

# E-BOOK OF EXTENDED ABSTRACT

## THE 14<sup>TH</sup> INTERNATIONAL INVENTION, INNOVATION & DESIGN COMPETITION 2025



14<sup>TH</sup> **INDES** 2025

ENVIRONMENTAL • SOCIAL • GOVERNANCE



# **E-BOOK OF EXTENDED ABSTRACT**

THE 14th INTERNATIONAL  
INVENTION, INNOVATION &  
DESIGN COMPETITION 2025

**Organized by:**

Office of Research, Industry,  
Community & Alumni Network  
UiTM Perak Branch

**© Unit Penerbitan UiTM Perak, 2025**

All rights reserved. No part of this publication may be reproduced, copied, stored in any retrieval system or transmitted in any form or by any means; electronic, mechanical, photocopying, recording or otherwise; without permission on writing from the director of Unit Penerbitan UiTM Perak, Universiti Teknologi MARA, Perak Branch, 32610 Seri Iskandar Perak, Malaysia.

Perpustakaan Negara Malaysia

Cataloguing in Publication Data

No e- ISBN: 978-967-2776-52-9

Cover Design: Dr. Mohd Khairulnizam Ramlie

Typesetting : Georgia

# **EDITORIAL BOARD**

## **Editor-in-Chief**

MUHD SYAHIR ABDUL RANI

## **Managing Editors**

NUR FATIMA WAHIDA MOHD NASIR

SYAZA KAMARUDIN

NORASYIKIN ABDUL MALIK

## **Copy Editors**

SHEEMA LIZA IDRIS

AZURAWATI ZAIDI

HALIMATUN SAADIAH ABD MUTALIB

HALIMATUSSAADIAH IKSAN

IZA FARADIBA MOHD PATEL

MOHAMAD SAFWAT ASHAHRI MOHD SALIM

MUHAMMAD WAJIHUDDIN JOHARI

NAZIRUL MUBIN MOHD NOOR

NORAZIAH AZIZAN

NOOR AILEEN IBRAHIM

NOOR FAZZRIENEE JZ NUN RAMLAN

NOORLINDA ALANG

NURAMIRA ANUAR

NURDIYANA MOHAMAD YUSOF

NURSHAHIRAH AZMAN

NURUL FARHANI CHE GHANI

NURUL MUNIRAH AZAMRI

ONG ELLY

PAUL GNANASELVAM

SITI SYAIRAH FAKHRUDDIN

WAN FARIDATUL AKMA WAN MOHD RASHDI

WAN NURUL FATIHAH WAN ISMAIL

ZARLINA MOHD ZAMARI

AMIRUL FARHAN AHMAD TARMIZI

IMRAN TORIQ

# A MACHINE-LEARNING-BASED RECOMMENDER SYSTEM FOR HOTEL REVIEWS

Nurul Hafizah Zainal Abidin<sup>1</sup>, Wong Wai On<sup>2</sup>, Norliana Muslim<sup>2</sup>, Rohani Bakar<sup>2</sup>,  
Yusrafizal Yusof<sup>3</sup>, Nur Idalisa<sup>4</sup>

<sup>1</sup>Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA,  
Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia

<sup>2</sup>Department of Computer and Communication Technology, Faculty of  
Information and Communication Technology, Universiti Tunku Abdul Rahman,  
31900 Kampar, Perak, Malaysia

<sup>3</sup>Unit of Mathematics, Perak Matriculation College,  
31600 Gopeng, Perak, Malaysia

<sup>4</sup>Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA,  
Terengganu Branch, Dungun Campus,  
23000 Dungun, Terengganu, Malaysia

*nurul354@uitm.edu.my*

## ABSTRACT

With the rapid growth of online review platforms, selecting suitable accommodations has become increasingly challenging due to the overwhelming volume of user-generated content. This study presents a machine learning-based hotel recommendation system that processes Google Reviews for hotels in Perak, Malaysia. The system aims to enhance decision-making by recommending hotels based on user preferences, such as star ratings and sentiment features extracted from textual reviews. A total of 2,993 reviews from 58 hotels were collected using a Python-based scraper and pre-processed data to remove noise and tokenize content. Three machine learning models, namely Naïve Bayes, Random Forest, and Support Vector Machine, were evaluated using standard performance metrics. The SVM model achieved the highest accuracy (74.46%) and was selected for integration into a web-based application developed using Flask. The final system allows users to filter hotel recommendations by name, rating, and custom keywords, providing personalised suggestions and valuable insights. This project demonstrates the potential of machine learning to improve user experiences and decision-making in the hospitality industry.

