справжньому зрозуміти клієнта, вкрай важливо зібрати все найважливіше "єдине уявлення клієнта" (або 360-градусне уявлення клієнта): єдиний, плоский, послідовний конгломерат всіх даних, що відносяться до кожного клієнта. Немає іншого способу поглянути на клієнта цілісно, включаючи його демографію, поведінку, переваги і інтереси. Це легше сказати, ніж зробити, і іноді може зайняти багато місяців зусиль.

Після того, як всі дані будуть об'єднані, очищені і синхронізовані, наступне завдання - отримання корисного значення з даних - інтелект клієнта-так що оптимальні та своєчасні бізнес-рішення можуть бути прийняті на основі даних. Основні програми бізнес-аналітики (BI) та аналітики клієнтів підходять до цього завдання за допомогою засобів для створення батареї запитів, панелей моніторингу та звітів для візуалізації того, що відбувається.

1.2. Етапи розвитку та технологічні тренди в обробці даних та аналітиці

В наступний час є величезний прогрес в області інформації і технологій. Дані та аналітика були найбільш часто використовуваними

словами в останнє десятиліття або два. Таким чином, важливо знати, чому вони взаємопов'язані, які ролі на ринку в даний час розвиваються і як вони змінюють бізнес.

Технологія часто розглядається як благо для тих, хто вже знає про її потенціал, а також може бути прокляттям для аудиторії, яка не може йти в ногу з її швидким ростом. Кожна епоха мала свої моменти прориву і рівну частку жертв (або, як я хотів би назвати їх побічний збиток). На сьогоднішній день кожна грошово-кредитна галузь повністю покладається на дані і аналітику для свого виживання.

Еволюція аналітики - це відносно короткий період, близько 30 років, і він все ще розгортається.

Аналітика 1.0 - епоха «бізнес-інформації». Це було повстання сховища даних, де клієнт (бізнес) та виробничі процеси (транзакції) були централізовані в одному величезному сховищі, такому як eCDW (Enterprise Consolidated Data Warehouse). Досягнутий реальний прогрес в об'єктивному, глибокому розумінні важливих бізнес-явищ, що дає менеджерам засноване на фактах розуміння, що виходить за межі інтуїції при прийнятті рішень.

Дані, пов'язані з eCDW, збиралися, перетворювалися і запитувалися з використанням інструментів ETL & BI. Тип аналітики, що використовується на цьому етапі, був в основному класифікований як описовий (що відбулося) і діагностичний (чому щось сталося).

Тим не менше, основні обмеження, що спостерігаються протягом цієї епохи, полягали в тому, що потенційні можливості даних використовувалися тільки всередині організацій, тобто діяльність з бізнес-аналітики стосувалася лише того, що сталося в минулому, і не давала жодних прогнозів щодо її тенденцій у майбутньому [10].

Аналітика 2.0 - епоха «великих даних». Певні недоліки попередньої епохи стали більш помітними з кожним днем, коли компанії вийшли із зони комфорту і почали прагнути до більш широкого (якщо не кращого) підходу до створення складної форми аналітики. Клієнти несподівано добре

відреагували на цю нову стратегію і зажадали інформацію з зовнішніх джерел (кліки, соціальні мережі, інтернет, громадські ініціативи і т.д). Потреба в потужних нових інструментах і можливість отримувати прибуток, надаючи їх - швидко стали очевидними. Неминуче, термін «великі дані» був придуманий для того, щоб відрізнити маленькі дані, які генеруються виключно внутрішніми системами транзакцій фірми.

Те, що компанії очікували від своїх співробітників, повинно було допомогти інженерним платформ обробляти великі обсяги даних за допомогою механізму швидкої обробки. Чого вони не очікували, так це величезного відгуку від формується групи людей або того, що сьогодні більш відоме як "співтовариство з відкритим вихідним кодом". Це було відмінною рисою Analytics 2.0.

Завдяки безпрецедентної підтримки спільноти, такі ролі, як інженери великих даних і адміністратори Hadoop вирости в робочому секторі і тепер були критично важливі для кожної ІТ-організації. Технологічні фірми поспішили побудувати нові фреймворки, які були б не тільки здатні поглинати, перетворювати і обробляти великі дані навколо eCDW / Data Lakes, а й інтегрувати прогностичну (що, швидше за все, відбудеться) аналітику над ним. Це дозволяє використовувати результати описової та діагностичної аналітики для виявлення тенденцій, кластерів і винятків, а також для прогнозування майбутніх тенденцій, що робить його цінним інструментом прогнозування [11].

Аналітика 3.0 - епоха даних, що генеруються споживчими товарами. Новаторські фірми big data почали інвестувати в аналітику для підтримки продуктів, послуг і функцій, орієнтованих на клієнтів. Вони залучали глядачів на свої веб-сайти за допомогою кращих алгоритмів пошуку, рекомендацій, пропозицій по продуктах для покупки і реклами, все це було викликано аналітикою, заснованою на величезних обсягах даних. Спалах феномена великих даних поширювався подібно вірусу. Таким чином, тепер не тільки технологічні фірми та онлайн-компанії можуть виробляти товари та

послуги на основі аналізу даних, це може практично кожна фірма в кожній галузі.

З іншого боку, широке визнання технологій обробки великих даних зробило неоднозначний вплив. У той час як технічно підковані гіганти просувалися вперед, роблячи більше грошей, більшість інших підприємств і нетехнічних фірм страждали за рахунок незнання про дані. В результаті була введена область дослідження Data Science, яка використовує наукові методи, дослідні процеси, алгоритми і системи для вилучення знань і уявлень з даних в різних формах [12].

Технологічна індустрія вибухнула перевагами впровадження методів науки про дані і використовувала всю міць прогностичної і предписательної аналітики (які дії зробити), тобто усунути майбутню проблему або в повній мірі скористатися перспективною тенденцією. Компанії почали конкурувати з аналітики не тільки в традиційному сенсі - покращуючи внутрішні бізнес - рішення, але і створюючи більш цінні продукти і послуги. В цьому і полягає суть Analytics 3.0 [13].

Звичайно, ці переваги не позбавлені своїх проблем, і з досвіду провідних компаній в цій галузі три основні проблеми обертаються навколо вивчення нових обчислювальних методів, нових статистичних методів і нового мислення.

Так як технологічний потенціал весь час зростає, що б не втратити конкурентну перевагу лідери даних і аналітики повинні вивчити потенційний вплив тенденцій на бізнес і, відповідно, скорегувати бізнес-модель.

Основна проблема, яка викликала цифровий збій (занадто багато даних), відкрила і безпрецедентну можливість. Величезна кількість даних укупі зі зростаючою міццю інструментів обробки, що надаються хмарними технологіями, дає чітке уявлення про те, чому можна зараз навчити і які алгоритми можна запуснути в великих масштабах, необхідні для реалізації всього потенціалу ІТ.

Компанія Gartner рекомендує лідерам в області обробки даних і аналітики обговорити з представниками бізнесу основні пріоритети розвитку компанії і подумати про те, як вони можуть інтегрувати її в роботу наступних напрямків.

Далі буде представлено список найкращих технологічних трендів в області даних і аналітики, які зображено на рисунку 1.2, до якого не включено тренди, які знаходяться менш ніж в трьох роках від основного впровадження (наприклад, self-service analytics і BI) або більш ніж в п'ять років (наприклад, квантові обчислення).

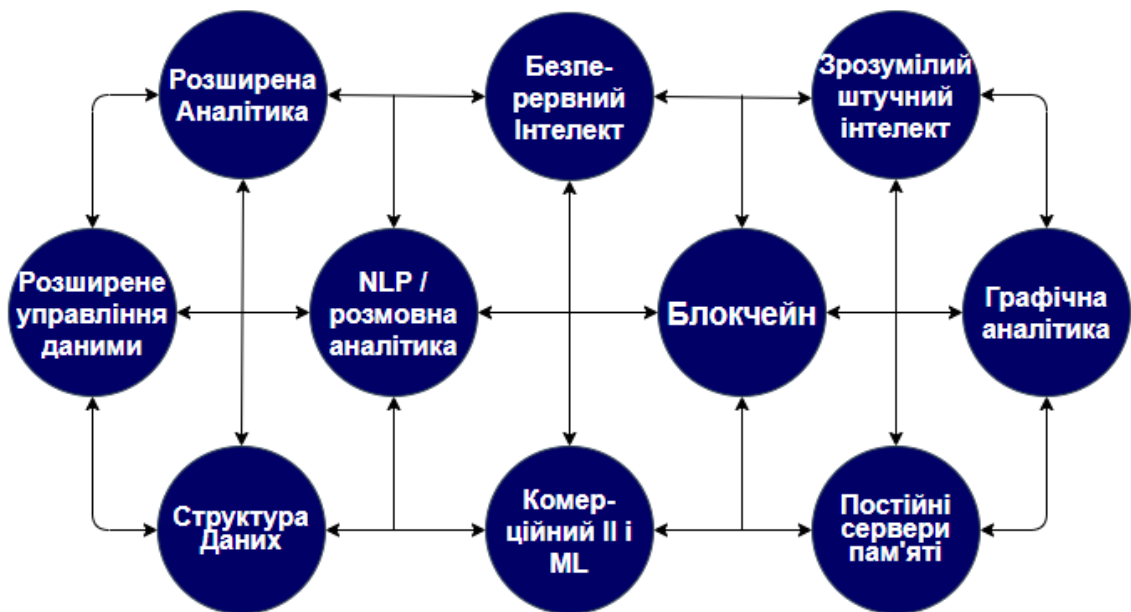


Рисунок – 1.2 Схема технологічних трендів в області даних і аналітики

1. Розширена Аналітика

Доповнена аналітика стане наступною хвилею збоїв на ринку даних і аналітики, оскільки вона буде використовувати методи машинного навчання (ML) і технології штучного інтелекту для перетворення методів розробки, споживання і спільного використання аналітичного контенту.

До 2020 року доповнена аналітика стане головним пунктом продажу для аналітичних і BI-рішень. Лідери даних і аналітики повинні планувати перехід на розширену аналітику в міру розвитку можливостей платформи.

2. Розширене управління даними

Розширене управління даними використовує можливості ML і технології штучного інтелекту для створення категорій управління даними, включаючи якість даних, управління метаданими, управління основними даними, інтеграцію даних, а також СУБД (системи управління базами даних) з автоматичним налаштуванням і самоналаштуванням.

Воно автоматизує багато ручних завдань і дозволяє менш технічно кваліфікованим користувачам бути більш автономними, використовуючи дані. Це також допомагає висококваліфікованим технічним ресурсам бути в змозі зосередитися на більш цінних завданнях. Ручні завдання в області управління даними будуть скорочені на 45% через ML і автоматизоване управління рівнем обслуговування до кінця 2022 року.

3. Безперервний інтелект

Безперервна аналітика використовує безліч технологій, таких як розширена аналітика, обробка потоку подій, оптимізація, управління бізнес-правилами і ML. Це шаблон проектування, в якому аналітика в реальному часі поєднується з бізнес-операціями, обробкою поточних і історичних даних для приписання дій у відповідь на події. Він забезпечує автоматизацію прийняття рішень або підтримку прийняття рішень.

До 2022 року більше половини великих нових бізнес-систем будуть включати безперервний інтелект, який використовує контекстні дані в реальному часі для поліпшення рішень.

4. Зрозумілий штучний інтелект

Моделі штучного інтелекту все частіше використовуються в управлінні даними і замінюють людське прийняття рішень. Однак більшість моделей ШІ не в змозі пояснити, чому вони прийшли до конкретної рекомендації або рішення. Ось тут-то і з'являється зрозумілий ШІ.

Зрозумілий ШІ в науці про дані і платформах ML полягає в створенні пояснення моделей даних з точки зору точності, атрибутів, статистики моделей і функцій на природній мові.

5. Графічна аналітика

Графічна аналітика - це набір аналітичних методів, який дозволяє досліджувати відносини між зацікавленими суб'єктами, такими як організації, люди і транзакції. Застосування графічної обробки і графічних СУБД буде зростати на 100% щорічно до 2022 року.

6. Структура Даних

Структура даних забезпечує єдину і узгоджену структуру управління даними, яка забезпечує безперешкодний доступ до даних і їх обробку в рамках розрізненого зберігання.

До 2022 року індивідуальні конфігурації структури даних будуть розгорнуті головним чином у вигляді статичної інфраструктури, що змусить організації перейти на нову хвилю витрат, щоб повністю змінити дизайн для більш динамічних підходів до сіток даних.

7. NLP / розмовна аналітика

З прогнозів можна сказати, що 50% аналітичних запитів будуть генеруватися за допомогою пошуку, обробки природної мови (NLP) або голосу, або будуть автоматично генеруватися до 2020 року. Необхідність аналізувати складні комбінації даних і робити аналітику доступною для всіх в організації приведе до більш широкого впровадження, дозволяючи аналітичним інструментам бути такими ж легкими, як інтерфейс пошуку або розмова з віртуальним помічником.

8. Комерційний штучний інтелект і Machine Learning

Згідно з прогнозами на найближче майбутнє 75% нових рішень для кінцевих користувачів, що використовують методи штучного інтелекту & ML, будуть побудовані з комерційними рішеннями, а не з платформами з відкритим вихідним кодом до 2020 року.

Комерційні постачальники вбудували з'єднувачі в екосистему з відкритим вихідним кодом, і вони надають підприємствам функції, необхідні для масштабування і демократизації II і ML, такі як управління проектами та моделями, передача даних, повторне використання, прозорість і цілісність платформи, інтеграція, які відсутні в технологіях Open Source.

9. Блокчейн

Основна цінність технологій розподіленої бухгалтерської книги, таких як блокчейн, полягає в забезпеченні децентралізованого довіри через мережу ненадійних учасників. Наслідки для випадків використання аналітики значні, особливо ті, які використовують відносини учасників і їх взаємодії.

Але, за словами Gartner, пройде ще кілька років, перш ніж чотири або п'ять основних блокчейн-технологій стануть домінуючими. До тих пір, поки це не відбудеться, підприємства замість цього будуть частково інтегруватися з блокчейн-технологіями і стандартами, які, ймовірно, будуть продиктовані їх домінуючими клієнтами або мережами [14]. Це включає інтеграцію з існуючою інфраструктурою даних і аналітики. Витрати інтеграції можуть переважувати будь-які потенційні вигоди.

Блокчейн є джерелом даних, а не базою даних, і не замінять існуючі технології управління даними.

10. Постійні сервери пам'яті

Нові технології постійної пам'яті допоможуть знизити витрати і складність впровадження архітектури з підтримкою обчислень в оперативній пам'яті (IMC). Постійна пам'ять представляє новий рівень пам'яті між флеш-пам'яттю DRAM і NAND, яка може забезпечити економічну масову пам'ять для високопродуктивних робочих навантажень.

Ця технологія має потенціал для підвищення продуктивності додатків, доступності, часу завантаження, методів кластеризації і методів забезпечення безпеки, зберігаючи при цьому витрати під контролем. Це також допоможе організаціям знизити складність їх прикладних програм і архітектури даних за рахунок зменшення необхідності дублювання даних [15].

Згідно досліджень, доповнена аналітика, безперервний інтелект і зрозумілий штучний інтелект (ШІ) є одними з головних тенденцій в області даних і аналітики, які мають значний руйнівний потенціал протягом

наступних трьох-п'яти років. Gartner рекомендує лідерам data and analytics обговорити зі старшими керівниками business C suite їх критичні пріоритети і вивчити, як ці тенденції можуть дозволити їм досягти успіху.

1.3. Інформаційні технології клієнтської аналітики

Насправді сьогодні кількість існуючих цифрових даних зростає швидкими темпами, подвоюється кожні два роки і змінює наш спосіб життя. За даними IBM, у 2012 році щодня генерувалося 2,5 мільярда гігабайт (ГБ) даних [16].

У статті Forbes говориться, що дані мають зростати швидше, ніж будь-коли раніше, і до 2020 року близько 1,7 мегабайта нової інформації буде створюватися кожну секунду для кожної людини на планеті [17]. Що робить надзвичайно важливим знати основи цієї області як мінімум. Зрештою, ось де наше майбутнє.

Сфера Big Data Analytics, яка знаходиться на перетині трьох напрямків: Big Data, Data Analysis та Methods and Algorithms (рисунок 1.3), ґрунтується на тому, що це таке, де вони використовуються, навичках, які вам необхідні, щоб стати професіоналом в цій області, і перспективи отримання заробітної плати в кожній області [18].

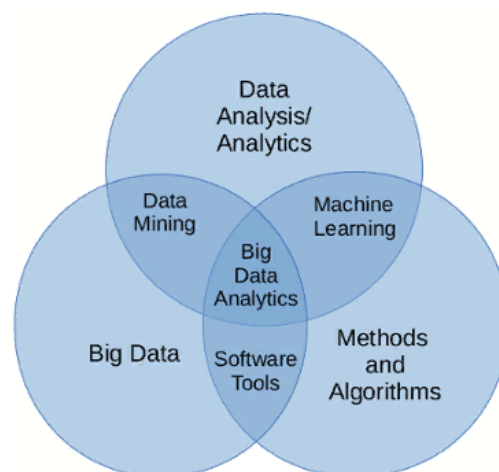


Рисунок – 1.3 Схема перетину термінів з Data Science

Перше, що потрібно розглянути – це поняття Big Data. Big Data - це в основному цифрові неструктуровані дані, які сучасне суспільство

намагається структурувати, об'єднувати і отримувати інформацію. Обсяг неструктурованих даних зростає в геометричній прогресії, і кошти для їх обробки повинні бути більш складними в порівнянні з інструментами аналізу даних, орієнтованими на невеликі набори даних. Великі дані мають на увазі набори даних, які занадто великі для зберігання в пам'яті одного комп'ютера і повинні зберігатися і оброблятися розподілено. Для останніх повинні бути застосовані нові алгоритмічні моделі розподілу [19].

Ще одна проблема, пов'язана з великими даними, полягає в тому, що більша частина цих даних недоступна. Велика частина сьогоденних даних зберігається в безпеці. Незалежно від цього ще залишилися загальнодоступні дані все ще величезні і підпадають під концепцію великих даних [20].

Застосування Big Data:

1. Великі дані для фінансових послуг: компанії, що випускають кредитні карти, роздрібні банки, приватні управляючі активами, страхові компанії, венчурні фонди і інституційні банки використовують великі дані для своїх фінансових послуг. Загальна проблема серед них усіх - це величезні обсяги багатоструктурних даних, які живуть в безлічі різнорідних систем, які можуть бути вирішені за допомогою великих даних. Таким чином, великі дані використовуються декількома способами, такими як:

- 1.1. Аналітика клієнтів
- 1.2. Аналітика відповідності
- 1.3. Аналітика шахрайства
- 1.4. Оперативна аналітика

2. Великі дані в комунікаціях. Залучення нових абонентів, утримання клієнтів і розширення бази існуючих абонентів є головними пріоритетами для постачальників телекомунікаційних послуг. Рішення цих проблем полягають у здатності комбінувати і аналізувати масу призначених для користувача даних і машинних даних, які створюються кожен день.

3. Великі дані для роздрібною торгівлі: полягає в тому, щоб бути конкурентоспроможними, краще зрозуміти клієнта. Для цього потрібна

здатність аналізувати всі розрізнені джерела даних, з якими компанії стикаються щодня, включаючи веб-журнали, дані транзакцій клієнтів, соціальні мережі, дані кредитних карт брендового магазину і дані програми лояльності.

Зрозумівши, що таке Big Data слід роздивитися Data Mining, який найчастіше має справу з великими даними.

Data Mining - це аналіз даних з метою виявлення непередбачених закономірностей або властивостей. Це перетворює неструктуровані дані в корисну інформацію.

Це обчислювальний процес виявлення закономірностей у великих масивах даних (включаючи абстракцію великих даних), що включає методи на перетині штучного інтелекту, машинного навчання та систем баз даних.

Data Mining є мультидисциплінарною областю, яка зображена на рисунку 1.4, яка виникла і розвивається на базі досягнень прикладної статистики, розпізнавання образів, методів штучного інтелекту, теорії баз даних та інших [21].

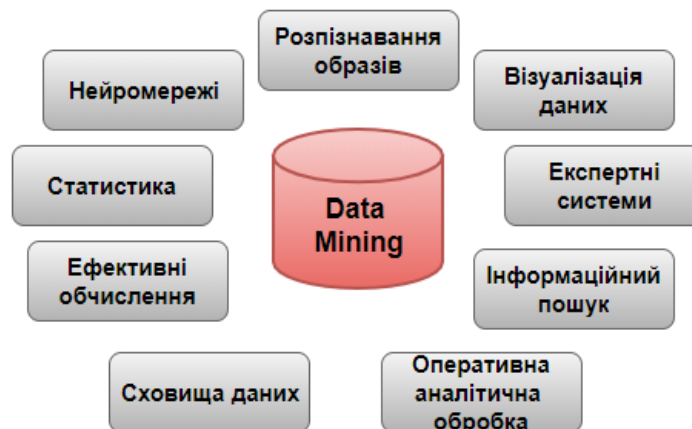


Рисунок – 1.4 Data Mining - мультидисциплінарна область

Кінцевою метою інтелектуального аналізу даних є прогнозування, а прогнозний інтелектуальний аналіз даних є найбільш поширеним видом інтелектуального аналізу даних, який має найбільші прямі бізнес-додатки. Процес інтелектуального аналізу даних складається з трьох етапів: початкове дослідження, побудова моделі або ідентифікація шаблону з перевіркою і

розгортання (тобто застосування моделі до нових даних для генерації прогнозу).

Етап 1: Розвідка. Цей етап зазвичай починається з підготовки даних, яка може включати очистку даних, перетворення даних, вибір піднаборів записів. Потім, в залежності від характеру аналітичної задачі, цей перший етап процесу інтелектуального аналізу даних може включати в себе будь-який простий вибір простих предикторів для регресійної моделі, щоб розробити дослідний аналіз з використанням широкого спектру графічних і статистичних методів з метою виявлення найбільш значущих змінних і визначення складності та / або загального характеру моделей, які можуть бути прийняті до уваги на наступному етапі.

