Understanding User Dissatisfaction with Penang Smart Parking: An Extended TAM-Based Analysis of App Store Reviews

Lim Pei Ying[1], Noor Hasliza Md Saad[2], Noor Azma Binti Ismail[1*], Farhad Nadi[1,3*]

[1]Faculty of Artificial Intelligence and Frontier Technologies, UNITAR International University, 47301 Petaling Jaya, Selangor, Malaysia

[2]School of Management, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia

[3]Centre for Innovation and Technology Adoption (CITA), UNITAR International University, 47301 Petaling Jaya, Selangor, Malaysia

***Corresponding authors**

DOI: https://doi.org/10.47772/IJRISS.2026.100601209

Received: 24 June 2026; Accepted: 29 June 2026; Published: 15 July 2026

ABSTRACT

The Penang Smart Parking (PSP) application was launched in August 2019 to modernise parking management in Penang, Malaysia, through a digital, sensor-based payment system. Despite more than six years of operation and a reported development investment of approximately RM115 million, the application continues to hold an average rating of only 1.8 out of 5 stars on the Google Play Store, reflecting persistent user dissatisfaction. This study investigates the root causes of that dissatisfaction by applying an Extended Technology Acceptance Model (TAM) that combines Perceived Ease of Use (PEOU), Perceived Usefulness (PU), and Perceived Security Risk (PSR) with computational text analysis. A dataset of 3,169 user reviews was analysed using Zero-Shot Text Classification, large language model (LLM)-based key-phrase extraction, and Term Frequency–Inverse Document Frequency (TF-IDF) analysis. The results show that usability problems dominate complaints, with PEOU accounting for 69.7% of all classified issues, driven mainly by application crashes, login failures, and registration errors. Security-related concerns, particularly payments deducted but not recorded and missing account balances, represent 8.8% of complaints and undermine user confidence in the system. PU complaints (5.7%) reveal a gap between user expectations and current functionality, including the absence of parking reminders and transaction receipts. A cross-version comparison indicates that these issues have persisted since the application's earliest releases rather than reflecting temporary resistance to change. The findings suggest that user dissatisfaction is systemic, and the study offers evidence-based recommendations across usability, transaction reliability, and version management. It also demonstrates how LLM-based review analysis can be integrated with an established theoretical framework for the post-adoption evaluation of smart-city applications.

Keywords: Technology Acceptance Model; smart parking; user dissatisfaction; zero-shot text classification; app store reviews.

INTRODUCTION

Urban parking in Penang, particularly in George Town, has long been a source of inefficiency and congestion. Before digitalisation, the parking system relied on manual methods such as paper-based parking coupons and tickets issued by local personnel. Drivers had to scratch the relevant year, month, day, and parking duration onto a coupon before displaying it; an incorrectly scratched coupon could not be reused and had to be discarded. Coupons were sold only by authorised agents, and where parking meters existed many were non-functional, leaving drivers exposed to the risk of a summons. These conditions, combined with drivers repeatedly circling to find available bays, contributed to traffic congestion and elevated vehicle emissions (Bernama, 2020). The

Consumers' Association of Penang further reported that public parking bays were being obstructed and illegally reserved with barrels, chairs, and flowerpots, reducing the supply of legal parking (NST, 2024).

To address these problems, the Penang government introduced a smart parking system as a Private Finance Initiative project, under which the contractor HeiTech Padu Berhad invested approximately RM115 million to build and operate the system under a seven-year concession (Mok, 2019). The Penang Smart Parking (PSP) application was officially launched in August 2019 (Mok, 2019). Using Internet of Things sensors and cloud technologies, the system was intended to let users check parking availability in real time, pay through an in-app wallet topped up via Touch 'n Go, credit and debit cards, and other online channels, and thereby reduce the time spent searching for parking (SUK Pulau Pinang, 2026).

Despite more than six years of operation, the application continues to hold an average rating of only 1.8 out of 5 stars on the Google Play Store, based on approximately 6,050 user reviews (SUK Pulau Pinang, 2026). In 2025 alone more than 350 one-star reviews were recorded, against only 15 five-star reviews. Given the scale of the public investment, this sustained negativity points to a substantial gap between system design and user expectations. The Technology Acceptance Model (TAM) is widely used to evaluate technology acceptance through Perceived Usefulness (PU) and Perceived Ease of Use (PEOU); in this study the model is extended with Perceived Security Risk (PSR) to capture how concerns about digital payment security shape satisfaction. Recent work indicates that LLM-based frameworks can enhance software usability analysis and identify user-reported issues in mobile reviews with high accuracy (Alsaleh et al., 2025; Eftee et al., 2025), motivating their use here.

Most existing studies of smart parking adoption rely on surveys and interviews, which are subject to low response rates and respondent bias and may not capture spontaneous, real-world experience. Systematic analysis of large volumes of app-store reviews using an established theoretical framework remains limited. This study therefore analyses Google Play Store reviews of the PSP application using computational text analysis within an Extended TAM framework. The objectives are: (i) to identify the primary factors contributing to user dissatisfaction; (ii) to categorise complaints under the TAM constructs (PU, PEOU, PSR) using Zero-Shot Classification; (iii) to quantify the most significant issues within each construct using TF-IDF; and (iv) to derive evidence-based recommendations for improvement. These objectives correspond to four research questions concerning the dominant factors of dissatisfaction, how complaints map to the extended constructs, which specific issues are most frequent, and what improvements the evidence supports.

LITERATURE REVIEW

The Technology Acceptance Model

The Technology Acceptance Model, originally developed by Davis in the mid-1980s (Dziak, 2024), has become a foundational framework for understanding how users adopt new technologies, with over a hundred related studies published between 1989 and 2001 alone (Ma & Liu, 2004). In TAM, Perceived Usefulness and Perceived Ease of Use shape users' attitudes and, in turn, their acceptance of a system. PEOU refers to the degree to which a user believes that interacting with a system requires minimal effort (Davis et al., 1989), whereas PU is the degree to which a user believes the system improves performance and productivity (Oematan et al., 2024). In short, PEOU concerns whether an application is easy to use, while PU concerns whether its functions are useful. Studies of smart parking applications confirm the relevance of both constructs: in Ningbo, China, perceived usefulness had the largest total effect on acceptance intention (standardised coefficient 0.984), followed by app attributes, ease of use, and trust (Yang et al., 2020), while Peng et al. (2021) found that ease of use and usefulness directly influence smart parking usage among nine identified factors.

Perceived Security Risk

Because the original TAM addresses only PEOU and PU, researchers have repeatedly extended it. Latif et al. (2025) argue that external factors such as trust, system quality, and service quality must be incorporated to explain digital-technology adoption, and Dong and Itoh (2025) integrate a user-experience model into TAM.

