International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

# Public Perception and Adoption Readiness for Machine Learning–Based House Price Prediction: An Empirical Survey Study with an Exploratory Predictive Analysis

**Mr. Yashkumar Ashok Koli,**
**Prof. Prasad Hiray**

## Abstract

Machine learning (ML) is now the dominant paradigm for automated property valuation, yet the technical literature rarely examines whether prospective users understand, trust, or intend to use such systems. This study analyses a questionnaire dataset of $N = 56$ respondents (76.8% aged 18–25; 76.8% students) that captures awareness of ML-based house price prediction, beliefs about price determinants, algorithm familiarity, perceived accuracy, intention to use, and principal concerns. Descriptive statistics, Wilson confidence intervals, Spearman rank correlations, chi-square tests with Cramér's $V$, Mann–Whitney tests, and cross-validated classifiers (logistic regression, decision tree, random forest) were applied. Location was the most frequently named price driver (53.6%) and was rated "very" or "extremely" important by 89.1% of respondents. Although 87.5% had at least heard of ML price prediction, only 37.5% believed it can predict prices accurately without qualification (44.6% answered "maybe"). Nonetheless, 78.6% rated ML as very or extremely useful and 67.9% would probably or definitely use an online predictor. Intention to use correlated with perceived usefulness ($\rho = 0.54$, $p < .001$) and awareness ($\rho = 0.35$, $p = .008$). Respondents who were at least somewhat aware reported higher intent than less-aware respondents (mean 4.26 vs. 3.39; $p = .013$). The leading concerns were outdated property data (29.1%) and incorrect predictions (25.5%). In exploratory cross-validation, predictive models only marginally exceeded a majority-class baseline, underscoring the limits of small convenience samples. The findings suggest that data freshness, location richness, and transparency are the main levers for adoption.

**Keywords:** house price prediction; machine learning; automated valuation models; technology acceptance; survey analysis; explainable AI

## 1. Introduction

### 1.1 Problem Statement

Residential property is typically the largest financial asset a household acquires, and errors in price estimation translate directly into financial loss for buyers and sellers [1]. Traditional valuation relies on appraiser judgment and comparable-sales analysis, which is slow, subjective, and difficult to scale. Machine learning-based *automated valuation models* (AVMs) learn a mapping from property attributes to market price,

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

$$\hat{y}_i = f(\mathbf{x}_i; \boldsymbol{\theta}) + \varepsilon_i, \qquad \mathbf{x}_i = \left(\text{location}_i, \text{area}_i, \text{bedrooms}_i, \text{age}_i, \text{amenities}_i, \ldots\right),$$

and have been shown to outperform classical hedonic regression in many settings [2], [3]. However, technical accuracy is only one precondition for real-world impact. Systems that people do not understand or trust will not be used, regardless of their error metrics [4].

### 1.2 Objectives

This study pursues four objectives:

1. To characterise how a sample of prospective home buyers perceive the determinants of house prices (location, size, bedrooms, age, nearby facilities).
2. To measure awareness of, familiarity with, and attitudes toward ML-based price prediction.
3. To identify the demographic and attitudinal correlates of intention to use an online prediction tool.
4. To assess, exploratorily, whether ML classifiers can predict adoption intent from the respondent attributes collected.

### 1.3 Research Questions

- **RQ1.** Which property attributes do respondents regard as the most influential on house price, and do these beliefs agree with the feature importance typically reported in the ML literature?
- **RQ2.** How aware are respondents of ML-based price prediction, and which algorithms do they recognise?
- **RQ3.** How much do respondents trust the accuracy of ML predictions, and what concerns do they hold?
- **RQ4.** Which variables are associated with intention to use a prediction system?

### 1.4 Significance

Understanding the user side of AVMs informs interface design, data-governance priorities, and communication strategy for proptech and mortgage platforms. Because the respondents here are largely young, prospective first-time buyers, their views indicate the expectations of the next generation of property-market participants.

### 1.5 Scope and Honest Framing

The dataset contains self-reported survey responses, not property transactions. It therefore contains **no price variable**, and no price-prediction model is trained here. The analysis addresses the *perceptual and adoption* dimension of ML-based house price prediction, complementing rather than replacing technical benchmarking studies.

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

## 2. Literature Review

### 2.1 Hedonic Pricing Foundations

The theoretical basis for modelling house prices is the hedonic price model of Rosen [5], which treats a dwelling as a bundle of characteristics, each carrying an implicit price:

$$P_i = \beta_0 + \sum_{k=1}^{K} \beta_k \, x_{ik} + \varepsilon_i.$$

Empirical hedonic studies consistently find that locational attributes, structural size, and neighbourhood amenities explain most price variation [6], [7]. Respondents' beliefs in the present survey can be read as lay expressions of this framework.

