Dr. Liang ZHAO

Predictive Modeling of Personality Traits from OpenRice Users Interaction Insights

BY

TAM Kai Tik

22239065

Information Systems and Business Intelligence Concentration

BBA Project Submitted to the

School of Business in Partial Fulfilment

of the Graduation Requirement for the Degree of

Bachelor of Business Administration (Honours)

Hong Kong Baptist University

Hong Kong

April 2024

TABLE OF CONTENT

1. ABSTRACT

The aim of this thesis is to examine the relationship between user interactions on the OpenRice and personality traits, particularly neuroticism and emotional stability. Utilising logistic regression models and machine learning techniques with artificial neural networks, the research investigates how different user interface (UI) features like photos, reviews, ratings, pricing information, and discounts can predict personality traits related to both neuroticism and emotional stability personality.

The findings indicate that higher engagement with user-generated content such as photos and reviews is connected to higher levels of neuroticism, whereas higher usage of objective restaurant information such as pricing and discounts is associated with emotional stability. The logistic regression model could forecast personality with a high degree of accuracy 82.93%. Combined with artificial neural networks further improved predictive performance to 87% accuracy.

Additionally, by analyzing user expectations, feedback, and restaurant ratings, it revealed potential reporting biases, which is users tend to provide higher ratings. The study highlights the value of analyzing user interface interactions to gain insights into underlying personality traits to enhance platform's user experiences.

2. INTRODUCTION

OpenRice is a popular restaurant rating website that plays a significant role in the context-aware interaction with digital entities in cities (Lee et al., 2021). By valuable and persuasive information on the platform, consumers' choices and dining experiences will be affected (Cho & Chan, 2019).

In the context of decision-making, it is essential to consider the impact of consumer-generated ratings and reviews. It has been found that having a significant influence on online restaurant popularity and visit intentions (Park & Nicolau, 2015). Consumer-generated ratings and reviews have positive relationships with online restaurant popularity (Park & Nicolau, 2015).

Moreover, the impact of electronic word-of-mouth (eWOM) through consumer-opinion platforms is evident. Tt helps to improve and develop the restaurants services by analyzing both positive and negative customer experiences (BOLANTE et al., 2022). In A business view, it is important to know the influence of the online commented platform to the company.

Additionally, the role of individual user personality traits, such as the Big Five personality traits, in designing review systems has been identified as an important but often neglected determinant of online behaviour and decision-making (Poniatowski & Neumann, 2020).

A platform's user interface undoubtedly influences user decision-making. Hence, by understanding the interaction between users and platform interface, it could gain an insight how platform’s user interface could enhance the user experience. From a company’s perspective, it could identify the personality traits of user that could provide best possible experience to users.

2.1 Literature review/ background research

Two major related issues to the project will be discussed in this session as follows:

I. User Interface (UI)
II. Personal traits
III. The importance of user-generated content (UGC) and its impact on online platforms
IV. Predictive Modelling and Machine Learning

I. User Interface (UI)

User Interface (UI) is significant in influencing consumer behaviour. In OpenRice features, it includes six crucial parts: Photo, Review, Rating, Discount, and Restaurant information.

User-generated content, particularly photos and reviews, has been identified as a significant factor in restaurant selection (Oliveira & Casais, 2019). As photograph from users shaping consumers' perception of a restaurant, it indicates the importance of photos from users (Liao et al., 2016). Additionally, the quality and quantity of consumer-generated ratings and reviews correlate with online restaurant popularity positively (Park & Nicolau, 2015). Therefore, it is essential to prioritize these elements in user interface.

In addition, factors such as taste, hygiene, food quantity, discounts, and offers significantly influence customers' restaurant selection decisions (Shanmugam et al., 2021). Research by Kim (2019) suggests that individuals often consider their own food preferences and dietary restrictions when choosing a restaurant. Hence, the basic restaurant information is important to individuals identifies their own restaurant preferences.

Considering above factors when measuring the impact of user interface to decision-making is effective to analysis the outcome.

II. Personal Traits

Personality traits have a substantial influence on user behaviour and engagement in online platforms, including those related to the restaurant industry.

Previous studies have demonstrated the effect of personality traits by using Big Five Model on online participation in various contexts (Badreddine et al., 2022; Poniatowski & Neumann, 2020).

According to John, O. P., & Srivastava, S. (1999) research, a comprehensive 44-item inventory has been developed to measure an individual on Big Five Factors of personality.

Figure 1. Neuroticism vs. emotional stability in Big Five Model

Neuroticism vs. emotional stability	Anxiety (tense) Angry hostility (irritable) Depression (not contented) Self-consciousness (shy) Impulsiveness (moody) Vulnerability (not self-confident)

The thesis has introduced 50 questions to rate an individual's personality. It is believed that is effective in measuring personality traits. Because of the limitation of this research, only questions related to neuroticism vs emotional stability will be included.

According to the economic theory of consumer behaviour, rational consumer would possess completed knowledge of all available options and prices, selecting products that are likely to maximize their utility within their budgetary limitations (Prebensen et al., 2015). Hence, neuroticism and emotional stability could typically view as a rational consumer and irrational consumer respectively in economics.

Hence, this research would identify rational and irrational consumer based on their personality traits.

The selected questions for each dimension are as follows:

Figure 2. Question design in Personality traits

Trait	**1**	**2**	**3**	**4**	**5**
Is depressed, blue	○	○	○	○	○
Is relaxed, handles stress well	○	○	○	○	○
Can be tense	○	○	○	○	○
Worries a lot	○	○	○	○	○
Is emotionally stable, not easily upset	○	○	○	○	○
Can be moody	○	○	○	○	○
Remains calm in tense situations	○	○	○	○	○
Gets nervous hardly	○	○	○	○	○

By calculating the difference between neuroticism scores and emotional stability scores, it is assumed that the higher marks tend towards that particular trait.