Етап 2: Побудова моделі та перевірка. Цей етап включає в себе розгляд різних моделей і вибір найкращої на основі їх прогностичної ефективності (тобто пояснення даної мінливості і отримання стабільних результатів по вибірках). Це може звучати як проста операція, але насправді це іноді вимагає дуже складного процесу.

Етап 3: Розгортання. Цей заключний етап включає використання моделі, обраної як найкращу на попередньому етапі, і її застосування до нових даних для генерації прогнозів або оцінок очікуваного результату [22].

Наступне і останнє поняття, яке тут розглядається - це Data Analytics.

Data Analytics наука вивчення необроблених даних, щоб зробити висновки про цю інформацію.

Аналітика даних включає застосування алгоритмічного або механічного процесу для отримання інформації. Наприклад, переглядаючи кілька наборів даних, щоб знайти значущі кореляції між собою.

Вона використовується в ряді галузей, щоб дозволити організаціям і компаніям приймати більш правильні рішення, а також перевіряти і спростовувати існуючі теорії або моделі. У центрі уваги Data Analytics лежить логічний висновок, який являє собою процес отримання висновків, заснованих виключно на тому, що дослідник вже знає.

Застосування Data Analytics:

1. Охорона здоров'я. Основна проблема для лікарень, в яких знижується вартість, полягає в тому, щоб ефективно лікувати якомога більше пацієнтів, з огляду на поліпшення якості медичної допомоги. Дані про інструменти і машини все частіше використовуються для відстеження, а також оптимізації потоку пацієнтів, лікування та обладнання, що використовується в лікарнях.

2. Ігри: Data Analytics допомагає збирати дані для оптимізації і витратити як всередині, так і між іграми. Ігрові компанії отримують уявлення про неприязні, відносини і симпатії користувачів.

3. Управління енергоспоживанням. Більшість фірм використовують аналітику даних для управління енергоспоживанням, включаючи управління інтелектуальними мережами, оптимізацію енергоспоживання, розподіл енергії і автоматизацію будівель в комунальних компаніях. Додаток тут зосереджено на контролі та моніторингу мережевих пристроїв, диспетчерських бригадах і управлінні перебоями в обслуговуванні [18].

Підводячи підсумок, якщо порівняти Big Data та Data mining згідно з проведеними вище декомпозиціями визначень - Data mining як би "виграє" у Big Data за рахунок демократичного підходу до обсягу даних.

Згідно зі списком завдань, що вирішуються за допомогою методів Big Data і Data Mining, "виграє" вже Big Data, так як вирішує завдання збору і зберігання даних.

Таким чином, якщо врахувати, що досліджувати малі обсяги даних в принципі не доцільно, то значення поняття Data Mining повністю включено в значення поняття Big Data.

1.4. Інструментальні засоби клієнтської аналітики

Сучасний напрямок в роботі з клієнтами використовує клієнтський аналіз. Ця сфера в нинішній час динамічно розвивається і має великий

потенціал. На IT ринку є достатня кількість інструментальних засобів, які використовуються в даній сфері. Серед них найбільш поширені це SPSS modeler, SAP Predictive Analytics, QLIKVIEW, Rapid Miner, STATISTICA Data Miner, Deductor, SAS Enterprise Miner.

Аналіз інформації про клієнтів (управління взаємовідносинами з клієнтами):

- Забезпечення кращого розуміння поведінки клієнтів за допомогою побудови прогностичних профілів;
- Прогнозування різних варіантів поведінки клієнтів;
- Максимізація віддачі від клієнтів за рахунок прогнозування їх інтересів за допомогою прогностичної аналітики [23].

Наприклад, які товари і послуги будуть необхідні клієнтам і, таким чином, максимізувати віддачу від клієнтів зображено на рисунку 1.5 [24].



Рисунок – 1.5 Архітектура аналітичної системи максимізації віддачі від клієнтів

Далі розглядається теоретико-методологічні і практичні можливості статистичних пакетів SPSS та SAP Predictive Analytics. На сьогоднішній день існує багато універсальних програм обробки і аналізу статистичної

інформації. Завдяки колу охоплених завдань, вони можуть бути корисні не тільки студентам, а й науковцям, економістам, вирішальним завданням аналізу і прогнозу з використанням статистичних даних.

Програмне забезпечення IBM SPSS Modeler дозволяє виявляти неочевидні закономірності в Ваших даних, будувати надійні моделі і оперативно впроваджувати отримані результати в процеси прийняття рішень.

Робоча панель IBM SPSS Modeler зображена на рисунку 1.6 інтуїтивно зрозуміла та легка в використанні.

SPSS представляє дружній призначений для користувача інтерфейс, який робить процес введення і статистичного аналізу доступним для початківця і зручним для досвідченого користувача. Редактор даних пакета дозволяє зручно (табличним способом) вводити і коректувати вхідні дані. SPSS дає можливість отримувати безліч високоякісних графіків і різних діаграм. За допомогою пакету, використовуючи таблиці, прості меню і діалогові вікна, можна виконувати, по-перше, аналіз величезних файлів даних з тисячами змінних, і по-друге, робити все це без рядкової записи команд в мові програмування.

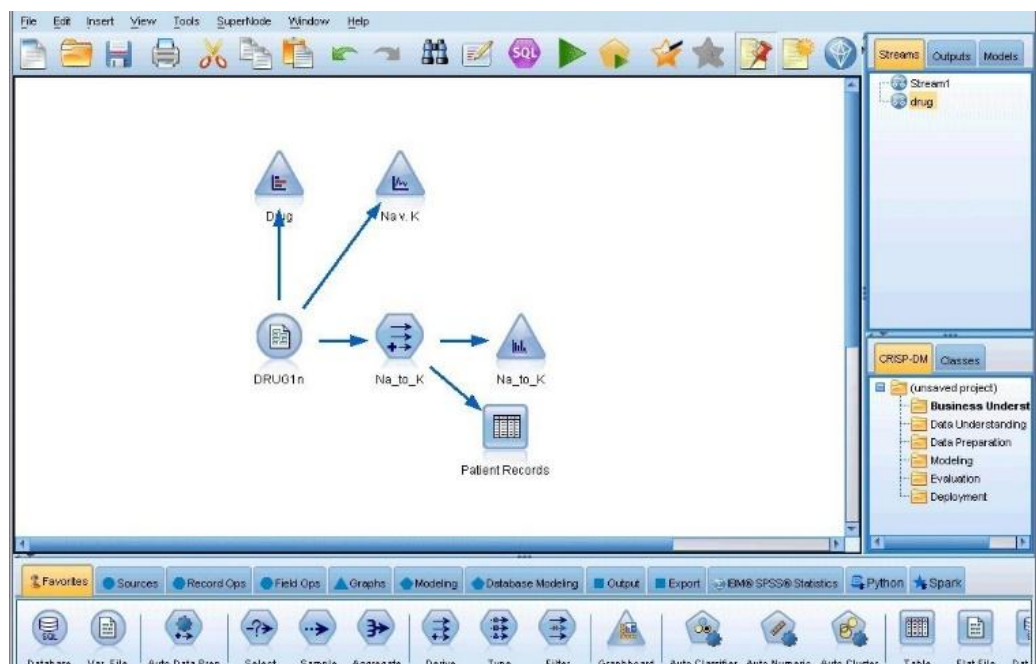


Рисунок 1.6 – Робочий простір IBM SPSS Modeler

IBM SPSS Modeler - програмне забезпечення для data mining, що поєднує в собі всі необхідні технічні та аналітичні інструменти для щоденної роботи з даними, розробки та впровадження ефективних прогностичних моделей. Завдання, які вирішуються за допомогою IBM SPSS Modeler, дуже різноманітні. Серед них:

- Підвищення ефективності політики утримання клієнтів;
- Стимуляція крос-продажів і повторних покупок;
- Сегментація клієнтів;
- Мінімізація кредитних ризиків;
- Виявлення та запобігання шахрайства [25].

Основні переваги даного ПО:

- простота доступу до даних, їх обробки, агрегування і зміни структури;
- швидке побудова та оцінка якості моделей на основі використання ефективних статистичних методів і алгоритмів машинного навчання;
- оперативне впровадження побудованих моделей в практику прийняття рішень;
- швидке отримання віддачі від інвестицій в програмне і апаратне забезпечення за рахунок високої продуктивності, інтегрованості і масштабованості;
- відповідність міжгалузевому стандарту data mining CRISP-DM.

Недоліки програми SPSS Modeler:

- високі вимоги до системи комп'ютера (потрібно 1GB оперативної пам'яті, 800MB пам'яті на жорсткому диску і процесор з частотою 1GHz і вище);
- висока ціна в порівнянні зі статистичними пакетами аналогічного рівня.

Можливі області застосування SPSS: збереження та аналіз даних опитувань, маркетингових досліджень і продажів, фінансовий аналіз та інше.

У соціології пакет дозволяє автоматизувати процес створення баз даних соціологічної інформації, їх зберігання та обробку.

Користувачами IBM SPSS Modeler є організації фінансової, страхової, телекомунікаційної галузей, підприємства роздрібного бізнесу та інших сфер діяльності [26].

Іншим безперечним лідером на ринку статистичних програм є система SAP Predictive Analytics, яка зображена на рисунку 1.7 що представляє собою рішення, що дозволяє здійснювати збір даних і будувати на основі цих даних прогнозні моделі, що дозволяють прогнозувати перебіг подальших подій. Побудова прогнозних моделей здійснюється шляхом пошуку закономірностей і взаємозв'язків в історичних і поточних даних.

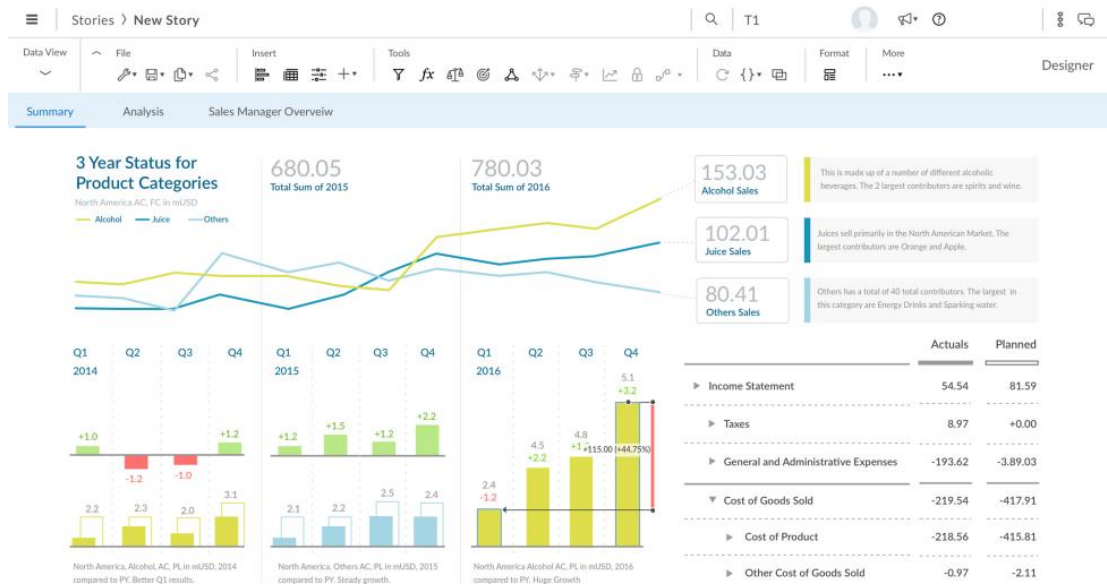


Рисунок 1.7 – Preview інтерактивного проекту SAP Predictive Analytics

Прогнозні моделі можуть бути побудовані за допомогою таких інструментів як Automated Analytics (дозволяє створювати прості моделі, наприклад моделі класифікації, регресії, кластеризації, часових рядів) і Expert Analytics (дозволяє використовувати складні алгоритми аналізу із застосуванням мови статистичного аналізу і відкритим вихідним кодом на мові R).

Особлива цінність цього рішення для бізнесу полягає в можливості зв'язати SAP Predictive Analytics з продуктивними БД, де зберігаються оперативні дані. А це дозволяє легко створювати актуальні прогнози даних. Зручний інтерфейс забезпечує можливість роботи з простими Інтелектуальне моделями, не вдаючись до допомоги фахівців. Тому процес прийняття рішень значно прискорюється. Як приклад можемо розглянути конкретний кейс, опублікований компанією SAP:

За допомогою SAP Predictive Analytics європейської спеціалізованої комісією з контролю перевезень тварин і продуктів тваринного походження була вирішена проблема виявлення шахрайства: раніше контроль здійснювався за індивідуальним підозрою інспектора, що дозволяло виявляти тільки невеликий відсоток неправомірних перевезень. Після впровадження SAP Predictive Analytics перевірка 30% контейнерів привела до виявлення 85% порушень.

Основні переваги програми:

- Підтримка алгоритмів на мові R;
- Зручний інтерфейс;
- Широкі можливості візуалізації даних;
- Інтеграція з продуктивними базами даних;
- Відстеження моделей протягом усього їх життєвого циклу;
- Можливість імпорту існуючих моделей пов'язаних з бізнес-контентом;
- Можливість працювати з різними обсягами даних.

SAP Predictive Analytics є універсальним інструментом для обробки, аналізу, прогнозування та візуалізації даних. Цей продукт дозволяє створювати і підтримувати прогнозні моделі будь-якої складності на основі даних будь-якого обсягу з різних джерел. Зрозумілий інтерфейс і широкі можливості візуалізації результатів прогнозування в самих різних формах сприяють швидкому і точному прийняттю оптимальних рішень на всіх рівнях

управління в компанії, що є запорукою успішного функціонування і розвитку компанії.

Коли потрібна SAP Predictive Analytics:

- Користувач має справу з багатовимірною проблемою: є безліч факторів, що впливають на об'єкт аналізу;
- В даних є пропуски або невірно заповнені поля;
- Не зовсім зрозуміло, чи підходять наявні дані для аналізу (первинна оцінка даних);
- Потрібно швидкий наочний результат, оскільки користувач не володіє навичками настройки моделі і її інтерпретації;
- Рішення потрібно «день-в-день»;
- Потрібно проаналізувати всі наявні дані (без ліміту на число змінних) [27].

Пакети прикладних програм SPSS та SAP є повністю русифікованими. У них реалізований основний набір популярних статистичних методів аналізу і випущено досить літератури для зручності роботи.

Програма статистичної обробки даних SPSS популярна серед фахівців з соціально-економічним ухилом. Дозволяє паралельно обробляти кілька підвбірок. Добре реалізовано прогнозування часових рядів. Простий в освоєнні. Є специфічні методи, націлені виключно на маркетингові та соціологічні дослідження. Якщо потрібен потужний, загально визнаний пакет з простим і зрозумілим навіть початківцям користувачам інтерфейсом, то краще скористатися SPSS [28].

Програма статистичної обробки даних SAP більш популярний серед бізнес-користувачів та спеціалістів з теорії та методам аналізу даних. SAP Predictive Analytics дозволяє працювати з даними, завантаженими з продуктивних баз даних (підтримуються таблиці і різні view SAP HANA, SQL view на SAP HANA, віртуальні таблиці Smart Data Access). Також існує можливість роботи з даними, завантаженими з плоских файлів (Microsoft

Excel і текстових файлів). Все це спрощує роботу з джерелами даних і робить роботу SAP Predictive Analytics більш універсальною.

Для побудови Інтелектуальне моделей даних використовуються інструменти Automated Analytics і Expert Analytics.

Automated Analytics є потужним інструментом аналізу інформації з соціальних мереж. У модулі Social присутні можливості вилучення і використання неявної структурної реляційної інформації, яка зберігається в наборах даних різних видів, в тому числі і з соціальних мереж.

Інструмент Expert Analytics є більш складним інструментом для поглибленого аналізу і більш точного прогнозування даних. Доступні можливості прогнозування часового ряду, виявлення сторонніх значень, аналізу трендів, класифікаційного аналізу, сегментного аналізу [29].

Так як для нашої задачі необхідно програма орієнтований, перш за все, на соціологію, тому було обрано для побудови моделі пакет SPSS Modeler. Дане ПО для вилучення цінної інформації з даних (data mining) і аналізу текстових даних (text analytics) дозволяє клієнтам здійснювати моніторинг змін, пов'язаних із клієнтами, персоналом та державними структурами, отримувати на основі цієї інформації глибокі і достовірні знання і потім прогнозувати ключові фактори, які будуть управляти майбутніми ініціативами з придбання і утримання клієнтів, говорить в повідомленні ІВМ. Зокрема, користувачі можуть отримувати інформацію про споживчі думки, настрої і сподівання з смайликами і сленговою термінологією, які люди часто використовують в описах свого ставлення до продукту або послуги.

2. АНАЛІЗ МЕТОДІВ ПРОГНОЗУВАННЯ ВІДГУКУ КЛІЄНТІВ НА МАРКЕТИНГОВУ КОМПАНІЮ

До того як роздивитися конкретні методи прогнозування необхідно дізнатися, що ж таке прогностичне моделювання. І так прогностичне моделювання відноситься до набору методів, які використовуються для обчислення ймовірностей різних результатів, як правило, з комбінацією статистики і логіки. Кожна прогностична модель призначена для використання з декількома припущеннями, які засновані на різних змінних. Ці змінні можуть бути зважені на основі того, наскільки ймовірний їх вплив на майбутні вихідні дані набору даних [30].

Оскільки ця область повністю заснована на прогнозах, у неї є свої плюси і мінуси. Незважаючи на те, що ви можете передбачити майбутні результати певної змінної, ви також можете зіткнутися з організаційними бар'єрами в доступі до необхідних даних.

Переваги прогностичного моделювання

Прогнозуюче моделювання знижує витрати, необхідні підприємствам для прогнозування результатів своїх рішень і використання конкурентної розвідки. Це також допомагає в:

1. Аналіз відтоку. Дослідження рівня відтоку клієнтів в компанії називається аналізом відтоку клієнтів. За допомогою інтелектуального аналізу ви можете реалізувати ефективні стратегії утримання і зберегти клієнтів, пов'язаних з вашим брендом;

2. Прогнозування зовнішніх факторів: компанії можуть прогнозувати майбутні зовнішні чинники і їх вплив на поточні послуги фірми, допомагаючи їм розробляти відповідні стратегії;

3. Кредити моделювання: прогнозне моделювання допомагає організаціям оцінювати ризик і вартість їх портфельних кредитів.

Мінуси прогностичного моделювання

Прогнозуюче моделювання відмінно справляється зі своїми завданнями, але воно пов'язане з рядом проблем:

1. Величезною проблемою є отримання правильних даних для використання при розробці алгоритму. На цьому етапі вчені витрачають приблизно 80% свого часу.

2. Це не просто математична проблема. Фірми повинні також підготуватися до організаційних і технічних бар'єрів, які можуть перешкодити їм реалізувати прогностичні моделі.

3. Для досягнення незмінно успішних результатів компаніям необхідно отримати достатній обсяг вибірки даних. Часто професіонали не мають достатньої кількості даних для побудови своїх моделей, що, очевидно, впливає на результати.

4. Найчастіше проблема полягає в тому, щоб проекти прогнозного моделювання вирішували реальні завдання бізнесу. Іноді розроблені моделі можуть не відповідати потребам бізнесу або фактично впливати на стратегію.

5. Нарешті, системи, що зберігають ці корисні дані, часто не підключені до централізованих сховищ даних, що ускладнює процес отримання доступу до даних одночасно [31].

2.1. Дерево рішень

Дерева рішення - один з найпопулярніших підходів до вирішення завдань Data Mining. Вони створюють ієрархічну структуру правил класифікації типу "ЯКЩО ... ТО ..." (if - then), яка має форму дерева. Щоб вирішити, до якого класу належить об'єкт або ситуацію, вам потрібно відповісти на запитання в вузлах цього дерева, починаючи з його кореня. Питання повинні мати вигляд "значення параметра А більше, ніж х". Якщо відповідь позитивна, тобто «так», виконується перехід до правого вузла наступного рівня, якщо негативна - то до лівого вузла; потім знову виникає питання, пов'язане з відповідним вузлом.

Популярність цього підходу як би пов'язана з наочністю, ясністю і зрозумілістю. Але дерева рішень принципово не можуть знайти "кращі" (найбільш повні і точні) правила в даних. Вони реалізують наївний принцип послідовного перегляду ознак і фактично "чіпляють" частини справжніх закономірностей, створюючи лише ілюзію логічного висновку [32].

Найчастіше дерево рішень служить узагальненням досвіду експертів, засобом передачі знань майбутнім співробітникам або моделлю бізнес-процесу компанії [33]. Наприклад, до впровадження масштабованих алгоритмів машинного навчання в банківській сфері завдання кредитного скорингу вирішувалася експертами. Рішення про надання позичальникові кредиту приймалося на основі деяких інтуїтивно (або з досвіду) виведених правил, які можуть бути представлені у вигляді дерева рішень, яке показано на рисунку 2.1.

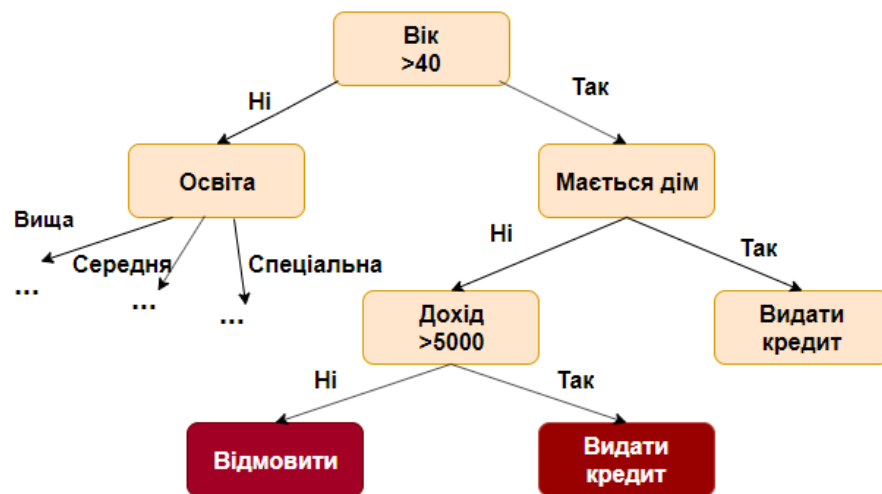


Рисунок 2.1 Дерево рішень надання кредитів

У цьому випадку можна сказати, що відповідно до ознак "вік", "місце проживання", "дохід" та "освіта" вирішується проблема бінарної класифікації (цільова категорія має два значення: "надання кредиту" та "відхилення").