### 2.2 Machine Learning for Price Prediction

Linear regression and its regularised variants remain common baselines. Non-linear methods, notably decision trees, random forests, gradient boosting, support vector regression (SVR), and neural networks, capture interactions between location and structure that linear models miss [2], [3], [8]. Random forests [9] aggregate $B$ decorrelated trees,

$$\hat{f}_{\mathrm{RF}}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} T_b \, (\mathbf{x}),$$

and gradient-boosted trees [10] have become the default for tabular valuation tasks. Comparative studies on public benchmarks such as Ames Housing [11] and Boston Housing report that ensemble methods typically reduce RMSE substantially relative to linear baselines, with the caveat that gains depend on feature engineering and data quality [3], [12]. Common evaluation metrics are

$$\mathrm{RMSE} = \sqrt{1/n \sum_i (y_i - \hat{y}_i)^2}\,, \qquad \mathrm{MAPE} = 100/n \sum_i |y_i - \hat{y}_i / y_i|\,, \qquad R^2 = 1 - \sum_i (y_i - \hat{y}_i)^2 \Big/ \sum_i (y_i - \bar{y})^2.$$

### 2.3 Data Quality, Location, and Spatial Effects

A recurring finding is that spatial and neighbourhood context carries much of the predictive signal; richer location information, including imagery, improves estimates [13]. Equally, stale or incomplete listing data produces biased estimates, a problem that motivates continuous data refresh in industrial AVMs [14].

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

### 2.4 Trust, Transparency, and Technology Acceptance

The Technology Acceptance Model (TAM) posits that *perceived usefulness* and *perceived ease of use* determine behavioural intention [15]; the extended UTAUT framework adds social influence and facilitating conditions [16]. Research on algorithmic decision aids shows that users often under-use accurate models when they cannot understand them ("algorithm aversion") [17], and that explanations improve calibrated trust [4], [18]. In real-estate contexts, concerns about accuracy and data provenance are salient after highly publicised AVM failures [19].

### 2.5 Research Gap

Although the algorithmic literature is large, few studies quantify lay awareness of ML valuation, the beliefs users hold about price determinants, and the link between perceived usefulness and intention to use. This study provides a small-sample empirical contribution to that gap, with the TAM logic (usefulness → intention) as its organising hypothesis.

## 3. Methodology

### 3.1 Data Source and Structure

The dataset is an export of an online questionnaire (file *Untitled form.csv*) comprising 56 responses and 16 columns: a timestamp and 15 closed-ended items. Responses were collected between 1 November 2026 (14:31 IST) and 2 November 2026 (10:42 IST), i.e., within roughly 20 hours, which is characteristic of a convenience sample distributed through personal networks. All fields are categorical or ordinal; there are no continuous variables. Missing values are rare (4 cells in 4 columns, 0.4% of all cells): home ownership ($n = 1$), location importance ($n = 1$), preferred nearby facility ($n = 1$), and main concern ($n = 1$). Missing values were excluded pairwise in statistical tests and mode/median-imputed inside classifier pipelines.

**Table 1.** Variables used in the analysis.

| ID | Variable | Type | Levels / Coding |
|---|---|---|---|
| Q1 | Age group | Nominal | Below 18; 18–25; 25–35; Above 35 |
| Q2 | Occupation | Nominal | Student; Private employee; Government employee; Business owner; Other |
| Q3 | House ownership | Nominal | Yes; No; Planning to buy |
| Q4 | Awareness of ML price prediction | Ordinal | 0 = Not aware … 3 = Fully aware |
| Q5 | Most influential price factor | Nominal | Location; Nearby facilities; Size; Age; Bedrooms |
| Q6 | Importance of location | Ordinal | 1 = Not important … 5 = Extremely important |
| Q7 | Effect of bedrooms | Ordinal | 1 = Does not … 4 = Strongly ("Not sure" set missing) |
| Q8 | Effect of house age | Nominal | Older cheaper; Newer costlier; No effect; Not sure |

