PREDICTING HOTEL RATINGS: THE DATA-DRIVEN APPROACH

PRESENTED BY : LOKESH KUMAR|KRITHIKA NATARAJAN|PARTHIBAN P|VIGNESHWARAN A|ANAND MOHAN

MENTOR: MR.NIMESH MARFATIA

PROBLEM STATEMENT

This project seeks to enhance hotel rating systems by incorporating numerical user ratings, sentiment analysis of text reviews, and essential amenity data. Our goal is to equip hotels with actionable insights to elevate customer satisfaction, foster loyalty, and boost performance in India's competitive Tourism and Travel sector.

DATA SOURCE

The study employed ethical and legally compliant web scraping techniques to gather data from www.booking.com. An automated Python script, utilizing BeautifulSoup4 and XPATH, extracted essential hotel details like ratings, amenities, and prices from the acquired URLs, resulting in 7,072 records with 23 variables. To obtain user text reviews, a separate automated Python script employing the Selenium library gathered 98,304 reviews from diverse hotels. This comprehensive data collection process forms the foundation for our analysis.

FEATURE SELECTION

We condensed the essential hotel amenities into 21 Boolean variables from the “Features” variable. Additionally, we have added columns such as “Accessibility”, “Type of City”, “Hotel Category” and “Polarity” by engineering the existing columns to add more value to the analysis.

HOTEL CATEGORIZATION

The three categories were named as “Thrift Haven”, “Tranquil Retreat” and “Grandiose Manor” based on star rating and room tariff.

TABLE 1 – HOTEL CATEGORIES

Category	Star	Price Range (in Rs)	Hotels in the Category
Thrift Haven	1,2	Less than 2500	3020
Tranquil Retreat	3	2500 to 5000	1618
Grandiose Manor	4,5	More than 5000	738

MODEL SUMMARY

From the analysis, it was found that Linear Regressions worked best for all the three hotel categories.

TABLE 2- THRIFT HAVEN – MODEL PERFORMANCE

	TRAIN			TEST		
Model	R2	MAPE	RMSE	R2	MAPE	RMSE
Linear Regression	**0.791**	**0.1**	**0.764**	**0.797**	**0.102**	**0.765**
Ridge	0.791	0.1	0.763	0.797	0.104	0.771
Lasso	0.7	0.123	0.915	0.694	0.13	0.948
Elastic Net	0.788	0.117	0.855	0.794	0.124	0.855
Decision Tree	0.973	0.03	0.273	0.972	0.031	0.273
Random Forest	0.976	0.037	0.264	0.88	0.07	0.57

TABLE 3 - TRANQUIL RETREAT – MODEL PERFORMANCE

	TRAIN			TEST		
Model	R2	MAPE	RMSE	R2	MAPE	RMSE
Linear Regression	**0.718**	**0.07**	**0.631**	**0.753**	**0.069**	**0.606**
Ridge	0.718	0.07	0.63	0.754	0.07	0.612
Lasso	0.543	0.092	0.802	0.587	0.093	0.792
Elastic Net	0.709	0.079	0.693	0.752	0.08	0.693
Decision Tree	0.947	0.029	0.276	0.951	0.028	0.259
Random Forest	0.968	0.025	0.217	0.798	0.058	0.529

TABLE 4 - GRANDIOSE MANOR – MODEL PERFORMANCE

	TRAIN			TEST		
Model	R2	MAPE	RMSE	R2	MAPE	RMSE
Linear Regression	**0.909**	**0.028**	**0.28**	**0.908**	**0.031**	**0.308**
Ridge	0.909	0.028	0.279	0.908	0.032	0.312
Lasso	0.87	0.034	0.334	0.858	0.041	0.387
Elastic Net	0.902	0.05	0.58	0.898	0.059	0.58
Decision Tree	0.984	0.009	0.11	0.98	0.011	0.13
Random Forest	0.954	0.021	0.203	0.785	0.042	0.443

CONCLUSION

In summary, our analysis reveals that different hotel customer segments in the Thrift Haven (Standard), Tranquil Retreat (Deluxe), and Grandiose Manor (Luxury) categories have distinct expectations regarding amenities and services. To enhance customer satisfaction and ratings, hotels serving the Thrift Haven segment should prioritize location, accessibility, air conditioning, free Wi-Fi, and quality television. Tranquil Retreat customers value location, aesthetics, sustainability, and added amenities like Currency Exchange, with a preference for Hindi-speaking staff. Grandiose Manor guests seek value for their higher rates, including room views, private smoking areas, express room service, laundry, and a multi-cuisine restaurant.

To leverage these insights, hotels should highlight praised amenities, engage in online reputation management, showcase positive reviews, use testimonials, and promote user-generated content. Addressing negative feedback and investing in employee training for exceptional service are essential for boosting ratings and guest loyalty.