Побудова дерева рішень

Припустимо, що з урахуванням деякого набору T , який містить об'єкти, кожен з яких має m атрибуту, і один з них вказує на те, що об'єкт належить до певного класу.

Р. Куінлан запропонував ідею побудувати дерево рішень із множини T , вперше представленої Хантом.

Якщо припустити, що класи представлені $\{C_1, C_2, \dots, C_k\}$, можуть виникнути такі ситуації:

Множина T містить один або більше об'єктів, що належать до одного класу C_k . Тоді, дерево рішень для T - це лист, що визначає клас C_k ;

Множина T не містить не одного прикладу, тобто порожня множина. Тоді це ще один лист, і категорія, пов'язана з листом, вибирається з іншої множини, яка відрізняється від T , наприклад, з множини, асоційованої з батьком;

Множина T містить об'єкти, що відносяться до різних категорій. У цьому випадку слід розділити безліч T на деякі підмножини. Для цього виберіть атрибут, який має два чи більше значень, які відрізняються один від одного O_1, O_2, \dots, O_n . T поділяється на підмножини T_1, T_2, \dots, T_n , де кожна підмножина T_i містить усі об'єкти, що мають значення O_i для вибраного атрибута. Цей процес триватиме рекурсивно, поки кінцева множина не складеться з об'єктів, що належать до одного і того ж класу.

Вищеописаний процес є основою багатьох сучасних алгоритмів побудови дерев рішень, цей метод також називається поділом і захопленням. При використанні цієї методики побудова дерева рішень буде проходити зверху вниз [34].

Оскільки всі об'єкти були попередньо віднесені до відомих нам класів, процес побудови дерева рішень називається навчанням з учителем. Процес навчання також називають індуктивним навчанням або індукцією дерева.

Дерева рішень мають певні особливості. Почнемо з переваг:

- простота інтерпретації;

- можливість роботи як з категоріями, так і з кількісними значеннями;
- універсальність в плані вирішення завдань і класифікації, і регресії;
- можливість роботи з пропусками в даних (порожніми значеннями атрибутів). Причому дерева рішень можна використовувати для заповнення пропусків найбільш імовірним значенням;
- хороша продуктивність в процесі класифікації по вже побудованому дереву (так як алгоритм пошуку в дереві дуже ефективний навіть для великих наборів даних).

Не обійшлося і без недоліків. Серед них:

- нестабільність процесу. Нерідко невеликі зміни в наборі даних можуть призводити до побудови абсолютно іншого дерева. Це пов'язано з ієрархічністю дерева. Зміни в вузлі на верхньому рівні ведуть до змін у всьому дереві нижче.
- складність контролю розміру дерева. Розмір дерева є критичним фактором, що визначає якість виконання завдання. При використанні простих критеріїв зупинки дерева часто ростуть або дуже короткими, або дуже великими.
- неадекватність поділу на класи в складних випадках. У найпростіших деревах рішень розбиття в вузлах відбувається за значенням одного атрибута, паралельно, так би мовити, осях координат. В даному випадку кожен атрибут - це вісь координат зі своїми значеннями. І дерево «нарізає» весь простір на «паралелепіпеди», всередині яких і групуються точки набору даних, які відповідають тому чи іншому класу. Іноді такий поділ не може точно описати складні області, утворені крапками, що належать певному класу.
- критерій приросту інформації характеризується схильністю надавати перевагу атрибутам, що мають велике число різних значень. У

граничному випадку у кожного рядка може бути своє значення атрибута. Тоді другий доданок в дорівнюватиме 0, і приріст буде максимальним.

- Більш просунуті алгоритми побудови дерев дозволяють вирішувати зазначені проблеми [35].

- Отже дерева рішень - цілком типовий приклад алгоритму так званого індуктивного навчання «по прецедентах», коли на основі значень атрибутів вихідних даних будується деяка вирішальна функція, оцінюються параметри моделі і т.д. Тобто в цьому випадку прогнозу модель конструює не користувач, а алгоритм - автоматично на основі вихідних даних.

Користувач ж, вибираючи метод (в нашому випадку дерева рішень), визначає якийсь клас моделей, які можуть бути побудовані і навчені за допомогою даного алгоритму [36].

2.2. Алгоритм списку рішень

Моделі списку рішень визначають підгрупи або сегменти, які показують більш високу або більш низьку правдоподібність даних бінарних (так чи ні) результатів в порівнянні з усією вибіркою. Наприклад, можна було б шукати клієнтів з низькою ймовірністю відпливу чи з високою ймовірністю відгуку на конкретну пропозицію або кампанію. Засіб перегляду списку рішень дає вам повний контроль над моделлю, дозволяючи редагувати сегменти, додавати свої власні бізнес-правила, задавати, як оцінюється кожен сегмент, а також налаштовувати модель багатьма іншими способами, щоб оптимізувати частку влучень по всім сегментам. В результаті він особливо добре підходить для створення списків розсилки чи іншого способу визначення цільових записів для конкретної кампанії [37]. Ви також можете вибрати різні завдання дослідження даних, щоб об'єднати підходи до моделювання - наприклад, шляхом ідентифікації добре і погано працюють сегменти в рамках однієї і тієї ж моделі, і включення або виключення кожного на етапі скорингу в міру необхідності.

Мета списків рішень - знайти групу осіб з чіткою моделлю поведінки, наприклад, з високою ймовірністю покупки продукту. Типовий перелік рішень складається з набору правил прийняття рішень. Правило прийняття рішення - це правило if-then, яке складається з двох частин: попередньої і наступної. Попереднім є булівський вираз предикторів, а наслідком є прогнозоване значення цільового поля, коли предиктор є істинним [38]. Найпростіша конструкція правила прийняття рішення - сегмент, заснований на одному ПРЕДИКТОРІ; наприклад, Пол = "Чоловічий" або $10 < \text{Вік} \leq$. Запис включається в правило, якщо попереднє правило є вірним. Якщо будь-яка справа підпадає під дію одного з правил в переліку рішень, то вважається, що воно підпадає під дію цього переліку. У переліку рішень порядок проходження правил має важливе значення; якщо справа підпадає під дію правила, то воно буде ігноруватися наступними правилами.

Алгоритм списку рішень можна розглянути на прикладі класифікації квітів ірису відповідно до довжини і ширини їх пелюсток (рисунок 2.2).

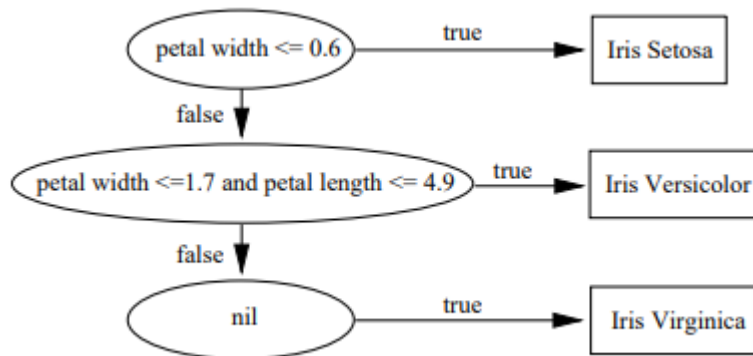


Рисунок 2.2 Список рішень для класифікації квітів ірису

Алгоритм складання списку рішень можна узагальнити наступним чином:

- Потенційні правила можна знайти в вихідному наборі даних;
- Кращі правила додаються до списку рішень;
- Записи, включені в список рішень, видаляються з бази даних;
- Нові правила можуть бути знайдені на основі скороченого набору даних;

- Процес повторюється до тих пір, поки не буде виконано один або кілька критеріїв зупинки.

Засіб перегляду списку рішень дозволяє переглядати, модифікувати і реорганізувати сегменти, а також вибирати, який сегмент слід включити або виключити для поставлених цілей. Наприклад, ви можете вибрати виключення однієї групи замовників з майбутніх пропозицій і включення інших груп і відразу побачити, як це впливає на загальний коефіцієнт попадання. Моделі списку рішень видають оцінку Yes для включених сегментів і оцінку \$null\$ для всіх інших, включаючи залишковий сегмент. Таке пряме управління робить моделі списку рішень ідеальними для генерування списків розсилки, і вони широко використовуються в організації відносин з клієнтами, в тому числі для call-центрів і програм маркетингу [39].

Первинний алгоритм

Для того, щоб розглянути основний алгоритм спочатку необхідно зрозуміти, які позначення будуть використовуватися при описі алгоритму списку рішень (таблиця 2.1).

Таблиця 2.1 – Позначення для формул алгоритму списку рішень

Позначення	Опис позначення
X	Матриця даних. Стовпці - це поля (атрибути), а рядки - записи (випадки)
L	Набір моделей списків
L_i	Модель і-го списку L
L_{null}	Модель списку, яка не містить правил
P_{L_i}	Оціночна ймовірність отримання відповіді від списку L_i
N	Загальна чисельність населення
$X_{m,n}$	Значення m-го поля (стовпця) для n-ого запису (рядки) X
X_{L_i}	Підмножина записів у X, які охоплені списком моделі L_i
Y	Цільове поле в X
Y_n	Значення цільового поля для n-ого запису

Закінчення таблиці 2.1

A	Колекція всіх атрибутів (полів) X
A _j	J-й атрибут X
R	Колекція правил для розширення попереднього списку правил
R _k	K-е правило в правилі колекції R
T	Набір моделей списку кандидатів
ResultSet	Набір моделей списку рішень

Основний алгоритм створення моделі списку рішень такий:

1. Ініціалізація моделі

- Нехай d – глибина пошуку, а w – ширина пошуку.
- Якщо $L = \emptyset$, додайте L_{null} до L .
- $T = \emptyset$.

2. Продивіться всі елементи L

- Оберіть записи \bar{X}_{L_i} , на які не поширюються правила L (формула

2.1):

$$X_{\bar{L}_i} = X - X_{L_i} \quad (2.1)$$

- Викличте алгоритм правила прийняття рішення для створення альтернативного набору правил R в $X_{\bar{L}_i}$.

- Побудуйте набір нових потенційних моделей, додавши кожне правило R в L_i .

- Збережіть розширений список в T .

3. Оберіть моделі із списку T .

- Розрахуйте оціночну ймовірність відгуку кожної з моделей списку в T у вигляді, як показано у формулі 2.2:

$$\hat{P}_{L_i} = \frac{N(Y_n=1, X_n \in X_{L_i})}{N(X_n \in X_{L_i})} \quad (2.2)$$

- Виберіть w -списки у форматі T з максимальним значенням \hat{P}_{L_i} . як L^* .

4. Додайте L^* в `ResultSet`

5. Якщо $d = 1$ або $L^* = \emptyset$ об'єднаний, поверніть `ResultSet` і

завершіть; інакше зменшіть значення d на одиницю та повторіть всі дії з 2-го кроку [37].

Завдання дослідження даних, вимірювання та відбору

Процес моделювання управляється завданнями дослідження даних.

Кожне завдання дослідження даних ефективно ініціює новий прохід моделювання і повертає новий набір альтернативних моделей, з яких можна робити вибір [40]. Завдання за замовчуванням засновано на ваших початкових специфікаціях в вузлі списку рішень, але ви можете визначити будь-яку кількість призначених для користувача завдань. Ви можете також застосовувати завдання інтерактивно - наприклад, запустити пошук записів з високою ймовірністю по всьому набору навчання, а потім пошук записів з низькою ймовірністю по залишку, щоб відсіяти погано працюючі сегменти.

Ви можете визначити відбір даних і призначені для користувача показники моделі для побудови моделі і її оцінки. Наприклад, можна задати відбір даних в задачі дослідження даних, щоб підігнати модель до конкретного регіону, і створити для користувача показник, щоб оцінити, наскільки добре ця модель працює по всій країні. На відміну від завдань дослідження даних, показники не змінюють базову модель, але вони дають іншу точку зору для оцінки, наскільки добре базова модель працює.

Додавання ваших знань бізнесу

За допомогою використання тонкого налаштування або розширення сегментів, ідентифікованих алгоритмом, засіб перегляду списку рішень дає можливість включити ваше знання бізнесу безпосередньо в модель. Ви можете редагувати сегменти, які було згенеровано цією моделлю, або додавати додаткові сегменти на основі заданих вами правил. Потім можна застосовувати ці зміни і попередньо переглядати результати [41].

Підсумовуючи, можна сказати, що алгоритм список рішень (також званий впорядкованим набором правил) являє собою набір окремих правил класифікації, які в сукупності утворюють класифікатор і має ряд своїх переваг:

- на відміну від неупорядкованого набору правил, списки рішень мають вбудований порядок, що робить класифікацію досить простою;
- простір введення повністю відкрито (тобто немає обмежень за кількістю і змістом введених правил);
- простір введення розділяється на основі необмеженої кількості вимірів, тому всі можливі логічні пропозиції можливі;
- простір введення розділяється довільна, а не автоматично виявленими умовами і не оптимізує функцію втрати деяких критеріїв [42].

2.3. Висновки

Підсумовуючи характеристики, переваги та недоліки всіх розглянутих методів прогнозування було обрано метод оснований на алгоритмі списку рішень.

Перш за все списки рішень використовуються в маркетингу для сегментації клієнтської бази. Наприклад, списки рішень дозволяють визначити, які демографічні групи мають максимальний показник відгуку. Цю інформацію можна використовувати, щоб максимізувати відгук при майбутньої прямий розсилці.

Дерева рішень і списки є двома тісно пов'язаними типами класифікатора. На відміну від більшості інших класифікаційних парадигм, наприклад, навчання на основі примірників, нейронні мережі, байєсовської мережі і логістична регресія, вони втілюють явне представлення всіх знань, які були отримані з даних навчання. З огляду на стандартне дерево рішень або список, користувач може вручну визначити, як виводиться конкретний прогноз і які атрибути мають відношення до висновку, без виконання будь-яких числових операцій (крім порівняння). Це дозволяє дуже легко

пояснити, як ці класифікатори повинні інтерпретуватися, і як вони генерують прогноз.

Списки рішень аналогічні деревам рішень в тому сенсі, що для отримання прогнозу потрібно послідовність рішень. Різниця в тому, що всі рішення мають двійковий результат, істинний або помилковий, і подальші тести на екземплярі потрібні тільки в тому випадку, якщо результат всіх попередніх рішень був негативним. Індивідуальні рішення приймаються відповідно до «правил», які складається з комбінацій тестів атрибут-значення і мітки класу.

Список рішень - це список правил, які оцінюються послідовно. Правило «спрацьовує», якщо екземпляр тесту проходить кожен тест значення атрибута, що міститься в правилі. В цьому випадку процес класифікації зупиняється, і мітка класу правила присвоюється тестовому екземплярі.

Також наявна, ще одна ключова відмінність між деревами рішень і списками рішень, яка полягає в тому, що перші можуть розглядатися як неупорядковані набори правил, де кожному листу дерева відповідає одне правило з умовною частиною, що складається із з'єднання всіх крайових міток на шляху від кореня до цього листа. Ієрархічна структура дерева гарантує, що правила в наборі не перекриваються, тобто кожен приклад покритий рівно одним правилом. Це додаткове обмеження спрощує класифікацію (відсутність конфліктів з декількох правил), але може призвести до більш складних правил. Наприклад, було показано, що списки рішень (впорядковані набори правил) з не більш ніж k умовами на правило строго більш виразні ніж дерева рішень глибини K .

І звичайно ж метод список рішень має низку переваг, які і вплинули на остаточній вибір методу прогнозування. Головне з них - це наочність представлення результатів (у вигляді списку сегментів). Дерева рішень дозволяють працювати з великим числом незалежних змінних. на вхід можна подавати всі існуючі змінні, алгоритм сам вибере найбільш значущі серед

них, і тільки вони будуть використані для побудови дерева (автоматичний відбір предикторів).

3. ПОБУДОВА МОДЕЛІ ДЛЯ УПРАВЛІННЯ ВІДГУКАМИ КЛІЄНТІВ НА МАРКЕТИНГОВУ КОМПАНІЮ

3.1. Постановка завдання

Поставлена задача розробки моделі для прогнозування відгуку клієнтів з ціллю підвищення рівня відгуку на нові товари буде націлена на рішення проблем підприємства і вирішена на прикладі ТОВ «Тубний Завод».

Для розуміння поставленої задачі та інформації, на якій буде будуватися майбутнє моделювання, далі буде приведено короткий опис підприємства. Тубний завод - одна з найбільших компаній в ніші виробництва Туб в Україні, яка займається виробництвом своєї продукції більш 18 років.

За час плідної роботи ними було освоєно виробництво всіх видів туб: ламінатних, поліетиленових і алюмінієвих. Гнучке виробництво в замкнутому циклі дозволило скоротити терміни виконання замовлень, збільшити асортимент, зменшити партію поставки і зробити продукцію заводу пріоритетною на ринку упаковки в цілому [43].

Маркетингова система є невід'ємною частиною ефективного управління виробничою і збутовою діяльністю будь-якої виробничої компанії, тому вона потребує налагодженої системи обміну інформацією між цими процесами [44]. Використовуючи сучасні інформаційні інструменти, ви можете застосовувати дані, накопичені в ході процесів продаж, виробництва, і взаємодії з клієнтами для формування різних аналітик, необхідних для розробки та коригування маркетингової стратегії.

Дана робота пов'язана з внесенням змін до існуючих систем. Для того, щоб внести зміни і отримати бажані результати, вам необхідно вивчити те, що вже існує. Незалежно від того, що ми робимо - налаштуємо або встановлюємо систему CRM з нуля або створюємо ефективну ERP-систему, інтегруємо різні систем для підвищення автоматизації роботи в цілому [45]. Але у будь-якому випадку, для початку, вам потрібно зрозуміти існуючий

план роботи, і тільки після цього ви можете пропонувати якісь зміни і продумувати варіанти вирішення поставленої задачі.

Моделювання реакції клієнтів – це складний процес, який включає в себе певну кількість етапів:

- Розуміння бізнес-процесів;
- Збір даних;
- Початкове дослідження даних;
- Підготовка даних;
- Моделювання процесу;
- Аналіз та оцінка результатів.

Почнемо з першого етапу – розуміння бізнес-процесу.

Для того щоб зрозуміти, як працювати з функціональним моделюванням, буде побудовано модель бізнес-процесу «Моделювання реакції клієнтів на маркетингову компанію підприємства».

Контекстна діаграма даного бізнес-процесу має два вихідних документа – звіт по результатам аналізу прогнозування та рекомендацій по змінам в маркетинговому плані (рисунок 3.1).

Також наявна вхідна інформація: інформація про клієнтів з БД, інформація про здійснені угоди з БД та результати опитування споживачів, яке було проведено перед початком роботи.

Учасники бізнес-процесу (механізми): спеціаліст по маркетинговим дослідженням, менеджер-маркетолог та аналітик.

Управління: правила збору та підготовки інформації, методологія побудови прогнозних моделей, методика аналізу.

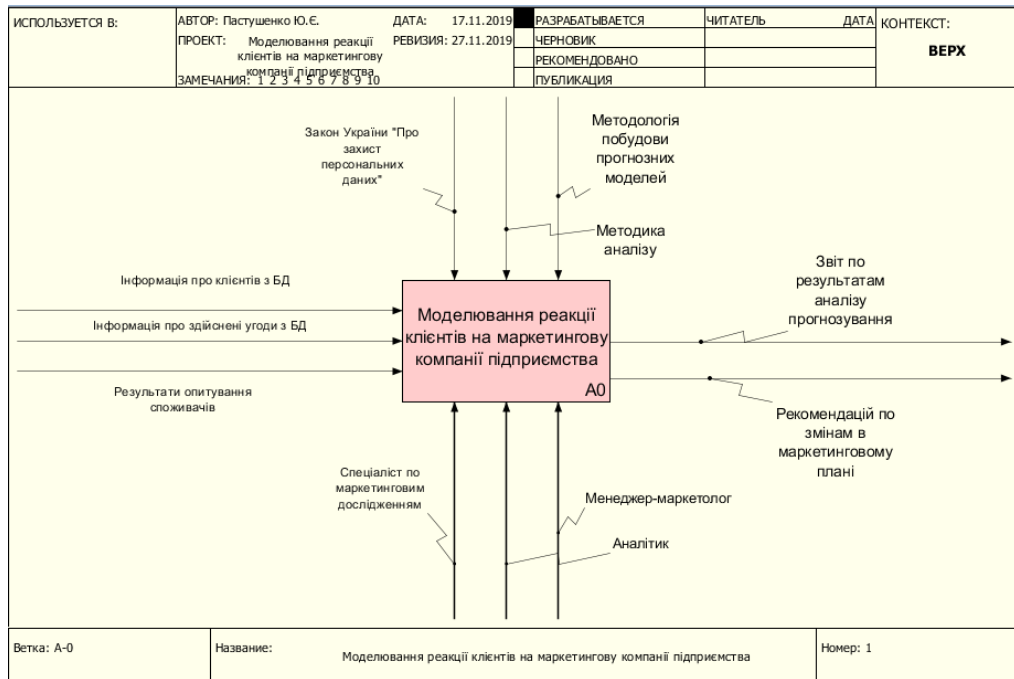


Рисунок 3.1 – Контекстна діаграма «Моделювання відгуку клієнтів на маркетингову компанію підприємства» в стандарті IDEF0

Декомпозиція контекстної діаграми «Моделювання реакції клієнтів на маркетингову компанію підприємства», що зображена на рисунку 3.2 має наступні елементи:

A1. «Збір інформації»

Інтерфейсні дуги:

Вхід: інформація про клієнтів з БД, інформація про здійснені угоди з БД та результати опитування споживачів, яке було проведено перед початком роботи.

Вихід: Відібрана інформація.

Управління: Закон України "Про захист персональних даних".

Механізми: Спеціаліст по маркетинговим дослідженням.

A2. «Систематизація та підготовка зібраної інформації»

Інтерфейсні дуги:

Вхід: Відібрана інформація.

