

Хмельницький національний університет
Факультет програмування та комп'ютерних і телекомунікаційних систем
Кафедра комп'ютерних наук та інформаційних технологій

ДИПЛОМНА РОБОТА МАГІСТРА

на тему Прогнозування завантаженості ресторану з використанням штучного інтелекту

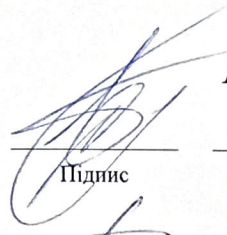
Галузь знань 12 – Інформаційні технології

Шифр і назва галузі знань

Спеціальність 122 – Комп'ютерні науки

Шифр і назва спеціальності

Виконав: студент 2 курсу, група КНМ-19-1



Підпис

Б.О.
Рибчинський

Ініціали, прізвище

Керівник: к.т.н., доцент кафедри КНІТ



Підпис

Р.О. Багрій

Ініціали, прізвище

Нормоконтроль: к.т.н., доцент кафедри КНІТ



Підпис

Р.О. Багрій

Ініціали, прізвище

До захисту допускаю:

Зав. кафедри КНІТ, д.т.н., професор



Підпис

О.В. Бармак

Ініціали, прізвище

7 12 2020 р.

Хмельницький 2020

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет програмування та комп'ютерних і телекомунікаційних систем

Кафедра комп'ютерних наук та інформаційних технологій

Освітній ступінь магістр

Галузь знань 12 – Інформаційні технології

Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук та інформаційних технологій

(підпис)

д.т.н., професор О.В.Бармак

« 7 » 12 2020 року

**ЗАВДАННЯ
НА ДИПЛОМНУ РОБОТУ МАГІСТРА**

1. Тема дипломної роботи магістра: «Прогнозування завантаженості ресторану з використанням штучного інтелекту»

2. Завдання видано студентці Рибчинський Богдан Олександрович

(прізвище, ім'я, по батькові)

3. Керівник роботи к.т.н., доцент Багрій Руслан Олександрович

(прізвище, ім'я, по батькові)

4. Затверджені наказом університету від « 9 » 9 2020 р. № 22

5. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Метою роботи – розробка методу прогнозування завантаженості ресторану та визначенню основних факторів впливу. Об'єктом дослідження є порівняння методів машинного навчання для предметної області, а саме прогнозування завантаженості ресторанів. Предметом дослідження є набір даних предметної області та методи розробки моделей машинного навчання.

Реферат

Дипломна робота магістра присвячена прогнозуванню завантаженості ресторану з використанням штучного інтелекту.

Актуальність теми. В магістерській роботі розроблена інформаційна система прогнозування завантаженості ресторану з застосування методів машинного навчання.

Збільшення попиту на туризм і конкуренція між напрямками сервісів були пов'язані з наслідками глобалізації. Сервіси конкурують на дедалі більш конкуруючих ринках та мало що відрізняє їх один від одного. Це змусило напрямки сервісів, що працюють в сфері обслуговування населення, зосередитися на особливостях їх місця розташування. Запуск нових та інноваційних продуктів та брендів має врахувати місцеві особливості для створення більш унікальної пропозиції. Однак, постала необхідність визначити фактори впливу на цільову функцію оптимізації.

Метою дослідження є розробка методу прогнозування завантаженості ресторану та визначенню основних факторів впливу.

Для досягнення зазначеної мети поставлені наступні задачі:

- показати, що застосування методів штучного інтелекту в ресторанній справі на сучасному рівні розвитку інформаційних технологій може в значній мірі підвищити бізнесові показники;
- провести дослідження ознак і класифікаторів застосованих в машинному навчанні;
- провести порівняння моделей машинного навчання та їх вплив на цільову функцію.

При цьому передбачається розв'язок таких підзадач, як

- попередня обробка даних та їх очищення;
- побудова та вибір факторів впливу на ефективність обраних моделей;
- розробка інформаційної системи прогнозування завантаженості ресторану;

- вибір моделей, виділення ознак і застосування методів машинного навчання;
- тестування методів на основі вибраних підходів;
- програмна реалізація системи прогнозування.

Об'єктом дослідження є процес машинного навчання для прогнозування завантаженості ресторанів.

Предметом дослідження є набір даних та методи розробки моделей машинного навчання.

Дослідження присвячено пошуку взаємозв'язків між важливістю і продуктивністю (задоволенням) місцевої кухні, а також вивчити вплив досвіду їдальні на задоволення клієнтів. Що стосується важливих факторів для їдальні, найбільш важливим фактором для клієнтів є продовольчі культурні аспекти. Зокрема, туристи очікують, що місцева їжа буде відображати унікальні культурні аспекти для приготування їжі, презентації та зовнішнього вигляду. Дотримуючись культурних аспектів, соціальних аспектів, які є важливими факторами та інші аспекти поведінки клієнтів, є наступним за важливістю фактором для клієнтів. Як третій важливий фактор, якість продуктів харчування та дотримувалася соціальних аспектів. Нарешті, фізичні властивості навколишнього середовища оцінюється як найменш важливий фактор для їдальні досвід. Однак ці оцінки базуються на досвіді, але аналіз даних може вказати на інші важливі спекти, які можуть вплинути на ефективність сервісу. Набір даних і є предметом дослідження для визначення.

Достовірність результатів забезпечується експериментальних дослідження ефективності застосовуваних підходів.

Рішення має теоретичний ландшафт, так як воно проявляється крізь призму безсимптомних рамок та досить просте: існує унікальне допустиме правило порогу, і, крім того, унікальне допустиме правило допуску, для значення. Це відрізняється від ситуації, з якою стикаються, наприклад, при оцінці складних систем, якою є система сервісу. Причина - крайня простота моделі, яку створює людина. Наприклад, замінили значення одних даних, які є випадковими для скінченного розміру матриці, з їх майже граничними межами, а також нехтували

випадковими коливаннями навколо цих меж. Ці коливання не завжди можна зрозуміти. Однак, включно з цими умовами другого порядку в безсимптомних розподілах, може досягти поліпшення другого порядку в оцінці середньої помилки над рекомендованим оптимальним порогом обмеження.

Крім важливості, респонденти також оцінювали своє задоволення кожним фактором. Як показує аналіз, заклади харчування демонструють високі показники якості продуктів харчування та продовольчих культурних аспектів. Однак якість продуктів харчування не оцінюється як така важлива клієнтом, як продовольчі культурні аспекти. Отже, продовольчі культурні аспекти являють собою точку диференціації для місцевих ресторанів. Тобто, як цільові організації, так і заклади харчування повинні зосередитися на культурних аспектах харчування та позиціонувати місцеву кухню, підкреслюючи унікальні аспекти харчування. Інша важлива оцінка цього дослідження стосується соціальних аспектів. Незважаючи на те, що респонденти вважають соціальні аспекти важливими, рівень задоволеності був низьким для цього фактора. Цей результат передбачає, що заклади харчування повинні більше зосередитися на службовій поведінці співробітників. Нарешті було встановлено що фізичне середовище, має низький пріоритет для цього виду сервісу.

Асимптотичні рамки, розглянуті тут, є найпростішою нетривіальною моделлю для матриці ознак. Це дозволяє забезпечити отримання необхідної точності. Нефундаментальні елементи матриці зменшують білий шум, який є структурними частинами складних моделей. Для прикладу, середня помилка будь-якого показника на основі зниження особливого значення містить компонент шуму в одиниці даних, і цей компонент визначає фундаментальну нижню межу. Результати обрахунків, розраховані в цій моделі, які не прикріплені до конкретного припущення про ранг (наприклад, константи, які визначають мінімакс) залишаються по суті правильними в більш складних моделях, що і визначає достовірність отриманих розрахункових даних.

Практична значимість дослідження полягає в тому, що запропоновані методи та підходи можна застосовувати в практичному руслі, а також в порівняльних сферах із предметною областю, тобто системах обслуговування та сервісів.

Утримання лояльних клієнтів дозволяє збільшити трафік. В певні періоди роботи ресторану може здатися, що розвиток сервісу є недостатній, але, використовуючи обробку даних про відвідування разом з даними про клієнтів, можна почати контролювати і збільшувати трафік та будувати більш ефективний сервіс з клієнтами. Крім того для прийняття оптимальних рішень необхідно використовувати засоби штучного інтелекту.

Машинне навчання має багато застосувань в ресторанах від допомоги кухарям на кухнях до прогнозування майбутніх продажів, що, в свою чергу, допомагає опрацювати вимоги до управління робочою силою та інвентаризацією.

В кінцевому рахунку, це може допомогти ресторанам працювати більш ефективно, наприклад, допомагаючи скоротити 600 мільйонів тонн продовольства даремно витраченого ресторанами по всьому світу і дозволяючи рестораторам зосередитися на тому, де вони можуть додати найбільшу цінність в їхній сервіс.

Апробація дипломної роботи

Основні положення і результати роботи опубліковані в збірнику наукових праць – Рибчинський Б. О. Прогнозування завантаженості ресторану з використанням штучного інтелекту / Б. О. Рибчинський, В. В. Доброловський, В. Ю. Медведчук // Збірник наукових праць за матеріалами Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук - 2020» Хмельницький, 2020, – С.300-302.

Структура та обсяг роботи. Дипломна робота магістра складається з завдання, реферату, змісту, вступу, 4 розділів, висновків, переліку посилань із 17 найменувань та додатків. Загальний обсяг дипломної роботи магістра становить 77 сторінок, з них 71 сторінка основного тексту та 2 сторінки додатків. в роботі наведено 19 рисунків.

Ключові слова: система прогнозування, машинне навчання.

Зміст

Вступ	9
Розділ 1	14
Аналіз сучасного стану предметної області	14
1.1 Опис предметної області	14
1.2 Дії для підвищення ефективності сервісу	14
1.3 Машинне навчання для прогнозування	19
1.4 Використання машинного навчання в предметній області	21
1.5 Постановка задачі	25
Висновок до розділу 1	26
Розділ 2	28
Аналіз моделі машинного навчання	28
2.1 Елементи контрольованого навчання	28
2.2 Впровадження загального принципу	30
2.3 Адитивна підготовка	32
2.5 Складність моделі	34
Висновки до розділу 2	37
Розділ 3	38
Розробки системи прогнозування	38
3.1 Обробка даних	38
3.2 Системні параметри	40
3.3 Показник оцінювання	42
3.4 Відвідувач і бронювання	50
Висновки до розділу 3	57
Розділ 4	58
Дослідження ефективності запропонованих підходів	58
4.1 Загальні спостереження з аналізу даних	58
4.2 Вибір функцій за допомогою рекурсивного усунення функцій	65
4.3 Пояснення моделі	66
4.6 Порівняння моделей	69

Висновки до розділу 4	70
Загальні висновки	72
Перелік посилань	74
Додатки	

Вступ

В останні роки програма продовольства отримала визнання урядами, бізнесом і вченими як невід'ємна частина туристичного продукту, і як засіб диференціації для напрямків. Є багато переваг, які повинні пов'язують продовольство і туризм для всіх зацікавлених сторін. Місцева їжа є життєво важливим елементом, який може допомогти створити відчуття "місця" і підвищити привабливість призначення. Місцеві продукти додають автентичності туристичному досвіду і забезпечує мотивацію для відвідувачів приїхати в місце. Туристи можуть навіть спокуситися довше залишатися в одному місці через наявність харчових продуктів і пов'язаних з ними заходів. Туризм забезпечує додатковий торговий центр для виробників продуктів харчування та витрат, пов'язаних з туризмом, на місцеві харчові продукти, що допомагає стимулювати та поживити місцеві економіки. Все частіше продовольство використовується в ініціативах розвитку для зміцнення туристичних напрямків, а також для створення тісного зв'язку як для харчової промисловості, так і для туристичної індустрії.

Їжа є вагомою складовою загальних витрат туристів. Є таке твердження, що "майже 100% туристів обідають під час подорожі, а їжа та напої стабільно займають перше місце за витратами відвідувачів".

Актуальність теми. В магістерській роботі розроблений і реалізований метод отримання інформації на підставі аналізу та застосування методів машинного навчання з метою прогнозування завантаженості ресторану у різні періоди функціонування.

Збільшення попиту на туризм і конкуренція між напрямками сервісів були пов'язані з наслідками глобалізації. Напрямки конкурують на дедалі більш конкуруючих ринках, мало що відрізняє їх один від одного. Це змусило напрямки сервісів обслуговування населення зосередитися на відмінності місця розташування, використовуючи місцеві атрибути при запуску інноваційних, нових продуктів і брендів, що може допомогти створити більш унікальну пропозицію продажу. Однак необхідно визначити фактори впливу на цільову функцію оптимізації.

Метою дослідження є розробка методу прогнозування завантаженості ресторану та визначенню основних факторів впливу.

Для досягнення зазначеної мети поставлені наступні задачі:

- показати, що застосування методів штучного інтелекту в ресторанній справі на сучасному рівні розвитку інформаційних технологій може в значній мірі підвищити показники бізнеса;
- провести дослідження ознак і класифікаторів застосування в машинному навчанні;
- провести порівняння моделей машинного навчання та їх вплив на цільову функцію.

При цьому передбачається розв'язок таких підзадач, як

- попередня обробка даних та їх очищення;
- побудова та вибір факторів впливу на ефективність обраних моделей;
- розробка інформаційної системи прогнозування завантаженості ресторану;
- вибір моделей, виділення ознак і застосування методів машинного навчання;
- тестування методів на основі вибраних підходів;
- програмна реалізація системи прогнозування.

Об'єктом дослідження є порівняння методів машинного навчання для предметної області, а саме прогнозування завантаженості ресторанів.

Предметом дослідження є набір даних предметної області та методи розробки моделей машинного навчання.

Дослідження присвячено пошуку взаємозв'язків між важливістю і продуктивністю (задоволенням) місцевої кухні, а також вивчити вплив досвіду їдальні на задоволення клієнтів. Що стосується важливих факторів для їдальні, найбільш важливим фактором для клієнтів є продовольчі культурні аспекти. Зокрема, туристи очікують, що місцева їжа буде відображати унікальні культурні аспекти для приготування їжі, презентації та зовнішнього вигляду. Дотримуючись культурних аспектів, соціальних аспектів, які є важливими факторами та інші аспекти по-

ведінки клієнтів, є наступним за важливістю фактором для клієнтів. Як третій важливий фактор, якість продуктів харчування та дотримувалася соціальних аспектів. Нарешті, фізичні властивості навколишнього середовища оцінюється як найменш важливий фактор для їдальні досвід. Однак ці оцінки базуються на досвіді, але аналіз даних може вказати на інші важливі спекти, які можуть вплинути на ефективність сервісу. Набір даних і є предметом дослідження для визначення

Достовірність результатів забезпечується експериментальних дослідження ефективності застосовуваних підходів.

Рішення має теоретичний ландшафт, так як воно проявляється крізь призму безсимптомних рамок та досить просте: існує унікальне допустиме правило порогу, і, крім того, унікальне допустиме правило допуску, для значення. Це відрізняється від ситуації, з якою стикаються, наприклад, при оцінці складних систем, якою є система сервісу. Причина - крайня простота моделі, яку створює людина. Наприклад, замінили значення одних даних, які є випадковими для скінченного розміру матриці, з їх майже граничними межами, а також нехтували випадковими коливаннями навколо цих меж. Ці коливання не завжди можна зрозуміти. Однак, включно з цими умовами другого порядку в безсимптомних розподілах, може досягти поліпшення другого порядку в оцінці середньої помилки над рекомендованим оптимальним порогом обмеження.

Крім важливості, респонденти також оцінювали своє задоволення кожним фактором. Як показує аналіз, заклади харчування демонструють високі показники якості продуктів харчування та продовольчих культурних аспектів. Однак якість продуктів харчування не оцінюється як така важлива клієнтом, як продовольчі культурні аспекти. Отже, продовольчі культурні аспекти являють собою точку диференціації для місцевих ресторанів. Тобто, як цільові організації, так і заклади харчування повинні зосередитися на культурних аспектах харчування та позиціонувати місцеву кухню, підкреслюючи унікальні аспекти харчування. Інша важлива оцінка цього дослідження стосується соціальних аспектів. Незважаючи на те, що респонденти вважають соціальні аспекти важливими, рівень задоволеності був низьким для цього фактора. Цей результат передбачає, що заклади харчування повинні більше зосередитися на службовій поведінці співробітників. Нарешті було

встановлено що фізичне середовище, має низький пріоритет для цього виду сервісу.

Асимптотичні рамки, розглянуті тут, є найпростішою нетривіальною моделлю для матриці ознак. Це дозволяє забезпечити отримання необхідної точності. Нефундаментальні елементи матриці зменшують білий шум, який є структурними частинами складних моделей. Для прикладу, середня помилка будь-якого показника на основі зниження особливого значення містить компонент шуму в одиниці даних, і цей компонент визначає фундаментальну нижню межу. Результати обрахунків, розраховані в цій моделі, які не прикріплені до конкретного припущення про ранг (наприклад, константи, які визначають мінімакс) залишаються по суті правильними в більш складних моделях, що і визначає достовірність отриманих розрахункових даних.

Практична значимість дослідження полягає в тому, що запропоновані методи та підходи можна застосовувати в практичному руслі, а також в порівняльних сферах із предметною областю, тобто системах обслуговування та сервісів.

Утримання лояльних клієнтів дозволяє збільшити трафік. В певні періоди роботи ресторану може здатися, що розвиток сервісу є недостатній, але, використовуючи обробку даних про відвідування разом з даними про клієнтів, можна почати контролювати і збільшувати трафік та будувати більш ефективний сервіс з клієнтами. Крім того для прийняття оптимальних рішень необхідно використовувати засоби штучного інтелекту.

Машинне навчання має багато застосувань в ресторанах від допомоги кухарям на кухнях до прогнозування майбутніх продажів, що, в свою чергу, допомагає опрацювати вимоги до управління робочою силою та інвентаризацією.

В кінцевому рахунку, це може допомогти ресторанам працювати більш ефективно, наприклад, допомагаючи скоротити 600 мільйонів тонн продовольства даремно витрачаємого ресторанами по всьому світу і дозволяючи рестораторам зосередитися на тому, де вони можуть додати найбільшу цінність в їхній сервіс.

Розділ 1

Аналіз сучасного стану предметної області

1.1 Опис предметної області

Запуск процвітаючого місцевого ресторану не завжди такий простий, як здається на перший погляд. Часто з'являються несподівані неприємності, які можуть зашкодити бізнесу.

Одним з поширених складних становищ є те, що ресторани повинні знати, скільки клієнтів очікувати кожен день, щоб ефективно придбати інгредієнти і запланувати співробітників. Цей прогноз нелегко зробити, тому що багато непередбачуваних факторів впливають на відвідуваність ресторану, як погода і місцева конкуренція. Це ще важче для нових ресторанів з невеликими історичними даними.

Як для прикладу використання даних для підвищення ефективності функціонування галузі можна навести: Recruit Holdings створив набір даних, які могли б зробити можливим автоматизоване прогнозування майбутніх клієнтів. Зокрема, Recruit Holdings володіє Hot Pepper Gourmet (сервіс огляду ресторану), AirREGI (ресторанна точка продажу) і Restaurant Board (програмне забезпечення для управління журналом бронювання) [1]. Це приклад комплексного підходу для визначення оптимальної стратегії розвитку базуючись на наборах даних.

1.2 Дії для підвищення ефективності сервісу

Вибираючи улюблений ресторан, клієнти шукають привабливу атмосферу, смачну їжу і відмінний сервіс. Хоча ці якості у ресторані мають вирішальне значення для підтримки постійного потоку клієнтів, дані є обов'язковими, коли справа доходить до збільшення трафіку і рентабельності інвестицій.

Як можна переконатися, що години спаду в сервісі існують, і потрібно отримати тих, хто готовий тільки відвідувати ресторан у вихідні дні, і тих які готові прийти протягом тижня. Окреслимо способи збільшення трафіку до ресторану.

1. Знати своїх клієнтів.

Можна знати деяких своїх клієнтів, але знання кожного з них і розуміння, що вони хочуть, має вирішальне значення для мотивації їх повернутися знову. Знаючи, чому клієнти приходять до ресторану, що вони думають про бізнес, і що їх купівельна поведінка, може вплинути на трафік.

Знаючи, що індивідуальні звички кожного клієнта можуть допомогти передбачити їх майбутні покупки. Це також може допомогти надати їм акції або стимули, щоб заохотити їх приходити до ресторану частіше [3]. Платформи Dynamics 365, такі як Thanx, надають дані клієнтів і допомагають інтерпретувати їх поведінку, щоб передбачити, що може бути далі. Ці дані можна використовувати для вимірювання середнього розміру перевірки, частоти відвідування, вартості споживання клієнтами та багато іншого.

Використання опитувань NPS (Net Promoter Score) для вимірювання задоволеності клієнтів - це спосіб зрозуміти, що відчувають клієнти, коли вони знаходяться у ресторан. Відгук дає уявлення про те, що робиться правильно, і що можна поліпшити. Дані показують, що навіть щасливі клієнти, які мають кілька негативних вражень, будуть відвідувати рідше, тому аналіз відгуків є ефективним способом боротьби із зменшенням трафіку та навіть збільшення доходу, запрошуючи їх повернутися з цільовою пропозицією [4].

Розуміння клієнтів і знання про те, які лояльні, і які збираються повернутись допомагає утримувати клієнтів і заохочувати інших клієнтів повернутися. Коли клієнти щасливі, вони, швидше за все, посилаються на своїх друзів, і трафік може зрости [4].

2. Використовувати цільові маркетингові кампанії.