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

| ID | Variable | Type | Levels / Coding |
|---|---|---|---|
| Q9 | Most value-adding facility | Nominal | Schools/colleges; Public transport; Malls; Hospitals |
| Q10 | Algorithm recognised (single choice) | Nominal | Linear regression; Decision tree; Random forest; SVR; None |
| Q11 | ML can predict accurately | Ordinal | 0 = No; 1 = Not sure; 2 = Maybe; 3 = Yes |
| Q12 | Would use online predictor | Ordinal | 1 = Never … 5 = Definitely |
| Q13 | Importance of accurate prediction | Ordinal | 1 … 5 |
| Q14 | Biggest concern | Nominal | Outdated data; Incorrect prediction; Lack of location info; Privacy; None |
| Q15 | Overall usefulness of ML | Ordinal | 1 = Not useful … 5 = Extremely useful |

### 3.2 Theoretical Framework

The analysis is organised around TAM [15]. *Perceived usefulness* (Q15) and *perceived accuracy/trust* (Q11) are treated as antecedents of *behavioural intention* (Q12), with *awareness* (Q4) and demographic variables as external variables. Beliefs about price determinants (Q5–Q9) are compared with the hedonic framework [5].

### 3.3 Analytical Methods

**Descriptive statistics.** Frequencies, percentages, means, medians, and standard deviations of ordinal codes. Proportions are accompanied by 95% Wilson score intervals,

$$\tilde{p} \pm \frac{z\sqrt{\hat{p}(1-\hat{p})/n + z^2/4n^2}}{1 + z^2/n}, \qquad \tilde{p} = \frac{\hat{p} + z^2/2n}{1 + z^2/n}.$$

**Association between ordinal variables.** Spearman's rank correlation,

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$ (with tie correction applied by the software).

**Association between nominal variables.** Pearson's $\chi^2$ test of independence,

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

with effect size Cramér's $V = \sqrt{\chi^2/[n(\min(r,c)-1)]}$.

**Group comparisons.** Mann–Whitney $U$ tests for ordinal outcomes between awareness groups (Fully/Somewhat aware vs. Heard-only/Not aware) and between students and non-students.

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

**Scale reliability.** Cronbach's $\alpha$ for a four-item attitude composite (Q11, Q12, Q13, Q15).

**Exploratory predictive modelling.** Adoption intent was binarised ($y = 1$ if Q12 is "Probably" or "Definitely"; 38 of 56, 67.9%). Logistic regression with $L_2$ penalty ($C = 0.5$), a depth-3 decision tree, and a random forest (150 trees, depth 4) were evaluated by 10× repeated stratified 5-fold cross-validation, with one-hot encoding of nominal predictors and standardisation of ordinal ones inside each fold. Accuracy, $F_1$, and ROC-AUC were reported against a majority-class baseline. Two feature sets were compared: (A) profile and awareness only (age, occupation, ownership, algorithm recognised, awareness) and (B) the full set, including perception items.

All computations used Python (pandas, SciPy, statsmodels, scikit-learn). Significance was assessed at $\alpha = .05$; because many tests were run, $p$-values are reported descriptively and should be treated as hypothesis-generating (Section 5.3).

## 4. Results and Findings

### 4.1 Sample Profile

**Table 2.** Demographic and ownership profile ($N = 56$).

| Variable | Category | $n$ | % |
|---|---|---|---|
| Age group | 18–25 years | 43 | 76.8 |
| | 25–35 years | 6 | 10.7 |
| | Below 18 years | 4 | 7.1 |
| | Above 35 years | 3 | 5.4 |
| Occupation | Student | 43 | 76.8 |
| | Private employee | 5 | 8.9 |
| | Government employee | 3 | 5.4 |
| | Other | 3 | 5.4 |
| | Business owner | 2 | 3.6 |
| House ownership | No | 27 | 49.1 |
| | Yes | 23 | 41.8 |
| | Planning to buy | 5 | 9.1 |
| | *(missing)* | 1 | — |

The sample is strongly skewed toward young students: 38 of 56 respondents (67.9%) are students aged 18–25. Roughly half do not own property, while 42% report owning a house, which is plausible if respondents answered on behalf of their family household.

### 4.2 Awareness and Algorithm Familiarity (RQ2)

**Table 3.** Awareness of ML-based house price prediction and recognised algorithms.

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

| Awareness level | $n$ | % | Algorithm recognised | $n$ | % |
|---|---|---|---|---|---|
| Yes, fully aware | 27 | 48.2 | Linear regression | 21 | 37.5 |
| Heard about it | 11 | 19.6 | Decision tree | 13 | 23.2 |
| Somewhat aware | 11 | 19.6 | None of these | 10 | 17.9 |
| Not aware at all | 7 | 12.5 | Random forest | 6 | 10.7 |
| | | | Support vector regression | 6 | 10.7 |

At least some awareness was reported by 87.5% (95% CI [76.4, 93.8]); 48.2% [35.7, 61.0] claimed full awareness. Linear regression was the most recognised algorithm (37.5%), followed by decision trees (23.2%). Ensemble methods, which dominate contemporary practice [3], [10], were recognised by only 10.7%. Because Q10 is single-choice, these figures give each respondent's *selected* algorithm rather than the full set they know; 82.1% selected at least one named algorithm. Respondents rating themselves fully aware nonetheless selected a named algorithm in 25 of 27 cases (chi-square association between awareness and algorithm: $\chi^2(12) = 12.59$, $p = .399$, $V = .27$), so self-reported awareness is not strongly tied to algorithm knowledge in this sample.