III. The importance of user-generated content (UGC) and its impact on online platforms

The previous study has demonstrated that UGC, such as electronic word-of-mouth (eWOM) on consumer-opinion platforms, significantly influences consumer behavior (Hennig-Thurau et al., 2004).

The credibility and influence of UGC on consumer behavior have been subjects of interest, with research delving into the impact of online conversations and impartial user-generated content on product sales (Danner & Thøgersen, 2021; Tang et al., 2014). UGC is viewed as reliable and trustworthy, making it a potent tool in shaping consumer attitudes and purchase choices (Tang et al., 2014). Additionally, UGC has been shown to enhance consumer trust in e-commerce platforms, fostering increased confidence and facilitating online transactions (Zandavalle et al., 2022).

Platforms that promote user-generated reviews and content creation aim to cultivate a community where users actively contribute valuable information for others (Goldsmith & Pagani, 2013). Given that platforms heavily rely on UGC for engagement and revenue, comprehending user motivations and behaviours in public sharing becomes imperative (Habib et al., 2019).

According to mentioned research, it is believed that OpenRice platform also focus on UGC content as a cornerstone of online platforms. That make OpenRice platform be a valuable digital community and fulfil all user needs.

In the context of online shopping, personality traits have also been found to influence consumer behaviour. Studies have highlighted associations between personality traits and online buying behaviours, with traits like neuroticism and openness to experiences being linked to the willingness to buy online (Lixăndroiu et al., 2021). Moreover, the relationship between personality traits and impulsive buying behaviour in online marketplaces has been explored, with variables like neuroticism, conscientiousness, and shopping enjoyment tendency significantly impacting impulsive buying behaviour (Erlangga et al., 2022). Additionally, studies have investigated the effect of consumer intrinsic factors, including personality traits, on impulsive buying behaviour in online marketplaces, highlighting the substantial influence of traits like neuroticism and conscientiousness on consumer behaviour (Erlangga et al., 2022). It is believed that it could apply the personality into the further study.

IV. Predictive Modelling and Machine Learning

In a study by (Lee et al., 2022), logistic regression was utilized alongside machine learning algorithms to predict user satisfaction in mobile healthcare services. The study emphasized the accuracy of logistic regression in this context. This suggests that logistic regression can be a reliable method for predicting user-related outcomes. It achieved a 90% accuracy rate in predicting user satisfaction with mobile healthcare services.

Furthermore, Bachrach et al. (2012) demonstrated the use of multivariate regression for predicting personality traits based on Facebook profiles, highlighting the potential of regression techniques in understanding user behaviour.

Hence, this research would deploy logistic regression model with neural networks to predict personality traits as the previous success prediction.

2.2 Objectives of the study

The thesis statement of the problem is: Predictive Modelling of Personality Traits from OpenRice Users Interaction Insights.

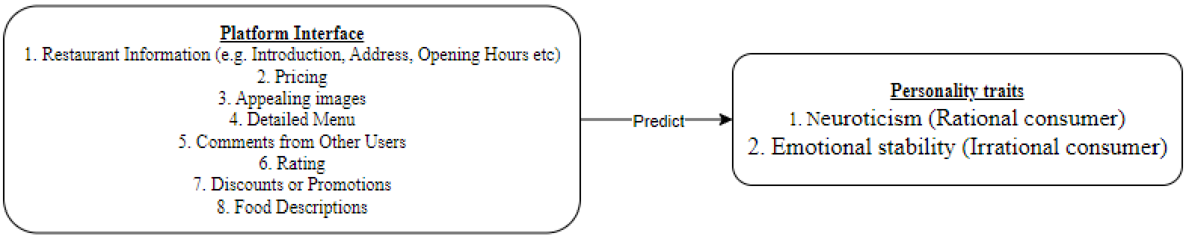
The user interface (UI) is a crucial element of any interactive software application to influence people's decision-making process (Yigitbas et al., 2020).

The Big Five Model has been found to organize broad individual differences into five categories that providing a comprehensive framework for understanding personality traits (McAdams & Pals, 2006).

Hence, this thesis aims to predict personality traits from the user interface interactions.

The model of this thesis as below:

Figure 3. Conceptual Model



By understanding the relationship between platform interface and personality traits, it would predict the personality traits from user interactions.

2.3 Statement of hypothesis

Hypothesis 1: The personality traits could be predicted by user interactions.

This hypothesis is the major assumption of this study, which suggests there is a relationship between an individual's personality traits and interactions on the OpenRice platform. If this hypothesis is true, it could predict the personality traits from user interactions. Hence, it is a foundation of this research and further prediction model assumption.

Hypothesis 2: Higher engagement with user-generated content features such as photos and reviews on the OpenRice is associated with higher levels of neuroticism.

The rationale behind this hypothesis is neuroticism individuals might seek more subjective and emotional information such as photos and reviews. It is because a appealing photos is a subjective feeling to users which could not be calculated and be compared to others restaurant. It is a personal experience to user. Hence, by knowing the relationship between photo and review interactions and user personality could enhance for platforms user experience as it could personalize the user interface for different users. that means it could fully fulfil to all users.

Hypothesis 3: Higher utilization of objective restaurant information features such as pricing and discounts on the OpenRice is associated with higher levels of emotional stability.

It is a conversely hypothesis to hypothesis 2. Emotional stability individuals are more likely to rely on objective and facts. Hence, they are prioritize practical and rational considerations over the subjective information namely photos and reviews.

Hypothesis 4: Specific combinations of UI feature usage such as high engagement with photos and reviews will be a stronger predictor of neuroticism.