Як тільки зрозуміло, хто клієнти і що вони хочуть, цільові маркетингові кампанії можуть впливати на трафік і поведінку. Наприклад, якщо відомо, що клієнт відвідує ресторан чотири рази на місяць у вихідні дні, можна відправити акцію

за знижками на фрикадельки понеділок [6]. Ця пропозиція може заохочувати відвідати додатковий час і керувати додатковим рухом протягом більш “повільного” дня.

Використання технологічної платформи, яка допомагає вимірювати та автоматизувати цей процес, гарантує, що клієнти націлені на персоналізовані кампанії та рекламні акції [7]. Після використання Thank для цільових кампаній, Hotze і Wellness Center побачили приріст в розмірі 33% на своїх спеціальних кампаніях, тому що змогли сегментувати своїх клієнтів і відправляти людям відповідні акції.

Коли клієнти підходять в потрібний час відповідно до їх поведінки, є більше шансів, що вони повернуться до ресторану.

3. Створення програми лояльності та переходи до клієнтів.

Лояльні клієнти хочуть відчувати себе цінними. Чим більше цінується клієнт, тим більше він хочуть повернутися і дати ресторану розвиватись.

Визначити, що для означає лояльний клієнт. Поглянути на те, як часто клієнт відвідує ресторан і скільки вони витрачають під час кожного візиту. Як тільки визначиться, хто найбільш лояльні клієнти, дати їм переваги і пільги, щоб заохотити їх відвідати більше і привести своїх друзів.

Деякі власники ресторанів відчувають вагання, коли мова йде про реалізацію програми лояльності клієнтів. Вони бояться пропонувати занадто багато і це негативно впливають на їх дохід і рентабельність інвестицій. Якщо зробили правильний шлях, використання цільових кампаній і поведінки клієнтів гарантує, що на зв'язку з потрібними клієнтами в потрібний час, і що програма підтримує позитивну рентабельність інвестицій.

Також можна відправити своїм клієнтам стимули, щоб привести більше людей у ресторан. Наприклад, можна повідомити своїм клієнтам, що для кожних двох людей, яких вони присилають в ресторан, вони отримають знижку від своєї наступної покупки. Побудова міцних відносин з найбільш лояльними клієнтами збільшує трафік, тому що, коли лояльні до них, вони захочуть бути лояльними до вас. Відправка їм пропозицій, коли вони незадоволені або нагородження їх за посилення друзів. Кроки, які можна зробити, щоб побудувати сприйняття бренду і поліпшити відносини у з клієнтами.

4. Приймати переконливі події.

Участь у місцевій громаді може отримати відгук про бізнес і заохочувати нових клієнтів відвідати. Зв'язатися з місцевими некомерційними організаціями, які шукають місце для свого наступного збору коштів, або зв'язатися з підприємствами, які шукають простір для групи місцевих підприємців [8].

Незалежно від того, чи співпрацюється з організацією або створюється спеціальна подія, побудова відносин з місцевою громадою збільшить трафік, залучення нових клієнтів.

Творчі підходи з подіями дозволяють переконатись в перевагах відносно конкурентів. Використання традиційних свят або подій може бути ефективним способом виділитися з натовпу [9].

5. Створити соціальні мережі.

Поширення слова про ресторан набагато простіше, коли є великі соціальні мережі. Соціальні медіа є одним з кращих способів побудувати відносини з клієнтами [11].

Соціальні повідомлення та події легко обмінюються, і вони забезпечують спосіб для клієнтів не тільки взаємодіяти з сервісом, але і один з одним. У міру зростання послідовників, трафік збільшиться. Чим більше людей, які знають про сервіс, тим більше клієнтів.

Наприклад, використати просування на Facebook, щоб керувати трафіком клієнтів.

Використовувати творчий контент, унікальні акції та конкурси в соціальних мережах для спілкування зі спільнотою. Навіть ті, хто не виграв конкурс, відвідають сторінку в соціальних мережах і, можливо, стежать за бізнесом і відвідають ресторан.

6. Використовувати творчі вивіски.

Хоча технології відіграють велику роль у збільшенні трафіку, відомо, що фізичний простір також привертає людей [12].

Наприклад, якщо магазин знаходиться на жвавому шосе, де люди їдять тільки і ніколи не ходять, і є великий знак, то водії можуть читати з далекої відстані. Мета полягає в тому, щоб заохотити людей зупинитися, і, можливо,

ніколи не чули про ресторан раніше, поки вони не побачили табличку і хотіли спробувати.

Знаки є ефективним способом підвищення обізнаності та збільшення трафіку, привертаючи увагу людей.

7. Мати правильний персонал.

У ресторанному бізнесі обслуговування - це все. Можна використовувати всі найкращі маркетингові інструменти, але якщо клієнти не на одній хвилині з персоналом, вони можуть не повернутися. Багато ресторанам важко найняти і утримувати якісний персонал через високі показники обороту, але є кроки, які можна зробити, щоб переконатися, що клієнти завжди отримують найкращий сервіс.

Один з найкращих способів переконатися, що персонал лояльний до клієнтів. Подобається це помешкання, але ще не впевнені. Наприклад, якщо клієнт залишає позитивний відгук про один із серверів, переконайтеся, що вони знають про це.

Маючи позитивний персонал, який присвячений задоволеності клієнтів також можна допомогти збільшити трафік і переконатись, що клієнти відчують себе важливими.

8. Підтримувати зв'язок.

Після того, як клієнт покине ресторан, переконатись, що є зв'язок. Незалежно від того, чи робиться це за допомогою цільових кампаній або взаємодії в соціальних мережах, можна переконатися, чи залишається взаємодія із клієнтами.

Розуміння клієнтів за допомогою даних про поведінку клієнтів покаже, коли залишатися на зв'язку і як часто. Надсилання рекламних акцій, пропозицій, повідомлень і заходів може спонукати клієнтів повернутися і направити своїх друзів.

9. Довести ефективність кампанії в реальних доходах.

Протягом багатьох років продавці та маркетологи були змушені спробувати виміряти ефективність своїх зусиль за допомогою показників, таких як електронна розсилка, кліки. Але відомо, що доказом є дохід, який генерують кампанії. Thank надає на зручній для розуміння прикладній системі, яка не вимагає маніпуляцій з даними. Ні електронних таблиць, ні звітів. Просто легко переварити дані,

які доводять цінність кампаній у доході на підставі підходів машинного навчання. Ці всі підходи є джерелами даних і мобуть бути проаналізовані.

1.3 Машинне навчання для прогнозування

Штучний інтелект (AI) і машинне навчання (ML) завоювали міцну позицію в різних галузях завдяки своїй здатності впорядкувати операції, заощадити витрати і зменшити людську помилку. Машинне навчання змінило сектори, такі як охорона здоров'я, фінанси та роздрібна торгівля, і ресторанна індустрія не є винятком.

Багато ресторанів вже використовували потужність AI і ML. З кіосків, які аналізують моделі покупки для створення пропозицій чат-ботам, які керують бронюваннями, приймають замовлення та відповідають на запити, оператори ресторанів знайшли безліч застосувань для ML.

Але додатки AI і ML не обмежуються управлінням рестораном і обслуговуванням клієнтів - вони також можуть генерувати дохід. На основі наявних даних можна коригувати рішення щодо замовлення та змінювати робочий процес. Зважаючи на це, кілька способів, в яких машинне навчання може зробити прогнозування продажів більш впливовим [13].

Згідно з даними досліджень,отриманими Forbes, погода має прямий вплив на сприйняття клієнтом ресторанного досвіду. Незалежно від того, наскільки бездоганною є їжа та сервіс, негода все ще може призвести до поганих відгуків.

За допомогою машинного навчання зможете прогнозувати продажі на основі погодних умов і минулих показників продажів, що дозволяє відповідним чином коригуватися. Скажімо, продали більше страв на основі супу в дощові дні минулого року. Платформа прогнозування на рівні ML може вказувати на попередні дані та інформувати про пункти меню, які продали більше, разом з погодними умовами в той час. ML також говорить, про можливість очікування сплеску або зниження клієнтів. Таким чином, менеджери будуть мати інформацію, щоб запланувати персонал, який може вмістити прогнозовану кількість клієнтів за зміну [14].

Скорочення харчових відходів у ресторані не тільки економить гроші, це допомагає в збереженні навколишнього середовища. Дослідження, проведене Інститутом світових ресурсів, показало, що за кожен долар ресторан інвестує в рішення зі скорочення харчових відходів, вони економлять.

Машинне навчання може узгодити минулі дані про продажі та погодні умови, щоб обчислити, скільки запасів потрібно для задоволення споживчого попиту. Якщо в змозі тільки купити те, що потрібно, управління запасами виключає ризик надмірного залишку товарів, тим самим скорочуючи харчові відходи і накладні витрати.

Існує величезна помилка, що свята і значні події можуть призвести до сплеску прибутку. На жаль, це явище стосується не кожного ресторану. Хоча багато хто відчуває сплеск продажів під час певних свят, є деякі, в першу чергу висококласні установи, які бачать зниження доходів до 60 відсотків за даними [12].

Можна пом'якшити втрати, покладаючись на ML, щоб проектувати продажі, інвентаризацію та кадрові потреби під час сезонних свят та значних подій. Замість того, щоб покладатися на інтуїцію, щоб зробити припущення, AI і ML можуть витягнути з минулих даних про продажі і дати точні прогнози щодо того, скільки продуктів харчування та напоїв і кількість персоналу потрібно буде задовольнити попит.

Акції та спеціальні пропозиції часто використовуються ресторанами для збільшення трафіку і збільшення продажів. Але не всі промо-ідеї працюють. Навіть якщо у є ідея, яка показує перспективи, не можна гарантувати, що це принесе користь ресторану.

Щоб зрозуміти, чи варто проводити акцію, можна покладатися на дані, надані AI та ML. Платформи з ML будуть враховувати поведінку клієнта щодо покупки, щоб передбачити, чи будуть споживачі сприйнятливі до ініціативи чи ні.

Оператори ресторанів не повинні знижувати сподівання на AI і ML в статистиці продажів, особливо в поєднанні з існуючою технологією ресторану. Приймаючи прогнозування та прогнозне моделювання, можна планувати краще, підвищити ефективність і відкрити бізнес для інших можливостей.

1.4 Використання машинного навчання в предметній області

Ні один день не проходить без машинного навчання, яке згадується в новинах. Технологія працює в транспортних засобах або перемогах чемпіона світу з шахів.

Опишено аспекти машинного навчання в ресторанній індустрії. По-перше, ML вже допомагає ресторанам багато в чому:

- Прогнозування продажів.
- Виявлення крадіжки персоналу.

Відповідність клієнтів ресторанам на основі попередньо обраного профілю смаку:

- Навчання роботів, як готувати.
- Визначення вмісту їжі на молекулярному рівні.

Надамо коротке визначення машинного навчання та обговоримо додаткові способи, якими воно може допомогти ресторанній індустрії, зокрема, прогножуючи продажі точніше та навчаючи роботів готувати [7].

Машинне навчання описує здатність алгоритмів вчитися при обробці даних. Тим більше даних вони обробляють, тим більше вони навчаються, і тим точніше вони відповідають на запитання.

Наприклад на зображено об'єкт автомобіль, пішохід або дерево. Коли алгоритми намагаються відповісти на цю проблему, вони оцінюють рішення з урахуванням декількох параметрів. У наведеному вище прикладі алгоритм може вивчити зображення та оцінити тип об'єкта на зображенні з урахуванням форми, кольору та руху.

Природно, що для більшості алгоритмів початкові прогнози будуть неправильними. Алгоритм дізнається, тестувавши те, що передбачалося проти того, що дійсно сталося - цей тип машинного навчання називається контрольованим навчанням. Для кожного раунду алгоритм змінює внутрішні параметри або частини його структури на основі початкових даних і намагається знову. Цей процес три-

ває, включає в себе відкидання змін, які знижують точність алгоритму і збереження змін, які підвищують точність. Алгоритм, як кажуть, "дізнався" коли нові зображення представлені і точно класифікуються.

Покращення точності прогнозу продажів за допомогою машинного навчання. Основні алгоритми, які не використовують машинне навчання також існують. Основні алгоритми можуть прогнозувати майбутні продажі на основі простих параметрів, таких як продажі за минулий тиждень і рік, а також з урахуванням свят, погоди тощо.

Як приклад, для прогнозування завтрашніх продажів основним алгоритмом буде:

- Обчислити середнє значення між минулорічними продажами того ж дня та минулого тижня продажів того ж дня.
- Збільшити продажі на 20%, якщо це святковий день.
- Збільшити продажі на 20%, якщо це сонячний день.

Однак всі ці параметри не мають однакового впливу на ресторани. Наприклад, погода може бути важливим фактором для ресторану морозива поруч з пляжем, але невеликий фактор для піцерії в торговому центрі. Базовий алгоритм не може персоналізувати прогноз і зрозуміти, що кожен ресторан відрізняється [10].

Алгоритми використання машинного навчання. Алгоритми можуть допомогти підвищити точність прогнозу з плином часу персоналізації кожного параметра для кожного місця розташування, що означає навчання під час обробки даних параметр, який має більший вплив на продажі для певного місця.

Досвід показав найбільш релевантні параметри прогнозування продажів ресторанів:

- Всі попередні дані (наприклад, сезонні дані, щотижневі дані, тенденція зростання).
- Погода (температура, опади, сонячні години).
- Святкові дні (тобто святкові дні, піврічний, День матері).
- Події, різного гатунку (футбольні матчі, театральна постановка).

Алгоритм поєднує в собі вищезгадані параметри для прогнозування майбутніх продажів. Тоді оцінюється точність, дивлячись на фактичні результати.

Для кожного нового прогнозу алгоритми змінюють вплив кожного параметра на основі початкових варіантів і намагаються ще раз. Як зазначалося раніше, алгоритм зберігає зміни, які підвищують точність при відкиданні змін, які знижують точність. Тим більше даних обробляємо, тим більше дізнаємося, і тим точніше з нашим наступним прогнозом.

MAPE (*Mean Absolute Percent Error*, середній абсолютний відсоток помилки) – показник того, що програмне забезпечення збільшить точність з плином часу.

Використання машинного навчання демонструє до 50% поліпшення точності прогнозування та базових алгоритмів прогнозування та до 30% проти типових прогнозів менеджерів.

Прогнозування може прогнозувати попит погодинно, що економить витрати на персонал ресторанів, оптимізуючи графіки праці. Також зменшує втрату запасів, готуючи оптимальний рівень інвентаризації.

Можна навчити роботів, як готувати за допомогою машинного навчання. Каліфорнія Miso Robotics - це стартап, орієнтований на роботизовані рішення на основі AI для кухонного середовища. Флагманський кухонний помічник компанії AI "Flippy" поєднує в собі 3D, термічний і регулярний зір, щоб допомогти з грилем, смаженням, приготуванням [15].

У цьому прикладі машинне навчання виристористористує для навчання робота:

- Розпізнати стейк, приготований стейк і сирний шматочок.
- Коли перевернути стейк і коли перестати його смажити.

Для процесу розпізнавання пояснюємо нижче, як кольори та фігури є параметрами, які використовуються для розпізнавання об'єкта з іншого. Кожен раз алгоритм змінює внутрішні параметри або частини його структури на основі початкових падінь (спалений або не приготований стейк) і намагається знову. Зрештою, алгоритм зрозуміє, що коричневий стейк - це приготований стейк.

Приклад того, як роботи можуть аналізувати гриль. Щоб зрозуміти, чи не настав час перевернути і якщо стейк приготований, алгоритм може вивчити його різні датчики і передбачити відповідь за цими параметрами.

У цьому прикладі найбільш релевантними параметрами є:

- Температура гриля.
- Температура стейку.
- Колір і форма стейка.
- Час, витрачений на гриль.

Потім під час контрольованого навчання людина оцінює точність, дивлячись на стейк.

Для кожного нового стейка алгоритми змінюють вплив кожного параметра на основі початкових одиниць і намагаються ще раз. Тим більше даних робит обробляє, тим більше він вчиться, і тим краще стейки готуються. Як можна побачити штучний інтелект має досить широкі межі практичного застосування, в тому числі в ресторанному бізнесі.

1.5 Постановка задачі

Для досягнення зазначеної мети поставлені наступні задачі:

- показати, що застосування методів штучного інтелекту в ресторанній справі на сучасному рівні розвитку інформаційних технологій може в значній мірі підвищити бізнесові показники;
- провести дослідження ознак і класифікаторів застосованих в машинному навчанні;
- провести порівняння моделей машинного навчання та їх вплив на цільову функцію.

При цьому передбачається розв'язок таких підзадач, як

- попередня обробка даних та їх очищення;
- побудова та вибір факторів впливу на ефективність вибраних моделей;
- розробка інформаційної системи прогнозування завантаженості ресторану;
- вибір моделей, виділення ознак і застосування методів машинного навчання;

- тестування методів на основі вибраних підходів;
- програмна реалізація системи прогнозування.

Висновок до розділу 1

Машинне навчання має багато застосувань в ресторанах від допомоги кухарям на кухнях до прогнозування майбутніх продажів, що, в свою чергу, допомагає опрацювати вимоги до управління робочою силою та інвентаризацією.

В кінцевому рахунку, це може допомогти ресторанам працювати більш ефективно, наприклад, допомагаючи скоротити витрати продовольства даремно ресторанами по всьому світу і дозволяючи рестораторам зосередитися на тому, де вони можуть додати найбільшу цінність.

У цій роботі будуть використовувати дані бронювання та відвідування, щоб передбачити загальну кількість відвідувачів ресторану на майбутні дати. Ця інформація допоможе ресторанам бути набагато ефективнішими і дозволить їм зосередитися на створенні приємного досвіду відвідування для своїх клієнтів.

Розділ 2

Аналіз моделі машинного навчання

2.1 Елементи контрольованого навчання

Моделі використовуються для контрольованих проблем навчання, де використовуємо навчальні дані (з декількома функціями) для прогнозування цільової змінної.

Перш ніж практично використаємо математичні моделі, оцінимо основні елементи у контрольованому навчанні.

Модель в контрольованому навчанні зазвичай відноситься до математичної структури, за допомогою якої передбачення робляться з вхідного сигналу.

Поширеним прикладом є *лінійна модель*, де передбачення надається як лінійне поєднання зважених вхідних функцій. Значення передбачення може мати різні інтерпретації, залежно від завдання, тобто регресії або класифікації. Наприклад, дані можуть бути логістично перетворені, щоб отримати ймовірність позитивного класу в логістичній регресії, і також можуть бути використані в якості рейтингового балу, коли необхідно оцінити результати.

$$y_i, x_i, y_i = \sum_j \theta_j x_i \quad (2.1)$$

Параметри є не визначеною частиною, яку потрібно вивчити з даних. У лінійних регресійних проблемах параметри є коефіцієнтами. Ці елементи будемо використовувати для позначення параметрів (в моделі багато параметрів).

Об'єктивна функція: Втрата навчання та регулярність. З розсудливим вибором для моделі, можна визначити різні завдання, такі як регресія, класифікація, і рейтинг.

Завдання навчання моделі полягає в пошуку найкращих параметрів, які найкраще відповідають навчальним даним і міткам.

Для того, щоб навчити модель, потрібно визначити об'єктивну функцію, щоб виміряти, наскільки добре модель відповідає навчальним даним.

Важливою характеристикою об'єктивних функцій є те, що вони складаються з двох частин: втрата навчання і термін узагальнення:

$$\text{obj}(\theta) = L(\theta) + \Omega(\theta), \quad (2.2)$$

де $L(\theta)$ є функція втрати навчання, і $\Omega(\theta)$ є терміном узагальнення.

Втрата навчання вимірює, наскільки передбачуваною є модель щодо навчальних даних. Поширеним вибором є середня квадратична помилка.

$$L(\theta) = \sum_i (y_i - \hat{y}_i)^2 \quad (2.3)$$

Іншою часто використовуваною функцією втрати є логістична втрата, яка буде використовуватися для логістичної регресії:

$$L(\theta) = \sum_i [y_i \ln(1 + e^{-y_i}) + (1 - y_i) \ln(1 + e^{y_i})] \quad (2.4)$$

Термін узагальнення - це те, що розробники зазвичай забувають додати. Термін узагальнення контролює складність моделі, що допомагає уникнути пере-навчання.

Це звучить трохи абстрактно, тому розглянемо наступну проблему на наступному рисунку. Запропоновано візуально помістити крок функції з огляду на вхідні точки даних у верхньому лівому куті зображення. Визначимо яке рішення серед них найкраще підходить.