Gantulga et al. (2022) extended TAM with perceived risk and found that usefulness, ease of use, and perceived risk all significantly affect attitudes toward adopting smart mobility systems. In mobile payment contexts, perceived security risk is a key adoption factor because users worry about privacy and security when transacting on mobile devices (Saprikis & Vlachopoulou, 2021), and perceived risk negatively affects both trust and attitude toward mobile banking, partly because some applications are vulnerable to fraud (Almaiah et al., 2023; Merhi et al., 2019). These concerns are pronounced in Malaysia, where a VMware consumer study reported that 53% of Malaysian consumers doubted the security of smart payment-enabled devices (Asia, 2020). PSR is therefore incorporated here to test whether dissatisfaction stems from insecurity or from the application itself.

User Resistance to Technology Adoption

User resistance refers to a reluctance or refusal to accept an innovation even when it offers clear benefits. Szmigin and Foxall (1998) distinguished three forms of innovation resistance, namely rejection, postponement, and opposition, and argued that such resistance can stem from rational judgement rather than from simple reluctance. Later research has shown that ease of use, usefulness, and trust shape attitudes toward digital payments, whereas concerns about risk and security tend to reduce the willingness to adopt them (Jain & Jain, 2024). The relationship has also been quantified. Technology adoption and user resistance are moderately and negatively correlated ($r = -0.274$, $p = .022$), so that more positive experience with a system tends to lower resistance (Durango & Monterola, 2025). In a similar vein, ease of use and usefulness both exert significant positive effects on adoption attitudes, with reported path coefficients of 0.33 and 0.54 respectively (Lin, 2022). Taken together, these findings frame a central question for the present study: whether the low ratings of the PSP application reflect user resistance or a genuine failure of the system itself.

Text Classification Methods

Early text classification relied on statistical approaches such as the Bag-of-Words model and TF-IDF, which represent text numerically but ignore word order and meaning (Said & Ismail, 2025). Machine-learning models such as Support Vector Machines improved on this but require manual feature engineering and labelled data (Nielbo et al., 2024). Deep-learning models, including Convolutional and Recurrent Neural Networks, capture patterns and word order (Mustafa & Abdulazeez, 2024) but depend on large labelled datasets and suffer from limited interpretability and vulnerability to adversarial inputs (Tsimenidis, 2020). These limitations motivate the use of Large Language Models (LLMs) with Zero-Shot Text Classification, which assign categories to new text without task-specific fine-tuning and are widely applied in sentiment analysis, recommendation, and automatic annotation (Y. Wang et al., 2024). Table 1 summarises this evolution.

Table 1. Comparison of text classification methods

Method	Technique	Strengths	Weaknesses
Statistical	TF-IDF	Identifies important terms	Ignores word order and context
Machine learning	SVM	Simple; effective on structured data	Needs labelled data
Deep learning	CNN, RNN	Captures word order and context	Needs large, labelled datasets
Large language model	GPT-4o mini	No training required; understands context	Sensitive to prompt design

Zero-Shot Classification, Prompting, and LLMs

Zero-shot classification draws on the general knowledge that a model acquires during pre-training to label new text without any further training (Zhang et al., 2025). In practice it can be implemented either by asking the model to choose among predefined labels or by prompting it to generate a label directly (Y. Wang et al., 2024).

Because performance is sensitive to the way prompts are written, with Mu et al. (2024) reporting accuracy variations of more than 10% across prompting strategies, this study adopts DSPy, an optimisation framework that programmatically constructs and refines prompts and that replaces hand-crafted prompt strings with concise, optimisable modules (Khattab et al., 2023; Lemos et al., 2025). LLMs have been shown to work effectively as zero-shot text classifiers (Z. Wang et al., 2023) and to serve as efficient, cost-effective tools for annotation, classification, and sentiment analysis (Carmona-Díaz et al., 2025; Törnberg, 2023). The earlier evidence is encouraging: GPT-3 produced promising zero-shot results, including 64.3% accuracy on TriviaQA and an 81.5 F1 score on CoQA (Brown et al., 2020). More recent work reports strong performance on app-review tasks specifically, with GPT-4o mini reaching an F1 of 0.842 in requirement classification (Chaudhary et al., 2025), ChatGPT outperforming a dedicated genre classifier on unseen data (Kuzman et al., 2023), and GPT-4 surpassing rule-based feature extraction by 23.6% in F1 (Shah et al., 2024). Although LLMs can lose information when processing very long texts (Mervaala & Kousa, 2025), this limitation is of little consequence for the short reviews examined here. GPT-4o mini was therefore selected to classify each review into PU, PEOU, PSR, or Other without any labelled training data.

Key-Phrase Extraction and TF-IDF

Key-phrase extraction identifies the most relevant terms in user-generated content, supporting deeper understanding of feedback (Shrestha & Mahmoud, 2025). Traditional term-frequency, co-occurrence, and statistical methods struggle to capture semantics and relationships among phrases (Jamil et al., 2017; Shi et al., 2017; Sung & Kim, 2019). LLM-based extraction is more effective for unstructured, informal review text and can improve downstream topic modelling (Bui & Nguyen, 2025; Shrestha & Mahmoud, 2025); it is therefore adopted here. The extracted phrases are weighted using TF-IDF, a standard technique in information retrieval that combines a term's frequency in a document with its inverse frequency across documents to measure importance (Alfaridzi et al., 2022; AlShammari, 2023; Naeem et al., 2022). TF-IDF has proven effective in related text-classification tasks such as SMS spam detection (Ahmed & Haruna, 2025; Dasgupta & Yavary Mehr, 2024; Sjarif et al., 2019). In this study TF-IDF is applied within each TAM category by treating each review's extracted key phrases as a document and averaging scores across documents in the same category, giving a representative measure of term importance.

Research Gap

There is little research evaluating the Penang Smart Parking system specifically, and existing smart parking studies tend to examine other regions or propose new applications (Mahat, 2025; Ting et al., 2024). Most rely on primary survey or interview data (Chan & Lee, 2021; Low & Poh, 2018; Ting et al., 2024), which is limited by low response rates and respondent bias (Andrade, 2020; Cronin et al., 2009). While text mining of app reviews has been applied in other domains (Permana et al., 2020; Tchakounté et al., 2020), no study has systematically analysed smart parking app reviews through a theoretical framework such as TAM. This study addresses that gap by applying Zero-Shot Classification, LLM-based key-phrase extraction, and TF-IDF analysis to Google Play Store reviews of the PSP application within an Extended TAM framework incorporating PSR.