As human behaviour is complex, it is hard to use a single factor to identify people personality. This hypothesis suggests certain combinations of UI feature usage patterns might provide higher accuracy for identifying neuroticism or emotional stability. For instance, individuals who engage with both photos and reviews might be more likely to predict as a higher levels of neuroticism.

The following analysis will examine all hypotheses by collected data. By statistical modelling and machine learning techniques, the study aims to uncover the underlying relationships between personality traits and user interaction in order to contribute valuable insights into the company user interface improvement and marketing strategies.

3. PROCEDURES & METHODOLOGY

3.1 Measurement Model

This study will employ a quantitative research design to examine relationship between UI interactions and personality. A cross-sectional survey approach will be used to collect data from online participants.

3.2 Participants

Since commenters on OpenRice are considered to be fully experienced with the platform's user interface, it is appropriate to select them as the sample. A purposive sampling technique will be used to ensure the sample of individuals who must have experience on OpenRice platform.

Power analysis will be conducted to ensure appropriate sample size. By considering factors such as statistical power and significance level (alpha), it can ensure the study has sufficient statistical power.

3.3 Data Collection

Data will be collected through an online survey questionnaire. The questionnaire will cover Target Audience Identification, User Interface (UI) Evaluation, Personal Factors Evaluation, Demographic Information and Last Experience Evaluation five parts. All participants could complete the survey at their own convenience that to increase the likelihood of participation and reduces potential barriers.

3.4 Research Design

Figure 4. Research Design Questions Table

Section	Purpose	Questions
0. Target Audience Identification	Identify level of engagement with OpenRice	1. Have you used OpenRice? 2. Have you commented on restaurants?
1. User Interface (UI) Evaluation	Understand importance of UI elements in decision-making	3. Features concerned during last visit 4. Rank importance of features 5. Rate likelihood of considering each UI feature
2. Personal Factors Evaluation	Assess personality traits using Big Five model	6. Rate agreement with statements about emotional stability, anxiety, moodiness
3. Demographic Information	Collect demographic data	7. Gender 8. Age group 9. Education level 10. Employment status
4. Last Experience Evaluation	Evaluate recent experience with OpenRice	11. Rate how well expectations were met 12. Rate overall satisfaction 13. Name of last restaurant visited

The research design would cover 5 parts that to investigate the interplay between user personality traits, user interface (UI) preferences, and dining experiences from OpenRice.

In section 0 serves as an initial screening phase, it would identify and filter the target participants for the study.

In section 1, it would delve into assessing the perceived importance of UI from the user's perspective. By understanding the level of the user satisfaction, it would help to gain insights into how UI interactions potentially shape an individual's personality.

In section 2, it would utilize Big Five Model that to identify the user's personality trait which focus on neuroticism or emotional stability. After pre-processed the data, it would assign a binary label to categorize the user's personality.

In section 3, it would collect demographic data such as gender, education level, and age group. The demographic characteristics will serve as control variables for the research. It might see the reason why similar personality types will have different UI preferences.

In section 4, it would focus to the user’s recent dining experiences after used OpenRice. It would gather data on user expectations, user feedback, and the names of recently visited restaurants. By using web scraping techniques on OpenRice.com, the recent visited restaurant name would be transferred to the OpenRice rating. The research aims to analyze the distribution of user experiences and to evaluate whether the platform meets the user’s needs effectively.

4. Data Analysis

4.1 Data Pre-processing

Handling Missing Values:

Respondents may have skipped certain questions or failed to meet the criteria for the target audience. To address this issue, the study would employ a listwise deletion approach for all incomplete or non-target individual's responses. For instance, the customer who has not used OpenRice before, would not be counted in dataset. Although this approach might reduce sample size, it ensures that all remaining data is completed and consistent. That could prevent potential biases caused by improper information.

Duplicate Removal:

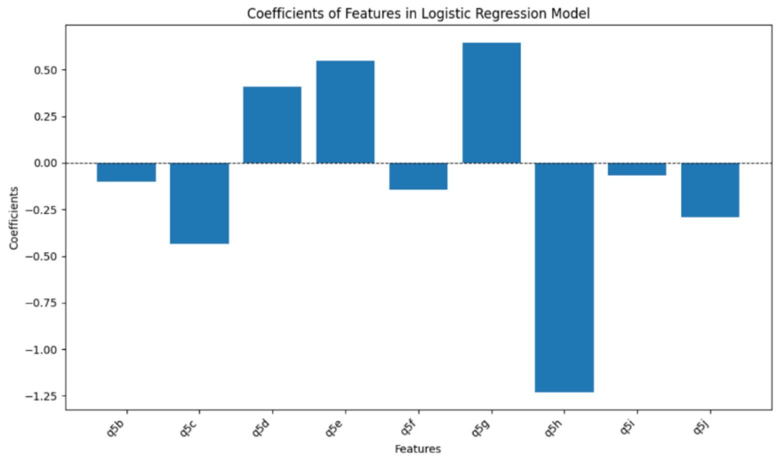
In online surveys, there is a possibility of respondents submitted multiple responses. To maintain data integrity, we employed a thorough check for duplicate rows. All duplicates were carefully inspected and removed that only retain the latest response for the study.

Data Standardization:

Standardizing data formats is essential for consistent and efficient data analysis. The study would utilize python for data analysis. Hence, all data would be coded consistently by using numeric or string data types. Additionally, the question would be converted in a consistent format for statistical analysis.

4.3 Data Analysis – Logistics Regression model

Figure 5. Graph of coefficients of features in logistics regression model



According to figure 5, it is a bar chart illustrating the coefficients of features in logistic regression model with features labelled on the x-axis and their corresponding coefficient values on the y-axis.