**Keywords:** data preprocessing, Google reviews, Naïve Bayes, random forest, support vector machine, machine learning, recommender system

## 1. INTRODUCTION

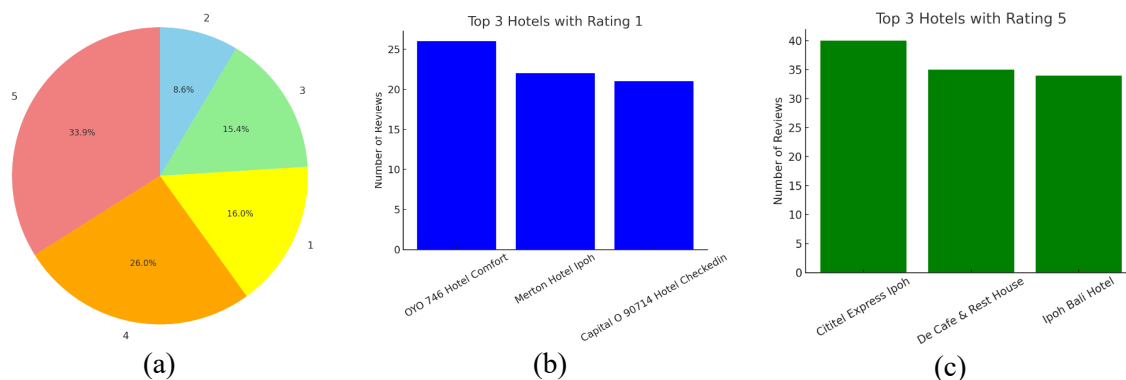
The surge in popularity of online review platforms has revolutionised how customers evaluate goods and services, particularly in the hospitality industry. As Malaysia gains momentum as a top tourist destination, hotel reviews on platforms like Google have become central to decision-making, both for travellers and hotel operators. These reviews encapsulate customer experiences, offer actionable feedback to hotel management, and significantly influence prospective customers' choices (Mehraliyev, 2022). Traditionally, travellers had to rely on static descriptions or word-of-mouth to find suitable accommodations. Today, they turn to real-time digital reviews. However, the overwhelming volume of this user-generated content makes manual analysis inefficient, costly, and prone to human error (Jain, 2021). Furthermore, hotels often underutilise this valuable feedback due to lack of tools for extracting insights from large-scale textual data (Zhang, 2020).

In addition, the evolution of big data analytics and machine learning provides new opportunities to harness customer feedback systematically. These technologies allow service providers to process vast review data, identify sentiments, classify trends, and deliver personalised experiences (Mehraliyev, 2022). Social media and collaborative platforms have further empowered customers to share nuanced reviews that reflect their service quality, location experience, and perceptions of price-value (Muslim, 2024).

This project introduces a hotel recommendation system that uses machine learning to analyse Google hotel reviews from Perak, Malaysia. It suggests hotels based on user preferences such as star ratings, review sentiments, and important features from the reviews. The use of machine learning improves the system’s accuracy, scalability, and overall performance compared to traditional methods.

## 2. METHODOLOGY

Hotel review data was collected for 58 hotels in Ipoh, Perak, Malaysia using a Python scraper on Google reviews, resulting in 2,993 reviews from 2,852 users. The data was pre-processed by removing stop words, cleaning text, and tokenising. Reviews were then labelled as positive, neutral, or negative based on their ratings. Using CountVectorizer, the cleaned text was converted into numerical features for machine learning. Most reviews were positive, with Rating 5 being the most frequent, followed by Rating 4. Very few users gave Ratings 1 or 2. Bar charts show that OYO 746 Hotel Comfort, Merton Hotel Ipoh, and Capital O 90714 Hotel Checkedin had the most negative reviews, while Cititel Express Ipoh, De Cafe & Rest House, and Ipoh Bali Hotel had the most positive feedback (see Figure 1).



**Figure 1** Hotels rating [6]

Three machine learning models were used to classify the dataset, namely Naïve Bayes, Support Vector Machine (SVM), and Random Forest. These models were selected after conducting literature reviews in text classification (Mariani, 2019). Then, their performance was measured using accuracy, precision, recall, and F1-score. SVM performed the best with 74.46% accuracy, followed by Random Forest at 73.29%, and Naïve Bayes at 62.60%.