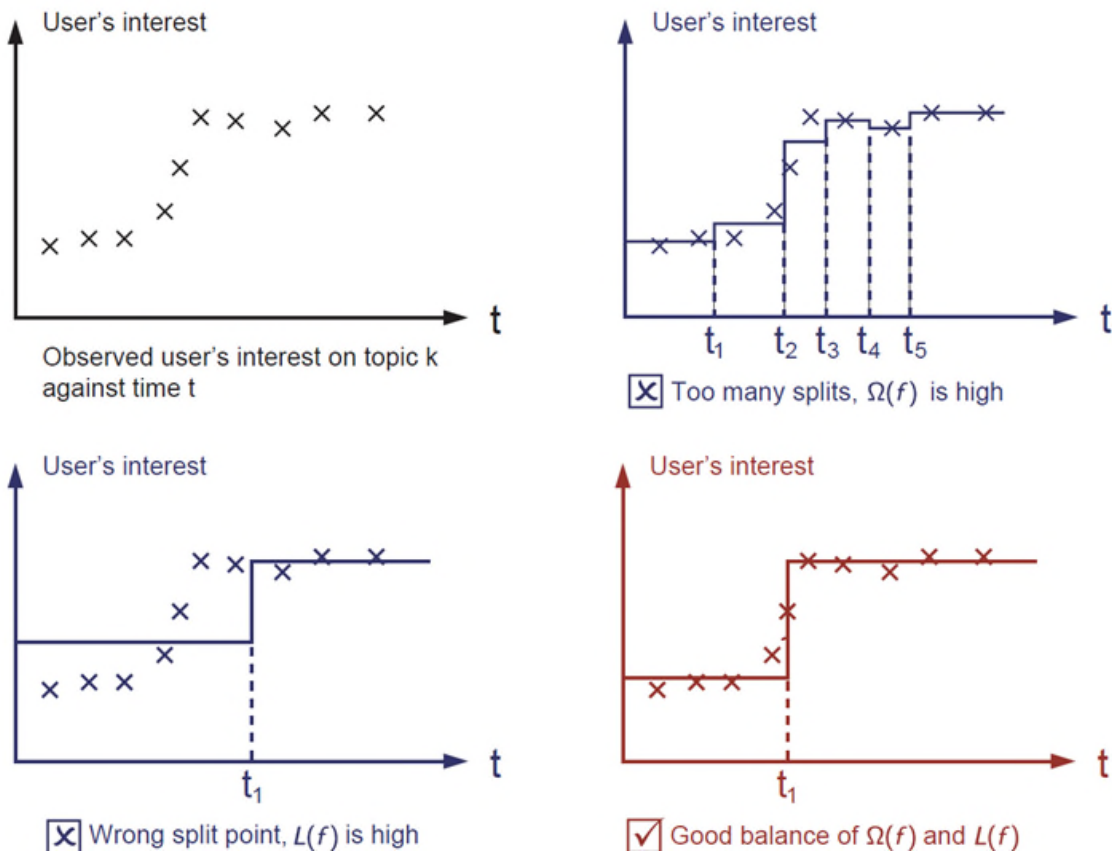


Рисунок 2.1 – Узагальнене рішення [14]

Загальний принцип полягає в тому, що хочемо як *простої*, так і *прогностичної* моделі. Компроміс між ними також називають упередженим компромісом у машинного навчання.

2.2 Впровадження загального принципу

Елементи, введені вище, утворюють основні елементи контрольованого навчання, і вони є природними будівельними блоками інструментів машинного навчання. Наприклад, повинні бути в змозі описати відмінності і спільності між градієнтом збільшених дерев і випадкових лісів. Розуміння процесу формалізованим способом також допомагає зрозуміти мету, яку вивчаємо, і причину, що стоїть за обробкою даних, такими як обрізка і згладжування.

Тепер, коли представили елементи контрольованого навчання, розглянемо моделі. Для початку спочатку визначимо модельний вибір XGBoost: рішення де-

ревних ансамблів. Модель ансамблю дерева складається з набору класифікаційних і регресійних дерев (CART). Ось простий приклад CART, який класифікує, чи сподобається комусь гіпотетична комп'ютерна гра X .

Класифікуємо членів сім'ї на різні вітки і призначаємо їм бал на відповідній вітці. CART трохи відрізняється від дерев рішень, в яких вітка містить лише значення прийняття рішень. У CART реальна оцінка пов'язана з кожною віткою, що дає кращі інтерпретації, які виходять за рамки класифікації. Це також дозволяє принциповий, уніфікований підхід до оптимізації, як побачимо нижче.

Як правило, одне дерево не є достатньо сильним, щоб бути використаним на практиці. Використаємо модель ансамблю, яка підсумовує передбачення декількох дерев разом.

Ось приклад ансамбль з двох дерев. Прогноз оцінки кожного окремого дерева підсумовуються, щоб отримати остаточний бал. Якщо подивитися на приклад, важливим фактом є те, що два дерева намагаються доповнити один одного. Математично можемо написати модель у формі

$$y_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (2.5)$$

де K знаходиться кількість дерев, f функцією в функціональному просторі, і F є набором всіх можливих варіантів. Об'єктивну функцію, яка буде оптимізована, надає F .

$$\text{obj}(\theta) = \sum_i^n (y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2.6)$$

Тепер тут виникають питання: яка модель використовується в випадкових лісах. Таким чином, випадкові ліси і підсилені дерева дійсно одні і тіж моделі; відмінність виникає від того, як їх тренувати. Це означає, що, якщо записати прогностику оцінку для ансамблів, потрібно лише написати одні і вона повинна працювати як для випадкових лісів, так і для градієнтних підсиленних дерев.

Тепер, коли представили модель, звернемося до навчання: як використувати дерева. Відповідь полягає в тому, як таку ж як для всіх контрольованих моделей навчання.

Нехай представимо модель об'єктивною функцією (, що вона завжди повинна містити тренування і узагальнення):

$$\text{obj} = \sum_{i=1}^n l(y_i, y_i^{(t)}) + \sum_{i=1}^t \Omega(f_i) \quad (2.7)$$

2.3 Адитивна підготовка

Визначимо параметри дерев. Можна виявити, що те, що потрібно дізнатися, це ті функції, кожна з яких містить структуру дерева і оцінки вітки. Вивчення структури дерева набагато важче, ніж традиційна проблема оптимізації, де можете просто взяти градієнт. Неможливо забезпечити побудову всіх дерев відразу. Замість цього, використовуємо адитивну стратегію: виправити те, що дізналися, і додати одне нове дерево за один раз. Пишемо значення прогнозування на наступному кроці. Тоді маємо $f_i y_i^{(t)}$

$$\begin{aligned} y^{(0)} &= 0 \\ y^{(1)}_i &= f_1(x_i) = y^{(0)}_i + f_1(x_i) \\ y^{(2)}_i &= f_1(x_i) + f_2(x_i) = y^{(0)}_i + f_2(x_i) \\ &\dots \\ y^{(t)}_i &= \sum_{k=1}^t f_k(x_i) = y^{(t-1)}_i + f_t(x_i) \end{aligned} \quad (2.8)$$

Залишається визначити: яке дерево записати на кожному кроці. Природна річ полягає в тому, щоб додати те, який оптимізує мету – цільову функцію.

$$\begin{aligned} \text{obj}^{(t)} &= \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^t \Omega(f_i) \\ &= \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + \text{constant} \end{aligned} \quad (2.9)$$

Якщо розглядати використання середнього квадрата помилки (MSE) як функції втрати, мета стає такою

$$\begin{aligned} \text{obj}^{(t)} &= \sum_{i=1}^n (y_i - (\hat{y}_i^{(t-1)} + f_t(x_i)))^2 + \sum_{i=1}^t \Omega(f_i) \\ &= \sum_{i=1}^n [2(\hat{y}_i^{(t-1)} - y_i) f_t(x_i) + f_t(x_i)^2] + \Omega(f_t) + \text{constant} \end{aligned} \quad (2.10)$$

Форма MSE є досить простою, в терміні першого наближення (називається залишковою) і квадратичним терміном.

Для інших втрат інформативності (наприклад, втрати логістики) отримати таку форму не так вже й просто. Таким чином, в загальному випадку, використано розширення функції втрати Тейлора до другого порядку:

$$\text{obj}^{(t)} = \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) + \text{constant} \quad (2.11)$$

де i визначаються як $g_i h_i$

$$\begin{aligned} g_i &= \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \\ h_i &= \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) \end{aligned} \quad (2.12)$$

Після того, як видалимо всі константи, конкретна мета на кроці стає t

$$\sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (2.13)$$

Це стає метою оптимізації для нового дерева. Однією з важливих переваг цього визначення є те, що значення об'єктивної функції залежить лише від i . Таким чином як модель підтримує функції втрати.

Можемо оптимізувати кожну функцію втрати, включаючи логістичну регресію і парне ранжирування, використовуючи точно такий же алгоритм.

2.5 Складність моделі

Потрібно визначити складність дерева. Для того, щоб зробити це, спочатку уточнимо визначення дерева як $\Omega(f)$

$$f_i(x) = w_{q(x)}, w \in R^T, q : R^d \rightarrow \{1, 2, \dots, T\}. \quad (2.14)$$

Вектор балів на вітках, є функція, що призначає кожну точку даних відповідній вітці, i є числом вітки. У моделі визначаємо складність як wqT

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (2.15)$$

Існує не один спосіб визначити складність, але цей добре працює на практиці. Узагальнення є однією з частин, більшість пакетів дерев будуються менш узагальнено.

Це сталося тому, що традиційне навчання дерев тільки підкреслює значення шумів, в той час як контроль складності був залишений для евристику. Визначивши це формально, можна краще зрозуміти, що вивчаємо тобто дані, і отримати моделі, які добре працюють на даних аналізу.

Оцінка структури. Після повторного формулювання моделі дерева можемо написати об'єктивне значення t

$$\begin{aligned}
\text{obj}^{(t)} &\approx \sum_{i=1}^n [g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \\
&= \sum_{j=1}^T [(\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2] + \gamma T
\end{aligned} \tag{2.16}$$

Тут знаходиться набір показників точок даних, призначених для вітки. В другому рядку змінили індекс підсумовування, оскільки всі точки даних на одному листку отримують однаковий бал. Могли б додатково спростити вираз, визначивши

$$\text{obj}^{(t)} = \sum_{j=1}^T [G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2] + \gamma T \tag{2.17}$$

У цьому рівнянні, елементи незалежні по відношенню один до одного, форма квадратична і краще для даної структури і скоротивши можемо отримати

$$\begin{aligned}
w_j^* &= -\frac{G_j}{H_j + \lambda} \\
\text{obj}^* &= -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T
\end{aligned} \tag{2.18}$$

Останнє рівняння вимірює, наскільки хороша структура дерева $q(x)$.

В основному, для даної структури дерева, використовуємо статистику віток, до дерев яких вони належать, підсумовуємо статистику разом, і використовуємо формулу, щоб обчислити, наскільки добре дерево. Ця оцінка схожа на міру значимості в дереві рішень, за винятком того, що вона також враховує складність моделі.

Тепер, коли у є спосіб виміряти, наскільки добре дерево, варто було б переназвали всі можливі дерева і вибрати краще. На практиці це складно, тому спробуємо оптимізувати один рівень дерева за раз. Зокрема, намагаємося розділити листок на два вітки, і оцінка, яку він отримує

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (2.19)$$

Ця формула може бути розкладається як 1) рахунок на новому лівому листі 2) рахунок на новому правому листку 3) рахунок на оригінальному листі 4) узагальнення на додатковому листі.

Висновки до розділу 2

Побудова моделі розроблена з урахуванням оптимізації систем, так і принципів машинного навчання. Мета цієї моделі полягає в тому, щоб підвищити нижню межу точності обчислення та забезпечити масштабовану, портативну і точну реалізацію при різних наборах даних.

Це дасть можливість проаналізувати дані із забезпеченням необхідної оцінки нижньої межі та забезпечити узагальнення з використання навченої моделі на наборах даних предметної області.

Також відмічено важливий факт: якщо приріст менший за попереднє значення, то краще не додавати таку гілку. Це головні обмеження в моделях на основі дерев.

Для реально корисних даних зазвичай необхідно знайти оптимальні параметри. Щоб ефективно зробити це, необхідно помістити всі екземпляри у відсортований рядок.

Підхід є достатнім для розрахунку оцінки структури всіх можливих розділених рішень, і можна ефективно знайти найкраще розділення.

Розділ 3

Розробки системи прогнозування

3.1 Обробка даних

Часто з'являються несподівані неприємності, які можуть зашкодити бізнесу. Одним з поширених скрутних становищ є те, що ресторани повинні знати, скільки клієнтів очікувати кожен день, щоб ефективно придбати інгредієнти і запланувати співробітників.

Цей прогноз нелегко зробити, тому що багато непередбачуваних факторів впливають на відвідуваність ресторану, як погода і місцева конкуренція. Це ще важче для нових ресторанів з невеликими історичними даними.

Маємо завдання використовувати дані бронювання та відвідування, щоб передбачити загальну кількість відвідувачів ресторану на майбутні дати. Ця інформація допоможе ресторанам бути набагато ефективнішими і дозволить їм зосередитися на створенні приємного досвіду їдальні для своїх клієнтів.

Наступні дії було виконано для подальшої роботи. Використання машинного навчання навчання. Обговорення, як ця проблема може бути поставлена як машинне навчання.

Надано джерело даних і поглиблене пояснення вмісту кожного файлу. Пояснено показник оцінки та його значення для даної проблеми. Розвідувальний аналіз даних. Проаналізовані дані дають інтуїтивне розуміння їх ширини та глибини. Роз'яснено наявні способи вирішення цієї проблеми.

Також пояснено подальші поліпшення до існуючих рішень. Дано пояснення кожної моделі з їх оптимізованими гіпер параметрами. З огляду на порівняння кожної моделі відповідно до їх продуктивності.

Використання машинного навчання. Маємо завдання використовувати дані про відвідування, щоб передбачити загальну кількість відвідувачів ресторану на майбутні дати, що пояснює той факт, що це проблема регресії, а отже, шляхом відповідної попередньої обробки та функції обробки даних та подачі їх до моделі машинного навчання, можемо вирішити цю проблему прогнозування.

Джерело даних. У роботі розглядається проблема прогнозування, яка була окреслена навколо відвідувачів ресторану. Дані надходять з двох окремих сайтів:

Hot Pepper Gourmet (hpg): подібно до Yelp, де користувачі можуть шукати ресторани, а також робити замовлення онлайн.

AirREGI / аналогічно Square, система контролю бронювання та касового апарату.

Крім вищезазначених двох джерел у є дані з метеостанцій, які можуть бути використані відповідно до довготи і широти ресторанів.

Повинні використовувати застереження, відвідування та іншу інформацію з цих сайтів, щоб прогнозувати відгуки майбутніх відвідувачів ресторану на дану дату. Дані тренування охоплюють певні року.

Тестовий набір розділений на основі часу і охоплює обрану підмножину ресторанів. Відзначимо, що тестовий набір навмисно охоплює святковий тиждень.

Є дні в тестовому наборі, де ресторани були закриті і не мали відвідувачів. Вони ігноруються в обробці.

3.2 Системні параметри

Навчальний набір провокує дні, коли ресторани були закриті. Це реляційний набір даних з двох систем. Кожен файл з передмовою про джерело (або `air_` або `hpg_`), щоб вказати його походження. Кожен ресторан має унікальнення. Не всі ресторани охоплені обома системами, і були надані дані за межами ресторанів, для яких повинні бути прогнози. Широти і довготи не є точними, щоб перешкодити деідентифікації ресторанів. `air_store_idhpg_store_id, air_reserve.csv`.

Цей файл містить застереження, зроблені в системі. Вказується час, коли було створено резервування, тоді коли настав час у майбутньому, коли відбудеться візит. `reserve_datetimevisit_datetime`

`air_store_id` - ідентифікатор ресторану в системі;

`visit_datetime` - час бронювання;

`reserve_datetime` - час оформлення бронювання;

`reserve_visitors` - кількість відвідувачів для цього бронювання;

hpg_reserve.csv;

Цей файл містить застереження, зроблені в системі hpg.

hpg_store_id - ідентифікатор ресторану в системі HPG;

visit_datetime - час бронювання

reserve_datetime - час оформлення бронювання;

reserve_visitors - кількість відвідувачів для цього бронювання.

air_store_info.csv

Цей файл містить інформацію про вибрані ресторани. Імена та вміст стовпців є зрозумілими.

air_store_id

air_genre_name

air_area_name

latitude

longitude

Примітка: широта і довгота - це широта і довгота області, до якої належить кухня.

hpg_store_info.csv.

Цей файл містить інформацію про вибрані ресторани HPG. Імена та вміст є зрозумілим.

hpg_store_id

hpg_genre_name

hpg_area_name

latitude

longitude

Примітка: широта і довгота - це широта і довгота області, до якої належить кухня.

store_id_relation.csv

Цей файл дозволяє приєднатися до вибраних ресторанів, які мають як air, так і hpg систему.

hpg_store_id

air_store_id

air_visit_data.csv

Цей файл містить історичні дані про відвідування air ресторанів.

air_store_id

visit_date - дата

visitors - кількість відвідувачів ресторану на дату

sample_submission.csv

У цьому файлі відображається надсилання у правильному форматі, включно з днями, які потрібно прогнозувати.

id - ідентифікатор формується шляхом об'єднання і з підкреслення
air_store_idvisit_date

visitors- кількість відвідувачів, прогнозованих для кухню і дат комбінації

date_info.csv

Цей файл надає основні відомості про дати календаря в наборі даних.

calendar_date

day_of_week

holiday_flg - це день відпочинку

3.3 Показник оцінювання

Показник оцінювання – це логарифмічна помилка Mean Squared, яка обчислюється як :

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2} \quad (3.1)$$

де: n - загальна кількість спостережень;

p_i - це прогноз відвідувачів;

a_i - це фактична кількість відвідувачів $\log(x)$ є природним логарифмом x.

У разі перевитрати, невеликі ресторани не підуть у програш, але в разі недооцінення, невеликі ресторани страждають від втрати, оскільки багато харчових матеріалів йдуть у відходи, тоді як великі ресторани мають стійкість до переоцінки, так і до недооцінювання, а RMSLE зменшить недопрогнозування більше, ніж надмірне передбачення, що є плюсом для невеликих ресторанів, оскільки є багато не малих ресторанів.

RMSLE зменшує вплив відносної величини чисел. Розвідувальний аналіз даних.

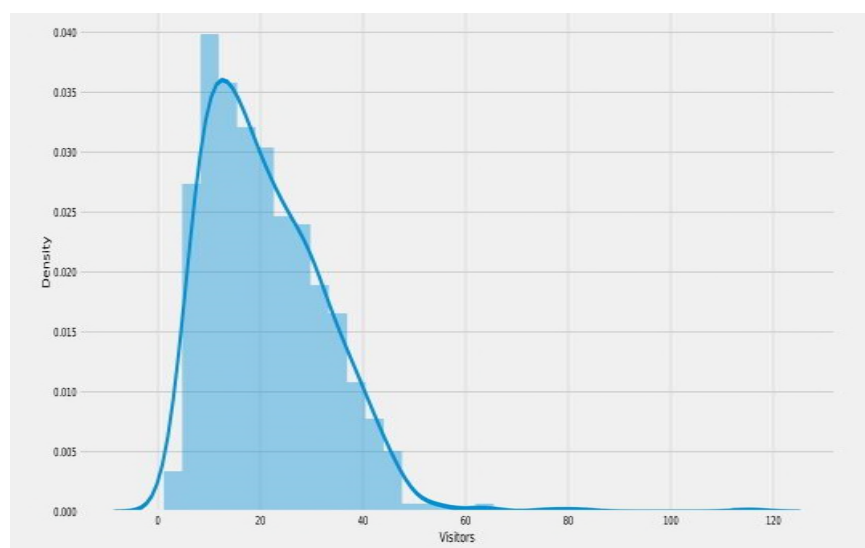


Рисунок 3.1 - Кількість середніх відвідувачів на ресторан

Спостереження, серед середніх відвідувачів на ресторан майже нормальний розподіл.

Можемо використовувати правило 68–95–99.7, щоб позбутися від крайніх аутсайдерів на основі стандартних відхилень. Є велика кількість ресторанів, які мають місткість менше 20. Показник середніх відвідувачів на ресторан.

Спостереження, у вищезгаданому представлені можна спостерігати, що майже 99% ресторанів мають менше 47. Майже 90% ресторанів мають менше 40 відвідувачів на день.

Це пояснює той факт, що існує велика кількість невеликих ресторанів. Існує незначна кількість ресторанів, які мають середню місткість відвідувачів більше 100.

Boxplot середніх відвідувачів на ресторан.

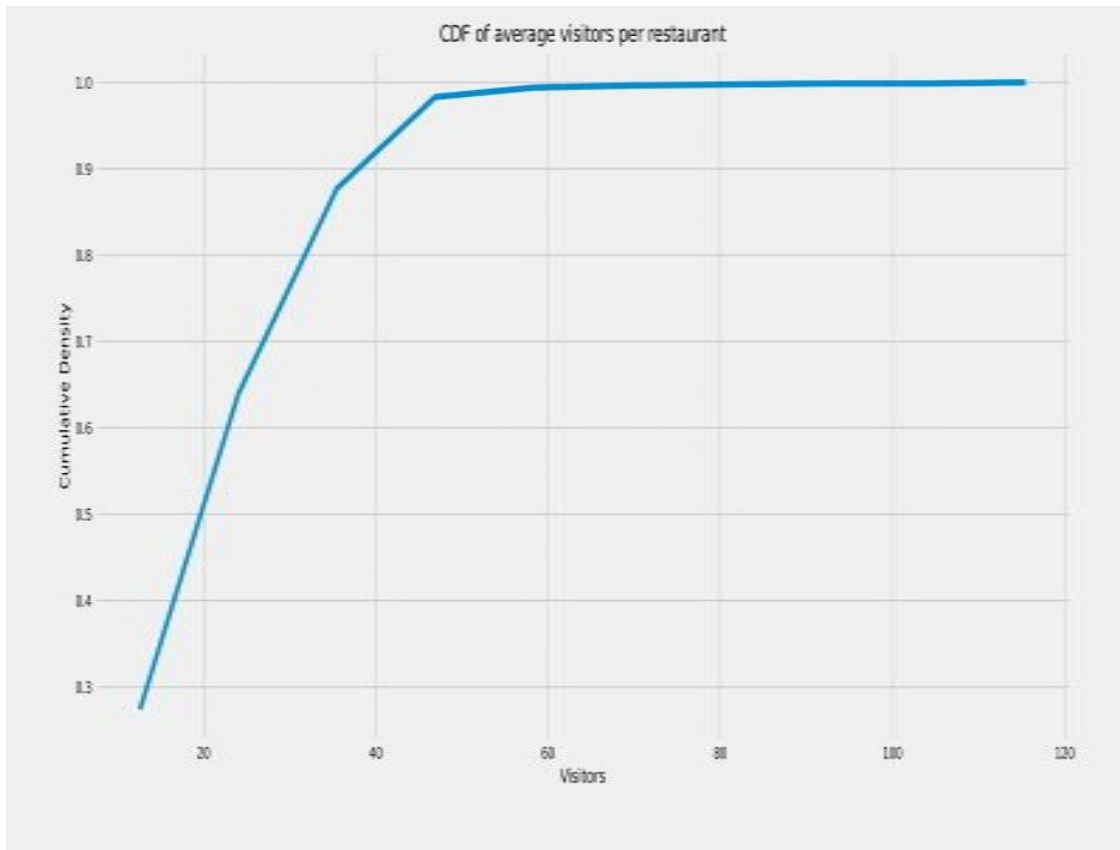


Рисунок 3.2 - Щільність середніх відвідувачів на ресторан

Спостереження, мінімум відвідувачів, які можемо спостерігати з цього графіку, майже досягає нуля. Середнє значення відвідувачів 20 відсотків.