Вихід: Оброблені (готові до аналізу) дані.

Управління: Правила збору та підготовки інформації.

Механізми: Спеціаліст по маркетинговим дослідженням.

А3. «Аналіз даних»

Інтерфейсні дуги:

Вхід: Оброблені (готові до аналізу) дані.

Вихід: Результат аналізу даних.

Управління: Методика аналізу даних.

Механізми: Спеціаліст по маркетинговим дослідженням.

А4. «Побудова прогнозної моделі»

Інтерфейсні дуги:

Вхід: Оброблені (готові до аналізу) дані.

Вихід: Результат аналізу даних.

Управління: Методика аналізу даних.

Механізми: Спеціаліст по маркетинговим дослідженням та менеджер-маркетолог.

А5. «Аналіз результатів прогнозування»

Інтерфейсні дуги:

Вхід: Модель прогнозу.

Вихід: Звіт по результатам аналізу прогнозування та рекомендацій щодо змін в маркетинговому плані.

Механізми: Менеджер-маркетолог.

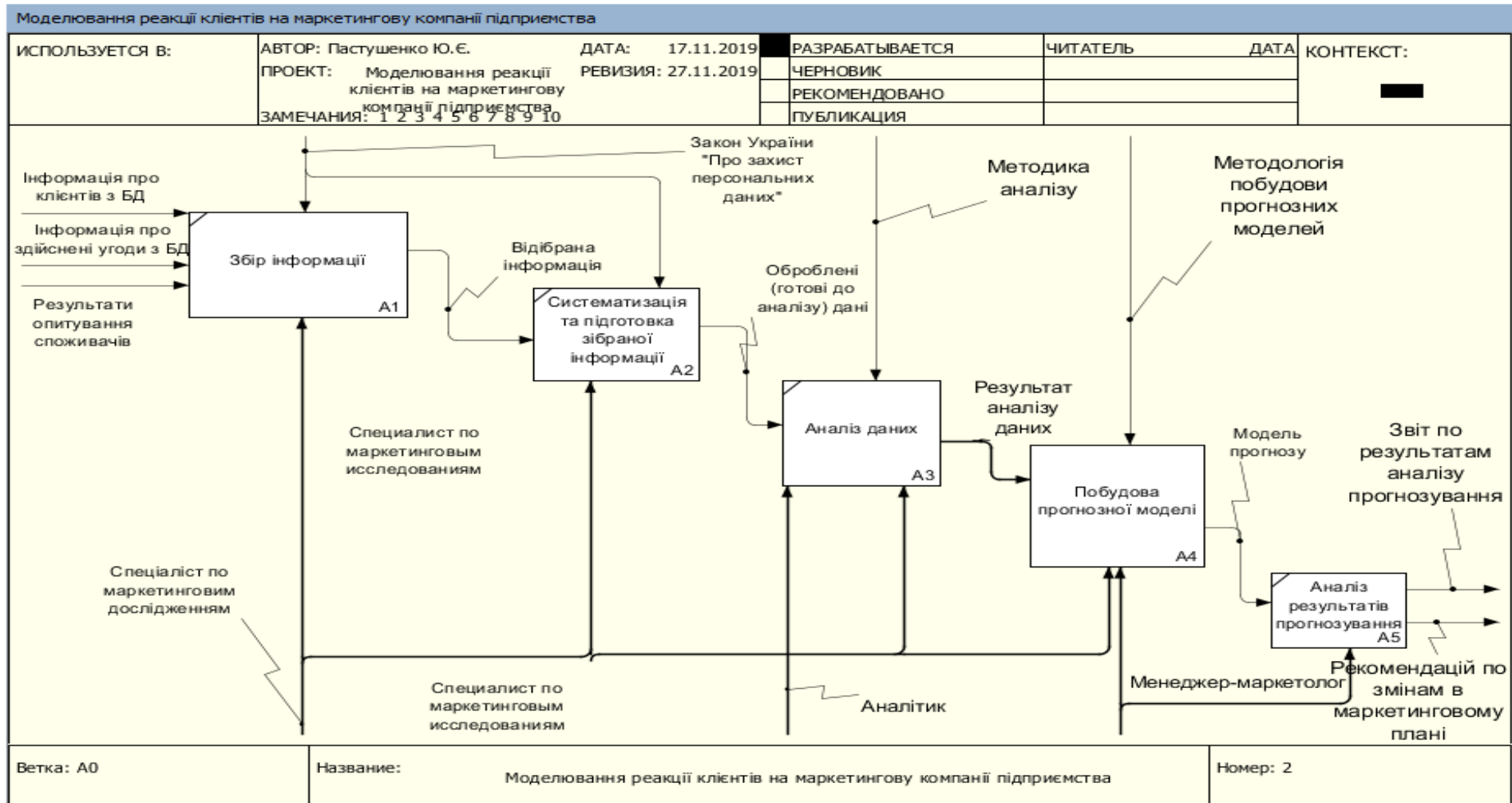


Рисунок 3.2 – Декомпозиція діаграми «Моделювання відгуку клієнтів на маркетингову компанію підприємства» в стандарті IDEF0

Декомпозиція контекстної діаграми «Побудова прогнозної моделі», що зображена на рисунку 3.3 має наступні елементи:

A1. «Виявлення та формування цілей прогнозування»

Інтерфейсні дуги:

Вхід: Результат аналізу даних.

Вихід: Цілі, яких потрібно досягти.

Управління: Методологія побудови прогнозних моделей.

Механізми: Спеціаліст по маркетинговим дослідженням.

A2. «Систематизація та підготовка зібраної інформації»

Інтерфейсні дуги:

Вхід: Цілі, яких потрібно досягти.

Вихід: Набор кінцевих показників.

Управління: Методологія побудови прогнозних моделей.

Механізми: Спеціаліст по маркетинговим дослідженням.

A3. «Аналіз даних»

Інтерфейсні дуги:

Вхід: Набор кінцевих показників.

Вихід: Модель прогнозу.

Управління: Методологія побудови прогнозних моделей.

Механізми: Спеціаліст по маркетинговим дослідженням та менеджер-маркетолог.

A4. «Побудова прогнозної моделі»

Інтерфейсні дуги:

Вхід: Модель прогнозу.

Вихід: Результати перевірки точності моделі.

Управління: Методологія побудови прогнозних моделей.

Механізми: Спеціаліст по маркетинговим дослідженням та менеджер-маркетолог.

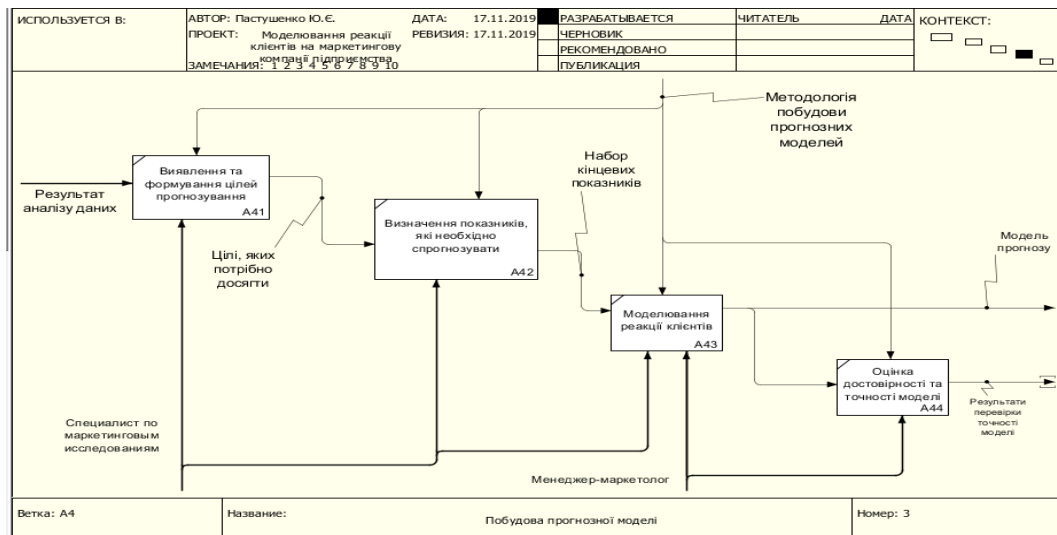


Рисунок 3.3 – Декомпозиція діаграми «Побудова прогнозної моделі»

Підсумовуючі результати моделювання бізнес-процесу можна виділити в ньому основні три етапи:

- підготовка даних про клієнтів;
- побудова моделі;
- аналіз результатів та підготовка рекомендацій щодо проведення маркетингових компаній.

Розглянемо далі більш детально основні етапи, виявлені в процесі декомпозиції процесу "Моделювання відгуку клієнтів на маркетингову компанію підприємства".

3.2. Збір , попередній аналіз та підготовка даних для моделювання

В даний час найбільш популярним методом збору необроблених (первинних) даних є метод опитування. Це опитування є досить ефективним способом отримання об'єктивної інформації (факти життєдіяльності людей), так і суб'єктивної інформації (про мотивацію діяльності, думки, оцінки і ціннісну орієнтацію). Опитування - це метод безпосереднього (інтерв'ю) чи опосередкованого (анкета) збору необробленої вербальної інформації шляхом запису відповідей респондентів на питання і угруповання їх у формі анкети відповідно до мети і завдань дослідження [46].

В ході дослідження використовувалися наступні загальні і приватні методи для збору інформації:

- анкетування, опитування [47];
- спостереження;
- узагальнення та аналіз результатів.

Від того, чи правильно вибрані методи дослідження, залежить повнота і об'єктивність результатів.

Підприємство ТОВ "Тубний завод" є дуже відомими заводом в Україні, країнах СНД і т.д.. Він як і будь-яка інша організація потребує маркетингових дослідженнях, спрямованих на виявлення загальної задоволеності клієнта товарами, на виявлення вражень споживача про покупку. Загальний вигляд створеної анкети і візуалізовані результати проведеного дослідження будуть представлені нижче на малюнках 3.4, 3.5, 3.6, 3.7, 3.8 та 3.9. Також всі питання наведені в опитуванні будуть в таблиці 3.1 та приклад файлу Excel буде приведено в додатку А. Результати дослідження можна використовувати для підвищення ефективності діяльності підприємства шляхом усунення тих недоліків, які виявляються після проведення дослідження.

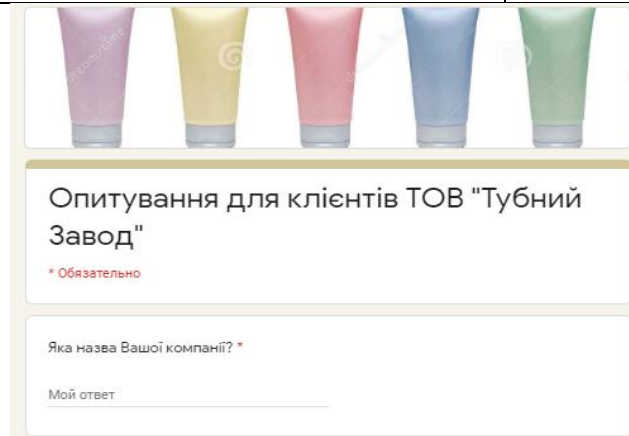
Таблиця 3.1 – Питання опитування для клієнтів ТОВ "Тубний Завод"

№	Питання	Можливі варіанти відповідей
1	Яка назва Вашої компанії?	Ручний введення
2	Розташування Вашого підприємства	- Області України; - Росія; - Інше.
3	Ваш вид діяльності?	- Виробництво косметики; - Фармацевтична промисловість; - Виробн. продукті харчування; - Виробництво побутової хімії; - Інше.

Закінчення таблиці 3.1

4	Протягом якого періоду часу ваш бізнес здійснює свою діяльність?	- Менше 1 року; - Від 1 до 3 років; - Від 3 до 5 років; - Більше 5 років.
5	До якого типу підприємства належить ваш бізнес?	- Мале підприємство; - Середнє підприємство; - Велике підприємство.
6	Вкажіть чисельність ваших найманих працівників	- від 1 до 10 осіб; - від 10 до 50 осіб; - від 50 до 200 осіб; - від 200 до 1000; - понад 1000.
7	Чи маєте власне виробництво?	Так або ні
8	З якого джерела дізналися про нашу компанію?	- Реклама в засобах масової інформації; - Реклама в інтернеті; - Зовнішня реклама; - Від знайомих.
9	Як довго ви з нами співпрацюєте?	- Менше 1 року; - Від 1 до 3 років; - Від 3 до 5 років; - Більше 5 років.
10	Яка продукція нашого виробництва цікавить Вас?	- Ламінатні туби; - Поліетиленові туби; - Алюмінієві туби; - Ковпачки для туб.
11	Ви вже купували наші товари?	Так або ні
12	Чи задоволені ви якістю нашої продукції?	Так або ні
13	Чи влаштовує Вас цінова політика нашого підприємства?	Так або ні
14	Чи будете Ви співпрацювати з нами надалі	Так або ні
15	Чи берете Ви участь в програмі лояльності?	Так або ні

16	Ви б рекомендували нашу компанію іншим людям?	Так, можливо або ні
----	---	---------------------



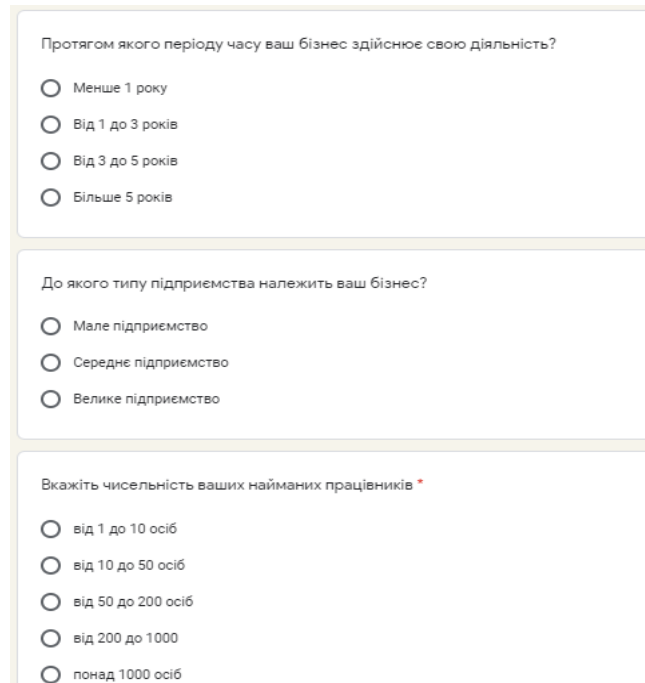
Опитування для клієнтів ТОВ "Тубний Завод"

* Обязательно

Яка назва Вашої компанії? *

Мой ответ

Рисунок 3.4 – Анкета для клієнтів ТОВ «Тубний Завод»



Протягом якого періоду часу ваш бізнес здійснює свою діяльність?

Менше 1 року

Від 1 до 3 років

Від 3 до 5 років

Більше 5 років

До якого типу підприємства належить ваш бізнес?

Мале підприємство

Середнє підприємство

Велике підприємство

Вкажіть чисельність ваших найманих працівників *

від 1 до 10 осіб

від 10 до 50 осіб

від 50 до 200 осіб

від 200 до 1000

понад 1000 осіб

Рисунок 3.5 – Анкета для споживачів та потенціальних клієнтів ТОВ «Тубний Завод»

Після збору необхідної кількості інформації було проведено початкове дослідження даних. Спочатку було розглянено результати опитування у формі діаграм у Google Forms.



Рисунок 3.6 – Результати опитування

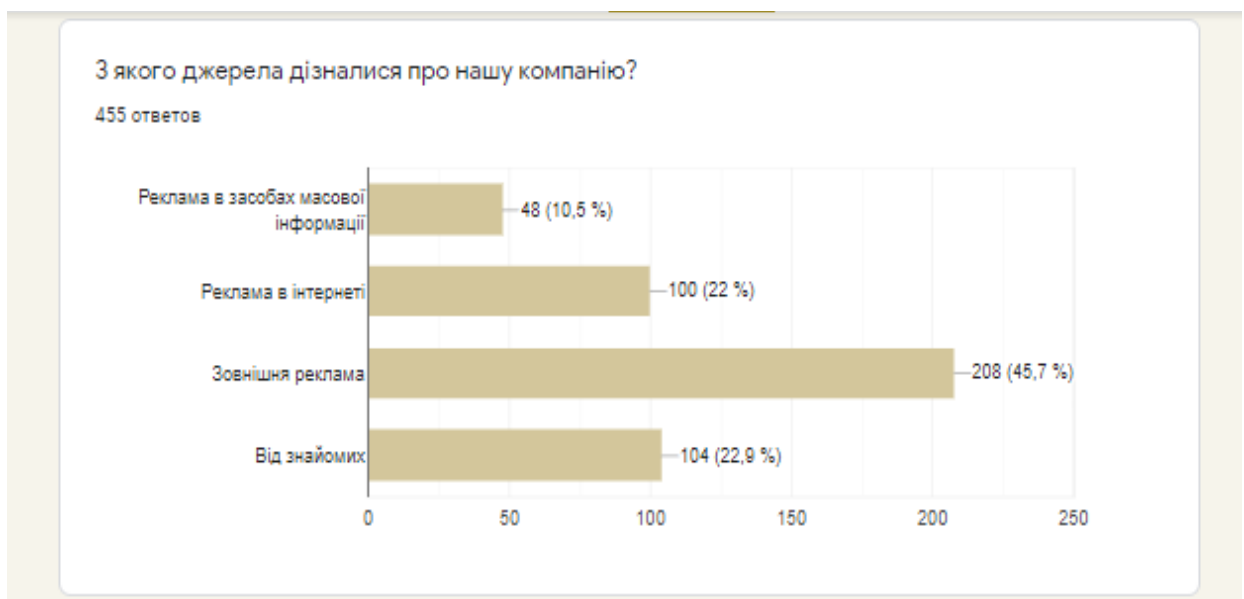


Рисунок 3.7 – Результати опитування

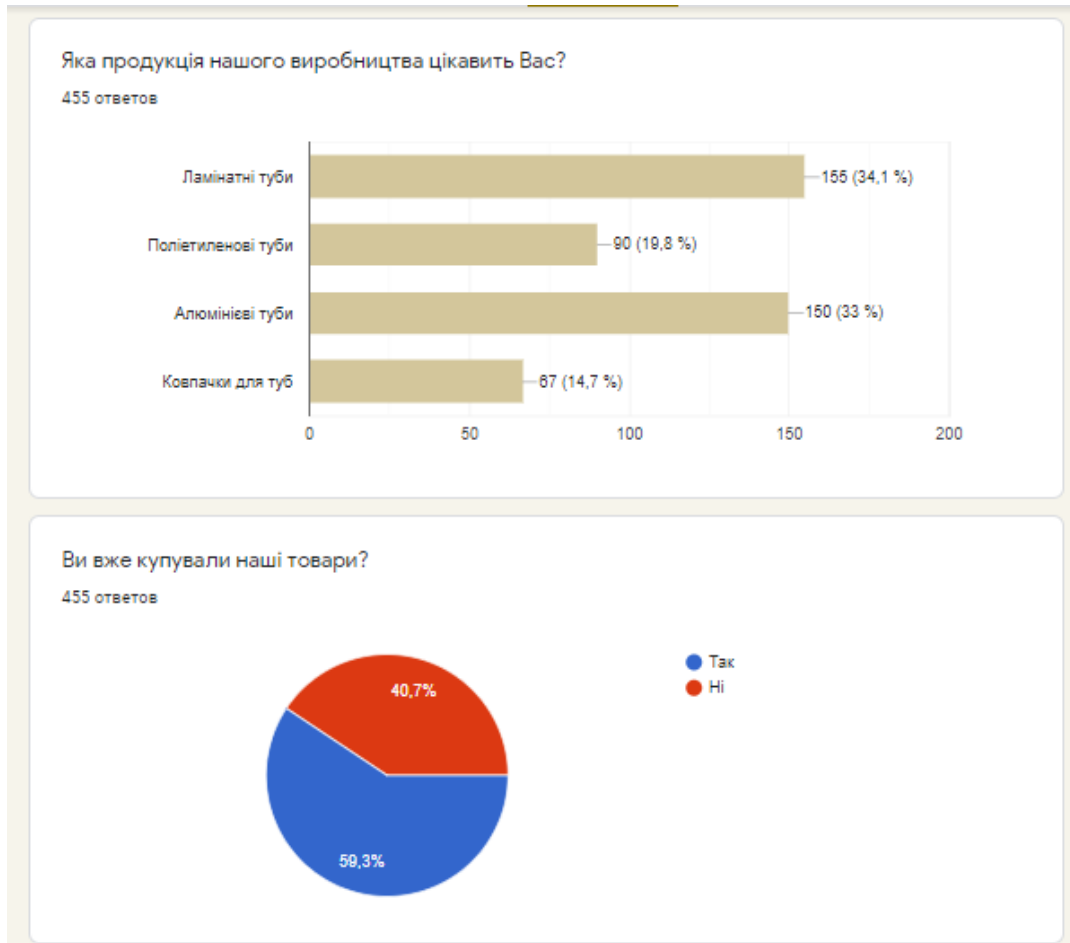


Рисунок 3.8 – Результати опитування

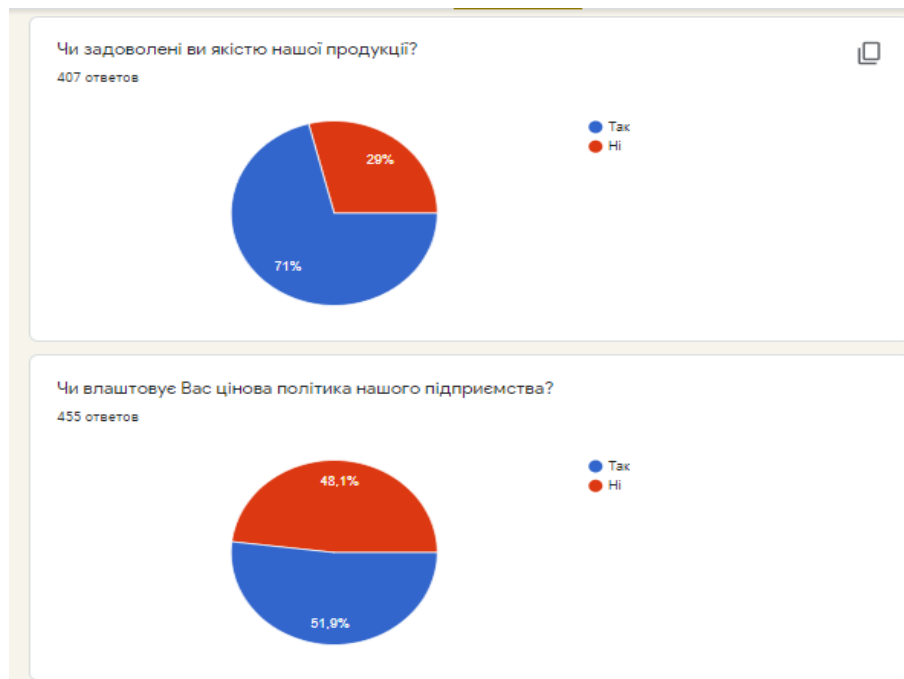


Рисунок 3.9 – Результати опитування

В ході проведеного дослідження було опитано 502 респондента-підприємства (39,9% - малі підприємства, 43,2% - середні підприємства, 17% - великі підприємства). З них 55,6% - розташовані на території України, 44,4% - на території Росії. Також підприємства поділяються по типу виробництва, 29,3% - виробництво продуктів харчування, 24,5% - фармацевтична промисловість, 24,1 – виробництво побутової хімії та 0,8% - інше.

