

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПЕРЕДБАЧЕННЯ РІВНІВ ЗАДОВОЛЕНOSTІ ПАСАЖИРІВ АВІАКОМПАНІЯМИ

Вінницький національний технічний університет

Анотація

В роботі розроблено технологію передбачення рівнів задоволеності пасажирів авіакомпаніями. Для виконання поставленої задачі було використано такі моделі машинного навчання: виконання поставленої задачі є Logistic Regression, Decision Tree, Random Forest, Extra Trees, KNeighbors.

Ключові слова: Python, передбачення, задоволеність пасажирів, машинне навчання, аналіз.

Abstract

The paper developed a technology for predicting the levels of passenger satisfaction with airlines. The following machine learning models were used to perform the task: Logistic Regression, Decision Tree, Random Forest, Extra Trees, KNeighbors.

Key words: Python, prediction, passenger satisfaction, machine learning, analysis.

Вступ

Щодня інформаційні технології продовжують розвиватися, що сприяє збільшенню обсягів даних у світі. Ці дані надають безліч можливостей для аналізу, класифікації та прогнозування в різних галузях, включаючи авіаційну індустрію. Використання даних може допомогти у проведенні комплексного аналізу для виявлення тенденцій та ідентифікації ключових факторів, що впливають на задоволеність пасажирів авіакомпаніями.

Враховання відгуків пасажирів та аналіз їх поведінки та уподобань є життєво важливим для авіакомпаній, щоб підвищувати рівень задоволеності клієнтів і оптимізувати послуги. Інформаційні технології дозволяють не тільки аналізувати поточний стан задоволеності, але й прогнозувати майбутні потреби та очікування пасажирів.

За допомогою засобів аналізу великих даних та машинного навчання авіакомпанії можуть розробляти персоналізовані пропозиції, покращувати якість обслуговування, і навіть передбачати потенційні проблеми, які можуть негативно вплинути на досвід пасажирів. Це дозволяє компаніям швидко реагувати на зміни у вимогах та бажаннях клієнтів, забезпечуючи їх максимальне задоволення і лояльність[1].

Таким чином, використання передових інформаційних технологій для аналізу та обробки даних є ключовим елементом в стратегії управління клієнтським досвідом в авіаційній індустрії, сприяючи підвищенню рівня задоволеності пасажирів та їхньої відданості авіакомпаніям.

Постановка задачі

Метою роботи є розроблення інформаційної технології передбачення рівнів задоволеності пасажирів авіакомпаніями з використанням методів машинного навчання.

Для досягнення цієї мети необхідно вирішити наступні завдання:

- повести огляд існуючих систем;
- підготувати дані для подальшої роботи;
- провести розвідувальних аналіз даних;
- побудувати моделі та виконати прогнозування;
- оцінити результати роботи моделей.

Результати дослідження

Даними для передбачення було обрано датасет «Airline Passenger Satisfaction» у середовищі Kaggle [3]. Приклад даних з цього датасету показано на рисунках 1–2.

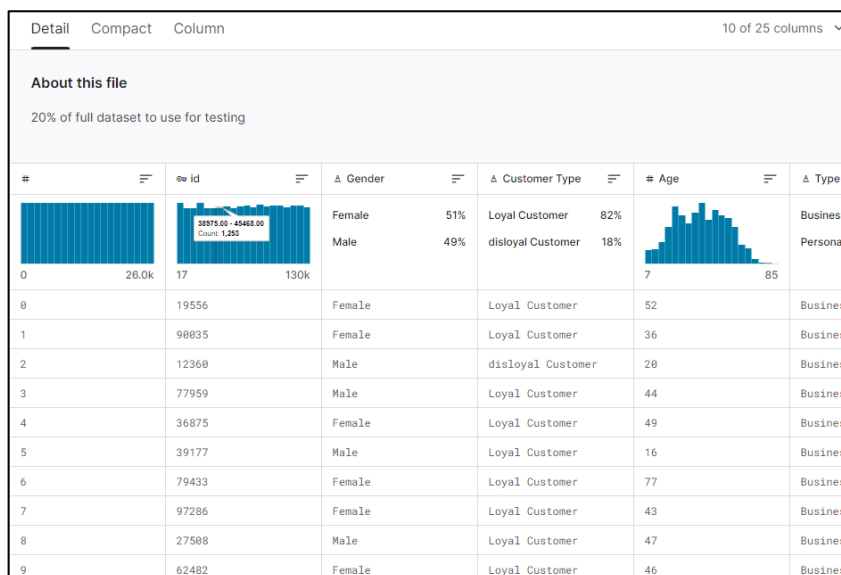


Рис. 1 – Набір даних «test.csv»