Based on the graph, the features 'q5d' (Appealing images), 'q5e' (Detailed menu), and 'q5g' (Ratings from other users) present the highest positive coefficient values. That indicates those features are strongly correlated with an "neuroticism" personality trait. This finding suggests that individuals with higher scores towards emotionality tend to place greater importance on appealing images, detailed menu and rating from other users when evaluating dining options. It also indicates that hypothesis 2 is true that higher level of neuroticism related to higher engagement of the user-generated content such as photo and reviews.

In opposition, the features 'q5c' (Pricing) and 'q5h' (Discounts or promotions) demonstrate negative coefficient values. That implied those features are strongly correlated with an "emotional stable" personality trait. This study suggests that individuals who are emotional stability, are likely to prioritize practical considerations such as pricing and discounts or promotional offers.

It also indicates that hypothesis 3 is accurate that higher level of neuroticism related to higher utilization of the objective restaurant information features namely pricing and discounts.

However, features 'q5b ', 'q5f ', 'q5i ', and 'q5j ' have coefficients close to zero, indicating a weaker relationship with either personality trait. This finding suggests that the detailed food descriptions and menus, as well as basic information about the restaurant, might not be significant factors for predicting personality. From a company perspective, it should improve the experience of other factors such as images and ratings that to fulfil neuroticism user's needs.

Figure 6. Results of model

```
Accuracy: 0.8292682926829268
Precision (emotional): 0.8947368421052632
Recall (emotional): 0.918918918918919
F1 Score (emotional): 0.9066666666666666
```

In figure 6, accuracy in neural networks refers to the ability of the network to make correct predictions or classifications. In this case, the model has an accuracy of approximately 82.93%. It is good for the first trial.

The precision of neural networks is crucial in various domains such as robotics, where precise control and improved performance are essential (Jiang et al., 2017). In this case, it is correct around 89.47% when predicts someone having an "neuroticism" personality.

The process of recall in neural networks has been extensively studied and optimized to improve the quality and capacity of retrieval (Perumal & Minai, 2009). In this case, it correctly identifies around 89.62% belong to the "neuroticism" personality class.

The F1 score is a widely used measure in the evaluation of machine learning models, particularly in the context of neural networks. In this case, the F1 score is 90.67% for the "neuroticism" class that indicated a balance between precision and recall.

To conclude, the accuracy of the model is high that proved hypothesis 1 is accurate. The personality traits could be predicted by user interactions as it is corelated.

Figure 7. Results of model after removed lower coefficients

```
Accuracy: 0.8780487804878049
Precision (emotional): 0.9
Recall (emotional): 0.972972972972973
F1 Score (emotional): 0.935064935064935
```

Given that having lower coefficients variables in recent model, it removed the lower coefficients and rebuild model. it increases a bit of the accuracy from 82% to 87%. It is great for predicting the result. To find out the best parameter of the model, it could combine with sigmoid function for further research that could build up a one-dimensional neural network for predicting personality.

4.4 Machine Learning – Artificial Neural Networks

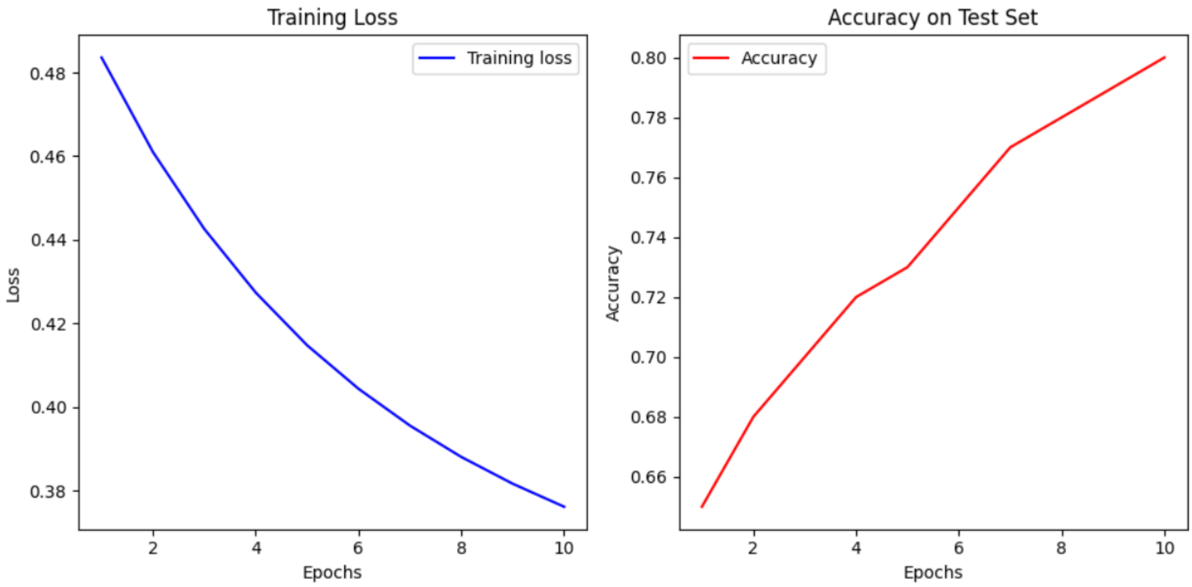
In neural networks, we combined both logistic regression model and sigmoid function to construct a one-dimensional artificial neural network. By using machine learning techniques, its able determinate the optimal parameters for predictive models by applying backpropagation. This approach is to learn complex patterns and relationships with neural network that enhancing the predictive ability of the model eventually.