## 3. FINDINGS

A Python script is used to train a hotel recommendation model based on review text. It imports key libraries like pandas, TfidfVectorizer, LinearSVC, and joblib. The script loads a cleaned CSV file of hotel reviews, removes rows with missing values, and converts review text into numerical features using TF-IDF while ignoring common stop words. It then trains a Linear Support Vector Classifier to predict hotel names based on the review content and saves the trained model for later use.

The recommendation system is built using Flask as the backend framework, with the app/ directory containing the core application logic. The best-performing model based on SVM, is stored in the model/ directory, ensuring a clear separation of concerns and reproducibility. The dataset/ directory houses crucial data such as user reviews and location details, including the cleaned dataset used during the training and evaluation phases. Frontend components are split into templates/ for HTML and static/ for styling and scripts, supporting a user-friendly interface. Additionally, requirement.txt lists the necessary packages, facilitating easy setup and replication, while the README.md provides comprehensive project documentation. The proposed system is illustrated in the user interface (as shown in Figure 2), where users can filter hotels based on name, star rating, minimum rating, and keywords such as "free Wi-Fi" or "non-smoking." The system displays hotel results with key details including star rating, review summaries, and available amenities, demonstrating how the integrated SVM model and cleaned dataset work seamlessly to provide relevant and personalised hotel recommendations.

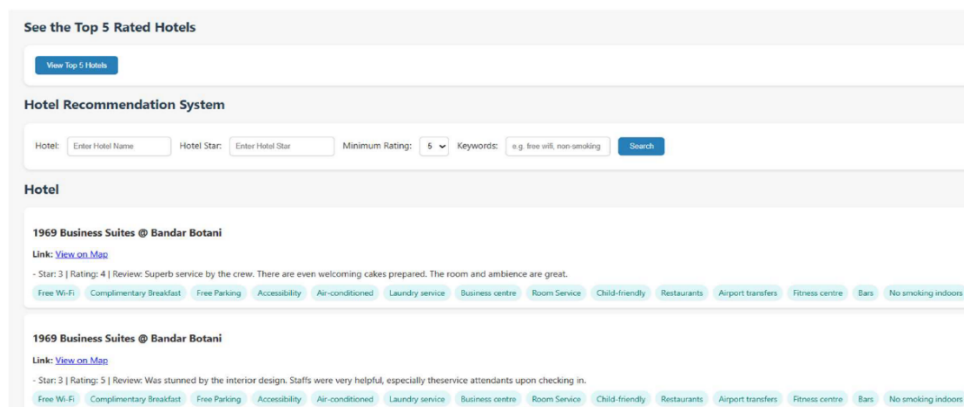


Figure 2 Proposed system

#### 4. CONCLUSION

This project successfully developed a machine learning-based hotel recommendation system that leverages Google review data from 58 hotels in Perak, Malaysia. Through data scraping and preprocessing, the system transformed raw review texts into meaningful features for classification. Among the three evaluated models, SVM demonstrated the highest accuracy at 74.46% and was selected as the core of the deployed system. Integrated into a Flask web application, the recommender system allows users to search for hotels based on star rating, sentiment, and keywords, delivering personalised recommendations. The project not only showcases the potential of machine learning in enhancing decision-making in the hospitality sector but also provides a scalable and user-friendly tool for travelers.

#### REFERENCES

- Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer science review*, *41*, 100413.
- Zhang, C., Wang, X., Cui, A. P., & Han, S. (2020). Linking big data analytical intelligence to customer relationship management performance. *Industrial Marketing Management*, *91*, 483-494.
- Mehraliyev, F., Chan, I. C. C., & Kirilenko, A. P. (2022). Sentiment analysis in hospitality and tourism: a thematic and methodological review. *International Journal of Contemporary Hospitality Management*, *34*(1), 46-77.
- Mariani, G. Di Fatta, and M. Di Felice (2019). Understanding customer satisfaction with services by leveraging big data: The role of services attributes and consumers' cultural background. *IEEE Access*, *7*, 8195–8208.

Muslim, N., On, W. W., Salh, A., Samsamnun, F. S. M., Zainuddin, Z., & Sabri, N. A. (2024). Machine learning models for hotel review: A case study of hotels in Perak, Malaysia. In A. J. Tallón-Ballesteros (Ed.), *Electronics, Communications and Networks* (p. 333). IOS Press.

E-Book of Extended Abstract THE 14th INTERNATIONAL INVENTION, INNOVATION &  
DESIGN COMPETITION 2025

e ISBN 978-967-2776-52-9



Unit Penerbitan UiTM Perak

(online)