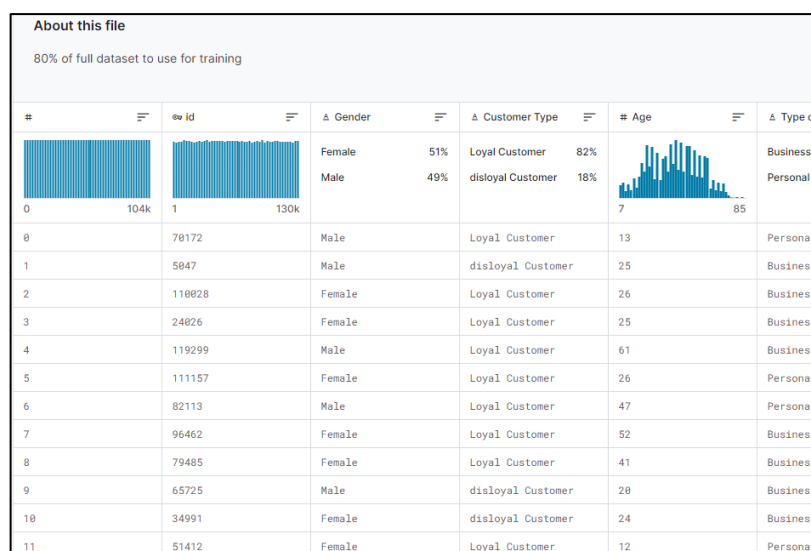


Рис. 2 – Набір даних «train.csv»

Набір даних, призначений для аналізу та передбачення рівнів задоволеності пасажирів авіакомпаніями, складається з двох частин: файлу для тренування train.csv і файлу для тестування test.csv. Файл для тренування містить 104,000 записів, а файл для тестування – 26,000 записів, кожен з яких має 25 стовпців. Ці стовпці включають такі дані, як вік і стать пасажирів, тип клієнта (наприклад, постійний клієнт чи випадковий пасажир), тип подорожі (ділова чи туристична), клас обслуговування на борту (економ, бізнес тощо), відстань польоту, якість служби Wi-Fi на борту, а також оцінка зручності часу виїзду та прибуття. Цей детальний опис дозволяє проводити глибокий аналіз задоволеності пасажирів, що є ключовим для підвищення якості обслуговування авіакомпаній.

Матриці плутанини для побудованих моделей машинного навчання показано на рисунках 3-7.