METHODOLOGY

Research Design

This study applies computational text analysis to user-generated Google Play Store reviews in order to identify the factors behind the application's low rating. Reviews of this kind offer spontaneous, real-world feedback on user complaints, expectations, and experiences. The Extended TAM is operationalised by classifying each review, through Zero-Shot Text Classification, into one of four constructs: Perceived Usefulness, Perceived Ease of Use, Perceived Security Risk, or Other. Whereas TAM is conventionally used to predict acceptance, its constructs are used here as an analytical lens for examining post-adoption dissatisfaction. TF-IDF is then applied in Python to surface the specific keywords and phrases that drive negative feedback within each construct.

Data Collection

According to the Google Play Store, the application had approximately 6,004 reviews, comprising about 5,930 mobile and 112 tablet reviews. A total of 3,178 reviews were collected manually; the full set could not be obtained owing to platform restrictions and anti-scraping mechanisms. To confirm representativeness, the rating distribution of the collected data was compared against the platform's overall distribution and found to align closely. Each review provided several attributes, including username, comment text, rating score, and review date; the comment text is the primary unit of analysis and the rating score indicates satisfaction. Using user-generated content in this way reduces common limitations of primary data collection such as response bias and limited sample size.

Ethical Considerations

Under Google Play's comment posting policy, most reviews are publicly visible and the platform controls their visibility (Google Play, 2026). Because the data are already public, individual reviewer consent was not required. The study analyses the content of comments rather than identifying individuals; although usernames appear among the collected attributes, they were not used in any part of the analysis. All data were collected manually for academic purposes, without automated scraping tools, to comply with platform policy.

Data Preprocessing

The unstructured data were transformed into a structured dataset containing the fields shown in Table 2. A unique user ID was generated for each record for management purposes only, without representing the reviewer's real identity. Because the application's version history is not shown on the Google Play Store, version numbers were obtained from external Android release-history sources and matched to review dates, enabling analysis of whether newer versions improved or worsened satisfaction. A comment-length field was also derived. Basic cleaning removed duplicates, converted text to lowercase, and stripped emojis and irrelevant characters; after cleaning, the dataset was reduced from 3,178 to 3,169 reviews because some reviews consisted only of emojis.

Table 2. Attributes of the structured dataset

No.	Field	Type	Description
1	User ID	Generated	Unique identifier assigned to each record during preprocessing
2	Username	Original	Reviewer name obtained from Google Play Store (not used in analysis)
3	Comment Year	Derived	Year extracted from the comment date
4	Comment Date	Original	Date the review was posted
5	User Comment	Original	Textual feedback provided by the user
6	Helpful Count	Original	Number of users who found the review helpful
7	Rating	Original	Rating score (1–5) provided by the user
8	Application Version	Derived	Version associated with the review
9	Comment Length	Derived	Number of characters in the comment

Zero-Shot Text Classification and Key-Phrase Extraction

Zero-Shot Text Classification was implemented with the OpenAI GPT-4o mini model through an API, so that each review could be assigned to the most appropriate construct on the basis of its meaning, without any labelled

training data. Three of the constructs derive from the Extended TAM (PEOU, PU, and PSR), and an additional Other category captures reviews that do not fit any of them, such as general remarks about parking or non-specific complaints. The DSPy framework structured the interaction with the model by defining both the input, the review text, and the expected output, the TAM category, which made the classification process consistent and repeatable. The same prompt also extracted three to five short key phrases per review, capturing issues such as system errors, login problems, or payment failures, and returned them in comma-separated form. Early runs showed that broad phrases such as 'usability issue' and 'payment issue' recurred across several categories, so the prompt was refined iteratively to improve specificity and reduce overlap, a pattern consistent with the prompt sensitivity reported by Mu et al. (2024).

TF-IDF Analysis

Once the reviews had been classified and their key phrases extracted, TF-IDF analysis was applied to the key phrases within each construct using Python and the scikit-learn library. By treating each review's phrases as a document and averaging the scores across all documents in the same category, the analysis identifies the terms most strongly associated with dissatisfaction, such as 'app crash', 'login issue', or 'payment deducted not recorded'. Pairing key-phrase extraction with TF-IDF allows the study to identify both the broad category of dissatisfaction and the specific issues raised within it.

Data Validation

Because some misclassification was evident, the accuracy of the LLM was validated for each construct using stratified samples proportional to category size (Table 3), with each sampled review checked manually against its assigned construct. Other achieved the highest accuracy at 92%, followed by PEOU at 89.33%, which suggests that clear linguistic signals such as 'good', 'app crash', or 'cannot login' support reliable classification. PSR was lower at 82%, largely because security-related concerns are often conveyed through context rather than through explicit security keywords, as when a user objects that a MyKad image is required simply to pay for parking. PU was the lowest at 77.14%, reflecting the overlap between usability, functionality, and payment issues within a single review. A complaint that the application paid the wrong council and caused a summons, for instance, could be read either as a usefulness failure (PU) or as a reliability concern (PSR). The mean accuracy across constructs was 86.67%, which indicates that the classification is generally reliable while identifying PU and PSR as the more challenging categories.

Table 3. Validation sample sizes and classification accuracy by construct

Construct	Reviews	Sample size	Accuracy
PEOU	2,208	150	89.33%
Other	503	50	92.00%
PSR	278	50	82.00%
PU	180	35	77.14%
Mean	n/a	n/a	86.67%