На питання «Яка продукція нашого виробництва цікавить Вас?» були отримані такі відповіді: 35,9% - ламінатні туби, 20,5% - поліетиленові туби, 31,7% - алюмінієві туби і 13,3% - ковпачки для туб. З усіх опитаних компаній 60,8 – вже купували товари нашого підприємства, а 39,2 – ще не купували.

Також респонденти в питанні «Чи влаштовує Вас цінова політика нашого підприємства?» поділилися майже навпіл, 51,2 % задоволені ціною, 48,8 – незадоволені. А в питанні про якість продукції, на користь підприємства, тільки близько чверті опитаних (27,6%) не задоволені.

Для того щоб побачити всі впорядковані відповіді та для подальшої обробки отриманої інформації результати опитування було експортовано з Google Forms у форматі Excel, що зображено на рисунках 3.10 та 3.11.

Яка назва Вашої компанії	Розташування Вашого підприємства	Ваш вид діяльності?	Протягом якого періоду	До якого типу підприємства	Вкажіть чисельність вашого підприємства	Чи маєте власне виробництво	З якого джерела дізналися про нас	Як довго ви з нами співпрацюєте
ОАО "Лубнифарм"	Полтавська обл.	Фармацевтична промис.	Менше 1 року	Середнє підприємство	від 50 до 200 осіб	Так	Реклама в засобах масо	Менше 1 року
ООО "Юрія-фарм"	Київська обл.	Фармацевтична промис.	Від 1 до 3 років	Мале підприємство	від 10 до 50 осіб	Так	Реклама в інтернеті	Менше 1 року
ООО «Жерминаль»	Росія	Виробництво косметики	Від 3 до 5 років	Велике підприємство	від 200 до 1000	Так	Від знайомих	Від 1 до 3 років
Компанія «ЕвроТек»	Росія	Виробництво побутової	Більше 5 років	Велике підприємство	понад 1000 осіб	Так	Реклама в інтернеті	Більше 5 років
АЛТЕЙ УКРАИНА, НПО	Харківська обл.	Виробництво косметики	Більше 5 років	Середнє підприємство	від 50 до 200 осіб	Так	Реклама в засобах масо	Від 3 до 5 років
БИОЛИК, ПАО	Вінницька обл.	Фармацевтична промис.	Від 1 до 3 років	Мале підприємство	від 10 до 50 осіб	Так	Від знайомих	Від 1 до 3 років
АЛТЕЙ УКРАИНА, НПО	Харківська обл.	Виробництво косметики	Від 3 до 5 років	Середнє підприємство	від 50 до 200 осіб	Так	Від знайомих	Від 1 до 3 років
Емі Люкс	Запорізька обл.	Виробництво побутової	Більше 5 років	Середнє підприємство	від 200 до 1000	Так	Реклама в інтернеті	Від 3 до 5 років
Евдемон	Одеська обл.	Виробництво побутової	Менше 1 року	Мале підприємство	від 10 до 50 осіб	Так	Реклама в засобах масо	Ще не співпрацювали
"Объединение" Gerard"	Донецька обл.	Виробництво побутової	Більше 5 років	Велике підприємство	понад 1000 осіб	Так	Від знайомих	Більше 5 років
АМАЛЬГАМА ЛЮКС, ОО	Миколаївська обл	Виробництво косметики	Від 3 до 5 років	Середнє підприємство	від 50 до 200 осіб	Так	Зовнішня реклама	Менше 1 року
АРОЗА, ПАРФЮМЕРНО-К	Київська обл.	Виробництво косметики	Від 3 до 5 років	Велике підприємство	понад 1000 осіб	Так	Реклама в засобах масо	Менше 1 року
Ріمال, ПрАТ (Ріمال, ЧАС	Рівненська обл.	Виробництво продуктів	Від 1 до 3 років	Мале підприємство	від 10 до 50 осіб	Так	Зовнішня реклама	Від 3 до 5 років

Рисунок 3.10 – Результати опитування в Microsoft Excel

З якого джерела дізнали	Як довго ви з нами спів	Яка продукція нашого ви	Ви вже купували наш	Чи задоволені ви якістю	Чи влаштує Вас цінов	Чи будете Ви співпрацю	Чи береже Ви участь в п	Ви б рекомендували нац
Реклама в засобах масо	Менше 1 року	Алюмінієві туби	Так	Так	Так	Так	Ні	Так
Реклама в інтернеті	Менше 1 року	Алюмінієві туби	Так	Так	Так	Так	Так	Так
Від знайомих	Від 1 до 3 років	Ламінатні туби	Так	Так	Так	Так	Ні	Так
Реклама в інтернеті	Більше 5 років	Ламінатні туби	Так	Так	Ні	Так	Так	Так
Реклама в засобах масо	Від 3 до 5 років	Поліетиленові туби	Так	Так	Ні	Ні	Ні	Ні
Від знайомих	Від 1 до 3 років	Ковпачки для туб	Так	Так	Так	Можливо	Ні	Так
Від знайомих	Від 1 до 3 років	Алюмінієві туби, Ковпач	Так	Так	Так	Так	Ні	Можливо
Реклама в інтернеті	Від 3 до 5 років	Ламінатні туби	Ні	Так	Так	Так	Ні	Так
Реклама в засобах масо	Ще не співпрацювали	Алюмінієві туби	Ні		Так	Так	Ні	Так
Від знайомих	Більше 5 років	Ковпачки для туб	Так	Так	Так	Так	Так	Можливо
Зовнішня реклама	Менше 1 року	Ковпачки для туб	Так	Ні	Так	Ні		Ні
Реклама в засобах масо	Менше 1 року	Алюмінієві туби, Ковпач	Так	Так	Ні	Ні		Можливо
Зовнішня реклама	Від 3 до 5 років	Поліетиленові туби	Так	Так	Ні	Ні	Ні	Можливо

Рисунок 3.11 – Результати опитування в Microsoft Excel

Наступний крок це підготовка даних до самого процесу моделювання. На основі отриманої інформації та даних з бази клієнтів про укладені договори (кількість попередніх покупок кожної компанії, дата останньої покупки та інше) можна створити прогностну модель.

Процес підготовки вхідної інформації передбачає подання її в формалізованому вигляді, тобто придатному для подальшого оброблення (рисунки 3.13, 3.14). Для цього отриманим показникам даємо умовні позначення, а символічні дані змінюємо на цифрові (табл. 3.2).

Для збільшення точності майбутнього прогнозу до сформованої в процесі опитування і формалізованої інформації треба додати дані з клієнтської БД підприємства, а саме про кількість угод укладених до останнього заказу кожною компанією, вартість останнього заказу та загальний дохід, який принесла кожна компанія за весь час співробітництва (додаток Б) та здійснити RFM аналіз.

Таблиця 3.2 - Опис та формалізація вхідної інформації

Назва поля	Показник	Умовні позначення (код)
Kind_of_activity	Вид діяльності	1 – виробництво косметики; 2 – фармацевтична промисловість; 3 – виробництво продуктів харчування; 4 – виробництво побутової хімії; 5 – інше.

Закінчення таблиці 3.2

Company_size	Розмір підприємства	В – велике підприємство; М – середнє підприємство; S – мале підприємство.
Period_of_existence / Collaboration_time	Період існування підприємства/період співробітництва	<1 – менше 1 року; 1 – 3 – від 1 до 3 років; 3 – 5 – від 3 до 5 років; >5 – більше 5 років; 0 – ще не співпрацювали.
Product	Назва товару	1 – алюмінієві туби; 2 – поліетиленові туби; 3 – ламінатні туби; 4 – ковпачки для туб.
Advertising	Джерело реклами	М – реклама в засобах масової інформації; І – реклама в інтернеті; ОА – зовнішня реклама; F – від знайомих.
Recommendations	Чи будуть клієнти рекомендувати дане підприємство	1 – так; 0,5 – можливо; 0 – ні.
Own_production/ Response/ Purchase/ Quality_Satisfaction/ Price_Satisfaction/ Future_Cooperation/ Loyalty_program	Власне виробництво/ відповідь на маркетингову компанію/ покупка/ задоволеність якістю/ задоволеність ціною/ майбутнє співробітництво/ участь в програмі лояльності	0 – ні; 1 – так.

Під RFM-аналізом мається на увазі аналіз цінності клієнта для компанії. По суті, злегка просунутий варіант ABC-аналізу, тільки з фокусом не на товари, а на клієнтах. На перше місце ставиться формалізація розміру користі кожного клієнта для бізнесу [48]. З метою виявлення цієї користі кожен клієнт розглядається за такими параметрами:

Recency - новизна (час з моменту останньої покупки);

Frequency - частота (частота покупок за період);

Monetary - монетизація (вартість покупок за період) [49].

Для визначення показника RFM необхідно спочатку розрахувати 3 проміжні параметри, які наведені вище.

Для того, щоб знайти Resency (обчислення часу з моменту останньої покупки) необхідна поточна дата (стандартна функція в Excel = ТДАТУ ()) і дата останньої покупки клієнта (від сьогоднішньої дати віднімається дата останньої покупки). Далі показник Frequency, який беремо зі зведеної таблиці (додаток Б) поле Number_prev_purch (кількість попередніх замовлень кожного клієнта) і переносимо в нову таблицю для розрахунку RFM. І останній показник Monetary також розраховується за допомогою поля Income (підсумовуються всі значення Income одного клієнта, тобто на яку загальну суму зробив замовлень клієнт за всі попередні замовлення).

Далі по кожному критерію відбувається ділення клієнтів на групи (в нашому випадку ділиться на 5 частин) Далі для вирахування підсумкових даних необхідно знайти мінімальне і максимальне значення у всіх стовпцях (Resency (min – сама недавня покупка, max – сама давня покупка), Frequency (min – найнижча кількість заказів, max – найбільша кількість заказів), Monetary (min – найменша сума заказу, max – найбільша сума заказу)). Між min та max кожного показника необхідно виділити п'ять ділянок (тобто між двома крайніми точками додати ще чотири точки, щоб вийшло п'ять рівних відрізків). Для цього від max віднімається min та ділиться на 5 і цим самим отримуємо крок (l) від однієї точки до іншої. Далі прораховуються ці самі точки. Перша точка x1 це до min додати l, друга точка x2 – x1 додати l і так далі (рисунок 3.12). На основі цих точок будуються ділянки (таблиця 3.3).

Наступним кроком кожному клієнту присвоюється числове позначення на кожний показник. Тобто, якщо значення показника Resency потрапляє в зону від min до x1, то йому присвоюється значення 5 (зовсім нещодавно).

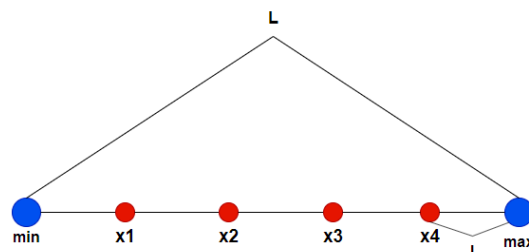


Рисунок 3.12 – Поділ клієнтів на сегменти

Розрахунки по RFM аналізу в цій роботі виконані в програмі Excel (рисунок 3.15).

Таблиця 3.3 – Сегментування по показникам Recency, Frequency, Monetary

Числове позначення групи	Recency, дні		Frequency, рази		Monetary, грн	
	1	Дуже давно	Від 788	Дуже рідко	1-8	Дуже мало
2	Давно	594-788	Рідко	8-15	Мало	3264155 - 6514688
3	Середньої давності	399-594	Середньо	15 – 21	Середньо	6514688 - 9765220
4	Нещодавно	205-399	Часто	21-28	Багато	9765220 - 13015753
5	Зовсім нещодавно	11-205	Дуже часто	Від 28	Дуже багато	Від 13015753

	A	B	C	D	E	F	G	H	I	J
1	ID_campaign	Campaign	Location	Kind_of_a	Period_of	Company	Own_proc	Advertisir	Response	Collaborat
2	1360	ОАО "Лубнифарм"	Полтавська обл	2	<1	M	1	M	1	<1
3	2601	ООО "Юрия-фарм"	Київська обл.	2	1 – 3	S	1	I	0	<1
4	815	ООО «Жерминаль»	Росія	1	3 – 5	B	1	F	0	1 – 3
5	1902	Компания «ЕвроТе	Росія	4	>5	B	1	I	1	>5
6	13	АЛТЕЙ УКРАИНА, Н	Харківська обл	1	>5	M	1	M	1	3 – 5
7	2890	БИОЛИК, ПАО	Вінницька обл.	2	1 – 3	S	1	F	0	1 – 3
8	2015	АЛТЕЙ УКРАИНА, Н	Харківська обл	1	3 – 5	M	1	F	1	1 – 3
9	930	Емі Люкс	Запорізька обл	4	>5	M	1	I	1	0
10	1167	Евдемон	Одеська обл.	4	<1	S	1	M	1	0

Рисунок 3.13 – Зведена таблиця

	K	L	M	N	O	P	Q	R	S	T	U
	Number_prev_purc	Product_ID	Purchase	Data_purchase	Income	Quality_S	Price_Sati	Future_Cc	Loyalty_pi	Recommendations	RFM
	2	1	1	15.11.2018	33900,00	1	1	1	0	1	6
	1	1	1	09.03.2019	56500,00	1	1	1	1	1	6
	6	3	1	11.12.2017	153000,00	1	1	1	0	1	6
	24	3	1	13.03.2019	45900,00	1	0	1	1	1	9
	12	2	1	16.08.2018	30900,00	1	0	0	0	0	8
	9	4	1	11.10.2019	15900,00	1	1	1	0	1	8
	5	1,4	1	31.01.2018	49800,00	1	1	1	0	0,5	6
	0	3	0			1	1	1	0	1	0
	0	1	0				1	1	0	1	0
	31	4	1	06.06.2018	53000,00	1	1	1	1	0,5	11
	2	4	1	28.01.2019	13250,00	0	1	0		0	6
	2	1,4	1	10.10.2018	179250,00	1	0	0		0,5	7

Рисунок 3.14 – Зведена таблиця

	B	C	D	E	F	G	H	I	J	K	L	
1												
2	Data_purchase	Recency	Income	Frequency	Monetary		Recency	Frequency	Monetary	RFM	RFM	
3	15.11.2018		370	33900	2	67800		4	1	1	6	6
4	09.03.2019		256	56500	1	56500		4	1	1	6	6
5	11.12.2018		344	153000	6	918000		4	1	1	6	6
6	13.03.2019		252	45900	24	1101600		4	4	1	9	9
7	16.08.2019		96	30900	12	370800		5	2	1	8	8
8	11.10.2019		40	15900	9	143100		5	2	1	8	8
9	31.01.2019		293	49800	5	249000		4	1	1	6	6

Рисунок 3.15 – RFM аналіз

Підводячи підсумки після проведення попередніх етапів можна сказати, що було відібрано тільки коректні дані, на основі яких можна отримати грамотно побудовану модель.

3.3. Побудова прогнозної моделі на основі IBM SPSS Modeler

Модель прогнозування є функціональним уявленням, яке адекватно описує досліджуваний процес і є основою для отримання його майбутніх значень.

В даному випадку буде використано алгоритм моделювання список рішень, як основний фрагмент.

Алгоритм Список рішень генерує правила, що визначають більш-менш високу ймовірність даного бінарного виходу (так чи ні). Моделі списку рішень широко використовуються в управлінні зв'язками з клієнтами (наприклад, в колл-центрах і в прикладних програмах маркетингу).

Сам побудований потік (рисунок 3.16) буде складатися з декількох послідовних вузлів (блоків).

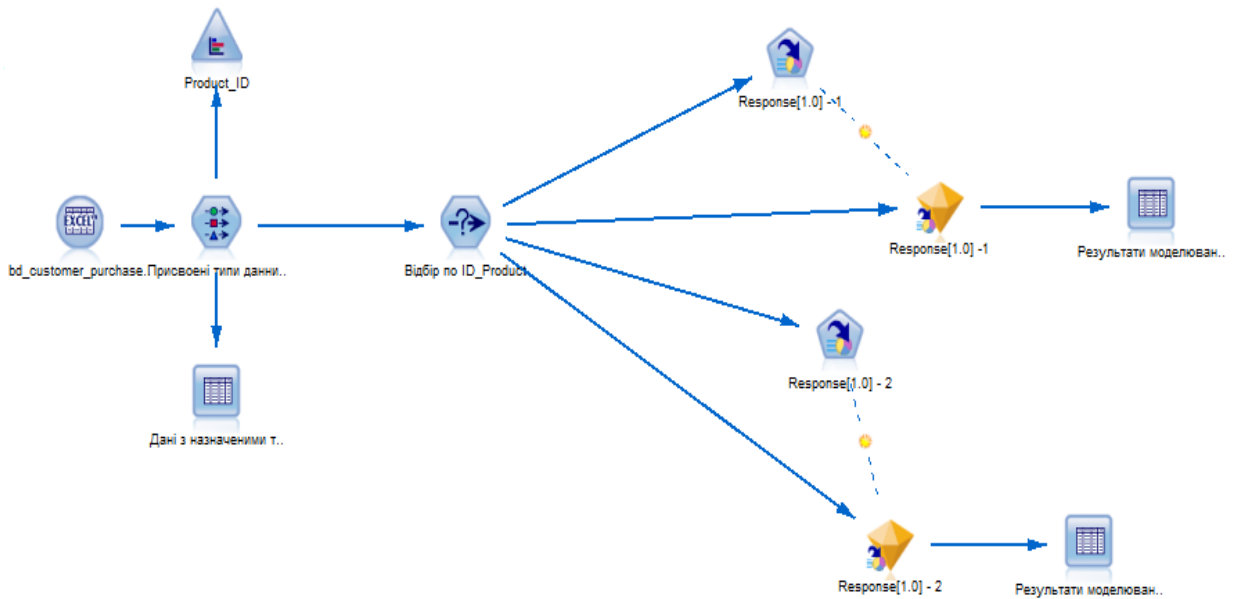


Рисунок 3.16 – Потік «Моделювання відгуку клієнтів»

Перший вузол – це джерело даних (рисунок 3.17). В цьому вузлі прикріпляється файл Excel з вихідними даними, в якому можна вказати дані з якого листа файлу потрібно підтягнути до моделі.

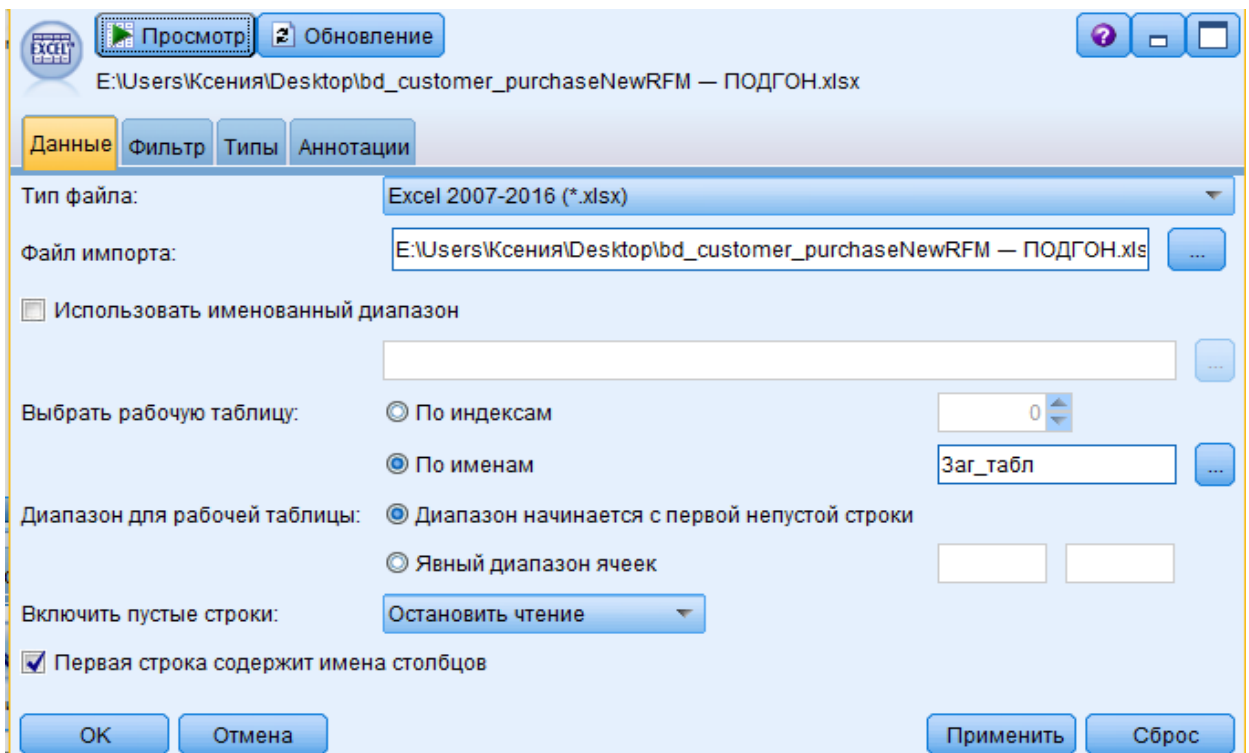


Рисунок 3.17 – Вкладка «Дані» у вузлі «Файл Excel»

На другій вкладці (рисунок 3.18) вказуємо тип даних для кожного стовпця та система сама зчитує варіацію значень даних, щоб гарантувати

ініціювання всіх значень.. Але програма автоматично визначає тип виміру категоріальний для таких полів як Location, Company_size, advertising, collaboration_time та product_id. Але цей тип значень не підходить, так як категоріальний тип використовується для строкових значень, коли точне число різних значень невідомо. А номінальний тип використовується для опису даних з декількома відмінними значеннями (кожне з яких обробляється як елемент набору), наприклад: малий / середній / великий. У номінальних даних може бути будь-який тип зберігання: числовий, строковий або дати-часу [50]. Також можна вказати ті поля, які не будуть відігравати роль в моделі. В цьому випадку вказуємо, що поле Campaign (Назва компаній) не має ролі.