Figure 8. performance of neural network

```
Epoch [100/1000], Loss: 0.49973756074905396
Epoch [200/1000], Loss: 0.49374398589134216
Epoch [300/1000], Loss: 0.48900941014289856
Epoch [400/1000], Loss: 0.4852710962295532
Epoch [500/1000], Loss: 0.48231592774391174
Epoch [600/1000], Loss: 0.4799737334251404
Epoch [700/1000], Loss: 0.47811028361320496
Epoch [800/1000], Loss: 0.4766206443309784
Epoch [900/1000], Loss: 0.47542300820350647
Epoch [1000/1000], Loss: 0.4744538366794586
Accuracy on test set: 0.8780487775802612
```

In figure 8, it illustrates that the performance of neural network after being trained over multiple epochs. The loss value is a metric that measures how well the model is performing during training. Hence, with a lower loss value indicates better performance in model as the model is made fewer errors in its prediction. As the loss value is remain decreasing after each epoch, it means the model is enhancing the performance each time. The following figure visualized a line chart of training performance.

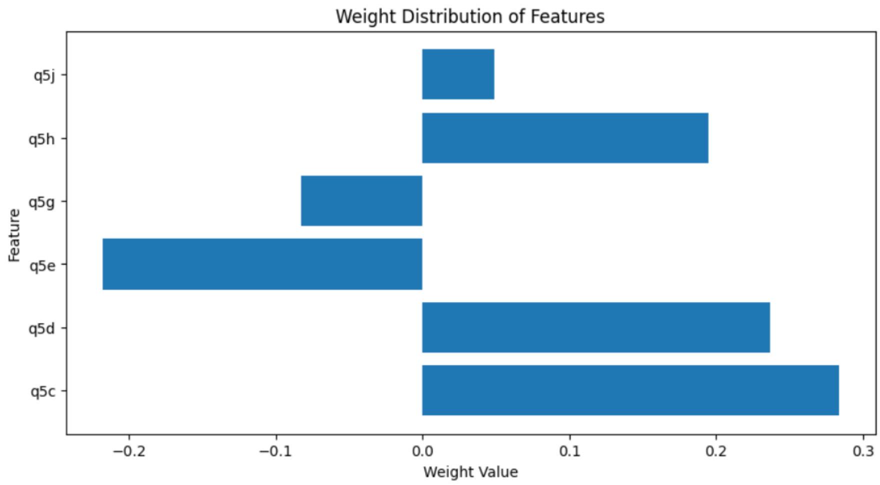
Figure 9. Line charts of training loss and accuracy when training model



The left plot shows the training loss curve of the model. The y-axis represents the loss value, and the x-axis represents the number of epochs. After each training, it started in 0.48 to 0.38. the downward trend indicated that the model is enhancing during training.

The right plot shows the accuracy of the model. The y-axis represents the accuracy value in test set, and the x-axis represents the number of epochs. During the training, it increased from 0.68 to 0.78. This upward trend suggested that the ability of the model is improving as trained.

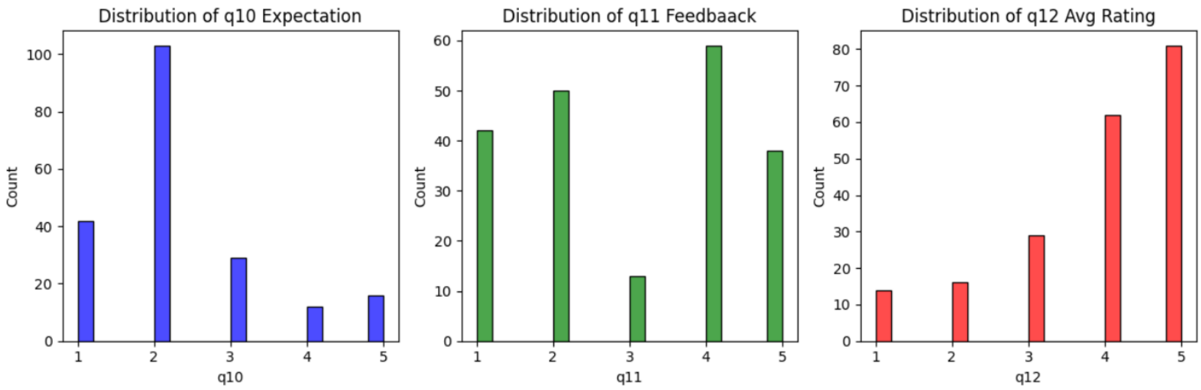
Figure 10. The weight distribution of features in neural networks



After constructing neural networks, it could remain the level of weight or coefficient of the features. The weight distribution of features provides insights into the influence of each feature in the underlying model. Features with larger positive weights are strongly associated to "emotional" personality class.

4.5 Data Analysis on Customers Expectation and Feedback

Figure 11. Distribution of user expectation, feedback and average rating



The first chart shows that the distribution of user expectation. The x-axis represents the value of the expectation and y-axis represents the frequency of each value. The data is heavily left skewed towards the value of 2 which has a significantly count compared to the other values. It is believed that user might tend to conservative or not highly expectation before dinning.

The second chart illustrates that the distribution of user feedback. The x-axis represents the value of the feedback and y-axis represents the frequency of each value. It is believed that the rating after dinning would increase that the distribution seems more spread out.

The third chart displays that the distribution of the average rating of the restaurant. The x-axis represents the value of the rating and y-axis represents the frequency of each value. This distribution appears to be right skewed towards the value of 5. The might appear to be reporting bias that OpenRice users tend to rate a higher marks in platform even the feedback is not that great.

5. Additional Analysis

Natural Language Processing of OpenRice Reviews

Natural Language Processing (NLP) is a critical component of various applications beyond speech recognition (Bengio et al., 2013). It encompasses a wide range of tasks, including part-of-speech tagging, chunking and named-entity recognition (Collobert & Weston, 2008). NLP has also been applied to information retrieval, text summarization, information extraction, and domain-specific applications (Chowdhury, 2003). Hence, the further research could combine with NLP to explore a deeper understanding of personality trait by OpenRice comments. By analyzing the comments, it could gain an insight into the restaurant's theme, restaurant target customers.