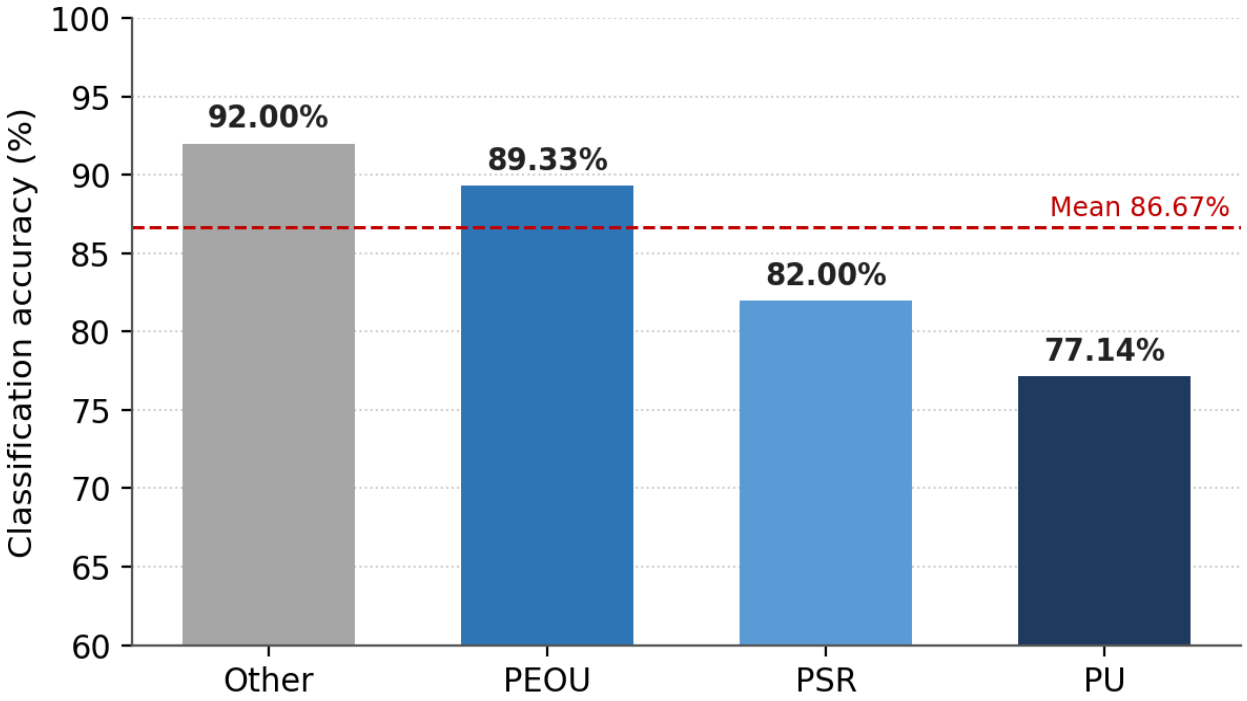

Figure 1. LLM classification accuracy by Extended TAM construct (derived from the reported validation results).

RESULTS AND ANALYSIS

Preliminary Temporal Analysis

Across the period from 2019 to 2026, low ratings dominate every year, and one-star ratings in particular, which confirms a general sense of dissatisfaction. One-star ratings rose sharply in 2020 to the highest value in the dataset; although they fell during 2021 and 2022, they climbed again in 2025, showing that dissatisfaction was never fully resolved. Four- and five-star reviews, by contrast, remained consistently low. When the data are examined by version, one-star ratings are the most prominent across almost every release, and dissatisfaction clusters around particular version-year combinations rather than spreading evenly. The clearest spike belongs to version 1.7.7 in 2020, which recorded 678 one-star ratings, the highest figure in the dataset, with version 2.0.3 in 2021 (193) and version 3.0.1 in 2025 (247) following. A few releases, such as version 2.0.4 in 2023, attracted very few negative reviews, but the recurrence of spikes around later major updates suggests that improvements were not sustained and that updates may introduce new problems rather than resolve existing ones.

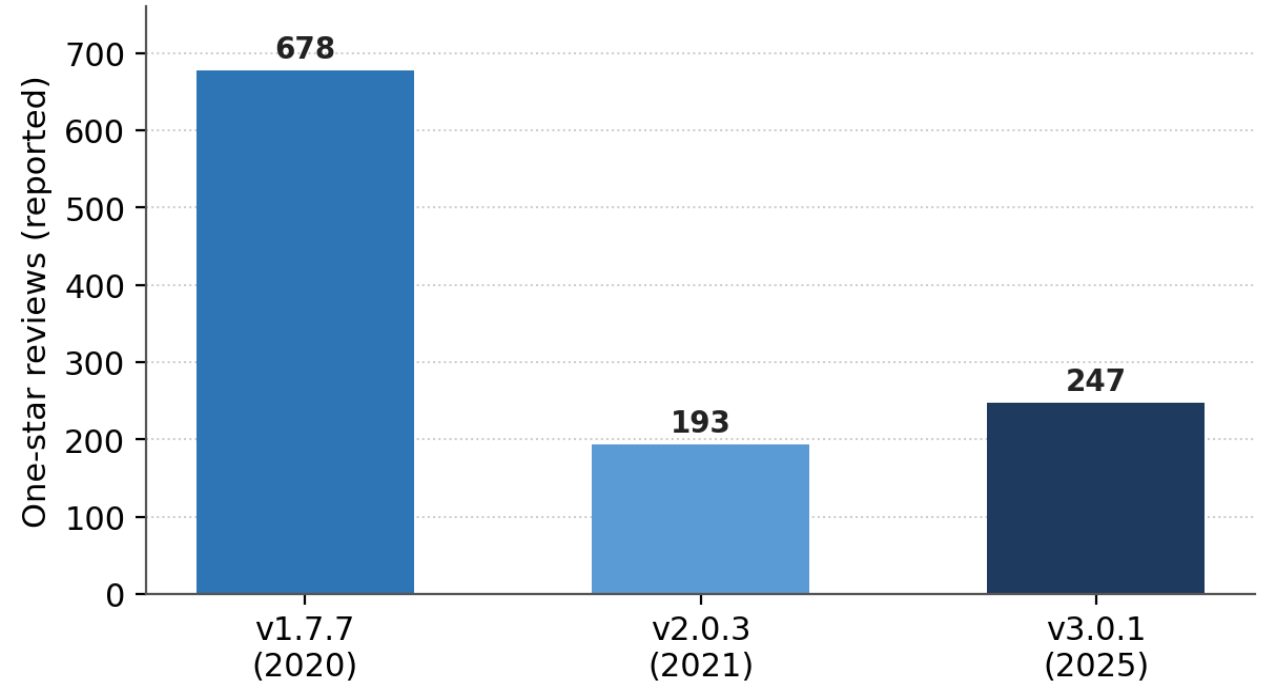

Figure 2. Reported one-star review counts for the most notable versions (selected versions as reported in the analysis).

Category Distribution

Zero-Shot Classification shows that Perceived Ease of Use is overwhelmingly the leading source of dissatisfaction, accounting for 69.67% of reviews, with examples such as 'cannot login after update' and 'app keeps crashing when opening'. Perceived Security Risk and Perceived Usefulness account for 8.77% and 5.68% respectively, indicating that payment-reliability and functionality concerns, while real, are reported far less often than usability problems. The remaining 15.87% fall under Other, mainly positive or non-specific comments. The dominance of usability issues is consistent with TAM, which holds that ease of use must be established before users can benefit from a system's usefulness (Davis et al., 1989; Ma & Liu, 2004); in a time-sensitive task such as parking payment, any usability barrier carries immediate real-world consequences.