Поле	Измерение	Значения	Пропущенные	Проверить	Роль
ID_campaign	Непрерывное	[4.0,3000.0]		Нет	Ввод
Campaign	Без типа			Нет	Нет
Location	Номинальное	"Івано-Франк...		Нет	Ввод
Kind_of_activity	Непрерывное	[1.0,5.0]		Нет	Ввод
Period_of_exist...	Номинальное	"", "1 — 3", "3 ...		Нет	Ввод
Company_size	Номинальное	"", "B,M,S		Нет	Ввод
Own_production	Непрерывное	[0.0,1.0]		Нет	Ввод
Advertising	Номинальное	F,I,"I, F", "I, OA"...		Нет	Ввод
Response	Непрерывное	[0.0,1.0]		Нет	Ввод
Collaboration_ti...	Номинальное	"0", "1 — 3", "3 ...		Нет	Ввод
Number_prev_p...	Непрерывное	[0.0,35.0]		Нет	Ввод
Product_ID	Номинальное	"1", "1, 4", "2", "..."		Нет	Ввод
Purchase	Непрерывное	[0.0,1.0]		Нет	Ввод
Data_purchase	Непрерывное	[2017-11-13,...		Нет	Ввод
Income	Непрерывное	[12431.0,697...		Нет	Ввод
Quality_Satisfact...	Непрерывное	[0.0,1.0]		Нет	Ввод
Price_Satisfaction	Непрерывное	[0.0,1.0]		Нет	Ввод
Future_Coopera...	Непрерывное	[0.0,1.0]		Нет	Ввод
Loyalty_program	Непрерывное	[0.0,1.0]		Нет	Ввод
RFM	Непрерывное	[0.0,14.0]		Нет	Ввод

Рисунок 3.18 – Вкладка «Типи» у вузлі «Файл Excel»

Виконавши всі налаштування вихідних даних можна відкрити попередній перегляд занесених даних (рисунок 3.19).

	ID_campaign	Campaign	Location	Kind_of_activity	Period_of_existence	Company_size	Own_production	Advertising
1	1360.000	ОАО "Лубнифарм"	Полтавська обл.	2.000	<1	M	1.000	M
2	2601.000	ООО "Юрия-фарм"	Київська обл.	2.000	1 – 3	M	1.000	I
3	815.000	ООО «Жерминаль»	Росія	1.000	3 – 5	B	1.000	F
4	1902.000	Компания «ЕвроТек»	Росія	4.000	>5	B	1.000	I
5	13.000	АЛТЕЙ УКРАИНА, НПФ	Харківська обл.	1.000	>5	M	1.000	M
6	2890.000	БИОЛИК, ПАО	Вінницька обл.	2.000	1 – 3	S	1.000	F
7	2015.000	АЛТЕЙ УКРАИНА, НПФ	Харківська обл.	1.000	3 – 5	M	1.000	F
8	930.000	Емі Люкс	Запорізька обл.	4.000	>5	M	1.000	I
9	1167.000	Евдемон	Одеська обл.	4.000	<1	S	1.000	M
10	2239.000	"Объединение" Gerard"	Донецька обл.	4.000	>5	B	1.000	F

Рисунок 3.19 – Попередній перегляд даних

Наступний крок – це проведення операцій з полями. В обраному вузлі типи даних присвоюємо кожному полю новий тип даних (рисунок 3.20), з яким буде зручно працювати надалі. Тепер змінюємо вимір даних з неперервного на тип прапор для полів `own_production` (власне виробництво), `response`, `purchase`, `quality_satisfaction`, `price_satisfaction` та `future_cooperation`. Так як тип прапор використовується для даних з двома різними значеннями, що вказують на наявність або відсутність тієї чи іншої особливості, наприклад: `true` і `false`, Та й Ні або 0 і 1. Використані значення можуть змінюватися, але одне завжди має визначатися як значення "true", а інше - як значення "false". Дані можуть бути представлені як текст, цілочисельне значення, дійсне число, дата, час або відмітка часу.

Також в стовпці роль знову змінюємо декілька значень вхід на відсутність ролей в таких полях, як `id_campaign`, `location`, `kind_of_activity`, `own_production`, `advertising` та `data_purchase`. Всі ці поля використовуються при обробці даних, але не будуть використовуватися в створенні фактичної моделі. І полю `response` назначаємо цільову роль, тому що на ньому ґрунтується вся подальша робота.

На рисунку 3.21 показано вкладку «Формат», на якій було змінено в стовпці формат розрядність значень, в яких повинно бути ціле число. І в кінці за допомогою вузла «Таблиця» можна побачити, те що наприклад в полі `id_campaign` і `own_production` змінилися формати значень (рисунок 3.22).

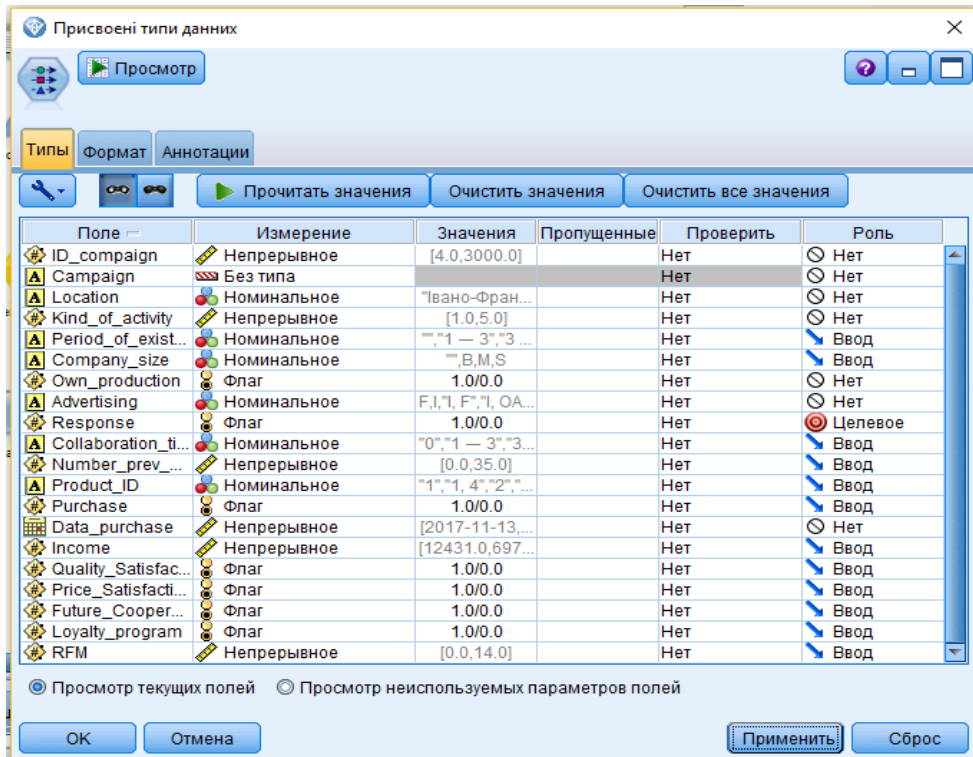


Рисунок 3.20 – Вкладка «Типы» у вузлі «Типи даних»

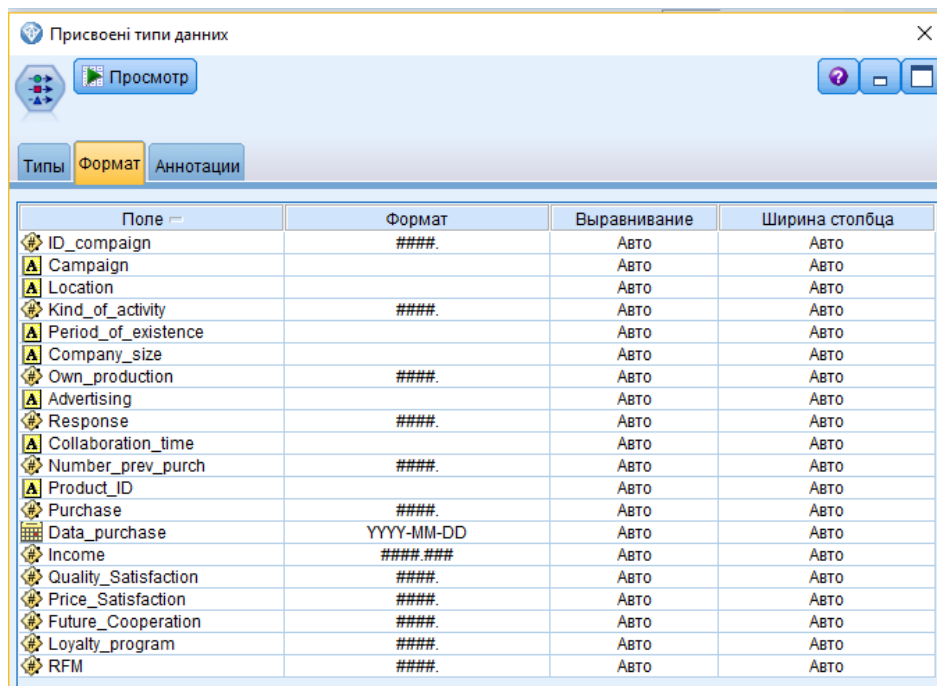


Рисунок 3.21 – Вкладка «Формат» у вузлі «Типи даних»

Дані з назначеними типами (20 полей, 502 записей)

Файл | Изменить | Генерировать

Таблица | Аннотации

	ID_campaign	Campaign	Location	Kind_of_activity	Period_of_existence	Company_size	Own_production	Advertising
1	1360	ОАО "Лубнифарм"	Полтавська обл.		2 <1	M		1 M
2	2601	ООО "Юрия-фарм"	Київська обл.		2 1 — 3	M		1 I
3	815	ООО «Жерминаль»	Росія		1 3 — 5	B		1 F
4	1902	Компания «ЕвроТек»	Росія		4 >5	B		1 I
5	13	АЛТЕЙ УКРАИНА, НПФ	Харківська обл.		1 >5	M		1 M
6	2890	БИОЛИК, ПАО	Вінницька обл.		2 1 — 3	S		1 F
7	2015	АЛТЕЙ УКРАИНА, НПФ	Харківська обл.		1 3 — 5	M		1 F
8	930	Емі Люкс	Запорізька обл.		4 >5	M		1 I
9	1167	Евдемон	Одеська обл.		4 <1	S		1 M
10	2239	"Объединение" Gerard™	Донецька обл.		4 >5	B		1 F
11	1937	АМАЛЬГАМА ЛЮКС, ООО	Миколаївська о...		1 3 — 5	M		1 OA
12	2246	АРОЗА, ПАРФЮМЕРНО...	Київська обл.		1 3 — 5	B		1 M
13	1257	Римал, ПрАТ (Римал, Ч...	Рівненська обл.		3 1 — 3	S		1 OA
14	2032	Римал, ПрАТ (Римал, Ч...	Рівненська обл.		3 >5	B		1 F
15	1873	ГОРОДЕНКОВСКИЙ С...	Івано-Франківс...		3 >5	M		1 OA
16	1297	БИОЛИК, ПАО	Вінницька обл.		1 1 — 3	S		1 F
17	2300	СВ, ООО	Харківська обл.		4 3 — 5	M		1 M
18	2945	Интер-Мол, ТОВ	Дніпропетровс...		3 1 — 3	S		1 OA
19	2288	СОУЛСБЕРІ. Пластико...	Росія		5 3 — 5	M		0 I
20	2221	Гемма Графикс, ЗАО	Росія		5 >5	M		0 M

OK

Рисунок 3.22 – Таблица даних з назначеними полями

Далі для наочності та для того щоб видалити з моделі непотрібні дані використовується вузол з блоку «Діаграми». Графік (рисунок 3.23) буде побудовано по полю product_id, де буде видно процент співвідношення купівлі кожного виду товарів. А за допомогою поля Response на графіку визначиться кількість відгуків на кожний тип товарі (рисунок 3.24).

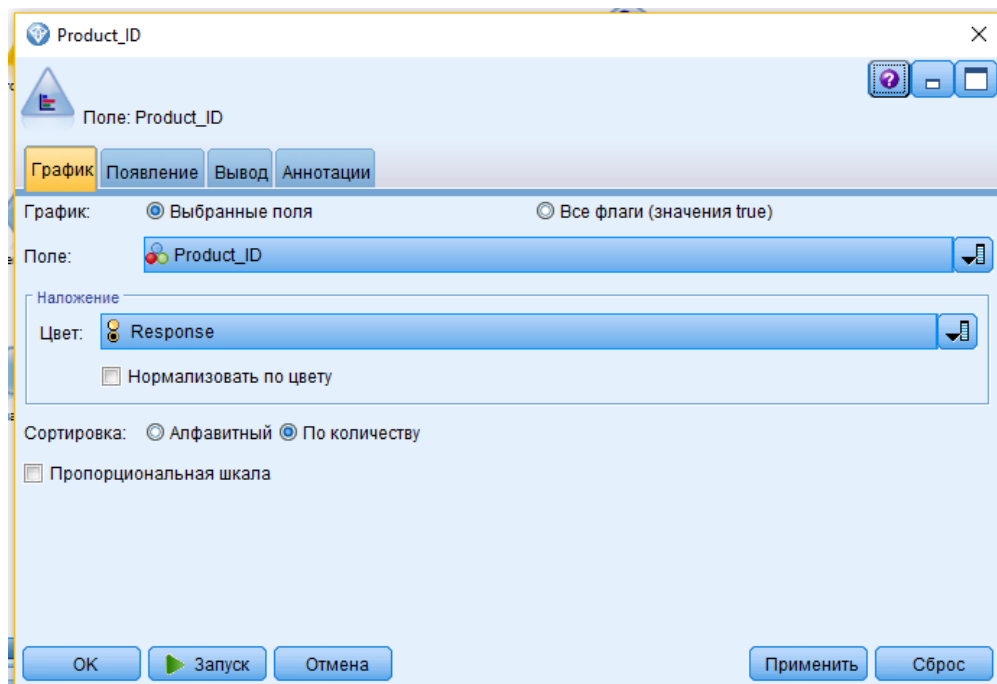


Рисунок 3.23 – Налаштування побудови графіку



Рисунок 3.24 – Графік «Розподіл відгуків по товарам»

Наступний крок – це вузол «Відбір записів» (рисунок 3.25). В цьому потоці такий вузол вставлений один, але можна поставити більше, якщо є декілька умов відбору даних. Хоча ці дані включають в себе інформацію про чотири різні типи товарів і ще чотири варіанти комбінацій товарів, ви будете фокусуватися тільки на одинарних покупка, так як в комбінованих дуже низький відсоток і не має сенсу їх розглядати.

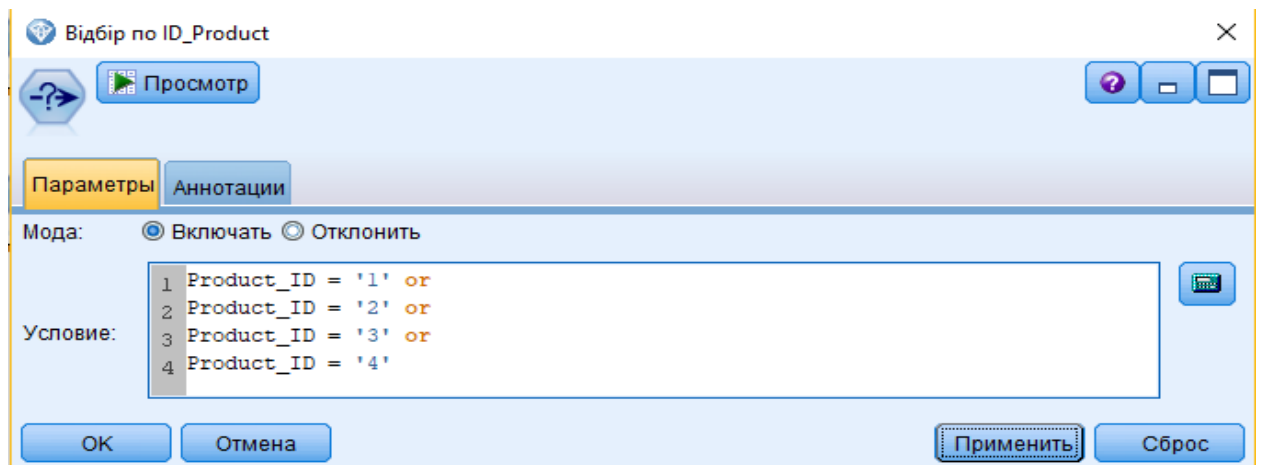


Рисунок 3.25 – Вибір записів

Після того, як було пройдено всі підготовчі етапи далі йде сам процес створення моделі.

В цьому випадку буде побудовано дві моделі з різними налаштуваннями, для того щоб або отримати кращий результат або

використати результати обох моделей для більшого покриття клієнтської бази.

В першому варіанті побудови моделі спочатку йде приєднання вузла Список рішень до потоку. На вкладці Модель (рисунок 3.26) задається 1 (True) для Цільового значення, щоб відзначити вихідні поля, для яких ви хочете виконати пошук [51]. В даному випадку ви будете шукати користувачів, які відповідали Так на попередню пропозицію. Також потрібно вказати, що модель повинна шукати сегменти, в яких висока ймовірність потрапляння в умови.

На цій же вкладці змінюємо максимальну кількість сегментів до 5, щоб охопити більшу кількість записів і модель була більш складною. Після встановлюється мінімальний розмір сегмента. У відсотках до попереднього сегменту задано 3%, а як абсолютне значення задається 20 мінімальний розмір групи, тому що загальна кількість записів в вставленому файлі не дуже велика.

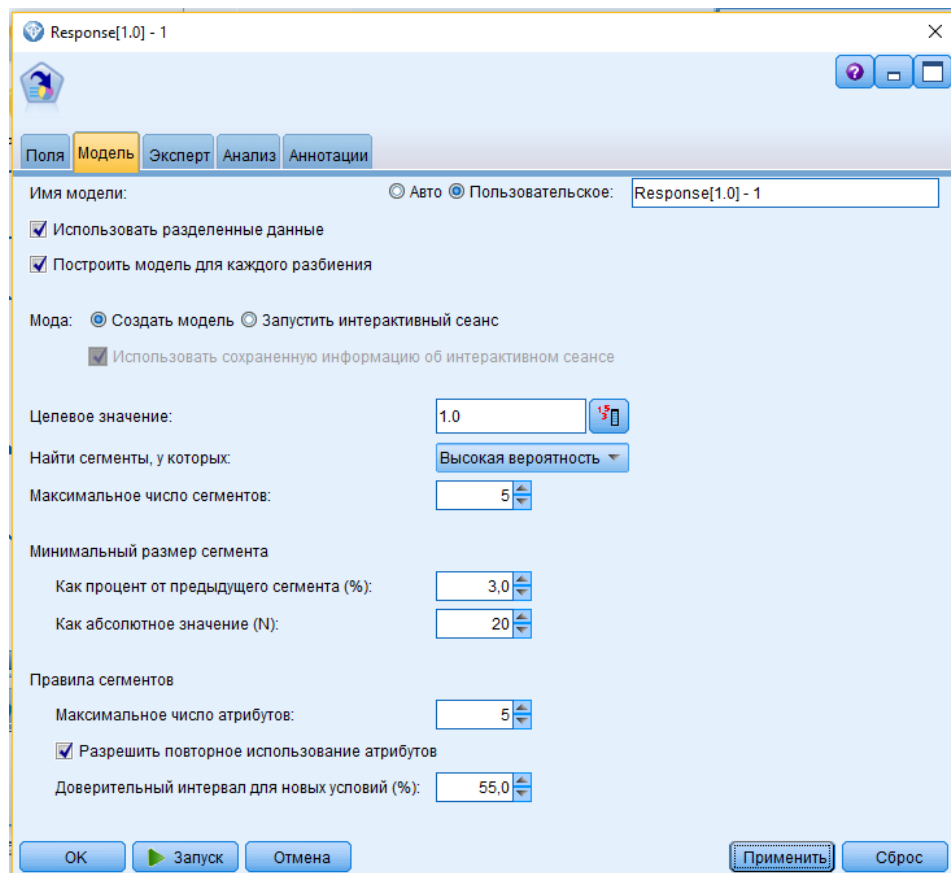


Рисунок 3.26 – Налаштування 1 моделі

Для того, щоб точно налаштувати процес побудови моделі зазвичай використовують параметри Експерт (рисунок 3.27).

Також необхідно збільшити максимальне число альтернативних варіантів до 4. Ця опція працює спільно з параметром Запустити інтерактивний сеанс, обраним на вкладці Модель.

Ще важливу роль відіграє мінімальна сума, на яку сегмент повинен вирости при злитті зі своїм сусідом. Для цієї моделі встановлюється значення 2. Всі інші поля залишаємо зі стандартними налаштуваннями.