Максимальна кількість відвідувачів становить від 55 до 60. спостерігаємо певні дуже високі значення (outlier) більше 60 і навіть більше, ніж 100 відвідувачів. 25 відсоткове та 75 відсоткові значення 13

Спостереження, поширення бронювання вище, ніж у резервації НРГ. Існує велика кількість бронювань в НРГ з відвідувачами розраховувати від 5 до 10. Є кілька застережень в НРГ, де кількість відвідувачів більше 20 або навіть досягнення 40.

Навіть в AIR, максимальна кількість зареєстрованих відвідувачів становить 40, але кількість реєстрацій більше, ніж у НРГ. В AIR максимальна кількість реєстрацій має кількість відвідувачів від 8 до 13.

Кількість незареєстрованих відвідувачів набагато більше, ніж кількість зареєстрованих відвідувачів. CDF середнього бронювання відвідувачів в ресторані

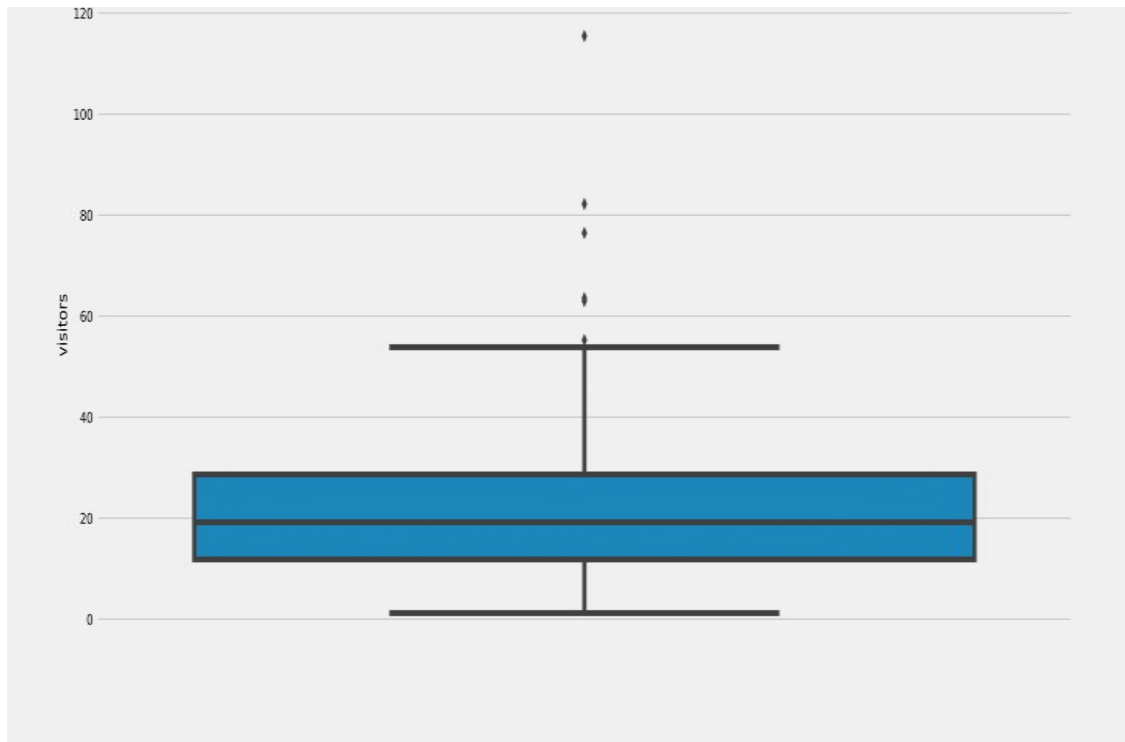


Рисунок 3.3 - Розподіл середніх відвідувачів на ресторан

Спостереження, сформувані вище середній показник відвідування, в AIR, майже 99% за бронювання ресторану становить менше 40.

У AIR незначна кількість бронювань перевищує 60.

У HPG майже всі бронювання рахуються менше 40.

У HPG близько 99% від кількості бронювань становить менше 20.

У HPG майже 90% бронювань менше або дорівнює 10.

Boxplot середнього бронювання відвідувачів в ресторані

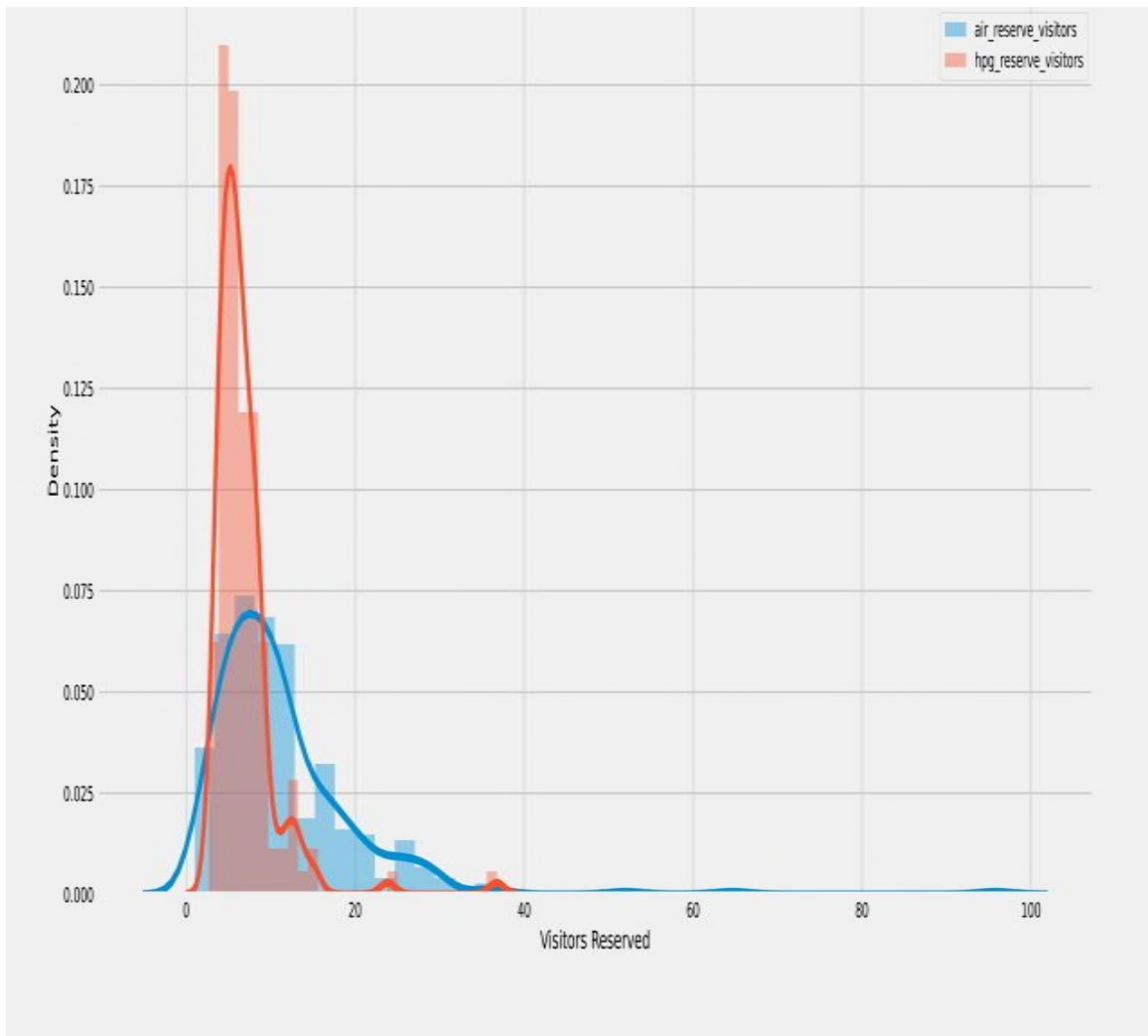


Рисунок 3.4 - Середнє бронювання відвідувачів в ресторані

Спостереження, з наведених вище рисунків, можемо спостерігати, що середня кількість бронювань не більше 10.

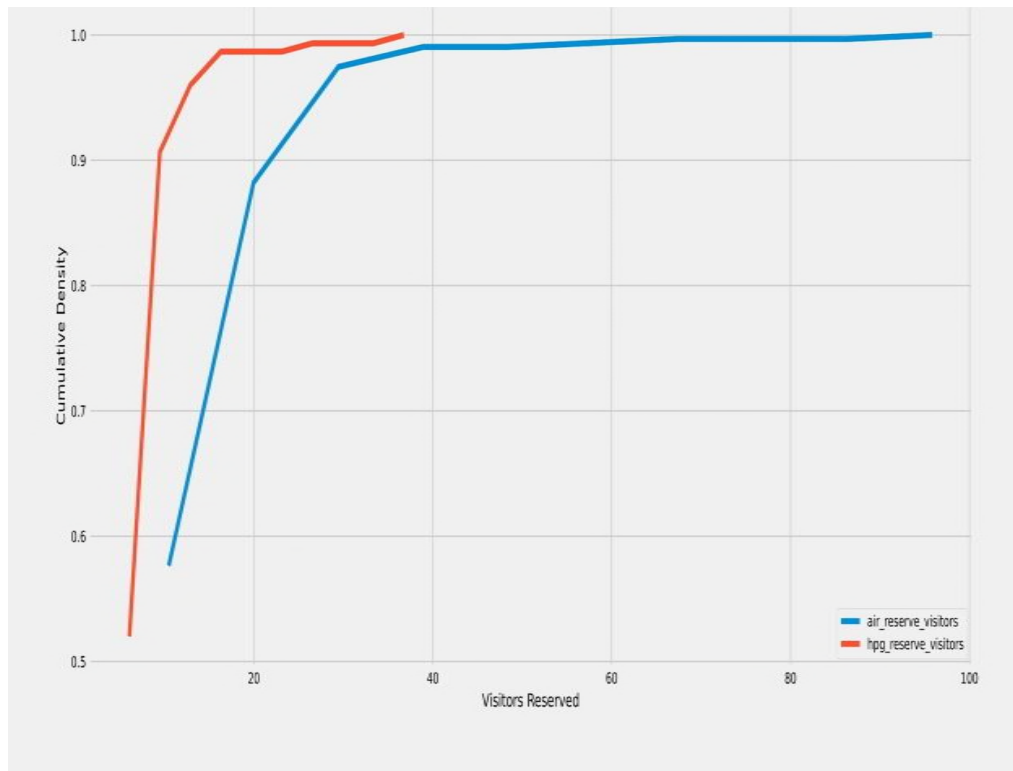


Рисунок 3.5 - Щільність середнього бронювання відвідувачів в ресторані

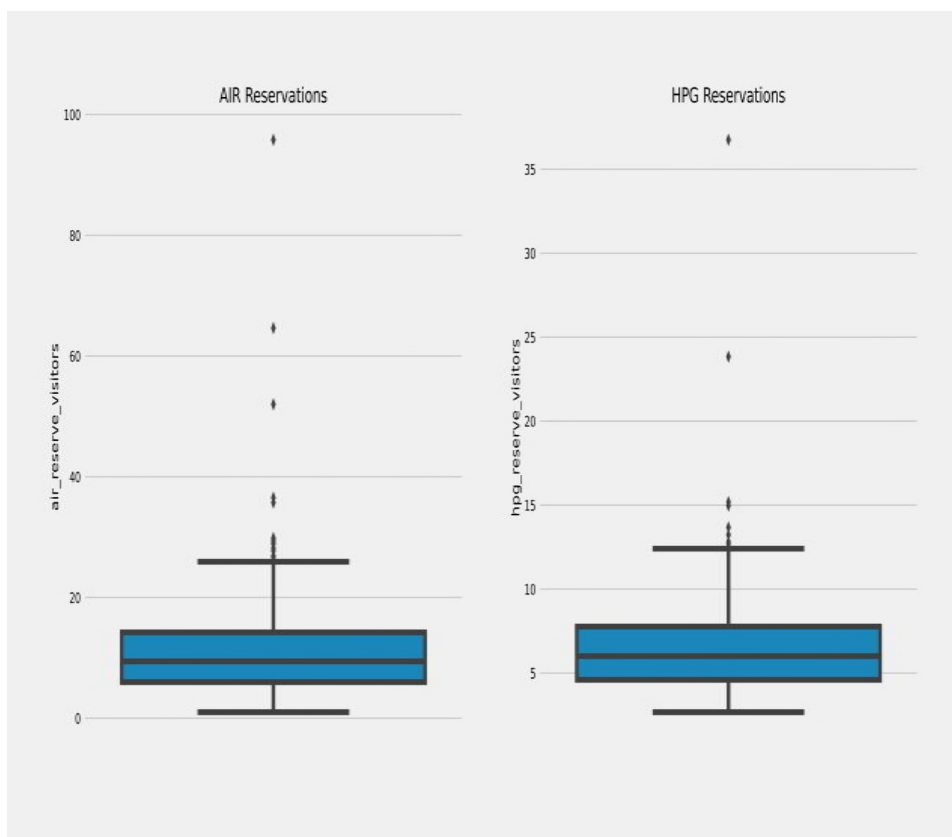


Рисунок 3.6 - Розподіл середнього бронювання відвідувачів в ресторані

У НРГ є певні високі значення (аутлети), які бачимо в діапазоні від 13 до 4.

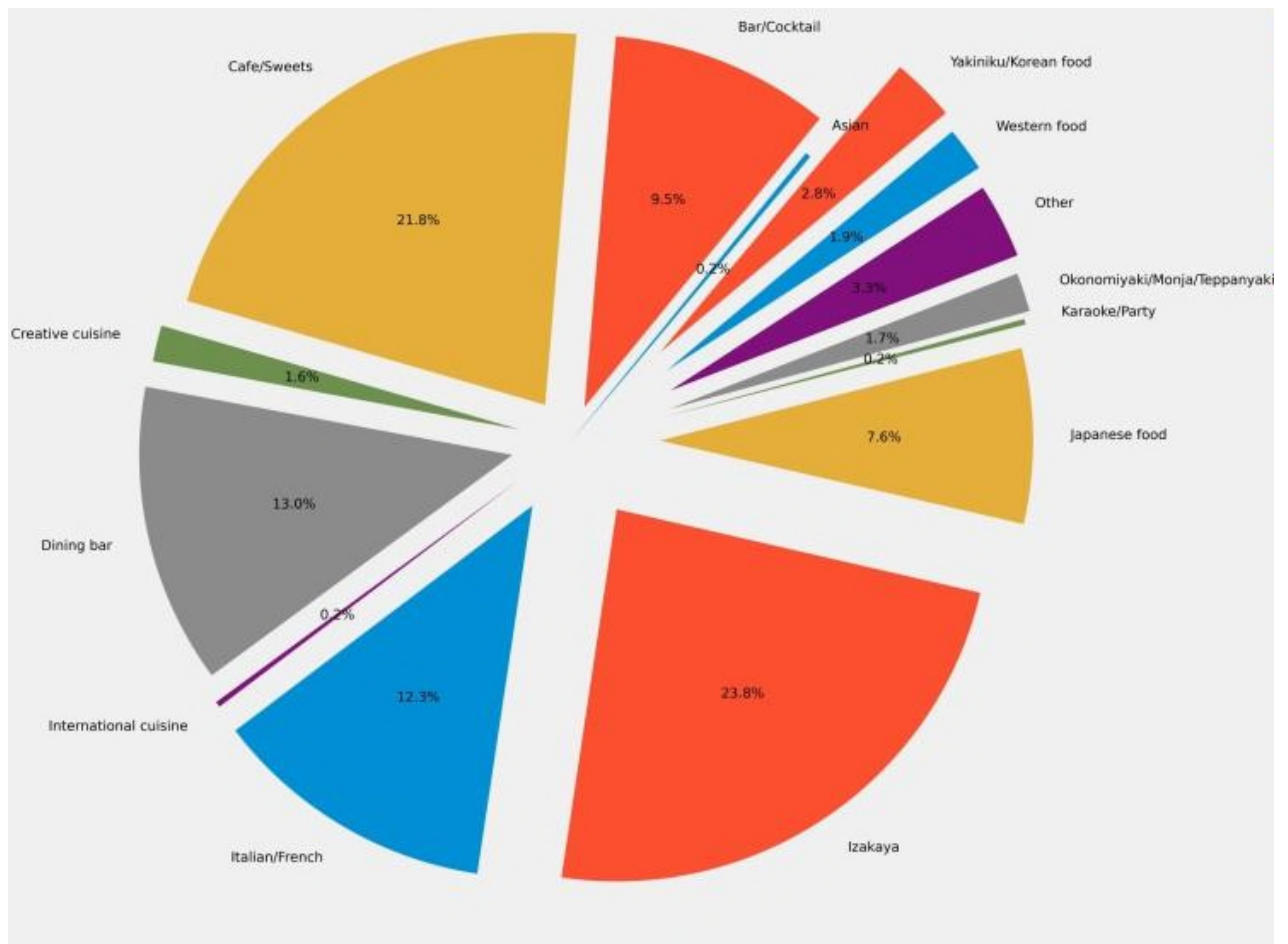


Рисунок 3.6 - Частки ринку

У AIR середня кількість бронювань становить 10(приблизно), а в НРГ середня кількість бронювань становить 6(прибл.

25-й і 75-й підрахунок процентиля в AIR становить 7 і 15 відповідно.

25-й і 75-й підрахунок процентильних бронювань в НРГ становить 4 і 8 відповідно.

У AIR є певні високі значення (аутлети), які бачимо в діапазоні від 40 до 100.

Спостереження, ресторани розділені на 14 жанрів харчування.

Ізакая є найпопулярнішим жанром , оскільки майже 23,8% ресторанів є жанром Ізакая. Другим за популярністю жанром є Cafe/Sweets, що має майже 21,8% частки ресторанного ринку. Міжнародна кухня, Азіатська, і Караоке / Party є найменш кращим жанром, що має тільки 0,2% кожної частки ринку.

Навіть західна і корейська їжа взагалі не користується популярністю.

Висновок – щоб почати ресторанний бізнес, вибір харчового жанру стане найважливішим рішенням.

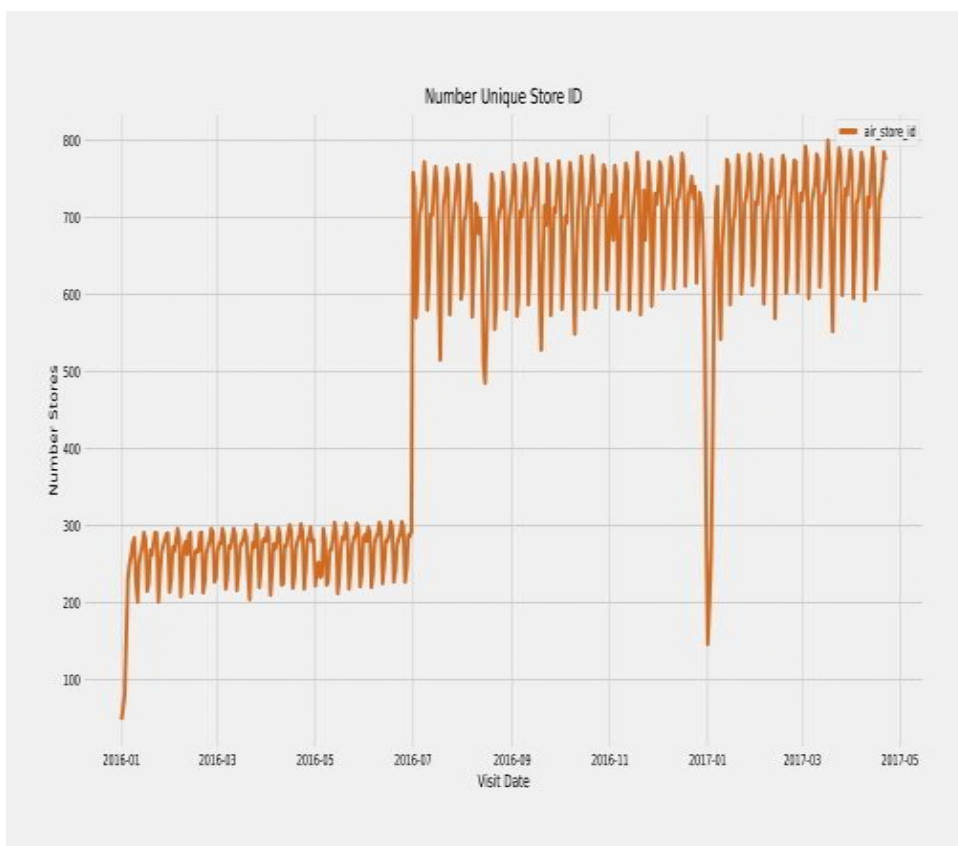


Рисунок 3.7 - Кількість унікальних кухонь

Спостереження, вище зазначена частина показує той факт, що протягом середини 2016 року відбувається майже 150% походу за кількістю ресторанів.