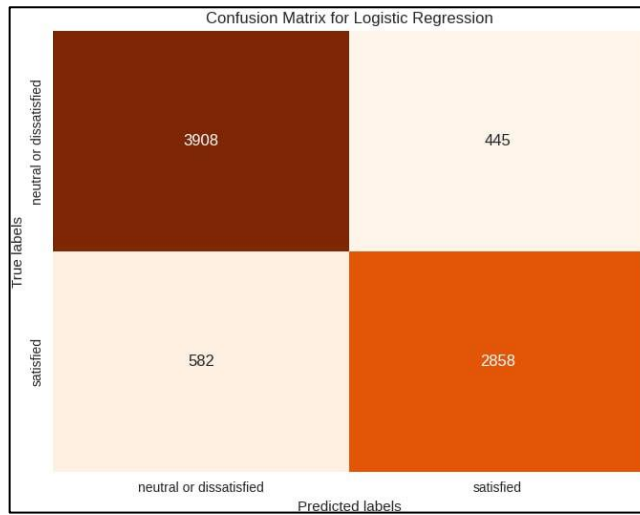


Рис. 3 – Матриця плутанини для моделі Logistic Regression

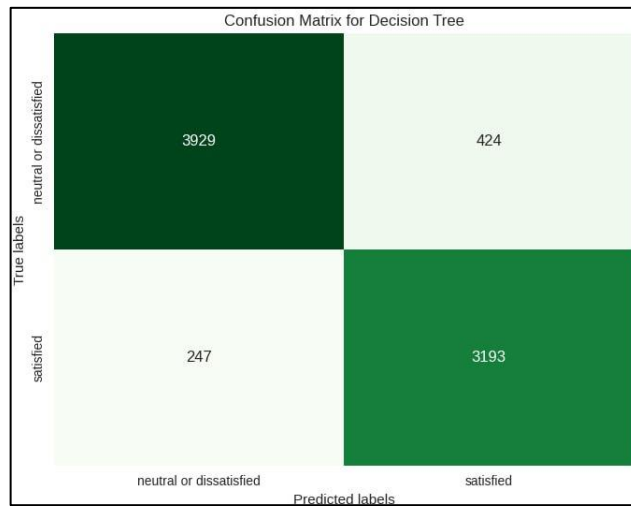


Рис. 4 – Матриця плутанини для моделі Decision Tree

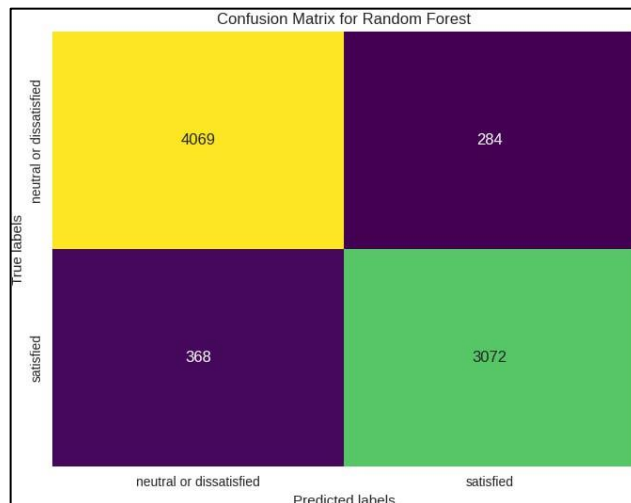


Рис. 5 – Матриця плутанини для моделі Random Forest

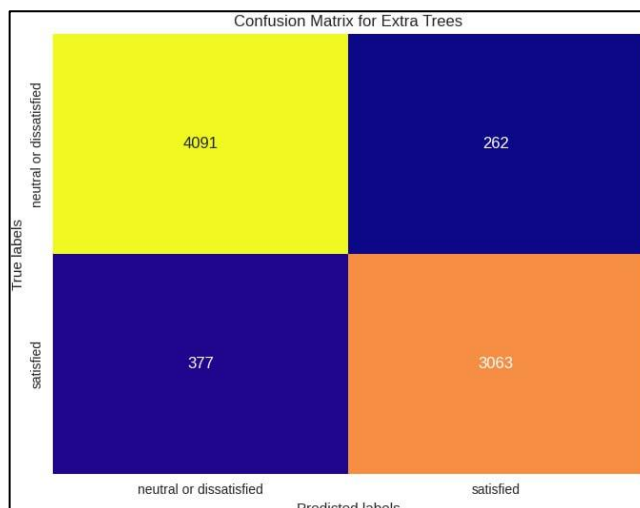


Рис. 6 – Матриця плутанини для моделі Extra Trees

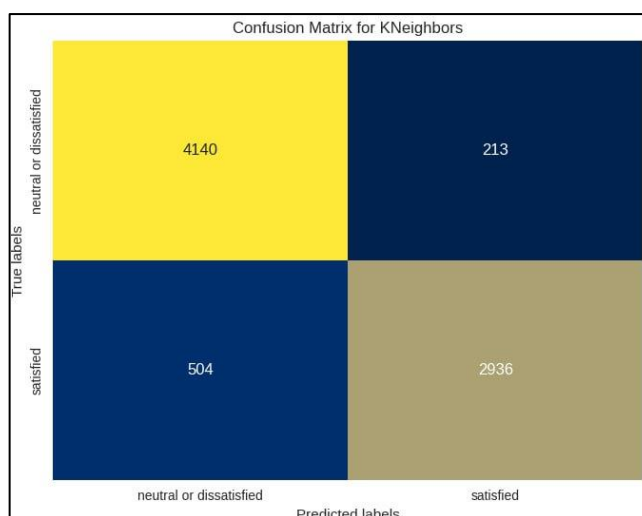


Рис. 7 – Матриця плутанини для моделі KNeighbors

На рис. 8 наведено результати усіх моделей.

Model	Scaling	Type	Precision
Naive Bayes	Normal Data	Gaussian	0.840626
Decision Tree	Normal Data	Gini	0.913897
Random Forest	Normal Data	Gini	0.916335
Extra Trees	Normal Data	Gini	0.918003
KNN	Normal	-	0.907994
Logistic Regression	Normal Data	-	0.868215
AdaBoost	StandardScaler	-	0.913127
GradientBoost	StandardScaler	-	0.915822
LGBM	StandardScaler	-	0.948929

Рис. 8 – Кінцеві результати моделей

Висновки

Під час виконання роботи було реалізовано інформаційну технологію для передбачення рівнів задоволеності пасажирів авіакомпаніями з використанням різних моделей машинного навчання. Результати їх роботи були порівняні між собою, і було визначено найбільш ефективну модель передбачення. У цілому, застосування різних моделей машинного навчання продемонструвало високий рівень точності в передбаченні рівнів задоволеності пасажирів авіакомпаніями. У дослідженні, було порівняно точність восьми різних моделей машинного навчання. Результати показали, що модель LGBM з масштабуванням StandardScaler є найкращою

моделлю для класифікації даних, які використовувалися в дослідженні. Вона має найвищу точність (0,948929).

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Predicting Air Passenger Satisfaction: A Machine Learning Approach [Електронний ресурс] – Режим доступу: [Predicting Air Passenger Satisfaction: A Machine Learning Approach | Estudo Geral \(uc.pt\)](#)
2. Matplotlib Pyplot Documentation. 2024 [Електронний ресурс] – Режим доступу: https://matplotlib.org/3.5.3/api/as_gen/matplotlib.pyplot.html
3. TJ Klein. Kaggle Dataset «Airline Passenger Satisfaction» [Електронний ресурс] – Режим доступу: [Airline Passenger Satisfaction \(kaggle.com\)](#)

Судець Анна Олександрівна – студентка групи 2ICT-206, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: annasudets2.0@gmail.com

Козачко Олексій Миколайович – к.т.н., доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, e-mail: pkom@vntu.edu.ua

Sudets Anna O. - student of Faculty of Intelligent Information Technology and Automation, 2IST-22m, Vinnytsia National Technical University, Vinnytsia, e-mail annasudets2.0@gmail.com

Kozachko Oleksii M. - Ph.D., Assistant Professor of the Department of Systems Analysis and Information Technology, Vinnytsia National Technical University, Vinnytsia, e-mail: pkom@vntu.edu.ua