


Hotel Price Prediction with Classification

GROUP 4

Team Members

6509566 - Sai Wai Yan Phone

6509871 - Ye Min Khant

6509437 - Saung Oo Lwin

6510008 - Sai Mwan Han Hlyan

6509391 - Nyan Lin Aung

6604310 - Kyaw Myo Htet

6508363 - Thuta Oo

6508442 - Htin Linn (Leader)

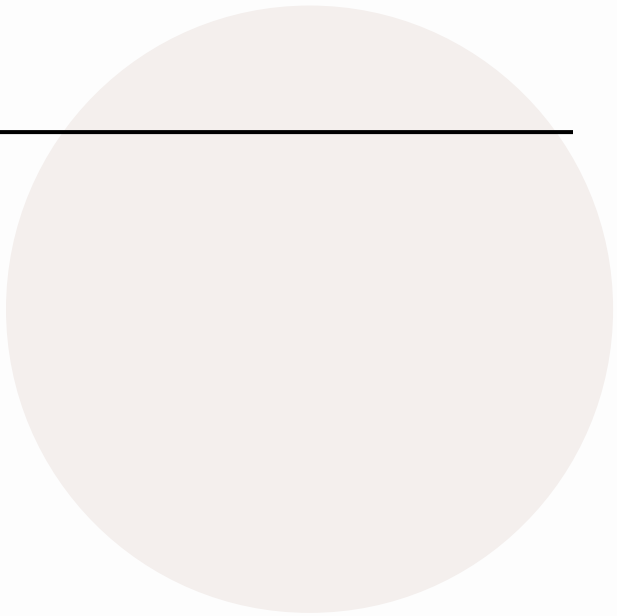
6508444 - May Khine Soe

6504936 - Su Yati

6508490 - Yee Mon Thu

GROUP 4

Problem Definition



Predict the price range (low, medium, high) of hotels based on various attributes, including star rating, location, room type, and available facilities, using a classification model.

Gathered Attributes

- | | | | |
|-----|---|-----|--------------------------------|
| 01. | Location (1 - 3, Major City, Beaches, Countryside) | 06. | Spa (1 or 0, Yes/No) |
| 02. | Star Rating (1 - 5) | 07. | Resturant/Bar (1 or 0, Yes/No) |
| 03. | Room Type (1 - 4, Budget/Small, Standard, Superior, Deluxe +) | 08. | Parking (1 or 0, Yes/No) |
| 04. | Swimming Pool (1 or 0, Yes/No) | 09. | Customer Reviews (0 - 10) |
| 05. | Fitness Center (1 or 0, Yes/No) | 10. | Price (Numeric) |
-

Data Gathering

Hotel price data will be collected exclusively from booking.com for hotels located in Thailand.

The chosen dates for data collection are on weekends, April 26th-27th, as this period avoids significant holidays in Thailand, ensuring that the prices reflect standard, non-peak rates and provide a more accurate baseline for analysis.

We will be gathering about 200 entries.