Причина походу полягає в тому, що в середині 2016 року додали нових ресторанів 500(прибл) до бази даних AIR. Зниження і падіння, які бачимо, можуть бути пов'язані з буднім і вихідними днями. Різке зниження на 1 січня пов'язано з новорічними святами, оскільки більшість ресторанів залишаються закритими в новий рік. Загалом маємо рекорди майже 800 японських ресторанів.

Спостереження, різке підвищення в середині 2016 року обумовлено додаванням нових ресторанів (як пояснюється в попередньому описі).

Можна легко спостерігати, що кількість не зареєстрованих відвідувачів набагато більше, ніж кількість зареєстрованих відвідувачів.

3.4 Відвідувач і бронювання

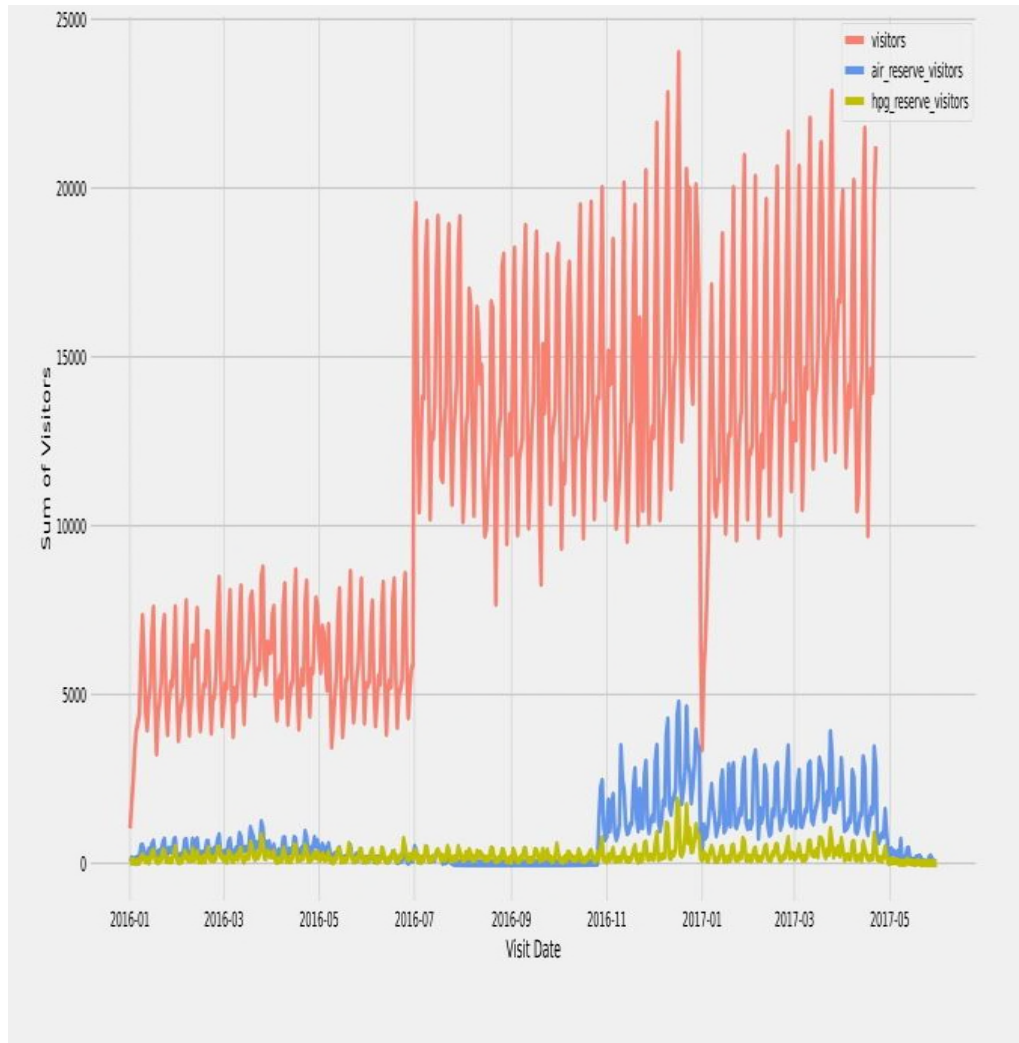


Рисунок 3.8 - Відвідувачі і бронювання

Спостерігається різке зниження напередодні Нового року, оскільки більшість ресторанів залишаються неактуальними напередодні Нового року.

Кількість реєстрацій в AIR більше, ніж у HPG. Максимальна кількість відвідувачів спостерігається в грудні місяці, як знаємо, є ряд фестивалів в грудні.

Спостереження, Існує близько 14 жанрів їжі, які подають ресторани.

Найпопулярнішим і найбільш вподобанням жанром є Ізакая, за яким слідує Cafe/Sweets, які подобаються максимальній кількості людей. Азіатська, караоке, і міжнародна кухня є новим жанром з найменшими клієнтами. Навіть західна їжа не дуже подобається .

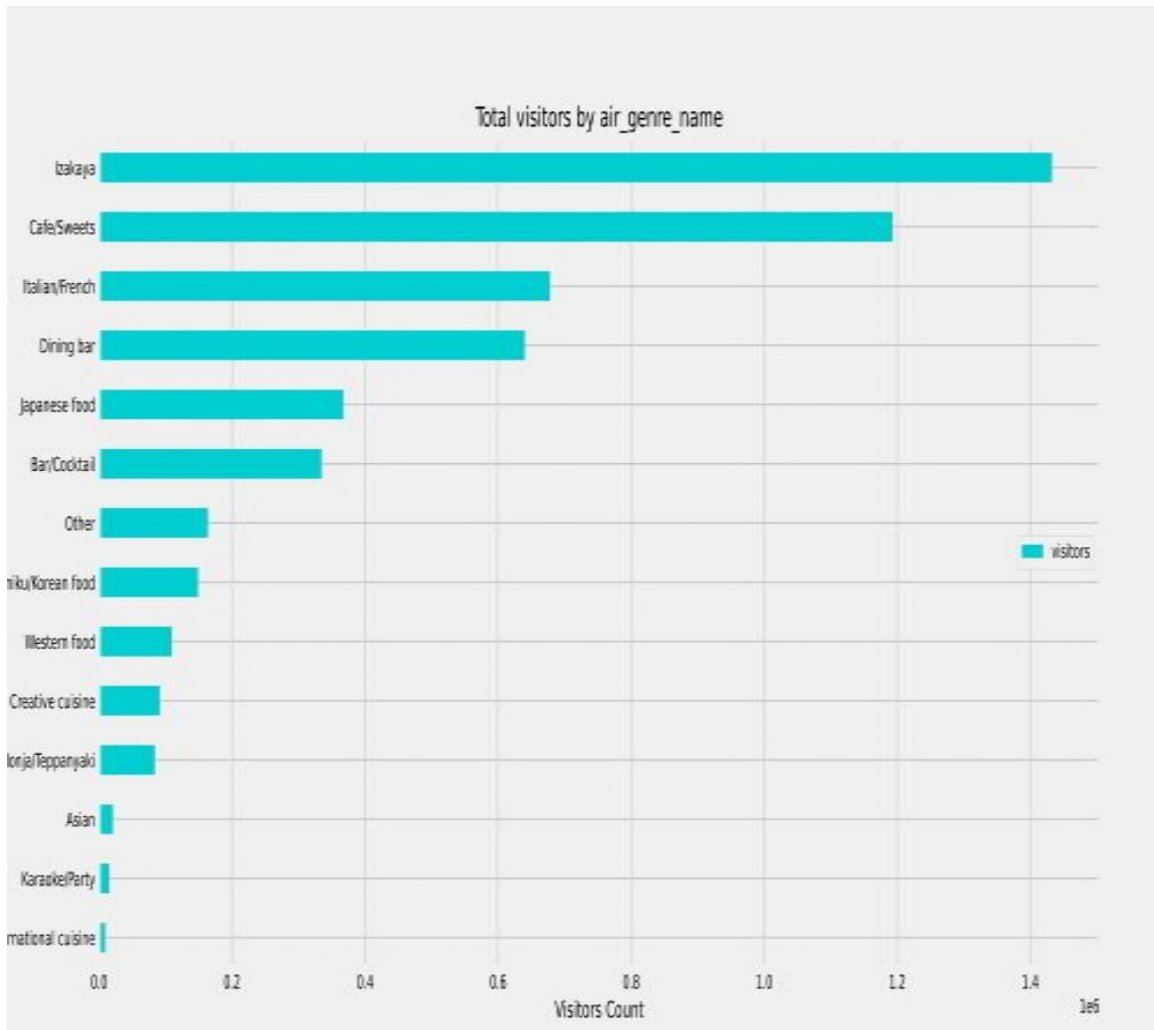


Рисунок 3.9 - Відвідувачі за типом кухні

Харчовий жанр є найважливішим фактором для зростання в японському ресторанному бізнесі.

Навіть для бронювання тенденцій, Ізакая є найпопулярнішим жанром. У незареєстрованих тенденціях відвідувача спостерігали, що Cafe/Sweets є другим за популярністю жанром, але тут італійський/французький є другим за популярністю жанром. Азіатська та міжнародна кухня є найменш популярною, як бачили в попередньому сюжеті. Японська їжа є 4-у за популярністю жанром.

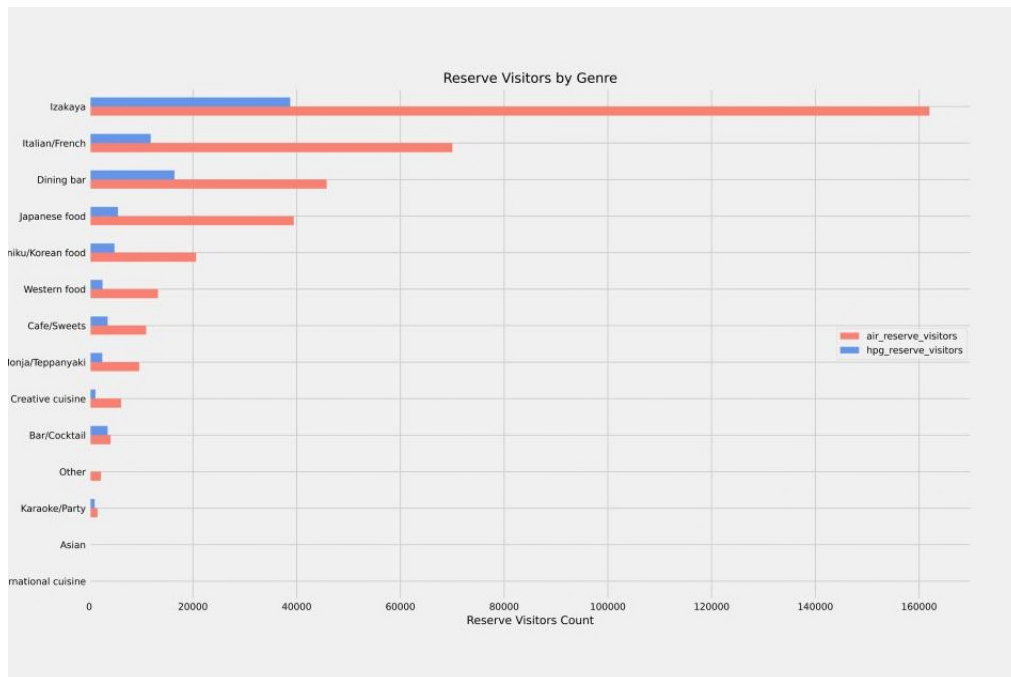


Рисунок 3.10 - Резервування відвідувачів за жанром

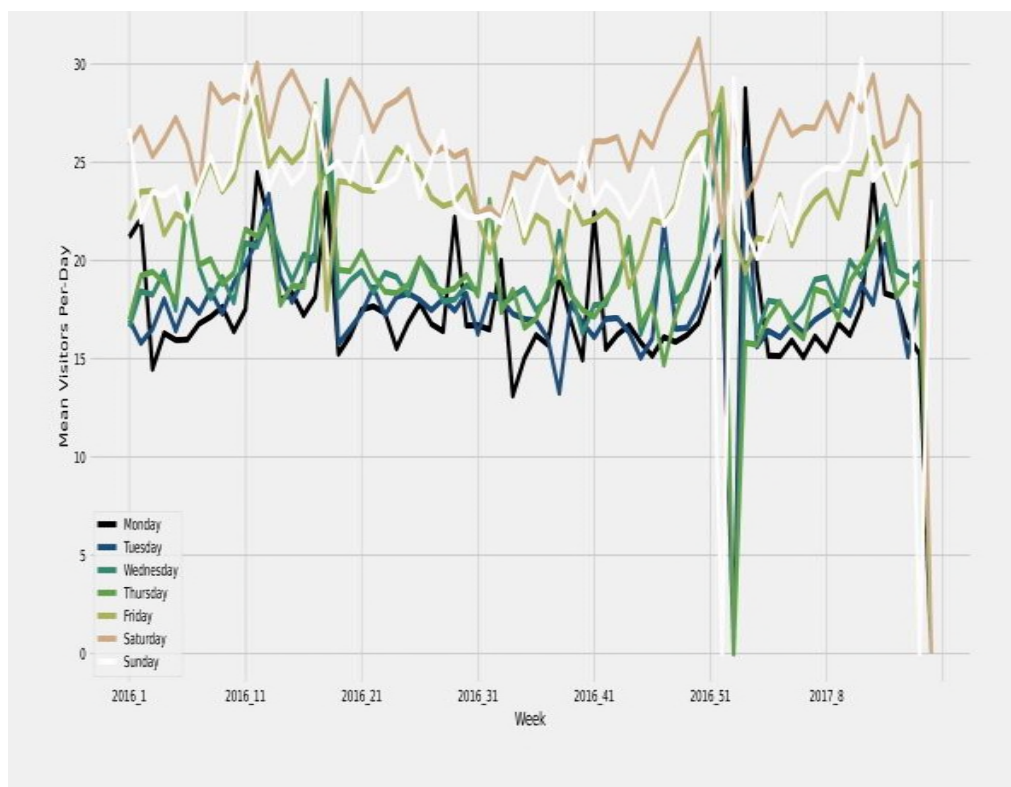


Рисунок 3.11 - Відвідувачі за днем тижня

Субота є днем, на який більшість людей вважають за краще виходити на вулицю, щоб поїсти, маючи найбільшу кількість відвідувачів уздовж усього року причиною того, що це вихідні дні.

Після суботи, навіть у неділю, спостерігається пік відвідувачів.

У понеділок найменша кількість людей виходить поїсти. Інші тижневі дні мають майже схожу тенденцію відвідувачів.

Різке зниження після 51-го тижня пов'язано з новорічною ніччю, як пояснювалося раніше.

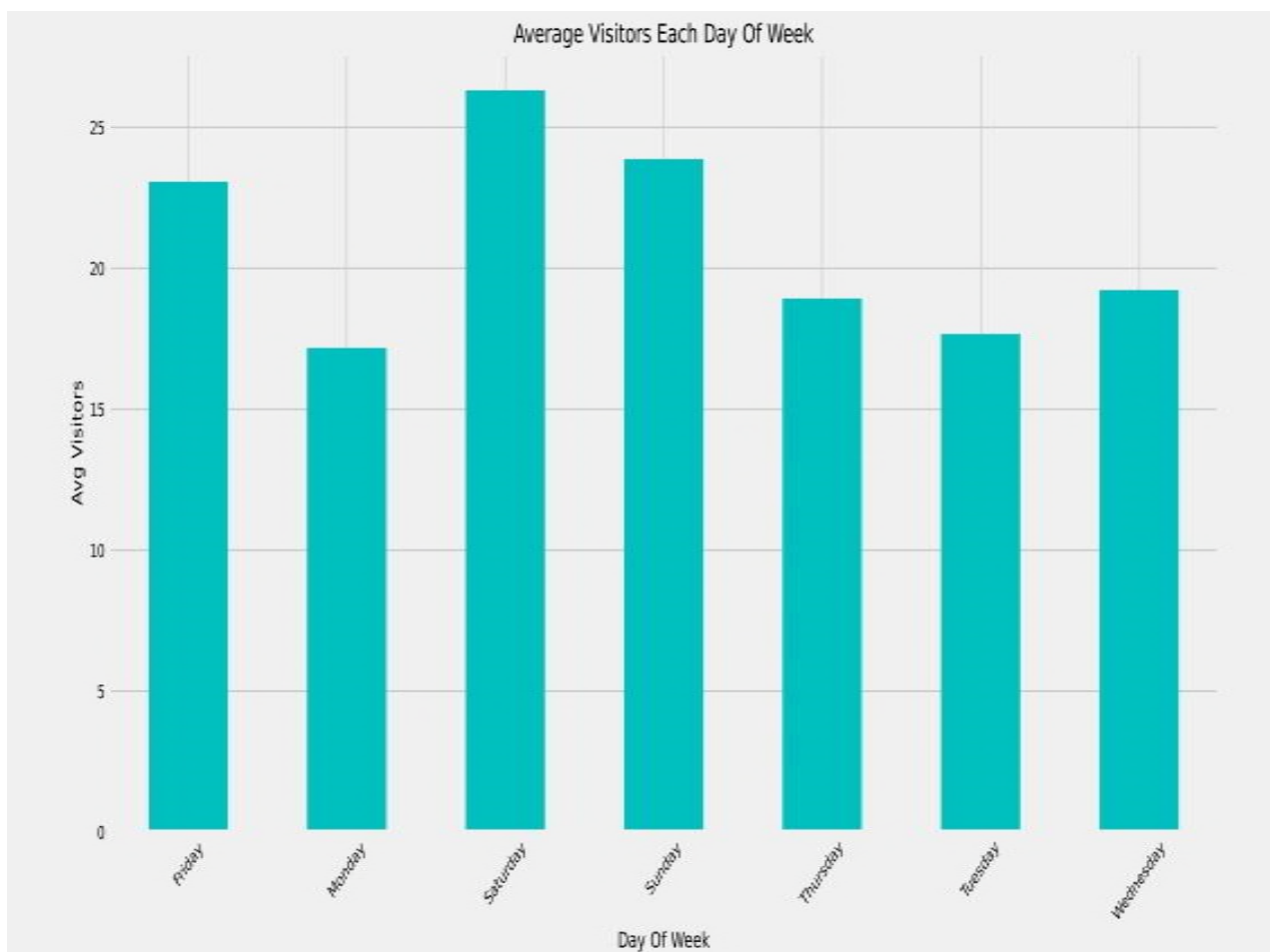


Рисунок 3.12 - Середні відвідувачі кожен день тижня

Навіть щодня ресторани спостерігають найбільшу кількість відвідувачів у суботу.

Другий за кількістю відвідувачів день – це неділя. У понеділок і вівторок найменше відвідувачів.

Четвер і середа мають майже однакові тенденції відвідувачів. Середні показники відвідувачів кожен день місяця.

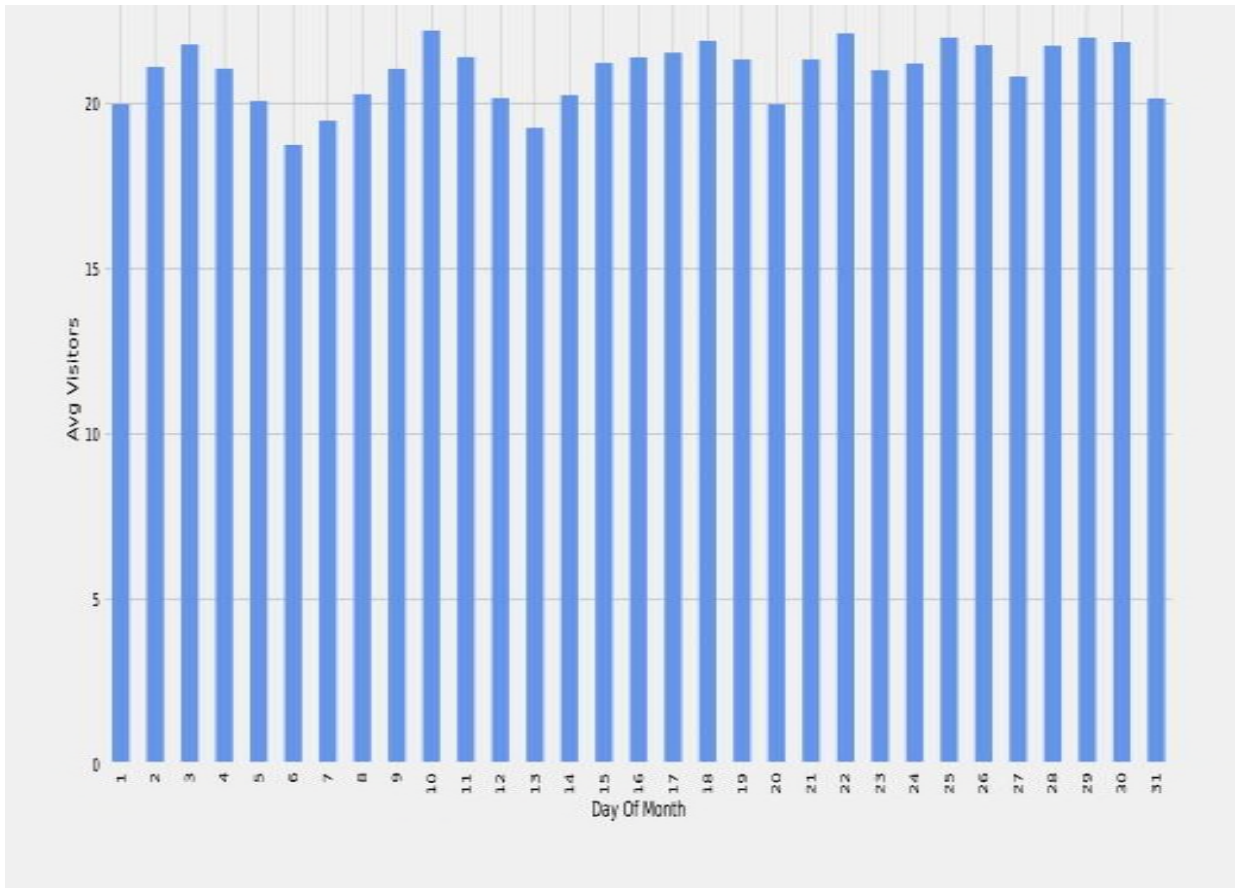


Рисунок 3.13 - Середні відвідувачі кожен день місяця

Сплеск відвідувача, пов'язано з суботнім ефектом, як описано в попередній частині.