Incorporating Geospatial Data

Geospatial data management and analysis have seen significant advancements with the integration of big data analytics and geospatial technologies. The utilization of cloud computing in geospatial sciences has been explored to enable the intensities of geospatial sciences (Yang et al., 2011). The research is able to map restaurant locations, identify hotspots of positive or negative opinion across different geographic patterns further. Hence, it could uncover the emotional stable personality people decision making that enhancing the platform user’s experience. For instance, stable people might consider the location of restaurant. However, because of the limitation in this study, the location data which is geospatial data are not counted. Combined with further research, it could highly integrate the model that to enhance user experience on OpenRice.

6. Conclusion

This study investigated the relationship between user interactions from an online restaurant review platform (OpenRice) and personality traits, specifically on neuroticism and emotional stability. By using logistic regression model and machine learning with neural networks, several key findings emerged:

1. User interactions with UI features on OpenRice, such as appealing images, detailed menus, ratings from other users, pricing information, and discounts/promotions were found to be significant predictors of personality traits.
2. Higher engagement with user-generated content like photos and reviews was associated with higher levels of neuroticism, supporting Hypothesis 2. Conversely, higher utilization of objective information like pricing and discounts was linked to greater emotional stability, supporting Hypothesis 3.
3. The logistic regression model achieved high accuracy (82.93%) in predicting personality traits based on UI feature usage, providing evidence for Hypothesis 1 that personality can be predicted from user interactions.
4. Incorporating the logistic model with a neural network further enhanced predictive performance to 82.93%.
5. Analysis of user expectations, feedback, and restaurant ratings uncover some reporting biases, which users tends to mark higher ratings on OpenRice instead of the true scores.

To conclude, this research demonstrates the value of analyzing user interface interactions to gain insights into underlying personality traits. For business like commented platform, understanding which user engage with specific UI features could help company to enhance user experience and satisfaction. Additionally, company also could do customer segmentation that providing to restaurant to identify their customers.

REFERENCES

1. Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). Personality and patterns of facebook usage..
2. Badreddine, B., Blount, Y., & Quilter, M. (2022). The role of personality traits in participation in an online cancer community. Aslib Journal of Information Management, 75(2), 318-341.
3. Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: a review and new perspectives. Ieee Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798-1828.
4. BOLANTE, C., SALCEDO, E., SUBTENIENTE, J., & ESPLANADA, D. (2022). Sentiment analysis on TripAdvisor reviews of customer satisfaction in Tagaytay City from 2017-2021. Quantum Journal of Social Sciences and Humanities, 3(5), 42-56.
5. Chan, D. (2019). How social influence through information adoption from online review sites affects collective decision making. Enterprise Information Systems, 15(10), 1562-1586.
6. Chowdhury, G. (2003). Natural language processing. Annual Review of Information Science and Technology, 37(1), 51-89.
7. Collobert, R. and Weston, J. (2008). A unified architecture for natural language processing..
8. Danner, H. and Thøgersen, J. (2021). Does online chatter matter for consumer behaviour? a priming experiment on organic food. International Journal of Consumer Studies, 46(3), 850-869.
9. Goldsmith, R. and Pagani, M. (2013). Social network activity and contributing to an online review site. Journal of Research in Interactive Marketing, 7(2), 100-118.
10. Habib, H., Shah, N., & Vaish, R. (2019). Impact of contextual factors on snapchat public sharing..
11. Hennig-Thurau, T., Gwinner, K., Walsh, G., & Gremler, D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?. Journal of Interactive Marketing, 18(1), 38-52.
12. John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), Handbook of personality: Theory and research (Vol. 2, pp. 102–138). New York: Guilford Press.

13. Lee, H., Lee, S., Nan, D., & Kim, J. (2022). Predicting user satisfaction of mobile healthcare services using machine learning. Journal of Organizational and End User Computing, 34(6), 1-17.
14. Lee, L., Braud, T., Hosio, S., & Hui, P. (2021). Towards augmented reality driven human-city interaction: current research on mobile headsets and future challenges. ACM Computing Surveys, 54(8), 1-38.
15. Liao, H., Yu-cheng, L., Hu, T., & Luo, J. (2016). Inferring restaurant styles by mining crowdsourced photos from user-review websites.
16. Liu, X., Xu, M., Teng, T., Huang, G., & Mei, H. (2019). Muit: a domain-specific language and its middleware for adaptive mobile web-based user interfaces in ws-bpel. IEEE Transactions on Services Computing, 12(6), 955-969.
17. McAdams, D. and Pals, J. (2006). A new big five: fundamental principles for an integrative science of personality. American Psychologist, 61(3), 204-217.
18. Oliveira, B. and Casais, B. (2019). The importance of user-generated photos in restaurant selection. Journal of Hospitality and Tourism Technology, 10(1), 2-14.
19. Park, S. and Nicolau, J. (2015). Asymmetric effects of online consumer reviews. Annals of Tourism Research, 50, 67-83.
20. Prebensen, N., Altin, M., & Uysal, M. (2015). Length of stay: a case of northern norway. Scandinavian Journal of Hospitality and Tourism, 15(sup1), 28-47
21. Poniatowski, M. and Neumann, J. (2020). You write what you are – exploring the relationship between online reviewers' personality traits and their reviewing behavior., 1609-1614.
22. Perumal, S. and Minai, A. (2009). Stable-yet-switchable (sys) attractor networks..
23. Jiang, G., Luo, M., Bai, K., & Chen, S. (2017). A precise positioning method for a puncture robot based on a pso-optimized bp neural network algorithm. Applied Sciences, 7(10), 969.
24. Shanmugam, S., Krishnan, S., & Tholath, D. (2021). A behavioral study on the factors influencing the selection of restaurants online during COVID-19 using multivariate statistical analysis.
25. Tang, T., Fang, E., & Wang, F. (2014). Is neutral really neutral? the effects of neutral user-generated content on product sales. Journal of Marketing, 78(4), 41-58.
26. Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., … & Fay, D. (2011). Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing?. International Journal of Digital Earth, 4(4), 305-329.