SELECTED DATE

Data Gathering Method

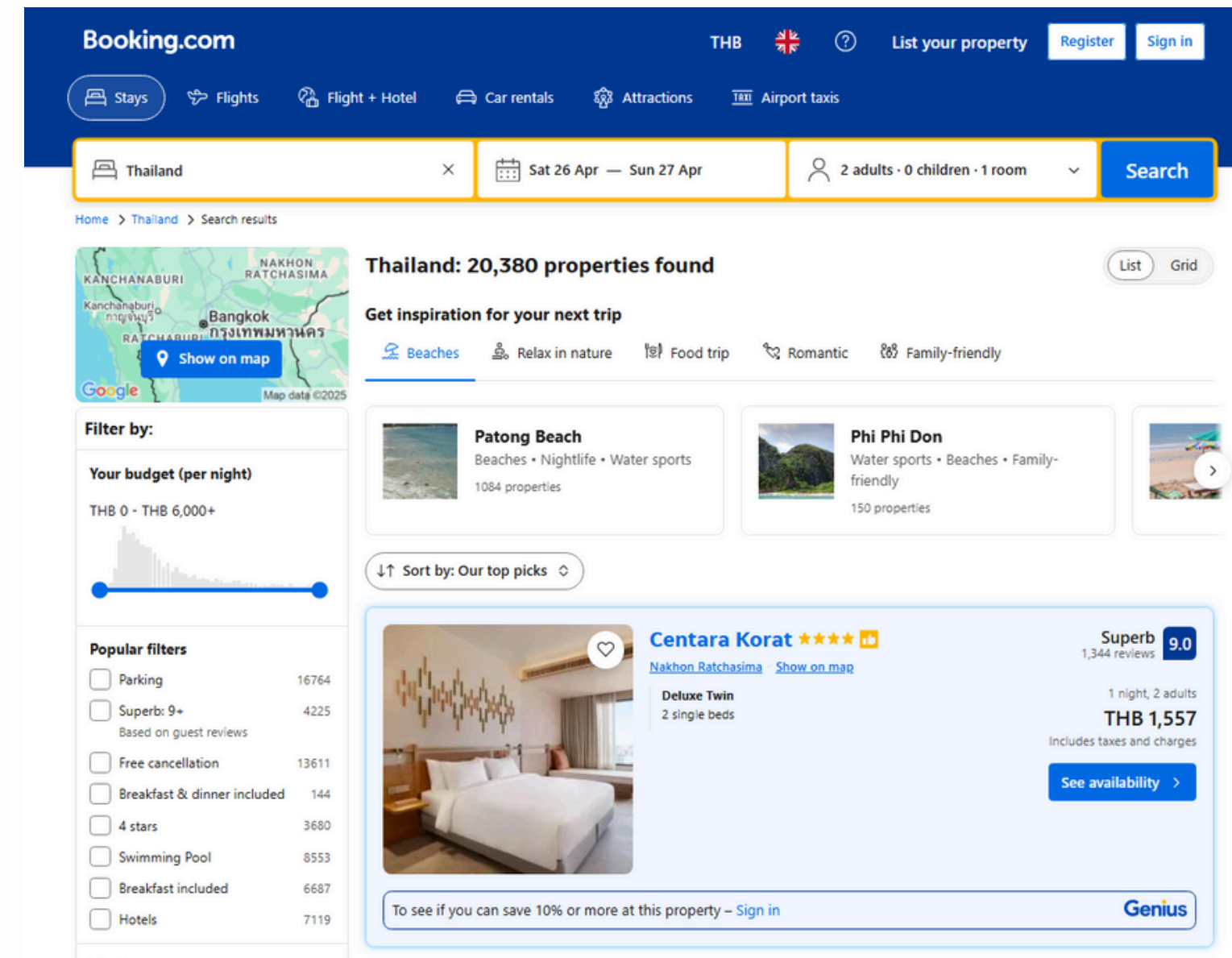
We selected 5 cities in Thailand.

- Bangkok
- Pattaya
- Nakhon Ratchasima
- Huahin
- Chiang Mai

Separated into 3 categories: Major City, Beach City and Countryside

With 1-3 different room types per hotel.

BOOKING.COM



The screenshot displays the Booking.com interface for a search in Thailand. The top navigation bar includes the Booking.com logo, currency (THB), and links for 'List your property', 'Register', and 'Sign in'. Below this, a search bar shows 'Thailand' as the destination, dates from 'Sat 26 Apr' to 'Sun 27 Apr', and '2 adults · 0 children · 1 room'. The search results page shows 'Thailand: 20,380 properties found'. A map of Thailand highlights the search area. The left sidebar contains filters for budget (THB 0 - THB 6,000+), popular filters (Parking, Superb: 9+, Free cancellation, Breakfast & dinner included, 4 stars, Swimming Pool, Breakfast included, Hotels), and a 'Filter by' section. The main content area features 'Get inspiration for your next trip' with categories like Beaches, Relax in nature, Food trip, Romantic, and Family-friendly. It also lists specific properties: Patong Beach (1084 properties), Phi Phi Don (150 properties), and Centara Korat (Nakhon Ratchasima) with a 'See availability' button. The Centara Korat listing shows a 'Deluxe Twin' room with 2 single beds, a 'Superb' 9.0 rating, and a price of THB 1,557 for 1 night, 2 adults. A 'Genius' banner at the bottom suggests a 10% discount.