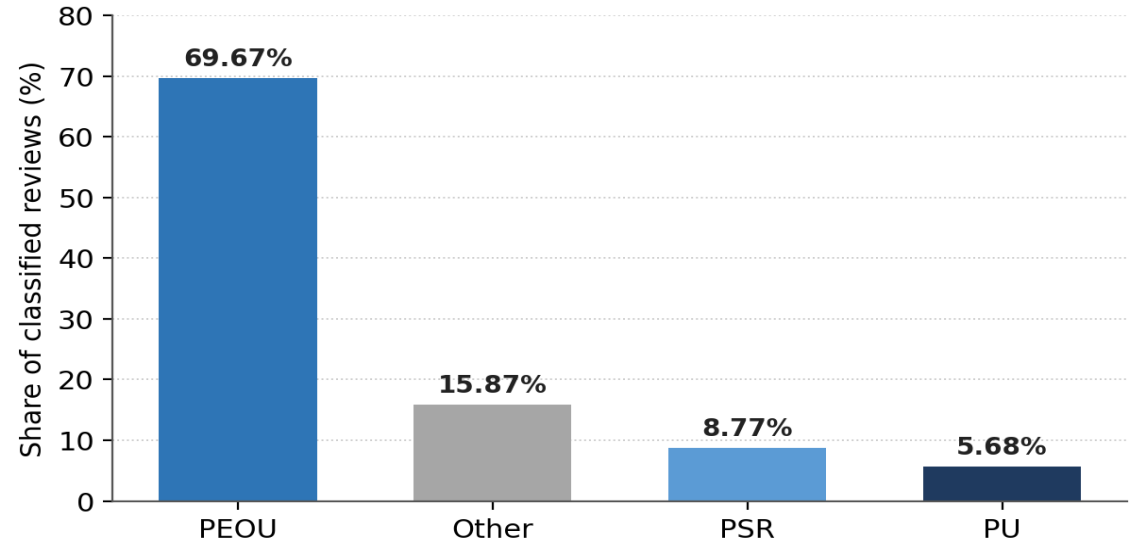

Figure 3. Extended TAM category distribution of classified reviews (derived from the reported percentages).

A closer look at the two largest spike versions reinforces this pattern. In version 1.7.7, 678 of 824 reviews (82.3%) were one-star, with PEOU dominating at 567 one-star reviews and PSR following close behind, which points to severe usability failures and financial-transaction concerns as the drivers of the spike. The effect was intensified by the phasing out of paper coupons, which made the application effectively mandatory during this period (Mutiara, 2019). In version 3.0.1, 247 of 283 reviews (87.3%) were one-star, and PEOU was again dominant with 203 one-star reviews. The absolute volume of negative feedback fell between the two versions, yet the structure of the complaints remained unchanged. Relating ratings to constructs tells a similar story. Most one-star reviews map to PEOU (about 1,970), PSR accounts for a smaller share (241 one-star reviews), and PU for fewer still, while Other contains the largest number of five-star reviews (272). Negative ratings, in other words, are tied to specific system problems, above all usability, whereas positive ratings tend to be general in nature.

TF-IDF Analysis by Construct

Within PEOU (Table 4), application crashes were the most frequent issue, with 940 occurrences (15.0%), followed by login problems (768; 12.26%), loading issues (9.72%), and registration errors (9.15%). Authentication-related difficulties also feature prominently, including password or PIN issues (5.24%) and OTP issues (3.86%), as do more specific failures such as payment not processing (1.39%) and connectivity issues (1.10%). Taken together, these findings indicate that users are frequently blocked at the most basic entry points of the application. The consequences are immediate, since a user who cannot log in while parking faces the risk of a summons for non-payment (Mok, 2019).

Table 4. TF-IDF results for Perceived Ease of Use (PEOU)

Rank	Keyword	Count	%	TF-IDF
1	app crash	940	15.00	0.1721
2	login issue	768	12.26	0.1696
3	registration issue	573	9.15	0.1507
4	loading issue	609	9.72	0.1317
5	usability issue	341	5.44	0.1058
6	password or pin issue	328	5.24	0.0892
7	technical issue	313	5.00	0.0861
8	otp issue	242	3.86	0.0743
9	navigation issue	153	2.44	0.0444
10	app update issue	131	2.09	0.0427
11	reinstallation issue	123	1.96	0.0388
12	app functionality issue	125	2.00	0.0374
13	payment not processing	87	1.39	0.0274
14	payment issue	71	1.13	0.0230
15	connectivity issue	69	1.10	0.0227

Within PSR (Table 5), complaints centre on financial transactions. The phrase 'payment deducted not recorded' ranked highest (18.27%), meaning that money is deducted without the parking transaction being registered, exposing users to fines. Reload issues (10.23%), payment failures (7.31%), and reload failures (2.19%) point to weaknesses in the payment gateway, while incorrect charges (1.22%), balance issues (10.48%), and wallet issues

(5.60%) indicate erroneous charges and missing balances that erode confidence in the system. These concerns are especially salient in Malaysia, where over half of consumers express doubts about smart payment-device security (Asia, 2020).

Table 5. TF-IDF results for Perceived Security Risk (PSR)

Rank	Keyword	Count	%	TF-IDF
1	payment deducted not recorded	150	18.27	0.2017
2	reload issue	84	10.23	0.1696
3	account access issue	96	11.69	0.1605
4	balance issue	86	10.48	0.1565
5	payment failure	60	7.31	0.1218
6	wallet issue	46	5.60	0.1068
7	refund issue	32	3.90	0.0775
8	transaction failure	24	2.92	0.0598
9	reload failure	18	2.19	0.0508
10	payment issue	19	2.31	0.0451
11	incorrect charge	10	1.22	0.0289
12	OTP issue	11	1.34	0.0283
13	password or pin issue	9	1.10	0.0222
14	loading issue	8	0.97	0.0208
15	login issue	8	0.97	0.0199

Within PU (Table 6), the dominant complaint is a lack of useful features (30.45%), reinforced by limited app features (5.59%), revealing a gap between expectations and functionality. Missing parking reminders (13.51%) show that users expect to be notified before a session expires, a feature the application lacks, while the absence of payment notifications (5.23%) and payment confirmations (3.42%) leaves users unsure whether transactions succeeded. Inaccurate parking availability (5.23%) indicates that the core real-time function is unreliable, and the absence of transaction history (3.42%) and receipts (1.98%) reflects unmet expectations for accessible financial records.

Table 6. TF-IDF results for Perceived Usefulness (PU)

Rank	Keyword	Count	%	TF-IDF
1	lack of useful features	169	30.45	0.2359
2	missing parking reminder	75	13.51	0.2063
3	no payment notification	29	5.23	0.1110
4	limited app features	31	5.59	0.1033
5	inaccurate parking availability	29	5.23	0.0966
6	notification issue	20	3.60	0.0827
7	no transaction history	19	3.42	0.0702

Rank	Keyword	Count	%	TF-IDF
8	parking function issue	18	3.24	0.0613
9	no payment confirmation	19	3.42	0.0603
10	receipt generation issue	12	2.16	0.0412
11	no payment receipt	11	1.98	0.0394
12	slow performance	6	1.08	0.0242
13	app functionality issue	6	1.08	0.0233
14	payment issue	6	1.08	0.0198
15	missing transaction history	4	0.72	0.0156

Cross-Construct and Cross-Version Observations

Despite iterative prompt refinement, some issue types appear across multiple constructs. Payment-related problems surface in PEOU (1.13%), PSR, and PU (1.08%), indicating that payment failure is a multi-dimensional problem affecting usability, financial security, and feature expectations simultaneously. Likewise, authentication issues such as login and password or PIN problems appear predominantly under PEOU but also within PSR at low frequencies (0.97% and 1.10%), suggesting that failed logins can also trigger security concerns when users cannot reach their account balance. A cross-version comparison (Table 7) shows that the leading issues persist: login problems and crashes top PEOU in both versions, 'payment deducted not recorded' is the leading PSR concern in both (18.1% then 15.9%), and 'lack of useful features' dominates PU and even rises over time (28.8% to 31%). The continuity of these issues from version 1.7.7 to version 3.0.1 indicates that the low satisfaction of the later release reflects unresolved long-standing problems rather than resistance to a new version.