Можемо спостерігати постійний похід в кінці місяця, можливо кредит заробітної плати в кінці місяця буде причиною цього походу.

Постійне зниження спостерігається 6-го,13-го,20-го і 31-го дня місяця. Постійний пік спостерігається 3-го,10-го і 22-го дня місяця. Середньомісячні відвідувачі.

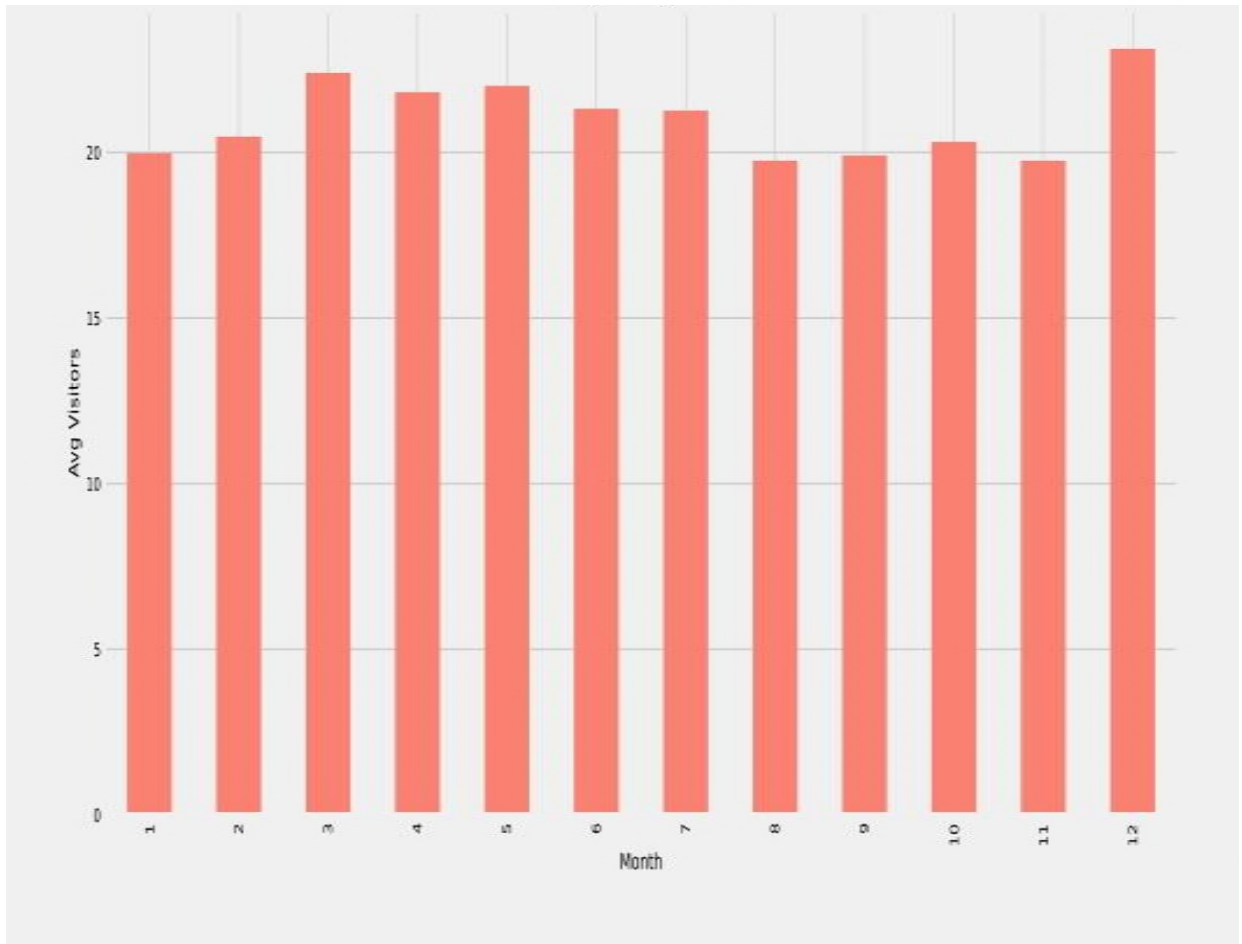


Рисунок 3.14 - Середньомісячні відвідувачі

Можемо спостерігати за походами відвідувачів у грудні місяці. Оскільки грудень місяць є святковим місяцем, можливо, це причина походу відвідувачів.

Після грудня березень - місяць, що має найвищих відвідувачів. Місяць серпень і листопад має найменше середніх відвідувачів. Середні відвідувачі у святкові дні та будні.

Це спостерігається з дослідження і очевидно більше відвідувачів у святкові дні, ніж робочі дні. Навіть тоді різниця між відвідувачами у святкові та робочі дні суттєво не велика, що пов'язано з ефектом вихідного дня.

Також з вищезазначеної ділянки спостерігається, що на більшу частину свят ресторани відкриті.

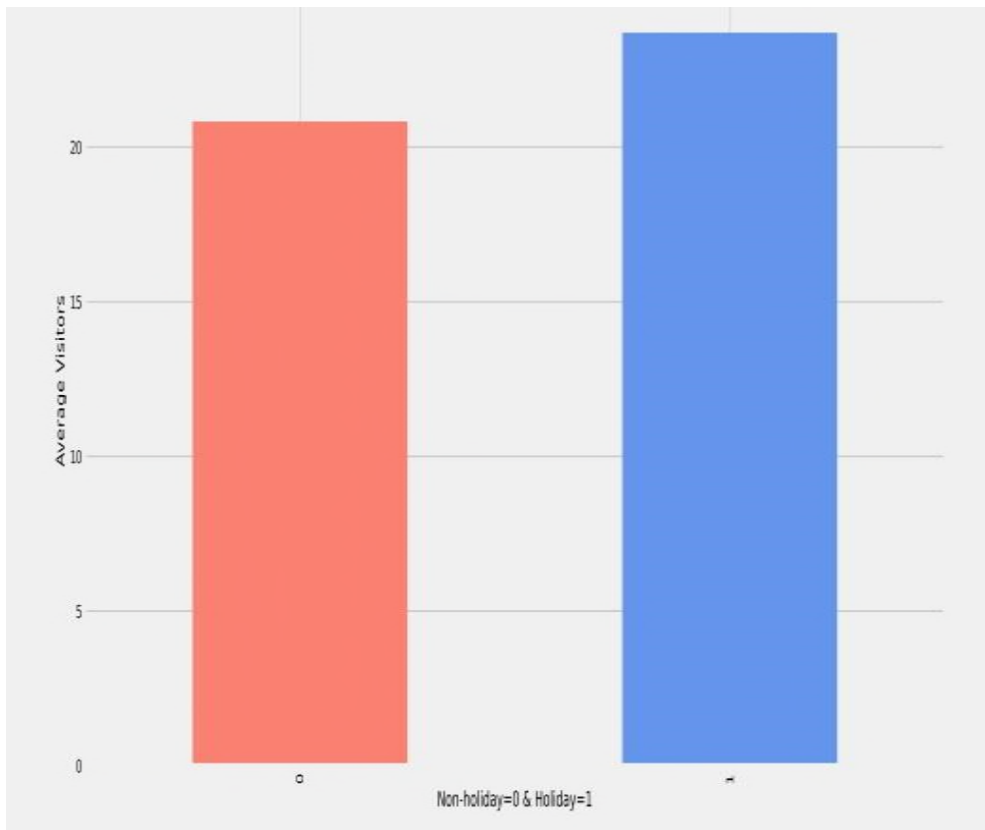


Рисунок 3.15 - Середні відвідувачі у святкові та не свята

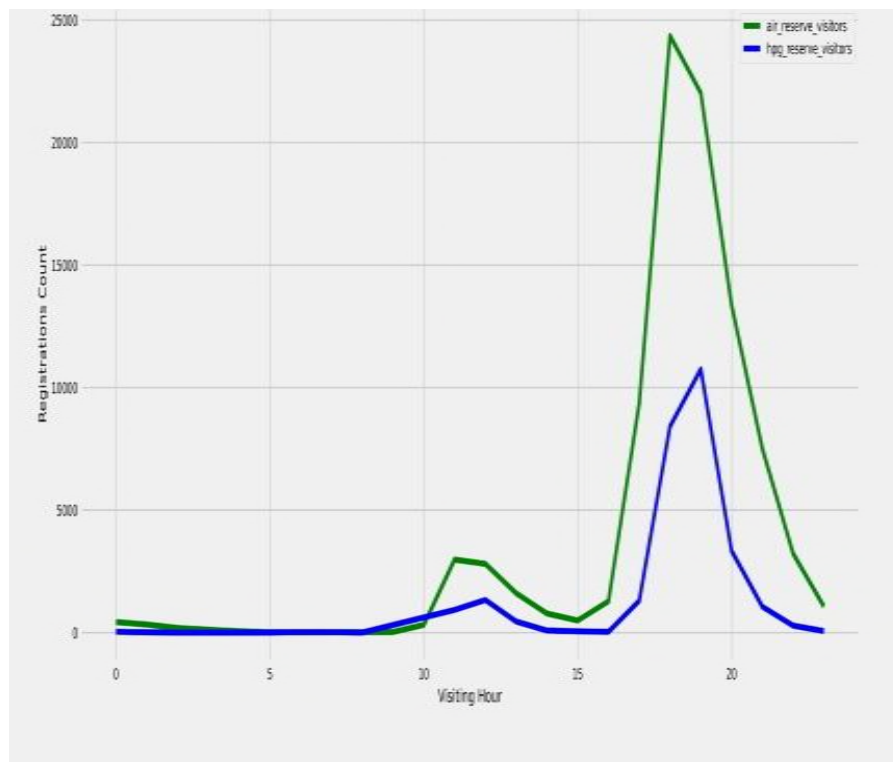


Рисунок 3.16 - Погодинна поведінка відвідувачів

При обробці даних повинні враховувати свята, які приходять у вихідні дні, такі свята слід розглядати тільки як вихідні дні не як свята просто для того, щоб врахувати ефект вихідних. Погодинна поведінка відвідувачів.

Вже спостерігається, що кількість реєстрацій в AIR більше, ніж HPG. Існує невеликий похід після 10:00 ранку. Це час, коли люди йдуть в офіс.

Найбільша кількість відвідувачів - з 17:30 до 19:00 (приблизно).

Після 19:00 спостерігається різке зниження кількості відвідувачів. Немає (майже нуль) відвідувачів між 12:00 ранку і 7:00 ранку (прибл.), це може бути тому, що ресторани залишаються закритими протягом ночі.

Висновки до розділу 3

Це дослідження намагається показати взаємозв'язок між важливістю і продуктивністю (задоволенням) досвіду місцевої кухні, а також вивчити вплив досвіду їдальні на задоволення призначення. Що стосується важливих факторів для їдальні, найбільш важливим фактором для мандрівників є продовольчі культурні аспекти. Зокрема, туристи очікують, що місцева їжа буде відображати унікальні культурні аспекти для приготування їжі, презентації та зовнішнього вигляду. Дотримуючись культурних аспектів, соціальних аспектів, які є співробітниками та поведінка клієнтів, є другим за важливості фактором для клієнтів. Третій важливий фактор, якість продуктів харчування дотримувалася соціальних аспектів. Нарешті, фізичні властивості навколишнього середовища оцінюється як найменш важливий фактор для їдальні досвід.

Для ефективної та економної роботи власникам ресторанів необхідно точно оцінити кількість майбутніх клієнтів. У цій роботі запропоновано підхід до прогнозування кількості майбутніх відвідувачів що відповідає ресторан з використанням великих даних і контрольованого навчання.

Великі дані включають інформацію про ресторан, історію відвідувань та історію бронювання. а основі цих даних, запропонований підхід генерує прогнози. Результати оцінки свідчать про ефективність такого підходу, а також корисні ідеї для майбутньої роботи.

Розділ 4

Дослідження ефективності запропонованих підходів

4.1 Загальні спостереження з аналізу даних

Огляд набору навчальних даних:

- Загальна кількість унікальних ресторанів AIR:- 829.
- Всього ресторанів, поширених в air і HPG:- 150.
- Всього унікального жанру в ресторанах AIR:- 14.
- Загальна кількість локацій ресторану AIR:- 103.
- Середньодобові відвідувачі:- 20.973761245180636.
- Тривалість навчальних даних:-2016-01-01-2017-04-22.

2. Огляд набору даних для тестування:

- Всього унікальних ресторанів:- 821.
- Тривалість тестових даних:- 2017-04-23-2017-05-31.

3. Є велика кількість ресторанів, які мають місткість менше 20.

4. Майже 90% ресторанів мають менше 40 відвідувачів на день.

5. Поширення бронювання AIR вище, ніж у резервації HPG.

6. Навіть в air, максимальна кількість зареєстрованих відвідувачів становить 40, але кількість реєстрацій більше, ніж у HPG.

7. Кількість незареєстрованих відвідувачів набагато більше, ніж кількість зареєстрованих відвідувачів.

8. Ізакая є найпопулярнішим жанром , оскільки майже 23,8% ресторанів є жанром Ізакая.

9. Другим за популярністю жанром є Cafe/Sweets, що має майже 21,8% частки ресторанного ринку.

10. Міжнародна кухня, азіатська та караоке/вечірка є найменш кращим жанром, що має лише 0,2% кожної частки ринку.

11. Протягом середини 2016 року відбувається майже 150% походу за кількістю ресторанів.

12. Причиною походу є додавання 500(приблизно) нових ресторанів до бази даних AIR в середині 2016 року.

13. Спостерігається різке зниження в новорічну ніч, оскільки більшість ресторанів залишаються близькими напередодні Нового року.

14. Максимальна кількість відвідувачів спостерігається в грудні місяці, Як відомо, є ряд фестивалів в грудні.

15. Субота - це день, в який більшість людей вважають за краще виходити на вулицю, щоб поїсти, маючи найбільшу кількість відвідувачів протягом усього року є причиною того, що це вихідні дні.

16. Можемо спостерігати постійний похід в кінці місяця, може бути, видача заробітної плати в кінці місяця буде причиною цього походу.

17. Очевидно, що у святкові дні більше відвідувачів, ніж у робочі дні.

18. Найбільша кількість відвідувачів - з 17:30 до 19:00.

З інформації календаря використовується функція під назвою `hour_gap`, яка дає проміжок між резервуванням ресторану та відвідуванням його в годинах, який знову поділяється на 5 категорій на основі довжини розриву (<12 годин, від 12 до 37 годин, від 37 до 59 годин, від 59 до 86 годин і більше 85 годин)

Середній, медіанний, максимальний параметр відвідувачів на ресторан враховується окремо для робочих днів і неробочих днів.

Густина загального підрахунку ресторанів також розраховується. З інформації про погоду, тут також використовується інформація про температуру та опади, але температура поділяється на низьку, середню та високу.

Також розраховується тижневий середній середній рівень відвідувачів за всі 7 днів всіх ресторанів.

Середньомісячна щільність кількості відвідувачів за всі 12 місяців для кожного ресторану також розраховується.

Замість того, щоб приймати дані відвідування, як є, повторно за день, так що протягом днів, коли немає значень даних (немає відвідувачів) кількість відвідувань 0, відстеження, якщо точка даних додається через повторне використання.

З інформації календаря, крім "дня тижня" та "вихідного дня", також використовуються дві додаткові функції, які вказують, чи є попередній і наступний день святом.

Що стосується інформації про кухню, то використовується попередньо підготовлена версія з даних про погоду. Оскільки попередньо підготовлений варіант містить дані метеостанції, важливі для приєднання до функцій погоди.

З інформації про погоду, тільки опади і температурні особливості використовуються там, де відсутні значення обробляються шляхом заміни їх на глобальний щоденний середній показник.

Розглядаючи відвідування як звичайний розподіл, значення 2.4 розглядається як кількісне значення нормального розподілу, значення, що має дисперсію більше $2,4 * \text{std-відхилення}$, вважається великим. Як тільки буде виявлено викид, нова функція під назвою `visitors_capped` використовується там, де значення викиду замінюються максимальними значеннями, що не є викидом.

Функція під назвою `день місяця` також використовується, що досить цікаво, тому що це може розглядатися як проксі, коли люди платять в місяць, припускаючи, що вони платять щомісяця.

Експоненціально зважені засоби - це спосіб зафіксувати тенденцію часового ряду, різних числових функцій використовується там, де альфа-значення оптимізовані за допомогою часової еволюції.

Середнє, медіанне, стандартне відхилення, кількість значень, мінімум, максимум також використовуються шляхом групування даних за допомогою дати відвідування.

Всі категоричні змінні закодовані міткою. Кілька стовпців у наборі даних потрібно скинути, оскільки вони непотрібні. Під час передбачення прогнозується журнал відвідувачів і для повернення до первісної величини просто експоненціальна функція застосовується до передбачення. Це працює, оскільки $\exp(\log(x)) = x$. Оскільки це допомагає дереву рішень упакувати значення в лист, оскільки значення "ближче" один до одного. Всього для підготовки фінальної моделі використовується 98 фінальних функцій.

LightGBM (Підсилення градієнта на основі дерева) використовується як остаточна модель, де 6 моделей навчаються на випадковій вибірці набору даних. Вибірка робиться без заміни. Середнє значення всіх прогнозів 6 моделей розглядається як остаточний прогноз. Самостійне налаштування гіпер параметра здійснюється без використання grid-search або випадкового пошуку.

Виявлення різнокольорової за допомогою VIF (коефіцієнт дисперсії). Різноколірність виникає, коли дві або більше незалежних змінних сильно корелюють один з одним у моделі регресії. VIF визначає силу кореляції між незалежними змінними. Це прогнозується шляхом прийняття змінної і регресування його проти будь-якої іншої змінної.

Оцінка VIF незалежної змінної представляє, наскільки добре змінна пояснюється іншими незалежними змінними. Щоб дізнатися оцінку VIF кожної функції, використали таку функцію:

```
from statsmodels.stats.outliers_influence import variance_inflation_factor
# function to calculate VIF
def calc_vif(X):
    '''calculates VIF and return a DataFrame containing features and their
    respective VIF'''
    vif = pd.DataFrame()
    vif["variables"] = X.columns
    vif["VIF"] = [variance_inflation_factor(X.values, i) for i in
    range(X.shape[1])]

    return(vif)
```

Функція розрахунку коефіцієнта кореляції. Вихід, який отримали після розрахунку VIF, такий:

	variables	VIF
3	longitude	5302.495301
7	year	2753.585287
2	latitude	781.280280
8	month	334.169366
9	week	245.856323
10	quarter	100.678712
1	air_area_name	7.994980
5	day	5.724121
0	air_genre_name	3.948267
6	dow	3.518413
11	air_reserve_visitors	1.105963
4	holiday_flg	1.093103
12	hpg_reserve_visitors	1.066087

Рисунок 4.1 - VIF для різних функцій

Розраховуючи оцінку VIF і видаляючи високі значення функції оцінки VIF ітераційно, поки всі функції не підйдуть під прийнятним діапазоном, можемо видалити різнокольорові функції і, отже, зменшити розмірність, зберігаючи інформацію.

Далі розраховали вектор TF-IDF, використовуючи назву області та назву жанру. Потім виконали розкладання особливого значення на отриманих значеннях TF-IDF. Але оптимальна усадка розкладання особливого значення важлива для збереження максимальної інформації при збереженні мінімальної розмірності. використали призначений метод оптимальної усадки однині значень в розкладання однині значення реалізовано наступним чином:

```
def del_zero_col(data):
    """Function to remove columns containing only zeros."""
    idx = np.argwhere(np.all(data[... , :] == 0, axis=0))
    data = np.delete(data, idx, axis=1)
    return data
```

```

def omega_approx(beta):
    """Return an approximate omega value for given beta. Equation (5) from
    Gavish 2014."""
    return (0.56 * beta**3 - 0.95 * beta**2 + 1.82 * beta + 1.43)

# https://arxiv.org/pdf/1305.5870.pdf (page no :- 3)
def Gavish_Donoho_Truncate(data):
    """ Function to find optimal SVD truncation using 'Gavish-Donoho'
    method"""
    # do SVD and find threshold
    U,sigma,Vt = svds(data, k = min(data.shape) - 1)
    beta = min(data.shape) / max(data.shape)
    threshold = np.median(sigma[np.nonzero(sigma)]) * omega_approx(beta)
    # reconstruct data (de-noised)
    sigma_dash = sigma.copy()
    sigma_dash[sigma < threshold] = 0
    X_dash = np.dot(U, np.diag(sigma_dash))
    return
del_zero_col(X_dash),sigma[np.nonzero(sigma)],sigma_dash[np.nonzero(sigma_
dash)]

```

Оптимальна усадка особливих значень у SVD. Вибір функцій за допомогою рекурсивного усунення функцій.