Data Gathering Method



Siri Hotel

688-690 Than Phoklang Amphur Muang, 30000 Nakhon Ratchasima, Thailand – [Great location - show map](#)

Room type	Number of guests	Today's price
Standard Double Room 1 large double bed 16 m ² Air conditioning Ensuite bathroom Flat-screen TV Free WiFi		THB 540 Includes taxes and charges

Most popular facilities

- Free parking
- Non-smoking rooms
- Free WiFi
- Family rooms
- Restaurant
- Room service
- Facilities for disabled guests
- Bar
- Superb breakfast

Very good **8.1**
550 reviews

Guests who stayed here loved

"Great hotel,good value near everything,only quip is no window,but that's why it was quiet,didnt drain in shower well ,but that's most hotels.decent..."

Miller United States

Staff

8.8

Data Preprocessing

The facilities are combined into “Amenity Score” in the end as well as separating the price into categories, utilizing equal-frequency binning, dividing values into "Low" (≤ 1170), "Medium" (1171–3200), and "High" (> 3200).

Final Attributes

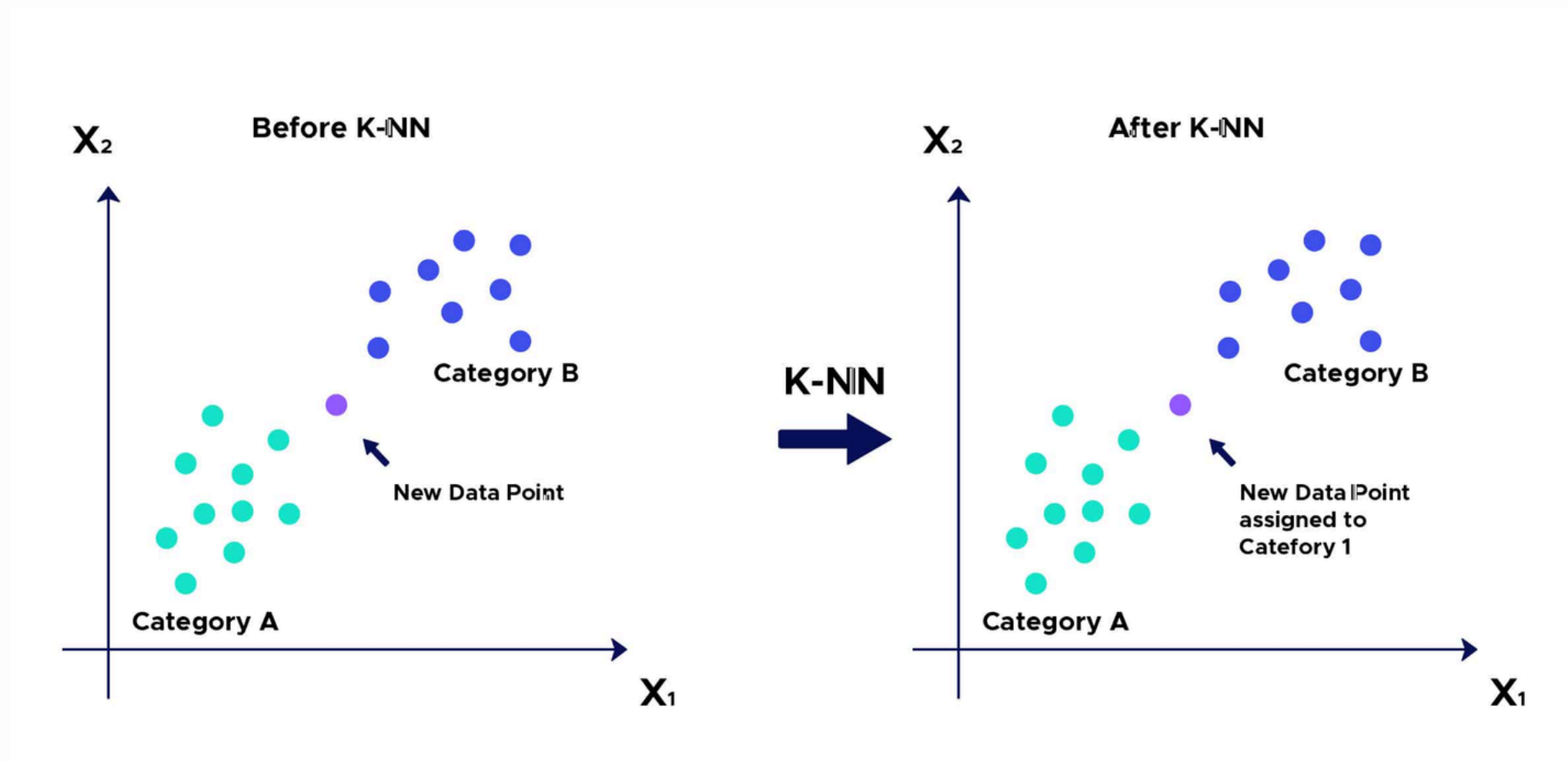
- | | | | |
|-----|---|-----|---------------------------------|
| 01. | Location (1 - 3, Major City, Beaches, Countryside) | 07. | Spa (1 or 0, Yes/No) |
| 02. | Star Rating (1 - 5) | 08. | Resturant/Bar (1 or 0, Yes/No) |
| 03. | Room Type (1 - 4, Budget/Small, Standard, Superior, Deluxe +) | 09. | Parking (1 or 0, Yes/No) |
| 04. | Swimming Pool (1 or 0, Yes/No) | 10. | Customer Reviews (0 - 10) |
| 05. | Fitness Center (1 or 0, Yes/No) | 11. | Price (Numeric) |
| 06. | Amenity Score (0-5) | 12. | Price Range (High, Medium, Low) |