Table 7. Key-phrase comparison across spike versions for each construct

Construct	Version 1.7.7 top issue	%	Version 3.0.1 top issue	%
PEOU	Login issue	14.8	App crash	17.7
PEOU	App crash	14.5	Login issue	15.4
PEOU	Loading issue	11.0	Registration issue	8.27
PSR	Payment deducted not recorded	18.1	Payment deducted not recorded	15.9
PSR	Balance issue	13.8	Reload issue	9.76
PSR	Account access issue	12.4	Account access issue	9.76
PU	Lack of useful features	28.8	Lack of useful features	31.0
PU	Missing parking reminder	17.5	Parking function issue	10.3
PU	Notification issue	7.5	Limited app features	10.3

DISCUSSION

Persistent, Systemic Dissatisfaction

The temporal pattern, in which one-star spikes cluster around major updates such as version 1.7.7 (2020) and version 3.0.1 (2025) with quieter periods between them, points to a reactive development process in which problems are addressed only after users complain rather than being prevented beforehand. The fact that complaints across all TAM categories follow the same temporal rhythm suggests that the issues originate in the overall development and testing process rather than in any single feature. For a mandatory public service with no alternative, this recurring instability is particularly serious. The early spike coincided with the phasing out of paper coupons, which ceased to be sold on 31 December 2019 and remained valid only until 31 December 2020 (Mutiara, 2019); once that alternative disappeared, the system's weaknesses became far more visible. This is consistent with innovation resistance theory, which holds that users postpone adoption while alternatives remain available (Szmigin & Foxall, 1998). Later dissatisfaction may also owe something to external policy change. In Seberang Perai, the MBSP raised parking charges from RM0.40 to RM0.80 per hour and daily rates from RM3.00 to RM6.00 with effect from 1 January 2025, the first revision in nearly three decades (Bernama, 2020). Users were thus asked to pay higher fees at the very time they continued to encounter persistent technical problems.

Dominance of Perceived Ease of Use

At 69.7% of complaints, PEOU is far more prominent here than in typical TAM-based smart parking studies, where usefulness and ease of use carry more comparable weight (Peng et al., 2021; Yang et al., 2020). The implication is that usability failures are not simply present but severe enough to overshadow every other source of dissatisfaction. When users cannot log in, register, or hold a stable session, they are effectively shut out of the system altogether, and its usefulness becomes irrelevant. This lends support to Davis's principle that ease of use must be established before a system can be judged useful (Ma & Liu, 2004).

Security Risk and User Confidence

Although PSR represents a smaller share (8.8%), its issues carry disproportionate weight. Payments deducted but not recorded, reload problems, and account-access failures are financial concerns likely to have a lasting effect on user confidence. This aligns with evidence that perceived security risk negatively affects trust and behavioural intention in mobile payments (Almaiah et al., 2023; Saprikis & Vlachopoulou, 2021) and that Malaysian consumers are particularly concerned about smart-payment security (Asia, 2020).

Expectation Gaps Under Perceived Usefulness

PU complaints account for the smallest share (5.7%), but they differ in kind from the others. Rather than describing failures, they describe missing capabilities, including a lack of useful features, limited functionality, and the absence of transaction history. This indicates that improving stability alone will not be enough to satisfy users; the application must also broaden its functional scope to keep pace with changing expectations (Yang et al., 2020). Since TAM defines usefulness as the degree to which a system improves a user's performance (Ma & Liu, 2004), the prominence of 'lack of useful features' shows that this threshold has not been met for a meaningful proportion of users.

Construct Overlap and the Resistance-versus-Failure Question

The recurrence of payment and authentication issues across PEOU, PSR, and PU shows that dissatisfaction with the application is systemic rather than confined to neat compartments. This is in line with extended-TAM studies in which the constructs interact and in which adding PSR improves the model's explanatory power for digital-

payment contexts (Gantulga et al., 2022; Latif et al., 2025). In practical terms, improving payment and authentication can strengthen several dimensions of the user experience at once. As for whether the low satisfaction reflects user resistance or system failure, the evidence points firmly to system failure. If resistance were the principal driver, complaints would ease as users grew accustomed to the system. Instead, the same issues, including login failures, crashes, and payments deducted but not recorded, persisted from 2020 to 2025, and the share of 'lack of useful features' actually rose from 28.8% to 31%. Given the negative relationship between ease of use and resistance ($r = -0.274$; Durango & Monterola, 2025), the high proportion of PEOU complaints indicates that usability has not improved sufficiently, which confirms that the principal cause lies with the system rather than with user attitudes.

RECOMMENDATIONS

The findings support a prioritised, evidence-based improvement programme.

- **Prioritise usability.** Because PEOU dominates complaints and the most critical issues are app crashes (15.0%), login problems (12.26%), and registration errors (9.15%), the login and registration flows should be redesigned for reliability across devices and network conditions, with clear error handling. Authentication mechanisms (OTP delivery, PIN management, password recovery) should be simplified and made fault-tolerant, and stability validated through comprehensive regression testing before each release.

- **Strengthen transaction reliability and security.** Given that payments deducted but not recorded (18.27%), reload issues (10.23%), and balance issues lead PSR complaints, the application should provide immediate transaction confirmation, handle failed transactions cleanly without erroneous charges, strengthen session management and OTP delivery, and communicate data-protection practices transparently to rebuild trust (Almaiah et al., 2023; Asia, 2020; Merhi et al., 2019; Saprikis & Vlachopoulou, 2021).

- **Expand functionality.** To close the gap revealed by PU complaints, the application should add reliable payment-success and failure notifications, accessible transaction receipts, and more accurate real-time parking information (Yang et al., 2020). As these are enhancements rather than fixes, they should follow the resolution of core usability and security issues.

- **Strengthen version management and continuous monitoring.** Because dissatisfaction rises around major updates, the development process should include realistic pre-release and regression testing and a staged rollout to a limited user group before full deployment. The LLM-based review-analysis method demonstrated here can be embedded in post-release monitoring to detect and respond to emerging issues systematically rather than reactively.