Вибір функцій – це методи вибору підмножини найбільш релевантних функцій (стовпців) для набору даних. Менше функцій може дозволити алгоритми машинного навчання працювати ефективніше (менше простору або складності часу) і бути більш ефективними. Деякі алгоритми машинного навчання можуть бути введені в оману неактуальними функціями введення, що призведе до гіршої прогнозовної продуктивності.

З'ясування оптимального параметра, який слід вибрати за допомогою рекурсивної функції усунення перехресної перевірки.

```

from sklearn.feature_selection import RFE,RFECV
# setting up XGBRegressor to instantiate RFECV
model = xgb.XGBRegressor(objective='reg:squarederror',

```

```

        eval_metric='rmse',
        tree_method='gpu_hist')
f, ax=plt.subplots(1,1, figsize=(15,8))

# Instantiate RFECV visualizer with a XGBRegressor with step=5
visualizer = RFECV(model,step=5,ax=ax)
visualizer.fit(X_train, y_train)      # Fit the data to the visualizer
visualizer.show()                    # Finalize and render the figure

# setting up Recursive Feature Eliminator with n_features_to_select=61 and
dropping off not selected features
rfe = RFE(estimator=model, n_features_to_select=61)
rfe.fit(X_train,y_train)
print("Selected Columns and their respective ranks:-\n\n")
for i,j in enumerate(X_train.columns):
    if rfe.support_[i]:
        print('Column Name: {}, Rank: {}'.format(j, rfe.ranking_[i]))
    else:
        X_train.drop(columns=j,inplace=True)
        X_test.drop(columns=j,inplace=True)
        X_val.drop(columns=j,inplace=True)

```

4.2 Вибір функцій за допомогою рекурсивного усунення функцій

Результатом того, що вийшли з рекурсивної функції усунення перехресної перевірки є наступний.

Оскільки це проблема прогнозування часових рядів, часові групування даних бронювання та дані про відвідування будуть корисними з точки зору вилучення поведінки відвідування ресторану, тому використали ряд функцій, що використовують групування на основі часу.

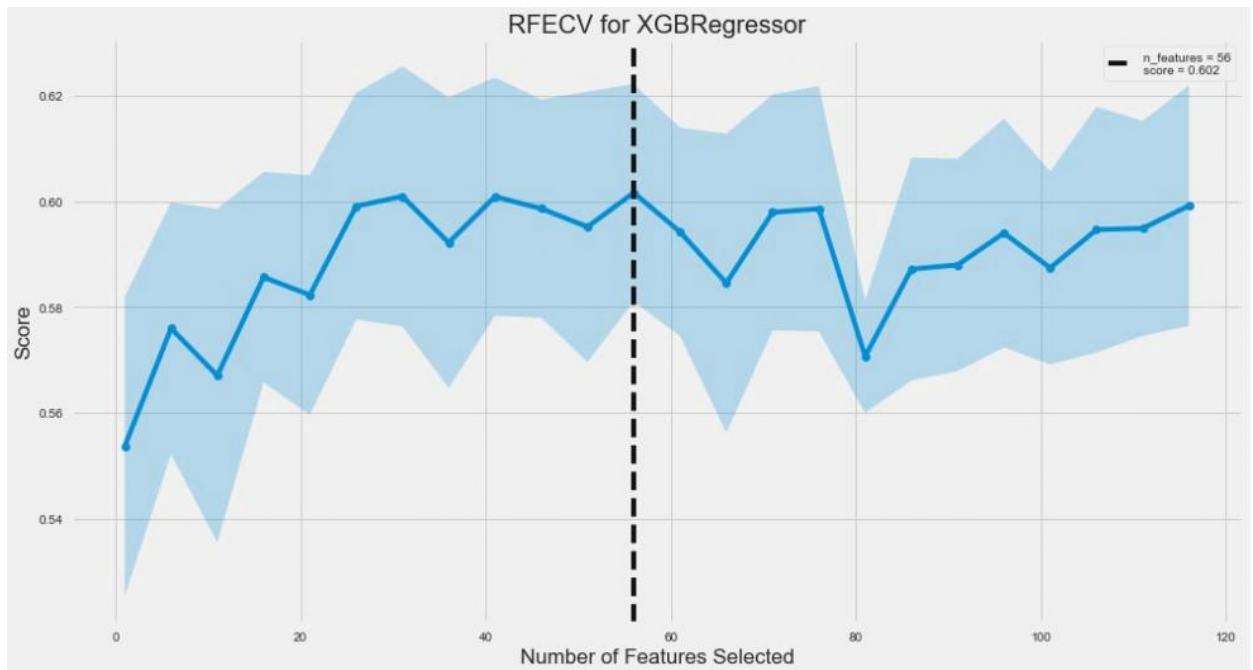


Рисунок 4.2 – Рекурсивна оптимізація функцій

З наведеного вище рисунку можна спостерігати кількість оптимальних функцій = 56 з рахунком=0,602.

Дані бронювання складаються з часу реєстрації та часу відвідування, за допомогою якого розраховали годинний проміжок між реєстрацією та відвідуванням, після чого розділили годинний проміжок на основі тривалості розриву.

Ресторани згруповані на основі жанру в кожній області, що дає підрахунок кількості ресторанів даного жанру в даній області. публікуємо всі відгуки та рейтинги, щодо району. Використовуючи календарні дані, з'ясували неробочі дні з урахуванням свят і вихідних, а також за день до свята.

Розраховали статистику відвідувача на основі робочих і неробочих днів, оскільки це дає інформацію про відвідування в робочий і неробочий день, як спостерігали раніше, що кількість відвідувачів більше у святкові дні.

Використовуючи дані, розраховали день тижня, день місяця, тиждень року та місяць року.

Також розраховали статистику відвідувача щодня, щотижня, щомісяця та щорічно.

Текстові дані з назви області та назви жанру для кожного ресторану використовуються для обчислювальних векторів TF-IDF, на яких SVD застосовується за допомогою методу описаного вище.

Остаточні функції вибираються за допомогою рекурсивної функції. Гіперпараметр налаштований та навчений KNeighbors Regressor, SGDRegression, Decision Tree Regressor, Random Forest Regressor та XGBoost (GBDT) та XGBoost виявився кращим.

4.3 Пояснення моделі

При налаштуванні гіпер параметра з'ясували, що $k = 19$ найкращий.

При налаштуванні позбавилися від ефекту викидів і для чисельної стабільності $\log(y)$. Модель навчання та прогнозування:

```
knn = KNeighborsRegressor(n_neighbors=19,n_jobs=-1)
knn.fit(X_train,np.log1p(y_train))
# saving model
pickle.dump(knn, open(generated + 'knn_model', 'wb'))

# loading model
knn = pickle.load(open(generated + 'knn_model', 'rb'))
# predicting
y_pred = knn.predict(X_val)
RMSLE = rmsle(y_true=y_val,y_pred=np.expml(y_pred))

print('RMSLE for KNeighborsRegressor:-',RMSLE)
```

RMSLE отримали для 0.52. Спочатку стандартизували дані, перш ніж передавати їх до моделі, як найкраще працює зі стандартизованими даними. SGDR, при налаштуванні гіпер параметра дізналися, що найкраще $\alpha=0.0001$. SGDR модель навчання та прогнозування:

```
# standardising data
std = StandardScaler()
```

```

X_train_std = std.fit_transform(X_train)
X_val_std = std.fit_transform(X_val)
# model
sgd = SGDRegressor(alpha=0.0001)
sgd.fit(X_train_std,np.log1p(y_train))
# saving model
pickle.dump(sgd, open(generated + 'sgd_model', 'wb'))

# loading model
sgd = pickle.load(open(generated + 'sgd_model', 'rb'))
y_pred = sgd.predict(X_val_std)
RMSLE = rmsle(y_true=y_val,y_pred=np.expml(y_pred))
print('RMSLE for SGDRegressor:-',RMSLE)

```

RMSLE отримали для 0.53. На налаштування гіпер параметра дізналися, `max_depth=10min_sample_split=500`. Тут також використовували дерево рішень упаковане значення в лист, тому що значення "ближче" один до одного $\log(y)$.

DecisionTreeRegressor модель навчання та прогнозування.

```

dt = DecisionTreeRegressor(max_depth = 10, min_samples_split = 500)
dt.fit(X_train,np.log1p(y_train))
# saving model
pickle.dump(dt, open(generated + 'dt_model', 'wb'))
# loading model
dt = pickle.load(open(generated + 'dt_model', 'rb'))
y_pred = dt.predict(X_val_after)
RMSLE = rmsle(y_true=y_val,y_pred=np.expml(y_pred))
print('Validation-RMSLE for DecisionTreeRegressor:-',RMSLE)

```

RMSLE отримали 0.49. При налаштуванні гіпер параметра отримали , `.max_depth=50n_estimators=500`.

RandomForestRegressor модель навчання та прогнозування:

```

rf = RandomForestRegressor(max_depth = 50, n_estimators = 500, n_jobs=-1)
rf.fit(X_train,np.log1p(y_train))
# saving model

```

```

pickle.dump(rf, open(generated + 'rf_model', 'wb'))
# loading model
y_pred = rf.predict(X_val)
RMSLE = rmsle(y_true=y_val, y_pred=np.expml(y_pred))
print('Validation-RMSLE of RandomForestRegressor:--', RMSLE)

```

RMSLE отримали 0.47. Оскільки це регресія проблемна використовували мета, як .reg: squarederror

Під час навчання використовували квадрат помилки як показник оцінювання. При налаштуванні гіперпараметра з'ясували, що .learning_rate=0.01 min_child_weight=0.8 subsample=0.7 colsample_bytree=0.5 max_depth=8

Параметр для графічного процесора оптимізований навчання tree_hist='gpu_hist'.

XGBRegressor модельного навчання та прогнозування.

```

# best parameters
parameters = {'objective': 'reg:squarederror', # objective for regression
              'eval_metric': 'rmse', # root mean squared error
              'learning_rate': 0.01,
              'min_child_weight': 0.8,
              'subsample': 0.7,
              'colsample_bytree': 0.5,
              'max_depth': 8,
              'tree_method': 'gpu_hist'} # GPU optimized training

# optimized xgb.DMatrix
train_matrix = xgb.DMatrix(data=X_train, label=np.log1p(y_train))
val_matrix = xgb.DMatrix(data=X_val, label=np.log1p(y_val))

# training model
model = xgb.train(params=parameters,
                  dtrain=train_matrix,
                  evals=[(train_matrix, 'train'), (val_matrix, 'test')],
                  num_boost_round = 1000)

```

```
# root mean squared logarithmic error
RMSLE = rmsle(y_val,np.expml(model.predict(val_matrix)))
print('Validation-RMSLE of GBDT (XGBoost):-',RMSLE)
```

RMSLE отримали 0.45.

4.6 Порівняння моделей

Порівняння всіх моделей полягає в наступному:

Model	RMSLE
KNeighbors	0.5257
SGD	0.5363
DecisionTree	0.4972
RandomForest	0.4757
XGB (GBDT)	0.4545

З наведених вище даних можна зробити висновок, що найкраща модель XGB.

Висновки до розділу 4

Як підтверджується висновками, соціальні аспекти являють собою важливий атрибут, на який клієнти звертають увагу при визначенні їх загального досвіду відвідування закладів харчування. Відповідно, ресторатори повинні зосередитися на навичках клієнтів і навичках своїх службових співробітників. Зокрема, працівники повинні бути хорошими з точки зору їхніх комунікативних навичок та їх службової поведінки. Крім того, обслуговуючий персонал повинен бути обізнаний про пункти меню щодо своїх інгредієнтів, або навіть їх історії. Оскільки їжа відображає культуру призначення, конкретні знання про місцеву кухню повинні

бути пріоритетними для співробітників служби. Крім того, що використовується як рекламний інструмент, місцеві кухні можуть допомогти напрямкам диверсифікувати туристичні пропозиції, а також створити привабливий економічний вплив. Тобто регіони в пункті призначення відомі кухні можуть бути пропаговані як гастрономічні регіони. При цьому напрямки можуть як збільшити свої туристичні пропозиції, так і сприяти економічному розвитку місцевих районах. Крім того, ці райони можна було б перетворити на основні зони атракціонів з подіями та фестивалями. Що стосується диверсифікації туристичних пропозицій, місцева кухня сприяє різноманітним зусиллям позиціонування місця призначення. Зокрема, місцева кухня може допомогти залучити клієнтів, чією основною мотивацією є досвід різних кухонь і культур. Таким чином, клієнти можуть бути орієнтовані як альтернативний сегмент ринку для призначення. У свою чергу, напрямки могли б збільшити як кількість клієнтів.

Загальні висновки

Крім важливості, респонденти також оцінювали своє задоволення кожним фактором. Як показує аналіз, заклади харчування демонструють високі показники якості продуктів харчування та продовольчих культурних аспектів. Однак якість продуктів харчування не оцінюється так важливо, як продовольчі культурні аспекти. Отже, продовольчі культурні аспекти являють собою точку диференціації для місцевих ресторанів. Тобто, як цільові організації, так і заклади харчування повинні зосередитися на культурних аспектах харчування та позиціонувати місцеву кухню, підкреслюючи унікальні аспекти харчування. Інша важливий елемент цього дослідження стосується соціальних аспектів. Незважаючи на те, що респонденти вважають соціальні аспекти важливими, рівень задоволеності був низьким для цього фактора. Цей результат передбачає, що заклади харчування повинні більше зосередитися на службовій поведінці співробітників. Нарешті, фізичне середовище було встановлено, що мають низький пріоритет для обідніх закладів.

Асимптотичні рамки, розглянуті тут, є найпростішою нетривіальною моделлю для матриці ознак. Це дозволяє обчислити необхідну точність. Нефундаментальні елементи матриці зменшують білий шум, який є структурними частинами складних моделей. Для прикладу, середнє значення помилки будь-якого показника на основі зниження особливого значення містить компонент шуму в одиниці даних, і цей компонент визначає фундаментальну нижню межу. Результати обрахунків, розраховані в цій моделі, які не прикріплені до конкретного припущення про ранг (наприклад, константи, які визначають мінімакс) залишаються по суті правильними в більш складних моделях.

Рішення, як теоретичний ландшафт, досить просте: існує унікальне допустиме правило жорсткого порогу, і, крім того, унікальне допустиме правило допуску для певного значення. Це, звичайно, зовсім відрізняється від ситуації, з

якою стикаються, наприклад, при оцінці нормальних розподілів даних. Причина - спрощення моделі. Наприклад, замінили значення одних даних, які є випадковими для скінченного розміру матриці, з їх майже фіксованими межами, а також нехтували своїми випадковими коливаннями навколо цих меж. Ці коливання можна пояснити зміною параметрів.. Однак, включно з цими умовами другого порядку в безсимптомних розподілах, можна досягти поліпшення другого порядку в квадратичній помилці над рекомендованим оптимальним порогом обмеження.

Перелік посилань

1. Hu, Yangzhou, and J. R. Brent Ritchie. 1993. Measuring destination attractiveness: A contextual approach. *Journal of Travel Research* 32: 25–34.
2. Ignatov, Elena, and Stephen Smith. 2006. Segmenting Canadian culinary tourists. *Current Issues in Tourism* 9: 235–55.
3. Jacobsen, Jens K. Steen, and Jan Vidar Haukeland. 2002. A lunch with a view: Motor tourists' choices and assessments of eating-places. *Scandinavian Journal of Hospitality and Tourism* 2: 4–16.
4. Jang, SooCheong Shawn, and Young Namkung. 2009. Perceived quality, emotions, and behavioral intentions:
Application of an extended Mehrabian–Russell model to restaurants. *Journal of Business Research* 62: 451–60.
6. Jenkins, Ian, and Andrew Jones. 2003. A Taste of Wales–Blas Ar Gymru': Institutional malaise in promoting Welsh food tourism products. In *Tourism and Gastronomy*. London: Routledge, pp. 129–45.
7. Josiam, Bharath M., Melissa Mattson, and Pauline Sullivan. 2004. The Histaunt: Heritage tourism at Mickey's dining car. *Tourism Management* 25: 453–61.
8. Kauppinen-Räsänen, Hannele, Johanna Gummerus, and Katariina Lehtola. 2013. Remembered eating experiences described by the self, place, food, context and time. *British Food Journal* 115: 666–85.
9. Kim, Jong-Hyeong, and J. R. Brent Ritchie. 2014. Cross-cultural validation of a memorable tourism experience scale (MTES). *Journal of Travel Research* 53: 323–35.
10. Kim, Yeong Gug, BoWon Suh, and Anita Eves. 2010. The relationships between food-related personality traits, satisfaction, and loyalty among visitors

attending food events and festivals. *International Journal of Hospitality Management* 29: 216–26.

11. Kivela, Jaksa, and John C. Crotts. 2006. Tourism and gastronomy: Gastronomy's influence on how tourists experience a destination. *Journal of Hospitality & Tourism Research* 30: 354–77.

12. Kivela, Jaksa J., and John C. Crotts. 2009. Understanding travellers' experiences of gastronomy through etymology and narration. *Journal of Hospitality & Tourism Research* 33: 161–92.

13. Kivela, Jakia, Robert Inbakaran, and John Reece. 2000. Consumer research in the restaurant environment. Part 3: Analysis, findings and conclusions. *International Journal of Contemporary Hospitality Management* 12: 13–30.

14. Kotler, Philip, and David Gertner. 2002. Country as brand, product, and beyond: A place marketing and brand management perspective. *Journal of Brand Management* 9: 249–61.

15. Lewin, Kurt. 1938. *The Conceptual Representation and Measurement of Psychological Forces*. Durham: Duke University Press.

16. López-Guzmán, Tomas, and Sandra Sánchez-Cañizares. 2012. Culinary tourism in Córdoba (Spain). *British Food Journal* 114: 168–79.

17. Mak, Athena H. N., Margaret Lumbers, and Anita Eves. 2012. Globalization and food consumption in tourism. *Annals of Tourism Research* 39: 171–96.

Додатки

```

# standardising data
std = StandardScaler()
X_train_std = std.fit_transform(X_train)
X_val_std = std.fit_transform(X_val)
# model
sgd = SGDRegressor(alpha=0.0001)
sgd.fit(X_train_std,np.log1p(y_train))
# saving model
pickle.dump(sgd, open(generated + 'sgd_model', 'wb'))

# loading model
sgd = pickle.load(open(generated + 'sgd_model', 'rb'))
y_pred = sgd.predict(X_val_std)
RMSLE = rmsle(y_true=y_val,y_pred=np.expml(y_pred))
print('RMSLE for SGDRegressor:-',RMSLE)

parameters = {'objective':'reg:squarederror', # objective for regression
              'eval_metric':'rmse', # root mean squared error
              'learning_rate':0.01,
              'min_child_weight':0.8,
              'subsample':0.7,
              'colsample_bytree':0.5,
              'max_depth': 8,
              'tree_method':'gpu_hist'} # GPU optimized training

# optimized xgb.DMatrix
train_matrix = xgb.DMatrix(data=X_train,label=np.log1p(y_train))
val_matrix = xgb.DMatrix(data=X_val,label=np.log1p(y_val))

```

```
knn = KNeighborsRegressor(n_neighbors=19,n_jobs=-1)
knn.fit(X_train,np.log1p(y_train))
# saving model
pickle.dump(knn, open(generated + 'knn_model', 'wb'))

# loading model
knn = pickle.load(open(generated + 'knn_model', 'rb'))
# predicting
y_pred = knn.predict(X_val)
RMSLE = rmsle(y_true=y_val,y_pred=np.expm1(y_pred))
```

Додаток Б

Ксерокопії наукових публікацій, виконаних при роботі над
дипломною роботою магістр

Міністерство освіти і науки України
Хмельницький національний університет



ЗБІРНИК НАУКОВИХ ПРАЦЬ
за матеріалами XII всеукраїнської науково-практичної конференції
«Актуальні проблеми комп'ютерних наук АПКН-2020»

9-10 листопада 2020

Хмельницький 2020

УДК 004:37:001:62

Збірник наукових праць за матеріалами XII всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2020». Хмельницький – 2020. – 365с.

У збірнику наукових праць подані перспективні практичні розробки аспірантів, студентів та здобувачів в області сучасних інформаційних технологій. Розглянуто актуальні проблеми комп'ютерних наук, комп'ютерної інженерії, прикладної математики й інженерії програмного забезпечення, приведено ряд робіт по впровадженню інформаційних технологій у виробництво та управління. Висвітлено перспективні розробки сучасних систем пошуку, обробки й захисту інформації, медійних та комунікаційних системи.

УДК 004:37:001:62

Матеріали конференції відтворені з авторських оригіналів. Оргкомітет конференції висловлює подяку учасникам конференції та сподівається на подальшу співпрацю.

З питань проведення конференції та подальшого обміну інформацією звертатись на e-mail конференції: apkt.khnu@gmail.com

АКТУАЛЬНІ ПРОБЛЕМИ КОМП'ЮТЕРНИХ НАУК - 2020

XII Всеукраїнська науково-практична конференція

Метою конференції є висвітлення актуальних проблем комп'ютерних наук, інформатики та інформаційних технологій.