Data

A	B	C	D	E	F	G	H	I	J	K	L	M
Hotel Name	Location	Star Rating	Customer Reviews	Swimming Pool	Fitness Center	Spa	Restaurant/B	Parking	Room Type	Accommodation S	Daily Price	Price Range
DusitD2 Hua Hin	2	4	8.7	2	1	0	2	0	4	4	4979	High
B2 Korat Boutique Budget Hotel	3	3	7.8	0	0	0	0	1	3	3	641	Low
The Beverly Hills Hotel	3	4	8.3	1	1	0	0	1	3	1	3108	Medium
The Siam Guest House	2	3	7.9	0	0	0	1	1	2	5	1200	Medium
Veranda Resort Pattaya	2	5	8.8	1	1	1	1	1	4	4	4500	High
Stardust Boutique Hotel	2	4	8.9	0	0	0	0	1	1	2	600	Low
Casa 17 Hotel Bangkok	1	4	8.7	1	1	0	1	1	2	1	2350	Medium
SHEIK ISTANA HOTEL CHIANGMAI	3	4	9.4	1	0	0	1	1	3	4	1121	Low
L Nimman	3	3	7.9	0	1	0	1	1	3	4	1135	Low
The Quarter Ratchathewi by UHG	1	4	8.7	1	0	0	1	1	3	1	2056	Medium
J Two S Pratunam Hotel	1	2	7.1	0	0	0	0	0	2	3	941	Low
We Briza Hotel Chiangmai	3	3	8	0	0	0	0	1	3	5	1400	Medium
Travelodge Nimman	3	4	8.6	1	1	0	1	1	3	5	1832	Medium
Oakwood Studios Sukhumvit Bangkok	1	5	8	1	1	1	1	1	3	1	3799	High
Mai Hotel	3	3	8.4	0	1	0	0	1	4	5	1521	Medium
Asia Hotels Group (Poonpetch Chiangmai)	3	4	7.8	0	0	0	1	1	3	5	600	Low

FINAL DATA

Algorithm



We utilized the KNN algorithm with the K value set to 5, and with the Euclidean distance formula.

We performed 5-fold cross validation to get the accuracy, with 160 training set and 40 testing set per iteration.