CONCLUSION

This study set out to understand why the Penang Smart Parking application continues to attract poor feedback, holding an average of just 1.8 out of 5 stars across approximately 6,050 reviews, despite more than six years of operation and an investment of roughly RM115 million. Drawing on 3,169 reviews analysed with Zero-Shot Text Classification, LLM-based key-phrase extraction, and TF-IDF within an Extended TAM framework, it found that usability problems dominate. PEOU accounts for 69.7% of complaints, driven by login failures, crashes, and registration errors that effectively lock users out of the system. Security concerns (8.8%) surrounding payment failures and missing balances are smaller in share but carry considerable weight for user trust, while PU complaints (5.7%) point to expected features, such as parking reminders, payment notifications, and receipts, that the application does not provide. That these issues have persisted since 2019 reflects a development process that reacts to problems rather than preventing them.

The study has two main limitations. First, the dataset of 3,169 reviews, while substantial, represents only part of the reviews available on the Google Play Store, owing to platform restrictions and anti-scraping mechanisms; although it closely matches the overall rating distribution, findings should be read with this constraint in mind. Second, classification accuracy for PU (77.14%) and PSR (82%) was lower than for PEOU and Other, reflecting semantic overlap among functionality, usability, and security complaints. Future work could refine the classification prompts, add subcategories, make fuller use of DSPy optimisation, and extend the analysis to Apple App Store reviews for a broader perspective. The LLM-based method demonstrated here also holds considerable promise as an ongoing monitoring tool. In the end, improving the PSP application is as much a matter of restoring public trust in government digital services as it is a technical challenge.

Ethical Approval

This study analysed user reviews that are publicly available on the Google Play Store and did not involve human participants, interviews, or the collection of private information. Usernames present in the raw data were excluded from the analysis, and no identifiable personal data were used. Formal institutional ethical approval was therefore not required. All data were gathered manually in accordance with Google Play's comment posting policy.

Conflict Of Interest

The author declares that there is no conflict of interest.

Data Availability

The user reviews analysed in this study were obtained from the publicly accessible Google Play Store listing of the Penang Smart Parking application. The compiled and preprocessed dataset is available from the author on reasonable request.

REFERENCES

1. Ahmed, A. B., & Haruna, K. (2025). Enhanced SMS spam detection using Bernoulli Naive Bayes with TF-IDF. FUDMA Journal of Sciences, 9(1), 393–399.
2. Alfaridzi, M. I., Syafaah, L., & Lestandy, M. (2022). Emotional text classification using TF-IDF and LSTM. JUITA: Jurnal Informatika, 10(2).
3. Almaiah, M. A., Al-Otaibi, S., Shishakly, R., Hassan, L., Lutfi, A., Alrawad, M., Qatawneh, M., & Abu Alghanam, O. (2023). Investigating the role of perceived risk, perceived security and perceived trust on smart m-banking application using SEM. Sustainability, 15(13), 9908.
4. Alsaleh, N., Alnanih, R., & Alowidi, N. (2025). Enhancing software usability through LLMs: A prompting and fine-tuning framework for analyzing negative user feedback. Computers, 14(9), 363.
5. AlShammari, A. F. (2023). Implementation of keyword extraction using term frequency-inverse document frequency (TF-IDF) in Python. International Journal of Computer Applications, 185(35), 9–13.
6. Andrade, C. (2020). The limitations of online surveys. Indian Journal of Psychological Medicine, 42(6), 575–576.
7. Asia, D. N. (2020). Almost 50% of Malaysian consumers distrust online payment security.
8. Bernama. (2020). CAP suggests two-hour limit for Penang public parking lots. Malay Mail.
9. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., & Dhariwal, P. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.
10. Bui, M. P., & Nguyen, M. T. N. (2025). How large language models enhance topic modeling on user-generated content. Journal of Physics: Conference Series, 3114(1), 012011.
11. Carmona-Díaz, G., Jiménez-Leal, W., Grisales, M. A., Sripada, C., Amaya, S., Inzlicht, M., & Bermúdez, J. P. (2025). An AI-powered research assistant in the lab: A practical guide for text analysis through iterative collaboration with LLMs. Behavior Research Methods.

12. Chan, W. M., & Lee, J. W. C. (2021). 5G connected autonomous vehicle acceptance: The mediating effect of trust in the technology acceptance model. Asian Journal of Business Research, 11(1).
13. Chaudhary, M., Jain, C., & Anish, P. R. (2025). Exploring zero-shot app review classification with ChatGPT: Challenges and potential.
14. Cronin, P., Coughlan, M., & Ryan, F. (2009). Survey research: Process and limitations. International Journal of Therapy and Rehabilitation, 16(1), 9–15.
15. Dasgupta, A., & Yavary Mehr, S. (2024). Enhanced MNB method for SPAM e-mail/SMS text detection using TF-IDF vectorizer. American Journal of Mathematical and Computer Modelling, 9(1).
16. Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. Management Science, 35(8), 982–1003.
17. Dong, Y., & Itoh, M. (2025). A modification of technology acceptance model for investigating driver-vehicle interaction systems usage. PLOS ONE, 20(4).
18. Durango, P. A. A., & Monterola, E. B. (2025). Exploring tech adoption and resistance to new systems at St. Michael's College, Iligan City. International Journal of Social Science and Human Research, 8(11), 8604–8611.
19. Dziak, M. (2024). Technology acceptance model (TAM. EBSCO Research Starters.
20. Eftee, S. Y., Khan, M. Y., Noor, R., Mahmud, H., & Hasan, M. K. (2025). Extraction of app problems and its corresponding user action from user review of apps using few-shot learning. SSRN.
21. Gantulga, U., Sample, B., & Tugsbat, A. (2022). Predicting RFID adoption towards urban smart mobility in Ulaanbaatar, Mongolia. Asia Marketing Journal, 24(1), 2.
22. Google Play. (2026). Google Play Community Post: Posting policy—Google Play Help. https://support.google.com/googleplay/answer/16386215?hl=en
23. Jain, V., & Jain, N. (2024). From cash to clicks: A systematic review of digital payment adoption using the ADO framework. SAGE Open, 14(4).
24. Jamil, N. S., Ku-Mahamud, K. R., Mohamed Din, A., Ahmad, F., ChePa, N., Wan Ishak, W. H., Din, R., & Ahmad, F. K. (2017). A subject identification method based on term frequency technique. International Journal of Advanced Computer Research, 7(30).
25. Khattab, O., Singhvi, A., Maheshwari, P., Zhang, Z., Santhanam, K., Vardhamanan, S., Haq, S., Sharma, A., Joshi, T. T., Moazam, H., Miller, H., Zaharia, M., & Potts, C. (2023). DSPy: Compiling Declarative Language Model Calls into Self-Improving Pipelines (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2310.03714
26. Kuzman, T., Mozetič, I., & Ljubešić, N. (2023). ChatGPT: Beginning of an end of manual linguistic data annotation? Use case of automatic genre identification.
27. Latif, I. S., Saputro, R. E., & Barkah, A. S. (2025). Improvement of Technology Acceptance Model (TAM) with PLS-SEM: A systematic literature review. Journal of Information Systems and Informatics, 7(2), 1376.
28. Lemos, F., Alves, V., & Ferraz, F. (2025). Is It Time To Treat Prompts As Code? A Multi-Use Case Study For Prompt Optimization Using DSPy (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2507.03620
29. Lin, C. (2022). Exploring transaction security on consumers' willingness to use mobile payment by using the Technology Acceptance Model. Applied System Innovation, 5(6), 113.
30. Low, Y. Y., & Poh, H. Z. (2018). The determinants of intention to use vehicle parking mobile payment application [[Bachelor's thesis,]. Universiti Tunku Abdul Rahman.
31. Ma, Q., & Liu, L. (2004). The technology acceptance model: A meta-analysis of empirical findings. Journal of Organizational and End User Computing, 16(1), 59–72.
32. Mahat, S. A. (2025). Assessing user satisfaction of smart parking payment system in Selangor [[Master's thesis,]. Universiti Utara Malaysia.
33. Merhi, M., Hone, K., & Tarhini, A. (2019). A cross-cultural study of the intention to use mobile banking between Lebanese and British consumers: Extending UTAUT2 with security, privacy and trust. Technology in Society, 59, Article 101151.
34. Mervaala, E., & Kousa, I. (2025). Out of context! Managing the limitations of context windows in ChatGPT-4o text analyses. Zenodo.