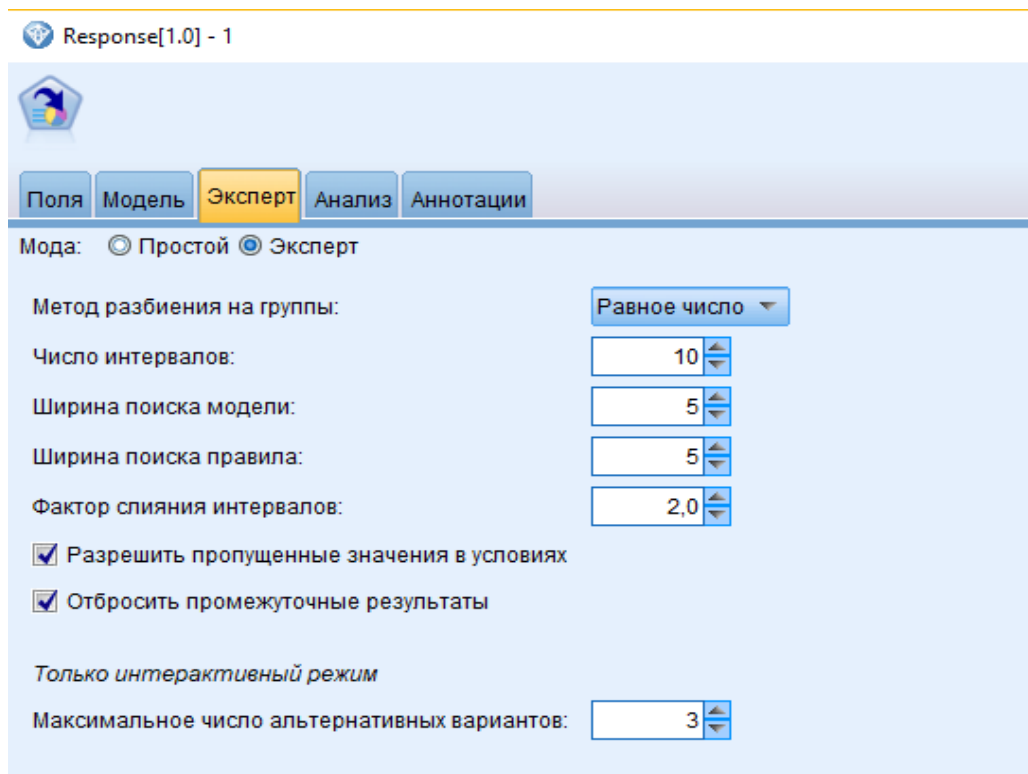


Рисунок 3.27 – Налаштування параметрів Експерт

Також, щоб модель не використовувала непотрібні поля для побудови сегментів у вкладці Модель обираємо вручну цільове поле Response та поля вводу RFM, income, future_cooperation, period_of_existence, company_size, Collaboration_time, number_prev_purc і purchase (рисунок 3.28).

На цьому етапі побудова першої моделі закінчується.

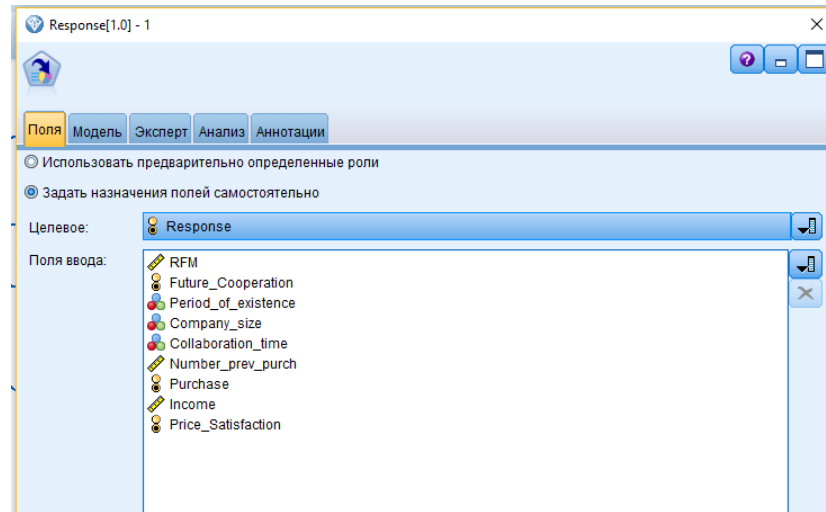


Рисунок 3.28 – Налаштування полів призначення

Другий варіант моделі складається з таких самих вузлів. У вузлі моделювання змінюються деякі параметри, для того щоб отримати нові сегменти. В поля абсолютне значення мінімального розміру сегмента вказується 30 та довірчий інтервал зменшуємо на 5% (рисунок 3.29).

Обираємо поля, які можуть бути використані у сегментах 2-ої моделі (рисунок 3.30).

Наступний етап аналіз виданих результатів.

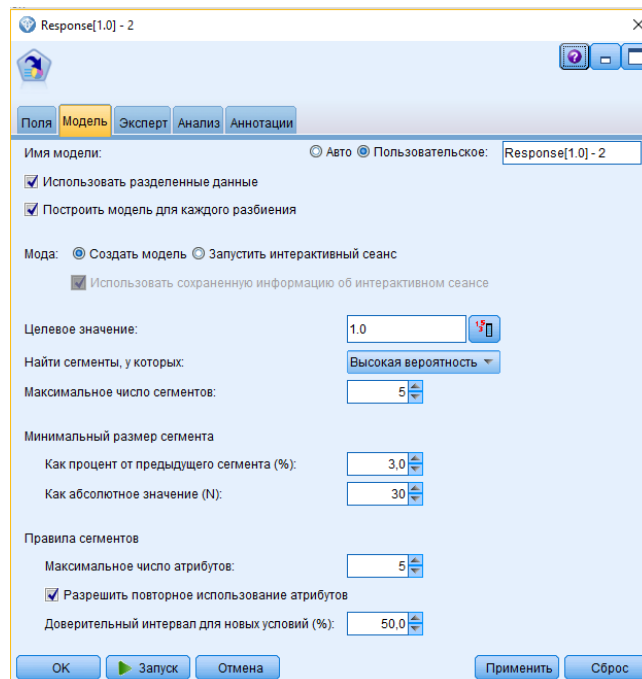


Рисунок 3.29 – Налаштування 2-ої моделі

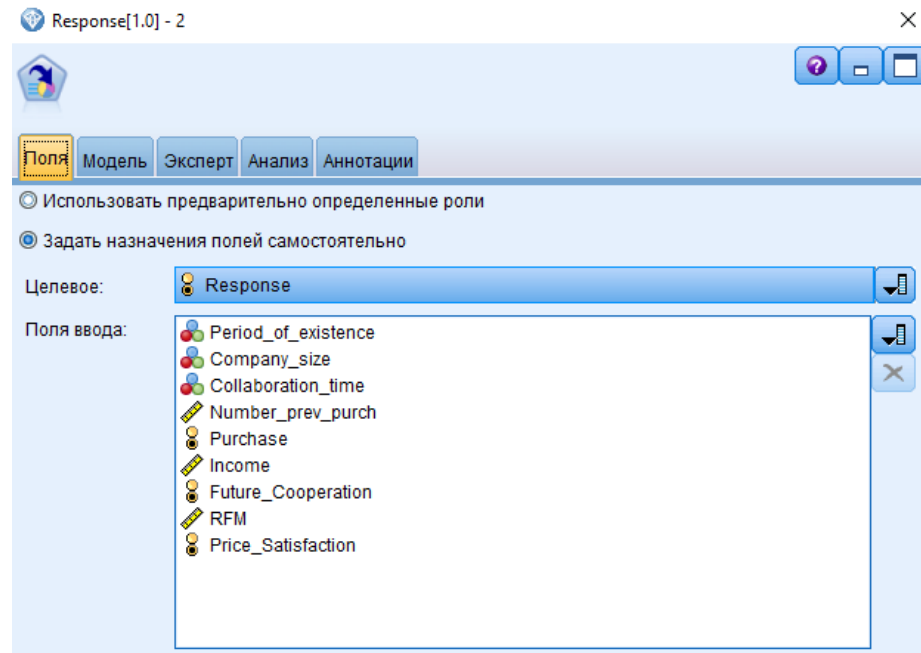


Рисунок 3.30 – Налаштування полів призначення

3.4. Аналіз отриманих результатів

Після налаштування обох варіантів вузлів моделей можна перейти до розбору результатів і виводу закономірностей.

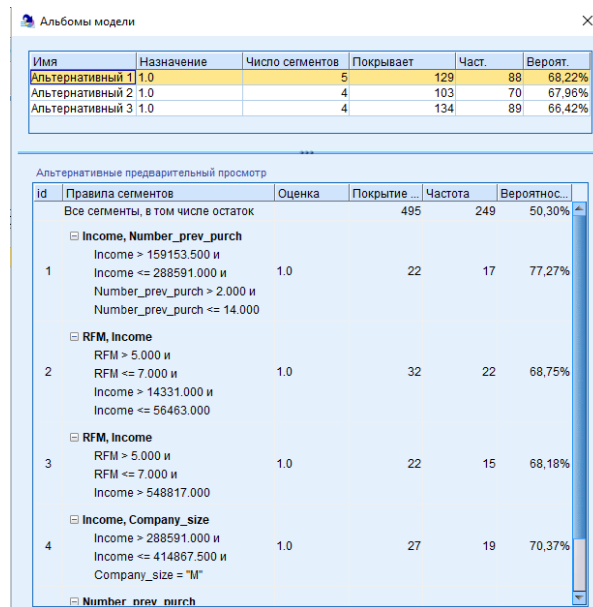
Спочатку аналізуємо результати інтерактивного сенсу (тобто модель одразу не створюється, а даються на вибір декілька альтернативних варіантів) (рисунок 3.31).

id	Правила сегментов	Оценка	Покрытие (n)	Частота	Вероятность
	Все сегменты, в том числе остаток		495	249	50,30%
	Остаток		495	249	50,30%

Рисунок 3.31 – Интерактивный список

Так як ніякі сегменти ще не визначені, всі записи потраплять в залишок. З 495 записів у вибірці є 249 відповідей Так, що дає коефіцієнт попадання 50,30%. Ви хочете збільшити цей коефіцієнт, визначивши сегменти покупців, які більш (або менш) ймовірно дадуть бажаний відгук.

Так як в процесі налаштування було обрано створення трьох альтернативних варіантів, тому система видала саме три варіанти (рисунок 3.32). Серед усіх виданих варіантів в першому - ймовірність попадання найбільша (68,2%), тому цей варіант і залишаємо.



Имя	Назначение	Число сегментов	Покрывает	Част.	Вероят.
Альтернативный 1	1.0	5	129	88	68,22%
Альтернативный 2	1.0	4	103	70	67,96%
Альтернативный 3	1.0	4	134	89	66,42%

id	Правила сегментов	Оценка	Покрытие	Частота	Вероятнос...
	Все сегменты, в том числе остаток		495	249	50,30%
	Income, Number_prev_purch				
1	Income > 159153.500 и Income <= 288591.000 и Number_prev_purch > 2.000 и Number_prev_purch <= 14.000	1.0	22	17	77,27%
	RFM, Income				
2	RFM > 5.000 и RFM <= 7.000 и Income > 14331.000 и Income <= 56463.000	1.0	32	22	68,75%
	RFM, Income				
3	RFM > 5.000 и RFM <= 7.000 и Income > 548817.000	1.0	22	15	68,18%
	Income, Company_size				
4	Income > 288591.000 и Income <= 414867.500 и Company_size = "M"	1.0	27	19	70,37%
	Number_prev_purch				

Рисунок 3.32 – Альтернативні варіанти моделей

Використовуючи правила на основі предикторів, таких як вхідні дані, прибуток підприємства від конкретного клієнта на останнє замовлення, кількість попередніх замовлень, розмір компанії клієнта і оцінка RFM (рисунок 3.33), модель ідентифікує п'ять сегментів (рисунок 3.34) з коефіцієнтом відгуку більшим, ніж в цілому по вибірці. Готова модель покриває 129 записів зі всієї вибірки. А кількість потраплянь по відношенню до покриття 88.

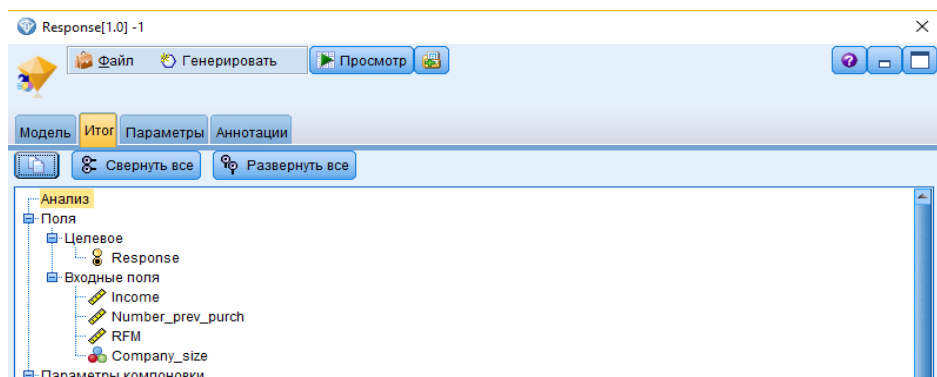


Рисунок 3.33 – Підсумки моделі

id	Правила сегментов	Оценка	Покрытие (n)	Частота	Вероятность
	Все сегменты, в том числе остаток		495	249	50,30%
1	Income, Number_prev_purch Income > 159153.500 и Income <= 288591.000 и Number_prev_purch > 2.000 и Number_prev_purch <= 14.000	1.0	22	17	77,27%
2	RFM, Income RFM > 5.000 и RFM <= 7.000 и Income > 14331.000 и Income <= 56463.000	1.0	32	22	68,75%
3	RFM, Income RFM > 5.000 и RFM <= 7.000 и Income > 548817.000	1.0	22	15	68,18%
4	Income, Company_size Income > 288591.000 и Income <= 414867.500 и Company_size = "M"	1.0	27	19	70,37%
5	Number_prev_purch Number_prev_purch > 0.000 и Number_prev_purch <= 1.000	1.0	26	15	57,69%
	Остаток		366	161	43,99%

Рисунок 3.34 – Результати моделі

З першого сегменту, який показано на рисунку 3.34 можна зробити висновок, що серед усіх клієнтів, яких ми аналізували в моделі потрапляють до списку тих хто може позитивно відповісти на майбутні маркетингові пропозиції опинилися ті у хто робить замовлення на суму від 159153,5 грн. до 288591 грн. та одночасно зробив від 2 до 14 замовлень в цій компанії.

На другому сегменті клєнти відібрані по іншим показникам, таким як RFM аналіз і сума останнього замовлення. Тобто цінність клієнта для підприємства (RFM аналіз) повинна бути від 5 до 7 одиниць (це нижче середини значення, так як найвища цінність 15, а найнижча 3) та сума останнього замовлення повинна складати від 14 331 до 56 463 грн..

Ще один цікавий сегмент де клієнти, які можуть гарно відреагувати на нові пропозиції – це сегмент №4. В ньому задіяно поле розмір компанії і поле дохід. Для більшого відгуку підприємство повинно орієнтуватися на компанії середнього розміру (M – medium) з замовленнями, які складають не менше 288 591 грн та не більше 414 867,5 грн..

Однак ця модель покриває тільки малу частину всієї вибірки, залишаючи 366 записів, серед яких кілька сотень влучень, в залишку. Хотілося б, щоб модель захоплювала більше таких влучень, як і раніше виключаючи незначні сегменти.

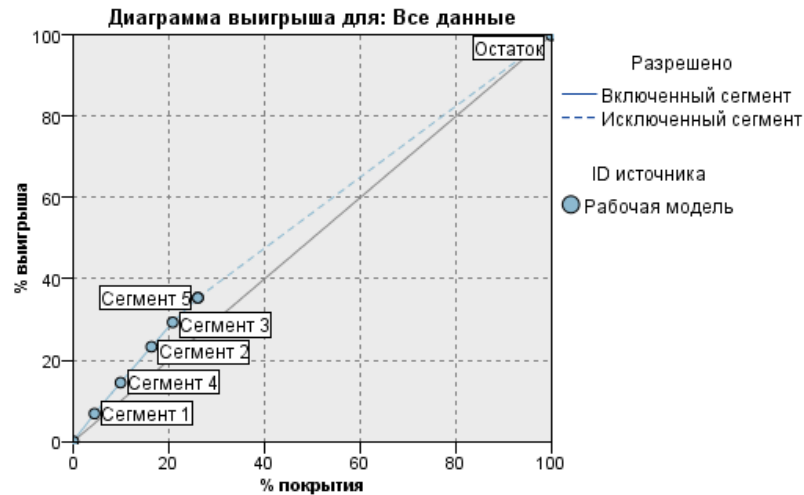


Рисунок 3.35 – Діаграма виграшу для 1 моделі

Діаграма виграшу представляє значення з стовпця % виграшу таблиці. Виграш визначається як відношення числа влучень на кожному діленні шкали до загальної кількості влучень в дереві з використанням такої формули:

$$\left(\frac{\text{Число влучень на сегмент}}{\text{загальне число влучень}} \right) \times 100\%$$

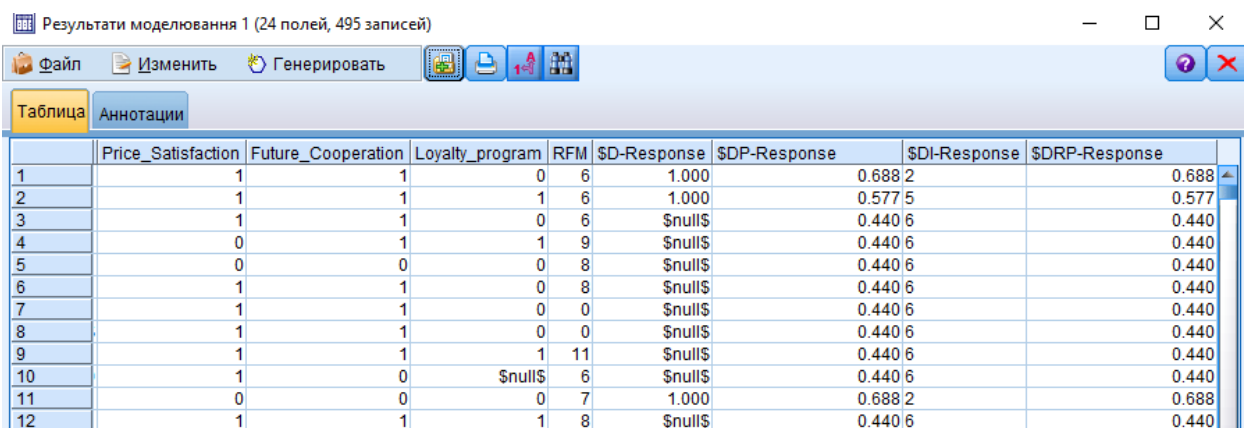
Діаграма виграшу наочно ілюструє, наскільки широко по дереву потрібно проводити пошук для набору даного відсотка всіх влучень. Діагональна лінія показує очікуваний відгук для всієї вибірки, якщо модель не використовується. В цьому випадку показник відгуку був би константою, так як ймовірність відгуку для одного клієнта така ж, як для іншого. Крива лінія позначає, наскільки можна підвищити відгук, включивши тільки тих клієнтів, які на підставі виграшу ранжуються в максимальних процентах.

Далі будуть представлені дані на основі, яких була побудована діаграма виграшу (таблиця 3.4). Номери сегментів розташовані в тому порядку, як і на діаграмі (рисунок 3.35).

Таблиця 3.4 – Дані для діаграми виграшу

№ сегменту	Покриття, %	Виграш, %
1	4,4	6,8
4	5,45	7,6
2	6,46	8,8
3	4,4	6
5	5,25	6
Разом	25,96	35,2

Для того, щоб побачити відібрані під час моделювання поля можна приєднати вузол Вихід даних у виді таблиці. В якому видно, що при запуску потоку, що містить модель списку рішень, вузол додає три нових поля, в тому числі оцінку (1, що означає так, для включених полів або \$null\$ для виключених полів), ймовірність (коефіцієнт попадань) для сегмента, в який потрапляє запис, і номер ID для сегмента. Імена нових полів виходять з імені вихідного поля, значення якого передбачається, додаванням префіксу \$D- для оцінки, \$ DP- для ймовірності і \$ DI- для ID сегмента (рисунок 3.36).



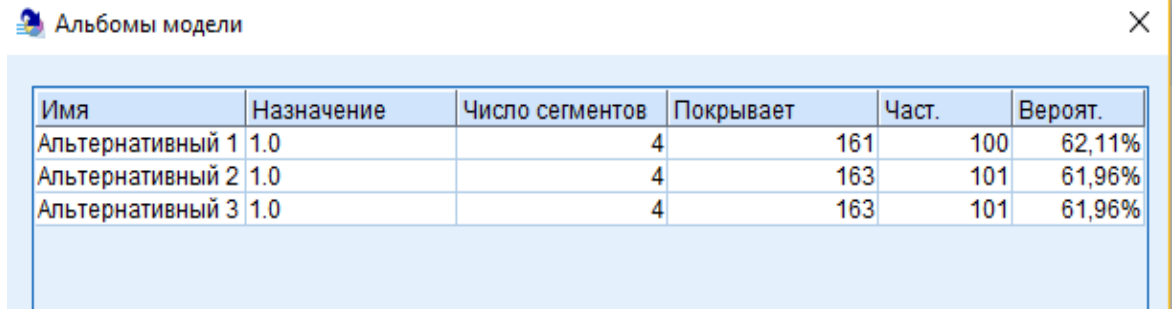
	Price_Satisfaction	Future_Cooperation	Loyalty_program	RFM	\$D-Response	\$DP-Response	\$DI-Response	\$DRP-Response
1	1	1	0	6	1.000	0.688	2	0.688
2	1	1	1	6	1.000	0.577	5	0.577
3	1	1	0	6	\$null\$	0.440	6	0.440
4	0	1	1	9	\$null\$	0.440	6	0.440
5	0	0	0	8	\$null\$	0.440	6	0.440
6	1	1	0	8	\$null\$	0.440	6	0.440
7	1	1	0	0	\$null\$	0.440	6	0.440
8	1	1	0	0	\$null\$	0.440	6	0.440
9	1	1	1	11	\$null\$	0.440	6	0.440
10	1	0	\$null\$	6	\$null\$	0.440	6	0.440
11	0	0	0	7	1.000	0.688	2	0.688
12	1	1	1	8	\$null\$	0.440	6	0.440

Рисунок 3.36 – Результат моделювання в табличному вигляді

Тепер можна роздивитися і проаналізувати результати другої моделі.