Result

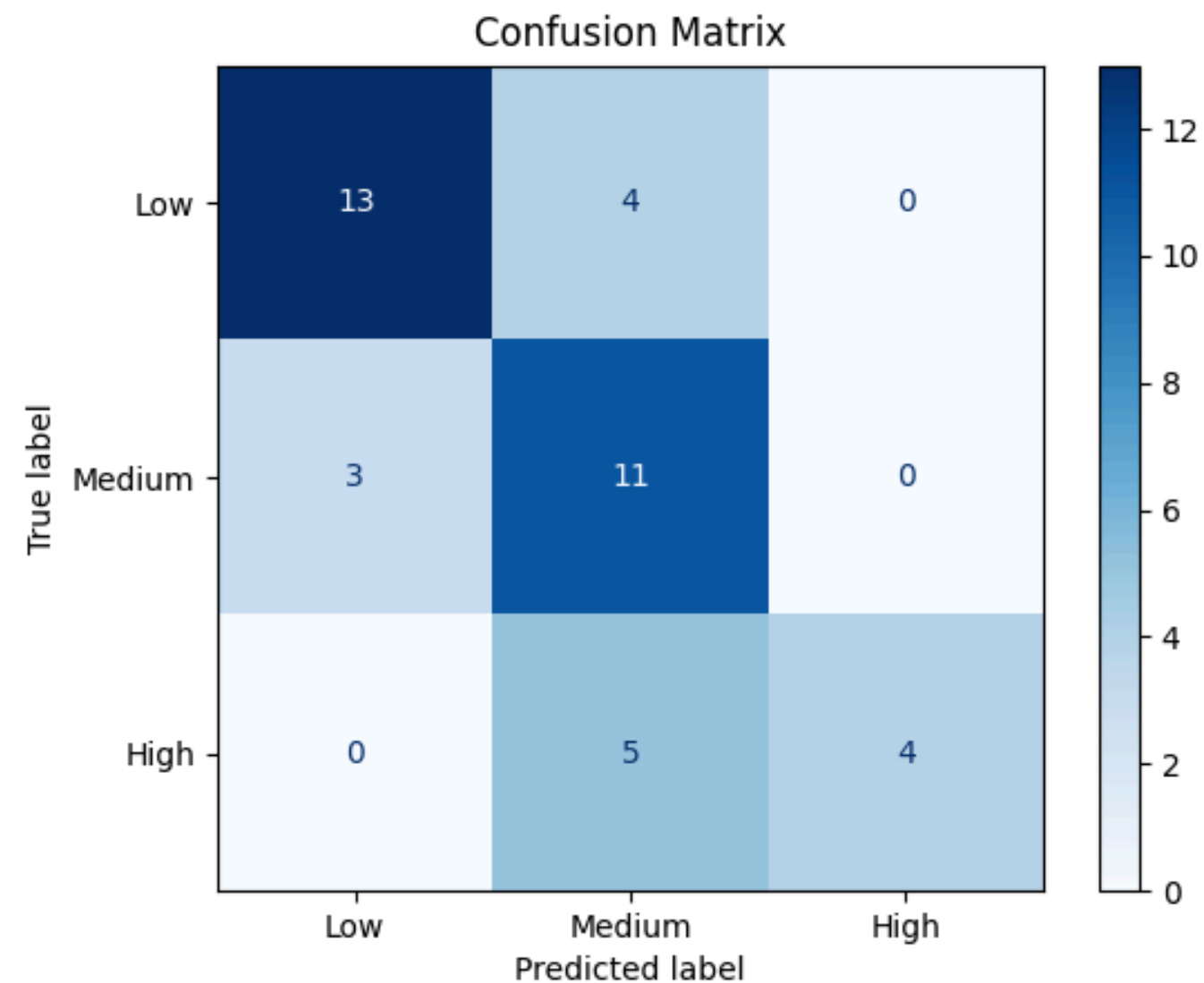
With KNN algorithm, it yields a prediction accuracy of ~70%.

In comparison, randomly selecting a class would result in ~33% accuracy.

However, our model might be lackluster when it compared to more advanced algorithms such as genetic algorithms or deep learning algorithms.