35. Mok, O. (2019). Penang's RM115m smart parking system hits snag, contractor told to show cause. Malay Mail.
36. Mu, Y., Wu, B. P., Thorne, W., Robinson, A., Aletras, N., Scarton, C., Bontcheva, K., & Song, X. (2024). Navigating prompt complexity for zero-shot classification: A study of large language models in computational social science.
37. Mustafa, A. A., & Abdulazeez, A. M. (2024). A review of text classification based on ML & data mining algorithms. Indonesian Journal of Computer Science, 13(3), 3896.
38. Mutiara, B. (2019, July 26). Penang Smart Parking (PSP) system starts next month.
39. Naeem, M. Z., Rustam, F., Mehmood, A., Mui-zzud-din, A., I., & Choi, G. S. (2022). Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms. PeerJ Computer Science, 8, Article e914.
40. Nielbo, K. L., Karsdorp, F., Wevers, M., Lassche, A., Baglini, R. B., Kestemont, M., & Tahmasebi, N. (2024). Quantitative text analysis. Nature Reviews Methods Primers, 4(1), 25. https://doi.org/10.1038/s43586-024-00302-w
41. NST. (2024). No one in the driver's seat: Penang's parking problem is out of control. https://www.nst.com.my/news/nation/2024/05/1050579/no-one-drivers-seat-penangs-parking-problem-out-control
42. Oematan, M. E., Rahayu, S., & Dyah, J. (2024). The effect of perceived usefulness and perceived ease of use on behavioral intention mediated by user satisfaction in e-commerce users. Jurnal Ekonomi, 13(3), 472.
43. Peng, G., Clough, P. D., Madden, A., Xing, F., & Zhang, B. (2021). Investigating the usage of IoT-based smart parking services in the Borough of Westminster. Journal of Global Information Management, 29(6).
44. Permana, M. E., Ramadhan, H., Budi, I., Santoso, A. B., & Putra, P. K. (2020). Sentiment analysis and topic detection of mobile banking application review. Proceedings of the 2020 International Conference on Informatics and Computational Sciences (ICICoS.
45. Said, A. J., & Ismail, A. M. (2025). Trends in natural language processing for text classification: A comprehensive survey. International Journal of Science and Research Archive, 14(2), 1540–1547.
46. Saprikis, V., & Vlachopoulou, M. (2021). Mapping mobile payment adoption: Customers' trends and challenges. International Journal of Business and Management, 16(9), 82.
47. Shah, F. A., Sabir, A., Sharma, R., & Pfahl, D. (2024). How Effectively Do LLMs Extract Feature-Sentiment Pairs from App Reviews? (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2409.07162
48. Shi, W., Zheng, W., Yu, J. X., & Cheng, H. (2017). Keyphrase extraction using knowledge graphs. In Lecture Notes in Computer Science (pp. 1–16). Springer.
49. Shrestha, S., & Mahmoud, A. (2025). Mobile application review summarization using chain of density prompting.
50. Sjarif, N. N. A., Mohd Azmi, N. F., Chuprat, S., Md Sarkan, H., Yahya, Y., & Mohd Sam, S. (2019). SMS spam message detection using term frequency-inverse document frequency and random forest algorithm. Procedia Computer Science, 161, 509–515.
51. SUK Pulau Pinang. (2026). Penang Smart Parking—Apps on Google Play. https://play.google.com/store/apps/details?id=com.psp.psppark&hl=en
52. Sung, Y. Y., & Kim, S. B. (2019). Topical keyphrase extraction with hierarchical semantic networks.
53. Szmigin, I., & Foxall, G. (1998). Three forms of innovation resistance: The case of retail payment methods. Technovation, 18(6/7), 459–468.
54. Tchakounté, F., Pagore, A. E. Y., Kamgang, J. C., & Atemkeng, M. (2020). CIAA-RepDroid: A fine-grained and probabilistic reputation scheme of Android apps based on sentiment analysis of reviews. Preprints.
55. Ting, C. P., Muniandy, M., & Batumalai, C. (2024). Smart parking mobile app system allocation and navigation. Journal of Engineering Science and Technology, 19(Special Issue on ICIT2022), 40–46.
56. Törnberg, P. (2023). How to use large language models for text analysis.
57. Tsimenidis, S. (2020). Limitations of deep neural networks: A discussion of G.

58. Wang, Y., Wang, W., Chen, Q., Huang, K., Nguyen, A., & De, S. (2024). Zero-shot text classification with knowledge resources under label-fully-unseen setting. Neurocomputing, 610, 128580.
59. Wang, Z., Pang, Y., & Lin, Y. (2023). Large language models are zero-shot text classifiers.
60. Yang, C., Ye, X., Xie, J., Yan, X., Lu, L., Yang, Z., Wang, T., & Chen, J. (2020). Analyzing drivers' intention to accept parking app by structural equation model. Journal of Advanced Transportation, 3051283.
61. Zhang, K., Zhang, Q., Wang, C.-C., & Jang, J.-S. R. (2025). Comprehensive study on zero-shot text classification using category mapping. IEEE Access, 13, 23526–23546.