Тому знову запускається інтерактивний сеанс і серед трьох виданих альтернатив обирається найліпша. В цьому випадку перша альтернатива хоча

і покриває менше записів, але має більшу ймовірність попадання (рисунок 3.37).

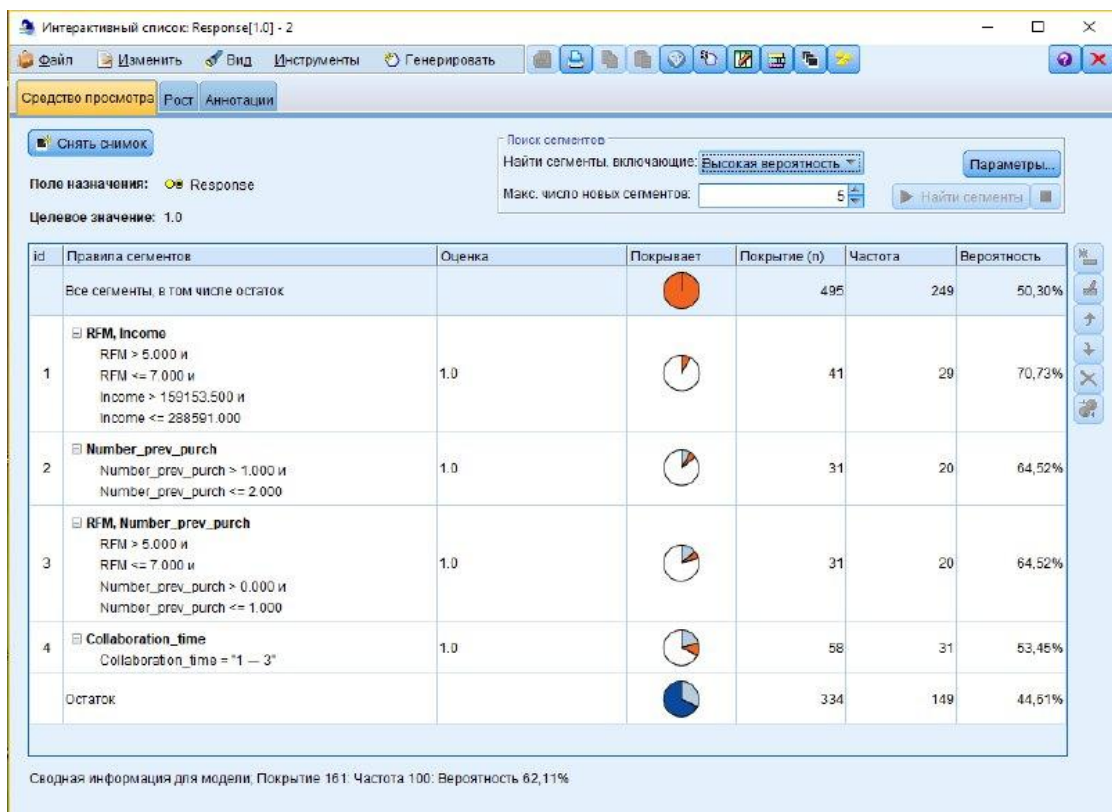


Имя	Назначение	Число сегментов	Покрывает	Част.	Вероят.
Альтернативный 1	1.0	4	161	100	62,11%
Альтернативный 2	1.0	4	163	101	61,96%
Альтернативный 3	1.0	4	163	101	61,96%

Рисунок 3.37 – Альтернативні варіанти 2-ої моделі

В результатах виконання другої моделі можна побачити 4 побудованих сегменти, які разом покривають 161 запис, з яких 100 частота потрапляння (рисунок 3.38).

В сегментах даної моделі задіяно 4 поля: RFM-аналіз, вартість останнього замовлення, загальна кількість замовлень кожним клієнтом та час співробітництва.



id	Правила сегментов	Оценка	Покрывает	Покрытие (n)	Частота	Вероятность
	Все сегменты, в том числе остаток			496	249	50,30%
1	RFM, Income RFM > 5.000 и RFM <= 7.000 и Income > 159153.500 и Income <= 288591.000	1.0		41	29	70,73%
2	Number_prev_purch Number_prev_purch > 1.000 и Number_prev_purch <= 2.000	1.0		31	20	64,52%
3	RFM, Number_prev_purch RFM > 5.000 и RFM <= 7.000 и Number_prev_purch > 0.000 и Number_prev_purch <= 1.000	1.0		31	20	64,52%
4	Collaboration_time Collaboration_time = "1 — 3"	1.0		58	31	53,46%
	Остаток			334	149	44,51%

Сводная информация для модели: Покрытие 161. Частота 100. Вероятность 62,11%

Рисунок 3.38– Результати 2-ої моделі

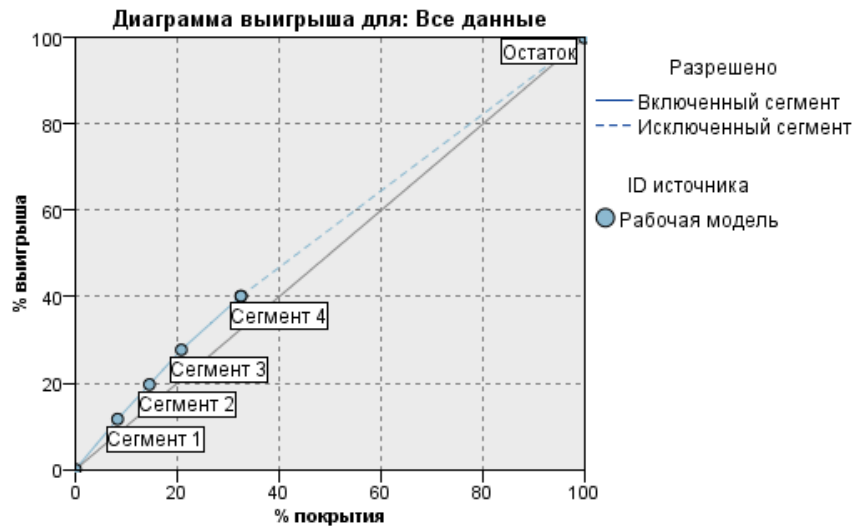


Рисунок 3.39 – Діаграма виграшу для 2 моделі

Для побудованої другої моделі теж є діаграма виграшу (рисунок 3.39) та дані на основі, яких була побудована діаграма виграшу (таблиця 3.5).

Таблиця 3.5 – Дані для діаграми виграшу

№ сегменту	Покриття, %	Виграш, %
1	8,2	11,64
2	6,26	8
3	6,26	8
4	11,7	12,4
Разом	32,42	40,04

Проаналізувавши результати обох моделей можна зробити висновок, що покриття записів в другій моделі трохи більше (перша модель - 129 , друга модель – 161) і виходячи з цього частота також в 2-ій моделі більше (перша модель – 100%, друга модель – 88%), але ймовірність першої моделі 68,22%, а другої – 62,11 і це компенсує невеликий розрив в покритті.

Процент виграшу в другій моделі на 5,02 % вище ніж в першій, тобто залучивши до використання другу модель можна очікувати більший відгук від клієнтів.

Але для прямого використання на підприємстві підходять обидві моделі і можна використовувати їх як окремо, так і побудувати майбутні маркетингові пропозиції на основі згрупованих результатів.

ВИСНОВКИ

У магістерській дипломній роботі були вирішені наступні задачі: розглянуто та проаналізовано цілі і завдання клієнтської аналітики; досліджені етапи розвитку аналітики і обробки даних; проаналізовані та класифіковані інформаційні технології клієнтської аналітики; розглянуто інструментальні засоби для вирішення проблем клієнтської аналітики; обґрунтований вибір методів для управління відгуками клієнтів на маркетингову компанію підприємств; розроблені пропозиції щодо практичного застосування вибраного методу та побудованої прогнозної моделі.

В результаті аналізу літературних джерел було визначено, що аналіз клієнта - це використання даних для розуміння складу, потреб і задоволеності клієнта. Крім того, технологія, що дозволяє розділити покупців на групи на основі поведінки, визначити загальні тенденції або розробити цілеспрямовану маркетингову і збутову діяльність. Також аналіз поведінки клієнтів значно впливає на просування товарів та пропозицій і збільшує відгук клієнтів на маркетингову компанію. Вимірюючи і аналізуючи дані за допомогою певних метрик, організації можуть створювати успішні взаємодії з клієнтами.

Було проведено порівняння двох аналітичних платформ: IBM SPSS Modeler та SAP Predictive Analytics. SPSS Modeler - це провідне рішення в сфері візуального аналізу даних і машинного навчання. Воно допомагає організаціям прискорити окупність інвестицій і досягти потрібних результатів за рахунок прискорення операційних завдань для фахівців з аналізу даних. Провідні організації по всьому світу користуються рішеннями IBM для підготовки і виявлення даних, прогнозного аналізу даних, управління моделями, впровадження моделей і машинного навчання з метою монетизації ресурсів даних. Дане рішення пропонує повні набори алгоритмів і моделей, готових до негайного використання. В свою чергу SAP Predictive

Analytics теж не відстає і дозволяє здійснювати збір даних і будувати на основі цих даних прогностні моделі, що дозволяють прогнозувати перебіг подальших подій. Побудова прогностних моделей здійснюється шляхом пошуку закономірностей і взаємозв'язків в історичних і поточних даних. Головні особливості даного рішення це широкі можливості візуалізації даних; відстеження моделей протягом усього їх життєвого циклу; можливість працювати з різними обсягами даних.

Але так як SAP Predictive Analytics призначений в основному працювати з великими даними, а база даних зібрана для вирішення проблеми цією роботи має всього 500 записів, то результати видані SPSS Modeler ,будуть більш точними і наглядними.

Проведено аналіз методів для управління відгуками клієнтів на маркетингову компанію підприємств дозволяє зробити наступні висновки: в цій роботі виділено і розглянуто два основні методи. По-перше – методи дерева рішень або дерево прийняття рішень - це інструмент підтримки прийняття рішень, який використовує деревовидний графік або модель рішень і їх можливі наслідки, включаючи випадкові події, витрати ресурсів і корисність. Це один із способів відображення алгоритму, який містить тільки умовні оператори управління. По-друге – алгоритм списку рішень надає корисний спосіб представлення булевих функцій (елементи булевої множини 1 і 0 зазвичай інтерпретують як логічні значення «істинно» і «хибно», хоча в загальному випадку вони розглядаються як формальні символи, що не несуть певного сенсу). Так само, як кожна булева функція може бути представлена формулою DNF, ми побачимо, що кожна булева функція також може бути представлена списком рішень. Це уявлення іноді представлено більш компактно.

Список рішень має ряд достоїнств перед методом дерев рішень, таких як: вбудований порядок наборів правил, що допомагає зробити процес класифікації простіше; користувач може сам обирати, які правила залишити в списку, а які видалити; правила, які вводять не мають обмежень за кількістю

або змістом; всі можливі логічні пропозиції можуть бути використані в просторі вводу.

Підсумовуючи виявлені переваги алгоритму списку рішень на основі, яких втілення теми даної роботи спрощується, а результати будуть більш точними та наглядними можна зробити висновок, що цей метод є найбільш прийнятним для прогнозного моделювання реакції клієнтів на маркетингову компанію.

Далі в практичній частині роботи було побудовано дві моделі на основі обраного алгоритму, але з різними налаштуваннями (різні показники мінімального значення розміру сегментів та довірливий інтервал для нових умов). В кожній моделі було відібрано серед альтернативних варіанті кращі сегменти (ті відібрані групи клієнтів, на яких краще орієнтуватися під час створення маркетингових компаній підприємства). В першій моделі відібрано 5 сегментів, а в другій – 4. Але хоча процент виграшу в другій моделі трохи більше, обидві моделі можуть бути використані і впроваджені в процес створення маркетингових заходів.

А для поліпшення маркетингових стратегії компанії можна згрупувати результати обох моделей, а не концентруватися на одних рішеннях і отримати більше покриття серед своїх клієнтів.

Результати дослідження, отримані в дипломній роботі, знайшли апробацію в таких статтях як: «Огляд і аналіз інструментальних засобів бізнес-аналітики» та «дослідження методів моделювання поведінки клієнтів на основі IBM SPSS Modeler».

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Customer Analytics [Electronic resource]. – Access mode: <https://www.educba.com/course/customer-analytics-course/>
2. Customer and marketing analytics: Integrating multichannel data to gain actionable business insights for greater profitability [Electronic resource]. – Access mode: [https://www.mindtree.com/about/resources/customer-and-marketing-analytics-integrating-multichannel-data-gain actionable](https://www.mindtree.com/about/resources/customer-and-marketing-analytics-integrating-multichannel-data-gain-actionable)
3. Customer Analytics vs Web Analytics – Know the Importance [Electronic resource]. – Access mode: <https://www.educba.com/customer-analytics-vs-web-analytics/>
4. What is customer analytics? [Electronic resource]. – Access mode: <https://mixpanel.com/topics/what-is-customer-analytics/>
5. What is Customer Analysis and how can it Help You? [Electronic resource]. – Access mode: <https://www.accountingdepartment.com/blog/what-is-customer-analysis-how-does-it-benefit-business-analytics>
6. Системи аналітики якості клієнтського сервісу: аналіз даних з метою поліпшення [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.apexberg.ru/kb/articles/sistemy-analitiki-kachestva-klientskogo-servisa-analiz-dannykh-v-tselyakh-uluchsheniya/>
7. What is Customer Marketing Analytics? Challenges, Best Practices, and More [Electronic resource]. – Access mode: <https://www.ngdata.com/what-is-customer-marketing-analytics/>
8. Complete Overview of Customer Data Analytics [Electronic resource]. – Access mode: <https://www.scnsoft.com/blog/customer-data-analytics>
9. Deep Customer Analytics [Electronic resource]. – Access mode: <https://www.optimove.com/resources/learning-center/deep-customer-analytics>

10. The Evolution of Analytics with Data [Electronic resource]. – Access mode: <https://towardsdatascience.com/the-evolution-of-analytics-with-data-8b9908deadd7>
11. Analytics 3.0 [Electronic resource]. – Access mode: <https://hbr.org/2013/12/analytics-30>
12. Управління інноваціями [Електронний ресурс]. – Режим доступу до ресурсу: <https://hbr-russia.ru/innovatsii/upravlenie-innovatsiyami/a13151>
13. The Evolution of Analytics 1.0, 2.0, 3.0 [Electronic resource]. – Access mode: <https://www.linkedin.com/pulse/evolution-analytics-10-20-30-shankar-meganatha>
14. Крива Gartner для блокчейн [Електронний ресурс]. – Режим доступу до ресурсу: <https://medium.com/orbita-center/blockchain-digest-60-f0d754ff55f>
15. Gartner top 10 data and analytics technology trends for 2019 [Electronic resource]. – Access mode: <https://www.ecloudvalley.com/gartner-top-10-data-and-analytics-technology-trends-for-2019/>
16. Big Data: Are you ready for blast-off? [Electronic resource]. – Access mode: <https://www.bbc.com/news/business-26383058>
17. Big Data: 20 Mind-Boggling Facts Everyone Must Read [Electronic resource]. – Access mode: <https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#62d1414317b1>
18. Data Science vs. Big Data vs. Data Analytics [Electronic resource]. – Access mode: <https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>
19. Поняття Big Data і для чого воно потрібне [Електронний ресурс]. – Режим доступу до ресурсу: <http://p-s.kz/big-data/>
20. Big Data, Data Analytics, Data Analysis, Data Mining, Data Science & Machine Learning [Electronic resource]. – Access mode:

<http://scraping.pro/data-analytics-data-analysis-data-mining-data-science-machine-learning-big-data/#machinelearning>

21. Data Mining – інтелектуальний аналіз даних [Електронний ресурс]. – Режим доступу до ресурсу: https://blog.iteam.ru/data-mining-intellektualnyj-analiz-dannyh/#45__decision_trees

22. What is Data Mining (Predictive Analytics, Big Data) [Electronic resource]. – Access mode: <http://www.statsoft.com/textbook/data-mining-techniques>

23. Завдання, які вирішуються за допомогою IBM SPSS [Електронний ресурс]. – Режим доступу до ресурсу: http://www.predictivesolutions.ru/solutions/data_analysis.htm

24. Золотарюк А. В. Професійні комп'ютерні програми: застосування пакету SPSS в соціології / А. В. Золотарюк Москва.: Фінуніверситет, кафедра «Інформаційні технології», 2015р. – 51с.

25. Можливості пакету SPSS для роботи з даними [Електронний ресурс]. – Режим доступу до ресурсу: <https://studfiles.net/preview/5862667/page:6/>

26. IBM SPSS Modeler: Що він може зробити для вашого бізнесу [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.ibm.com/ru-ru/marketplace/spss-modeler#product-header-top>

27. Неляпенко С. Н. Концепція безпеки SAP Business Objects Enterprise / С.Н. Неляпенко Київ.: - 2010р.

28. Сайт програми IBM SPSS Modeler [Електронний ресурс]. – Режим доступу до ресурсу: <http://www.predictivesolutions.ru/software/modeler.htm>

29. SAP ERP Управління ресурсами підприємства // Сайт компанії SAP GmbH. [Електронний ресурс]. – Режим доступу до ресурсу: http://www.sap.com/cis/pdf/mySAP_ERP.pdf

30. What is Predictive Modeling? [Electronic resource]. – Access mode: <https://medium.com/better-programming/what-is-predictive-modeling-d918a4cf178e> -
31. Моделі дерев рішень [Електронний ресурс]. – Режим доступу до ресурсу: https://www.ibm.com/support/knowledgecenter/ru/SS3RA7_sub/modeler_mainhelp_client_ddita/clementine/nodes_treebuilding.html
32. Деревя рішень [Електронний ресурс]. – Режим доступу до ресурсу: https://blog.iteam.ru/data-mining-intellektualnyj-analiz-dannyh/#45__decision_trees
33. Зайцева Т. В. Програмна реалізація методу дерев рішень для вирішення завдань класифікації та прогнозування / Т. В. Зайцева, Н. В. Васіна, О. П. Пусна, Н. Н. Смородіна Тула.: Тульський державний університет, 2013 – 127с.
34. Дерево рішень [Електронний ресурс]. – Режим доступу до ресурсу: http://help.prognoz.com/ru/mergedProjects/Lib/06_datamining/lib_decisiontree.htm
35. Використання дерев рішень в задачах прогнозової аналітики [Електронний ресурс]. – Режим доступу до ресурсу: <http://www.prognoz.ru/blog/platform/decision-tree-in-predictive-analytics/>
36. Класифікація, дерева рішень і метод найближчих сусідів [Електронний ресурс]. – Режим доступу до ресурсу: <https://habr.com/ru/company/ods/blog/322534/#derevo-resheniy>
37. IBM SPSS Modeler 17 Algorithms Guide, 2015. – 519р.
38. Decision list [Electronic resource]. – Access mode: https://en.wikipedia.org/wiki/Decision_list
39. Decision lists [Electronic resource]. – Access mode: <http://www.cdam.lse.ac.uk/Reports/Files/cdam-2005-23.pdf>
40. Список рішень: Засіб перегляду списку рішень: Вкладка Альтернативи // Сайт компанії IBM [Електронний ресурс]. – Режим доступу

до ресурсу: https://www.ibm.com/support/knowledgecenter/ru/SS3RA7_18.1.1/modeler_mainhelp_client_ddita/clementine/dec_list_view_sessionresults_tab.html

41. IBM SPSS Modeler Help: Modeling Nodes: Decision List // IBM Website [Electronic resource]. – Access mode: https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.modeler.help/decisionlist_node_general.htm

42. Encyclopedia of Machine Learning [Electronic resource]. – Access mode: https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-30164-8_199

43. Технології / Офіційний сайт Тубного заводу [Електронний ресурс]. – Режим доступу до ресурсу: http://tube-plant.com/?page_id=526

44. Роль маркетингу в діяльності підприємства [Електронний ресурс]. – Режим доступу до ресурсу: <https://e-koncept.ru/2016/56771.htm>

45. Data mining techniques for customer relationship management [Electronic resource]. – Access mode: <https://www.sciencedirect.com/science/article/pii/S0160791X02000386>

46. Анкетне опитування як інструмент діагностики проблеми на підприємстві [Електронний ресурс]. – Режим доступу до ресурсу: https://knowledge.allbest.ru/marketing/3c0b65625a3bc79a5d43a89521316c37_0.html

47. Опитування для клієнтів ТОВ "Тубний Завод" [Електронний ресурс]. – Режим доступу до ресурсу: <https://docs.google.com/forms/d/e/1FAIpQLSecuf2PJRccUzTioOwJxynXIsdQyrtEglPyeOhnLUpNLxopg/viewform>

48. RFM-аналіз. RFM-аналіз від А до Я [Електронний ресурс]. – Режим доступу до ресурсу: <https://proektrt.ru/rfm-analiz-rfm-analiz-ot-a-do-ya.html>

49. RFM-аналіз на коліні (Excel) [Електронний ресурс]. – Режим доступу до ресурсу: <https://habr.com/ru/post/243219/>

50. SPSS Modeler 18.1.1: Рівні виміру // Сайт компанії IBM [Електронний ресурс]. – Режим доступу до ресурсу: https://www.ibm.com/support/knowledgecenter/ru/SS3RA7_18.1.1/modeler_mainhelp_client_ddita/clementine/typenode_kindsotypes.html

51. Опції моделі списку рішень // Сайт компанії IBM [Електронний ресурс]. – Режим доступу до ресурсу: https://www.ibm.com/support/knowledgecenter/ru/SS3RA7_sub/modeler_mainhelp_client_ddita/clementine/decisionlistnode_modeltab.html