Test Set	Training Set	Accuracy
Fold 1	Fold 2,3,4,5	0.75
Fold 2	Fold 1,3,4,5	0.675
Fold 3	Fold 1,2,4,5	0.7
Fold 4	Fold 1,2,3,5	0.725
Fold 5	Fold 1,2,3,4	0.625
	Total Accuracy	0.695

Analysis



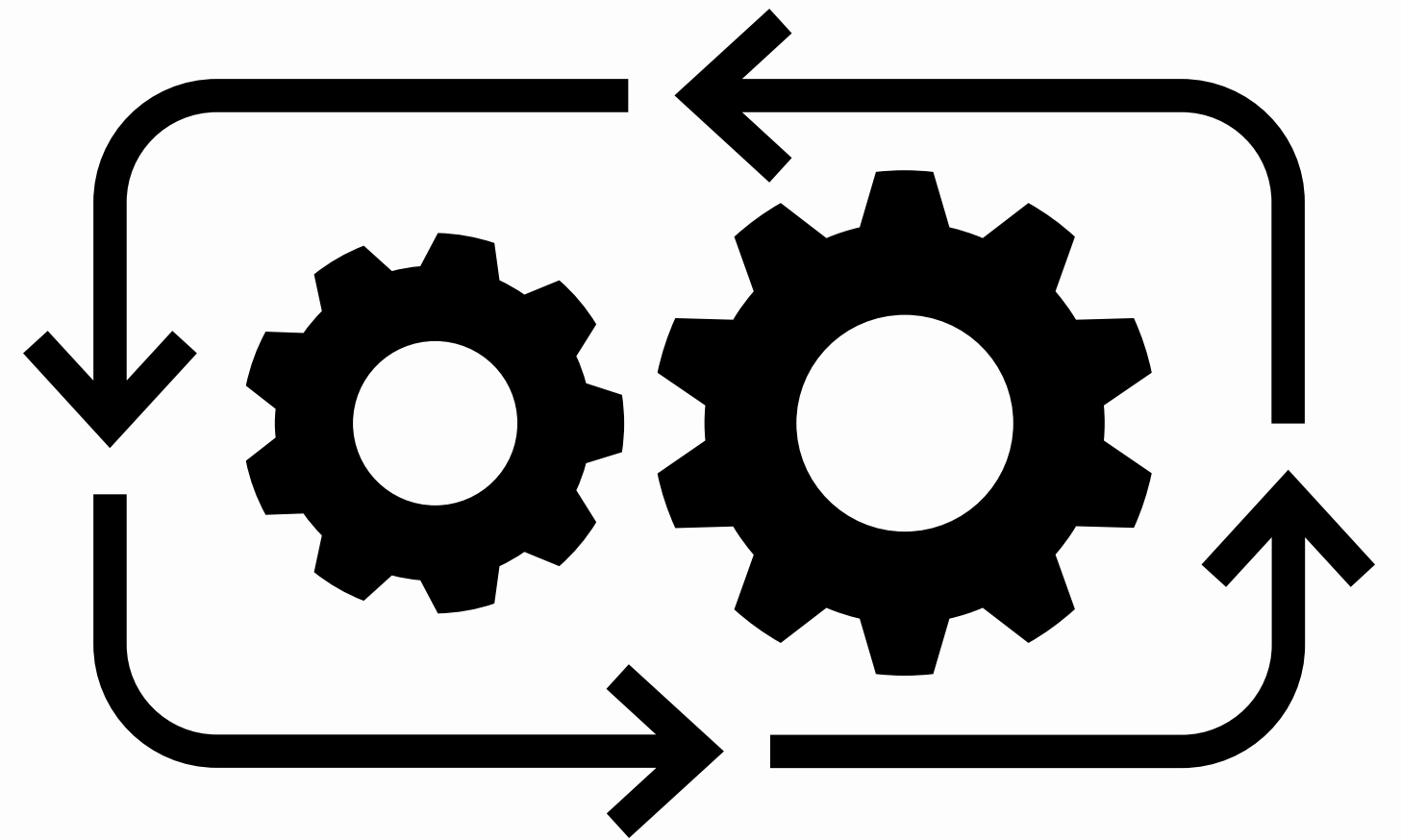
The image to the left is the confusion matrix for Fold 3 of cross-fold validation.


It has an accuracy of about 70%.

Upon examination, we found that the algorithm is able to accurately classify low and medium classes yet fails predict those for the High price-point class.

Ideas for Improvements

- Better results can be yield by using more complex algorithms such as deep learning algorithms.
- By collecting more data or adding more attributes maybe also improve the prediction's accuracy.





Thank You