27. Yigitbas, E., Jovanovikj, I., Biermeier, K., Sauer, S., & Engels, G. (2020). Integrated model-driven development of self-adaptive user interfaces. Software & Systems Modeling, 19(5), 1057-1081.
28. Zandavalle, A., Nascimento, V., Gadelha, C., Gama, T., Zagatti, F., Nildaimon, L., … & Real, L. (2022). Automated content moderation in a brazilian marketplace..

Appendix

Questionnaire question

Questionnaire: Predictive Modeling of Personality Traits from OpenRice Users Interaction Insights

Thank you for participating in our study! This questionnaire is designed to gather valuable insights into your experiences with OpenRice, a popular online restaurant commented platform. We are particularly interested in understanding the influence of User Interface (UI) elements and personal factors on decision-making during your interactions with the platform.

Section 0: Target Audience Identification

This section aims to identify your level of engagement with OpenRice, whether you have only used the platform for restaurant information or have actively participated by providing reviews.

1. Please select the option that best describes your experience with OpenRice: *

○I have used OpenRice.

○I have NOT used OpenRice. (end the questionnaire)

2. Please select the option that best describes your experience with OpenRice: *

○I have commented restaurants.

○I have NOT commented restaurants. (end the questionnaire)

Section 1: User Interface (UI) Evaluation

In this section, we want to understand your perception of the importance of various User Interface elements on OpenRice in influencing your decision-making.

3. Which feature(s) did you concern during your last visit? *

□Restaurant Information (e.g. Introduction, Address, Opening Hours etc)

□Pricing

□Appealing images

□Detailed Menu

□Comments from Other Users

□Rating

□Discounts or Promotions

□Food Descriptions

4. Rank the following features according to your level of concern during your last visit. *

[ ]Restaurant Information (e.g. Introduction, Address, Opening Hours etc)

[ ]Pricing

[ ]Appealing images

[ ]Detailed Menu

[ ]Comments from Other Users

[ ]Rating

[ ]Discounts or Promotions

[ ]Food Descriptions

5. On a scale of 1 to 5, how likely are you to consider **the following UI features** when making a choice? *

	Not Important at All	Slightly Important	Moderately Important	Very Important	Extremely Important
Restaurant Information (e.g. Introduction, Address, Opening Hours etc)	○	○	○	○	○
User interface elements (price, images, menu)	○	○	○	○	○
Pricing	○	○	○	○	○
The Appealing images	○	○	○	○	○
The detailed menu	○	○	○	○	○

The comments from other users	○	○	○	○	○
The rating from other users	○	○	○	○	○
The discounts or promotions	○	○	○	○	○
The detailed food descriptions	○	○	○	○	○

Section 2: Personal Factors Evaluation

This section aims to assess your personality traits using the Big Five Model. Please indicate your agreement or disagreement with each statement.

6. To what extent do you agree with the statement about your personalities? **Use a scale of 1 to 5, where 1 indicates "Strongly Disagree" and 5 indicates "Strongly Agree."** *

	Strongly Disagree	Disagree	Normal	Agree	Strongly Agree
Is depressed, blue	○	○	○	○	○
Is relaxed, handles stress well	○	○	○	○	○
Can be tense	○	○	○	○	○
Worries a lot	○	○	○	○	○
Is emotionally stable, not easily upset	○	○	○	○	○
Can be moody	○	○	○	○	○

Remains calm in tense situations	○	○	○	○	○
Gets nervous hardly	○	○	○	○	○

Section 3: Demographic Information

This section aims to know your demographic information.

7. What is your gender? *

○Male ○Female

8. Which age group do you belong? *

○0 - 18 ○28 - 24 ○25 - 34 ○35 - 44

○45 - 54 ○55 - 64 ○Above 64

9. What is the highest degree or level of education you have completed? *

○Primary school

○Secondary school

○Associate degree / Diploma

○Bachelor's degree

○Master's degree

○Doctoral degree (Ph.D., Ed.D., etc.)

10. What is your current employment status? *

○Employed Full-Time ○Employed Part-Time

○Seeking opportunities ○Retired

○Student ○Others _________________

Section 4: Last Experience Evaluation

In this section, we aim to gather insights into your recent experiences and perceptions.

11. On a scale of 5, how well your **expectations** were met after reviewing OpenRice. *

Very Unsatisfied	○1	○2	○3	○4	○5	Very Satisfied

12. On a scale from 1 to 5, how would you rate your recent experience after perusing OpenRice? *

Very Unsatisfied	○1	○2	○3	○4	○5	Very Satisfied

13. Please provide the name of the last restaurant you dined at? *

____

Code in Python

Logistic Regression Model	import pandas as pd import numpy as np import matplotlib.pyplot as plt from sklearn.linear_model import LogisticRegression from sklearn.model_selection import train_test_split df = pd.read_csv('merged_dataset.csv') X = df[['q5b', 'q5c', 'q5d', 'q5e', 'q5f', 'q5g', 'q5h', 'q5i', 'q5j']] y = df['personality'] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42) print("Training set shape:", X_train.shape, y_train.shape) print("Testing set shape:", X_test.shape, y_test.shape) logreg_model = LogisticRegression() logreg_model.fit(X_train, y_train) print("Coefficients:", logreg_model.coef_) print("Intercept:", logreg_model.intercept_)