СЕКЦІЇ КОНФЕРЕНЦІЇ:

1. Комп'ютерні науки та прикладні інформаційні технології.
2. Комп'ютерна інженерія та системи захисту інформації.
3. Математичне моделювання та інженерія програмного забезпечення
4. Телерадіокомунікації, медійні та комунікаційні системи.
5. Проблеми впровадження інформаційних технологій у виробництво та управління.

Робочі мови конференції: українська, англійська

ОРГКОМІТЕТ:

СИНЮК О. М. голова оргкомітету, проректор Хмельницького національного університету з наукової роботи, доктор технічних наук, професор
СОРОКАТИЙ Р. В. заступник голови оргкомітету, завідувач кафедри Комп'ютерних наук та інформаційних технологій ХНУ, доктор технічних наук, професор

БАРМАК О. В. заступник голови оргкомітету, доктор технічних наук, професор

САВЕНКО О. С. декан Факультету програмування та комп'ютерних і телекомунікаційних систем ХНУ, доктор технічних наук, професор

ВИСОЦЬКА О. В. доктор технічних наук, завідувач кафедри радіоелектронних та біомедичних комп'ютеризованих засобів і технологій Національного аерокосмічного університету ім. М. Є. Жуковського «Харківський авіаційний інститут», професор

ЛАВРОВ Є. А. доктор технічних наук, професор (Сумський державний університет)

ТІМОФЄЄВА Л. В. відповідальна за студентську науково-дослідну роботу ХНУ

МАЗУРЕЦЬ О. В. секретар конференції, старший викладач кафедри комп'ютерних наук та інформаційних технологій ХНУ

КОНТАКТНА ІНФОРМАЦІЯ:

e-mail для листування: apkt.khnu@gmail.com

УДК 004.4

Рибчинський Б. О., Доброловський В. В., Медведчук В. Ю.

Хмельницький національний університет

ПРОГНОЗУВАННЯ ЗАВАНТАЖЕНОСТІ РЕСТОРАНУ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ

Розроблений і реалізований метод отримання інформації на підставі аналізу та застосування методів машинного навчання з метою прогнозування завантаженості ресторану у різні періоди функціонування. Запропоновані методи та підходи можна застосовувати в практичному руслі а також в порівняних сферах із предметною областю.

Developed and implemented a method of obtaining information based on the analysis and application of machine learning methods to predict the load of the restaurant in different periods of operation. The proposed methods and approaches can be applied in practice, as well as in comparable areas with the subject area.

Збільшення попиту на туризм і конкуренція між напрямками були пов'язані з наслідками глобалізації. Напрямки конкурують на дедалі більш конкуруючих ринках, мало що відрізняє їх один від одного. Це змусило напрямки зосередитися на відмінності місця, використовуючи місцеві атрибути при запуску інноваційних, нових продуктів і брендів, що може допомогти створити більш унікальну пропозицію продажу.

Асимптотичні рамки, розглянуті тут, є, мабуть, найпростішою нетривіальною моделлю для матриці ознак. Це дозволяє обчислити, в основному будь-яку кількість відсотків, для будь-якого денозера інтересу. Недаментальні елементи матриці денозуючий білий шум, який є менш складними моделями, присутні ще зрозумілі і кількісні. Для наприклад, AMSE будь-якого показника на основі зниження особливого значення містить компонент через шумове забруднення в однині даних, і цей компонент визначає фундаментальну нижню межу. Кон'юнктура, як результати обрахунків, розраховані в цій моделі, які не прикріплені до конкретного припущення про ранг (наприклад, константи, які визначають мінімакс) залишаються по суті правильними в більш складних моделях.

Спостереження, мінімум відвідувачів, які можемо спостерігати з цього сюжету, майже досягає нуля. Серед відвідувачів 20 відсотків.

Максимальна кількість відвідувачів становить від 55 до 60. спостерігаємо певні дуже високі значення (outlier) більше 60 і навіть більше, ніж 100 відвідувачів. 25-і процентильне та 75-і процентильні значення 13. Спостереження, поширення бронювання вище, ніж у резервації HPG.

Існує велика кількість бронювань в НРГ з відвідувачами розраховувати від 5 до 10. Є кілька застережень в НРГ, де кількість відвідувачів більше 20 або навіть досягнення 40.

Це дослідження намагається зрозуміти взаємозв'язок між важливістю і продуктивністю (задоволенням) досвіду місцевої кухні, а також вивчити вплив досвіду їдальні на задоволення призначення. Що стосується важливих факторів для їдальні, найбільш важливим фактором для мандрівників є продовольчі культурні аспекти. Зокрема, туристи очікують, що місцева їжа буде відображати унікальні культурні аспекти для приготування їжі, презентації та зовнішнього вигляду. Дотримуючись культурних аспектів, соціальних аспектів, які є співробітниками та інші поведінки клієнтів, є другим за важливості фактором для іноземних мандрівників.

Крім важливості, респонденти також оцінювали своє задоволення кожним фактором. Як показує аналіз, заклади харчування демонструють високі показники якості продуктів харчування та продовольчих культурних аспектів. Однак якість продуктів харчування не оцінюється як така важлива, як продовольчі культурні аспекти. Отже, продовольчі культурні аспекти являють собою точку диференціації для місцевих ресторанів. Тобто, як цільові організації, так і заклади харчування повинні зосередитися на культурних аспектах харчування та позиціонувати місцеву кухню, підкреслюючи унікальні аспекти харчування. Інша важлива знахідка цього дослідження стосується соціальних аспектів. Незважаючи на те, що респонденти вважають соціальні аспекти важливими, рівень задоволеності був низьким для цього фактора. Цей результат передбачає, що заклади харчування повинні більше зосередитися на службовій поведінці співробітників. Нарешті, фізичне середовище було встановлено, щоб мають низький пріоритет для обідніх стабілізацій.

Машинне навчання має багато застосувань в ресторанах від допомоги кухарям на кухнях до прогнозування майбутніх продажів, що, в свою чергу, допомагає опрацювати вимоги до управління робочою силою та інвентаризацією.

Перелік посилань

1. Jang, SooCheong Shawn, and Young Namkung. 2009. Perceived quality, emotions, and behavioral intentions.
2. Application of an extended Mehrabian–Russell model to restaurants. *Journal of Business Research* 62: 451–60.
3. Jenkins, Ian, and Andrew Jones. 2003. A Taste of Wales—Blas Ar Gymru': Institutional malaise in promoting Welsh food tourism products. In *Tourism and Gastronomy*. London: Routledge, pp. 129–45.



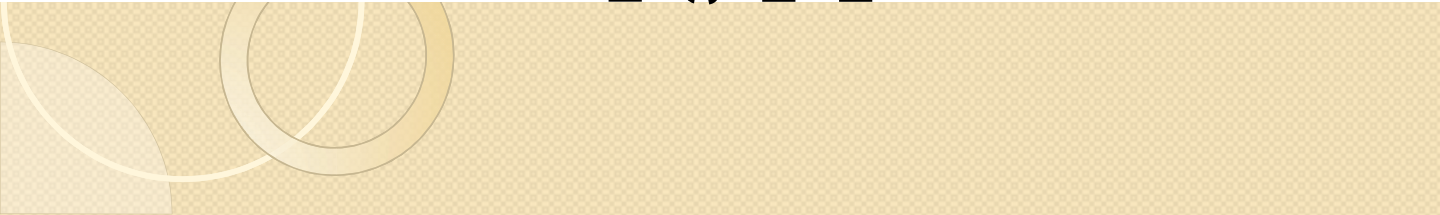
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ДИПЛОМНА РОБОТА
МАГІСТРА

Прогнозування завантаженості ресторану
з використанням штучного інтелекту

Розробив
ст. гр. КНМ-19-1-1: Рибчинський Б.О.

Хмельницький - 2020



В магістерській роботі розроблений і реалізований метод отримування інформації на підставі аналізу та застосування методів машинного навчання з метою прогнозування завантаженості ресторану у різні періоди функціонування.



Метою дослідження є розробка методу прогнозування завантаженості ресторану та визначенню основних факторів впливу.

Задачі дослідження:


- показати, що застосування методів штучного інтелекту в ресторанній справі на сучасному рівні розвитку інформаційних технологій може в значній мірі підвищити бізнесові показники;
- провести дослідження ознак і класифікаторів застосованих в машинному навчанні;
- провести порівняння моделей машинного навчання та їх вплив на цільову функцію.




Об'єктом дослідження є порівняння методів машинного навчання для предметної області, а саме прогнозування завантаженості ресторанів.

Предметом дослідження є набір даних предметної області та методи розробки моделей машинного навчання.

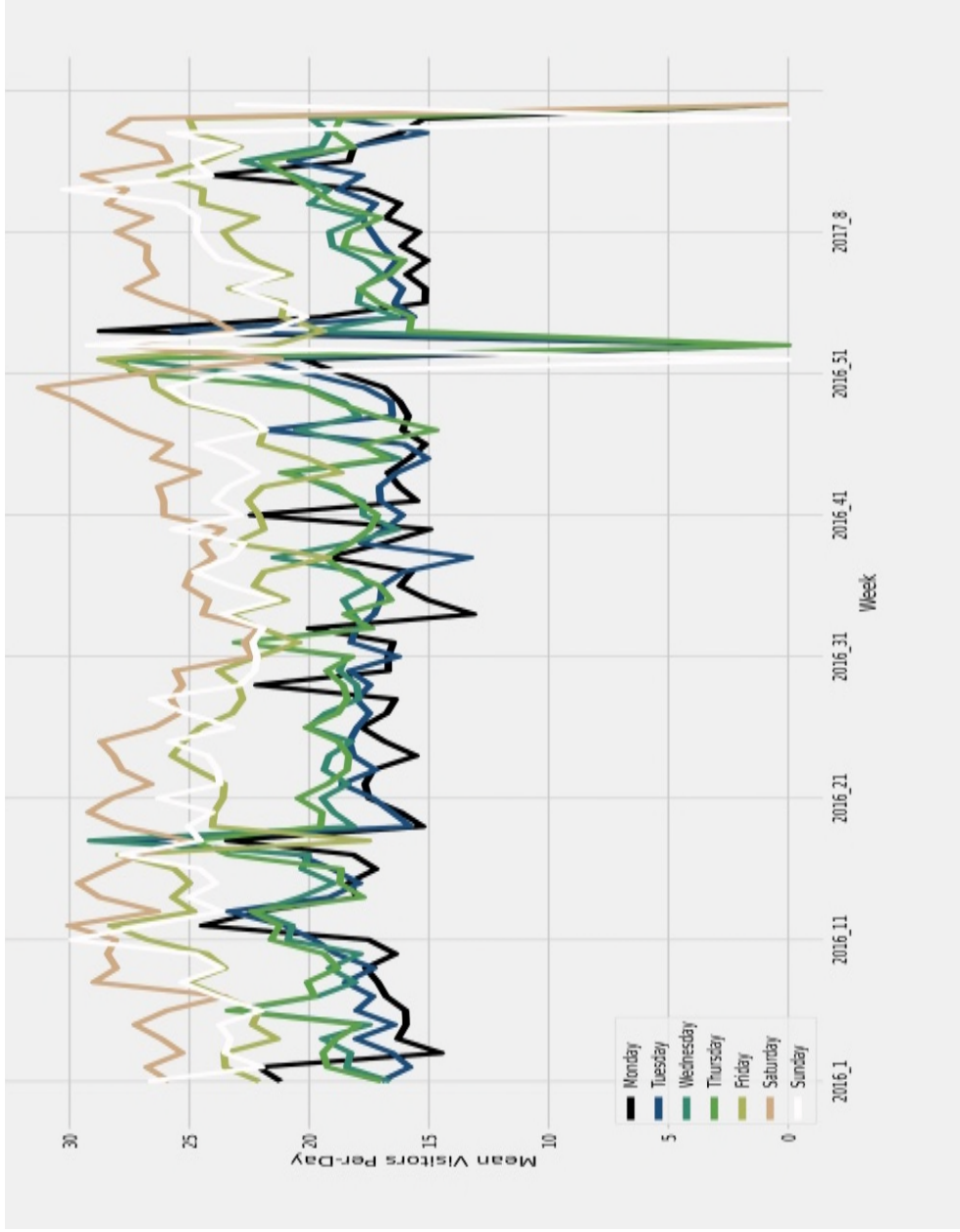
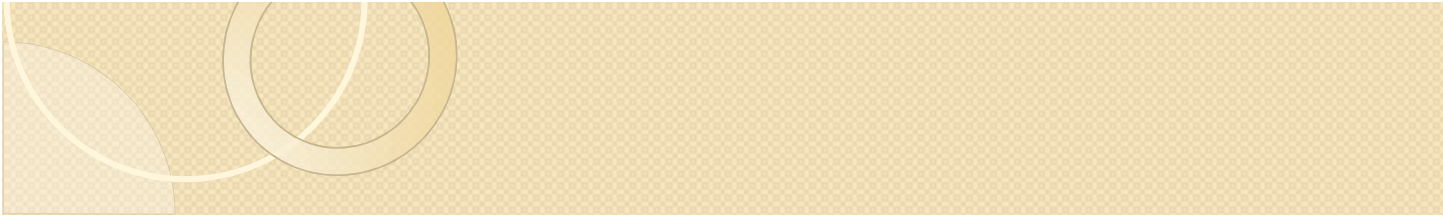
Дослідження присвячено пошуку взаємозв'язків між важливістю і продуктивністю (задоволенням) місцевої кухні, а також вивчити вплив досвіду надання сервісу кухні на задоволення клієнтів. Що стосується важливих факторів для їдальні, найбільш важливим фактором для клієнтів є продовольчі культурні аспекти. Зокрема, туристи очікують, що місцева їжа буде відображати унікальні культурні аспекти для приготування їжі, презентації та зовнішнього вигляду. Дотримуючись культурних аспектів, соціальних аспектів, які є важливими факторами та інші аспекти поведінки клієнтів, є наступним за важливістю фактором для клієнтів. Як третій важливий фактор, якість продуктів харчування та дотримання соціальних аспектів.



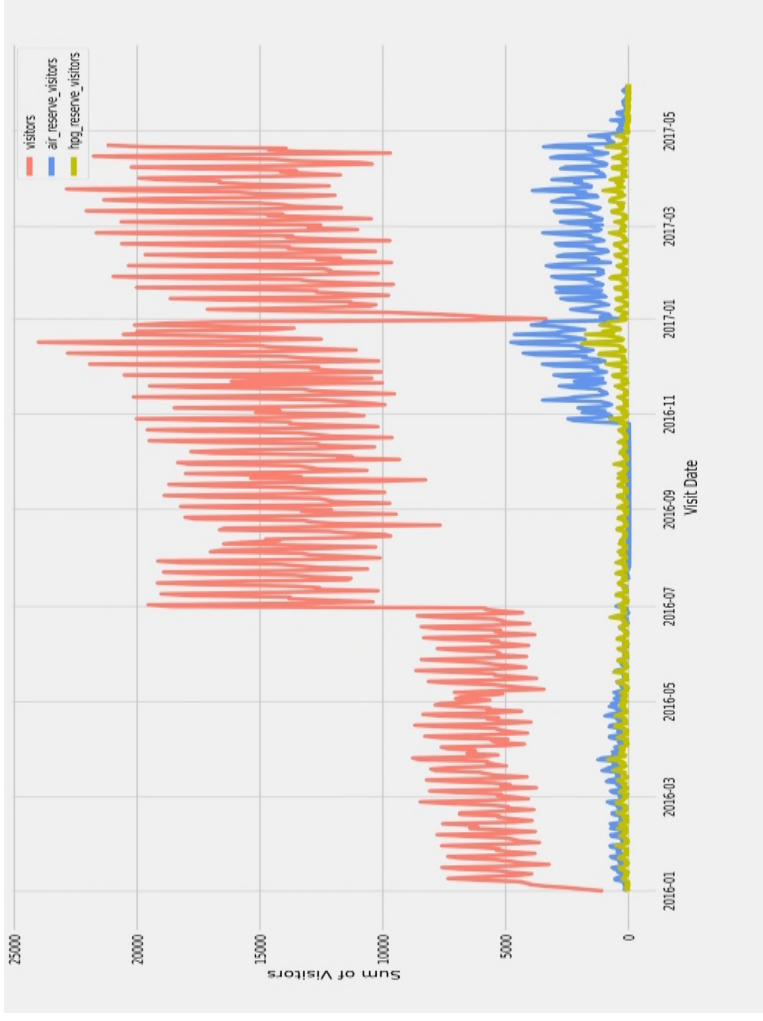
Небазові елементи матриці ознак зменшують білий шум, який є структурними частинами складних моделей. Для прикладу, середня помилка будь-якого показника на основі зниження особливого значення містить компонент шуму в одиниці даних, і цей компонент визначає фундаментальну нижню межу. Результати обрахунків, розраховані в цій моделі, які не прикріплені до конкретного припущення про ранг (наприклад, константи, які визначають мінімакс) залишаються по суті правильними в більш складних моделях, що і визначає достовірність отриманих розрахункових даних.



Практична значимість дослідження полягає в тому, що запропоновані методи та підходи можна застосовувати в практичному руслі, а також в порівняльних сферах із предметною областю, тобто системах обслуговування та сервісів.



Відвідувачі за днем тижня

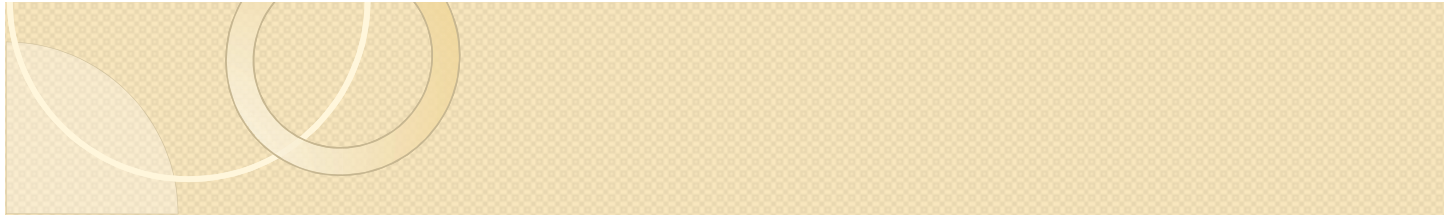


Відвідувачі і бронювання

Висновки

Як показує аналіз, заклади харчування демонструють високі показники якості продуктів харчування та продовольчих культурних аспектів. Однак якість продуктів харчування не оцінюється так важливо, як продовольчі культурні аспекти. Отже, продовольчі культурні аспекти являють собою точку диференціації для місцевих ресторанів. Тобто, як цільові організації, так і заклади харчування повинні зосередитися на культурних аспектах харчування та позиціонувати місцеву кухню, підкреслюючи унікальні аспекти харчування. Інша важливий елемент цього дослідження стосується соціальних аспектів. Незважаючи на те, що респонденти вважають соціальні аспекти важливими, рівень задоволеності був низьким для цього фактора. Цей результат говорить про те, що заклади харчування повинні більше зосередитися на службовій поведінці співробітників. Нарешті, було встановлено що фізичне середовище має низький пріоритет для закладів.

Дякую за увагу



Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 1.0%

Словники перевірки: en_US, ru_RU, ua_UA. Помилоч в документах: 5%

ID: 81735 Назва: Прогнозування завантаженості ресторану з використанням штучного інтелекту Додано в БД: 2020-11-30 Автора: Рибчинський Богдан Олександрович Керівники: Багрий Р.О. Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	64231	586	723 (1%)	9 (2%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

РІШЕННЯ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Прогнозування завантаженості ресторану з використанням штучного інтелекту

Автор: Рибчинський Б.О.

Спеціальність: 122 Комп'ютерні науки

Науковий керівник: к.т.н., доц. Багрій Р.О.


Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних). Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи	-
3	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	-
4	Інше:	-

Підтвердження: Виявленні запозичення не є плагіатом т.я. розміщені в розділах, які не описують безпосередньо авторське дослідження, складають 4,6% та мають посилання на приведений список літературних джерел.

01.11.2020

Дата


Підпис керівника


Підпис завідувача кафедри

ВІДГУК ОЦОНЕНТА
на дипломну роботу магістра

Магістра *гр. КНМ-19-1* Рибчинського Богдана Олександровича

На тему: Прогнозування завантаженості ресторану з використанням штучного інтелекту

1. Актуальність і значення теми

В магістерській роботі розроблений і реалізований метод отримання інформації на підставі аналізу та застосування методів машинного навчання з метою прогнозування завантаженості ресторану у різні періоди функціонування. Проведені дослідження використовуючи набори даних та дозволяють оптимізувати діяльність закладів що має практичну значимість.

2. Оцінка якості та достовірності проведених досліджень.

Достовірність результатів забезпечується експериментальними дослідженнями та методами статистичної перевірки. Також було проведено серію експериментальних досліджень на різноманітних наборах даних.

3. Оцінка запропонованих заходів та пропозицій, практичної цінності та ефективності.

Практична значимість дослідження полягає в тому, що запропоновані методи та підходи можна застосовувати в практичному руслі, а також в порівняльних сферах із предметною областю, тобто системах обслуговування та сервісів.

4. Загальний висновок та оцінка

Робота виконана на належному науково-практичному рівні. Пояснювальна записка оформлена в відповідності з необхідними нормами. За структурою, описом завданнями та методами їх вирішення відповідає вимогам вищої школи і вимогам, що пред'являються до освітньо-кваліфікаційного рівня «магістр», а її автор заслуговує присвоєння кваліфікації магістра з комп'ютерних наук та інформаційних технологій.

Робота заслуговує на оцінку « 30006 ».

Опонент Духа О.В., д.т.н., проф., зав. кафедр ТАН ХНУ