Visualize the data	import matplotlib.pyplot as plt coefficients = [-0.09987072, -0.43566674, 0.41118754, 0.55041272, -0.14543133, 0.64706894, -1.23116558, -0.06885745, -0.29230219] intercept = 1.25795034 feature_names = ['q5b', 'q5c', 'q5d', 'q5e', 'q5f', 'q5g', 'q5h', 'q5i', 'q5j'] plt.figure(figsize=(10, 6)) plt.bar(feature_names, coefficients) plt.xlabel('Features') plt.ylabel('Coefficients') plt.title('Coefficients of Features in Logistic Regression Model') plt.xticks(rotation=45, ha='right') plt.axhline(y=0, color='black', linestyle='--', linewidth=0.8) plt.tight_layout() plt.show()

Print the result of the model	from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score y_pred = logreg_model.predict(X_test) accuracy = accuracy_score(y_test, y_pred) precision_emotional = precision_score(y_test, y_pred, pos_label='emotional') recall_emotional = recall_score(y_test, y_pred, pos_label='emotional') f1_emotional = f1_score(y_test, y_pred, pos_label='emotional') print("Accuracy:", accuracy) print("Precision (emotional):", precision_emotional) print("Recall (emotional):", recall_emotional) print("F1 Score (emotional):", f1_emotional)
Reconstruct model after removed lower coefficients	from sklearn.linear_model import LogisticRegression from sklearn.model_selection import train_test_split X = df[['q5c', 'q5d', 'q5e', 'q5g', 'q5h', 'q5j']] y = df['personality'] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42) print("Training set shape:", X_train.shape, y_train.shape) print("Testing set shape:", X_test.shape, y_test.shape) logreg_model = LogisticRegression() logreg_model.fit(X_train, y_train)

	`print("Coefficients:", logreg_model.coef_)` `print("Intercept:", logreg_model.intercept_)`

<table>
<tr>
<td>Construct a neural networks</td>
<td>

```
import torch
import torch.nn as nn
import torch.optim as optim

df['personality'] = df['personality'].replace({'emotional': 1, 'stable': 0})
X = df[['q5c', 'q5d', 'q5e', 'q5g', 'q5h', 'q5j']].values
y = df['personality'].values

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

X_train_tensor = torch.FloatTensor(X_train)
y_train_tensor = torch.FloatTensor(y_train)
X_test_tensor = torch.FloatTensor(X_test)
y_test_tensor = torch.FloatTensor(y_test)

class NeuralNetwork(nn.Module):
    def __init__(self, input_size):
        super(NeuralNetwork, self).__init__()
        self.fc1 = nn.Linear(input_size, 1)
        self.sigmoid = nn.Sigmoid()

    def forward(self, x):
        x = self.fc1(x)
        x = self.sigmoid(x)
        return x

input_size = X_train.shape[1]
model = NeuralNetwork(input_size)

criterion = nn.BCELoss()
optimizer = optim.SGD(model.parameters(), lr=0.01)
```

</td>
</tr>
</table>

	`` num_epochs = 1000 for epoch in range(num_epochs): optimizer.zero_grad() outputs = model(X_train_tensor) loss = criterion(outputs, y_train_tensor.view(-1, 1)) loss.backward() optimizer.step() if (epoch+1) % 100 == 0: print(f'Epoch [{epoch+1}/{num_epochs}], Loss: {loss.item()}') with torch.no_grad(): model.eval() test_outputs = model(X_test_tensor) predicted_labels = (test_outputs >= 0.5).float() accuracy = (predicted_labels == y_test_tensor.view(-1, 1)).float().mean() print(f'Accuracy on test set: {accuracy.item()}') ```

Visualize the loss value and accuracy	import matplotlib.pyplot as plt loss_values = [0.4836, 0.4610, 0.4426, 0.4274, 0.4148, 0.4044, 0.3956, 0.3881, 0.3817, 0.3762] accuracy_values = [0.65, 0.68, 0.70, 0.72, 0.73, 0.75, 0.77, 0.78, 0.79, 0.80] epochs = range(1, len(loss_values) + 1) plt.figure(figsize=(10, 5)) plt.subplot(1, 2, 1) plt.plot(epochs, loss_values, 'b', label='Training loss') plt.title('Training Loss') plt.xlabel('Epochs') plt.ylabel('Loss') plt.legend() plt.subplot(1, 2, 2) plt.plot(epochs, accuracy_values, 'r', label='Accuracy') plt.title('Accuracy on Test Set') plt.xlabel('Epochs') plt.ylabel('Accuracy') plt.legend() plt.tight_layout() plt.show()

Visualize the weight in neural networks	import matplotlib.pyplot as plt weights = model.fc1.weight.data.squeeze().numpy() feature_names = ['q5c', 'q5d', 'q5e', 'q5g', 'q5h', 'q5j'] plt.figure(figsize=(10, 5)) plt.barh(feature_names, weights) plt.xlabel('Weight Value') plt.ylabel('Feature') plt.title('Weight Distribution of Features') plt.show()

Visualize the distribution of feedback, expectation, average rating	import matplotlib.pyplot as plt import seaborn as sns fig, axes = plt.subplots(1, 3, figsize=(12, 4)) sns.histplot(df['q10'], bins=20, ax=axes[0], color='blue', alpha=0.7) sns.histplot(df['q11'], bins=20, ax=axes[1], color='green', alpha=0.7) sns.histplot(df['q12'], bins=20, ax=axes[2], color='red', alpha=0.7) axes[0].set_title('Distribution of q10 Expectation') axes[1].set_title('Distribution of q11 Feedbaack') axes[2].set_title('Distribution of q12 Avg Rating') plt.tight_layout() plt.show()