### 4.3 Beliefs About House Price Determinants (RQ1)

**Table 4.** Perceived determinants of house price.

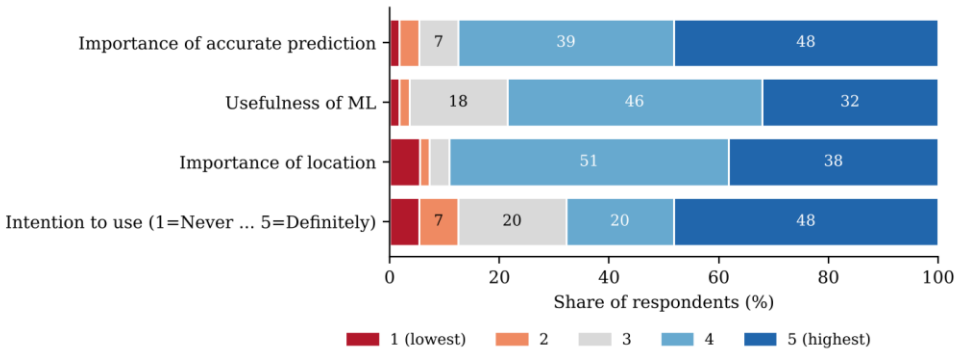
| Item | Response | $n$ | % |
|---|---|---|---|
| Most influential factor (Q5) | Location | 30 | 53.6 |
| | Nearby facilities | 11 | 19.6 |
| | House size (area) | 9 | 16.1 |
| | Age of the property | 4 | 7.1 |
| | Number of bedrooms | 2 | 3.6 |
| Importance of location (Q6, $n = 55$) | Extremely important | 21 | 38.2 |
| | Very important | 28 | 50.9 |
| | Moderately / Slightly / Not | 2 / 1 / 3 | 3.6 / 1.8 / 5.5 |
| Effect of bedrooms (Q7) | Strongly | 20 | 35.7 |
| | Moderately | 19 | 33.9 |
| | Slightly | 9 | 16.1 |
| | Not sure | 5 | 8.9 |
| | Does not | 3 | 5.4 |
| Effect of age (Q8) | Older houses cost less | 21 | 37.5 |
| | Newer houses cost more | 19 | 33.9 |
| | Age has no effect | 11 | 19.6 |
| | Not sure | 5 | 8.9 |

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

| Item | Response | $n$ | % |
|---|---|---|---|
| Facility adding most value (Q9, $n = 55$) | Schools and colleges | 22 | 40.0 |
| | Public transport | 16 | 29.1 |
| | Shopping malls | 12 | 21.8 |
| | Hospitals | 5 | 9.1 |

Location is the dominant belief: 53.6% [40.7, 66.0] named it the top factor, and 89.1% [78.2, 94.9] rated it very or extremely important (mean 4.15, median 4, skewness −1.85). Adding "nearby facilities", which is itself locational, raises the share attributing price mainly to *where* a house is to 73.2%. Bedrooms were seen as influential by 69.6% (moderate or strong), yet only 3.6% chose them as the *primary* driver, indicating that respondents treat bedrooms as a secondary attribute. On age, 71.4% held a directional view (newer costs more or older costs less), consistent with depreciation in hedonic models, but 28.6% saw no effect or were unsure. Schools and colleges were the most valued amenity (40.0%), which may reflect the student-dominated sample.

These beliefs are broadly consistent with the literature that identifies location and structural size as leading price determinants [6], [7], [13], supporting a positive answer to RQ1. However, the survey measures *beliefs*, not model-derived importances.




*Distribution of five-point responses (percentage labels shown for segments ≥ 7%).*

### 4.4 Trust, Intention, and Perceived Usefulness (RQ3)

**Table 5.** Attitudes toward ML-based price prediction.

| Item | Mean | SD | Median | Top-2 / favourable share (95% CI) |
|---|---|---|---|---|
| Perceived usefulness (Q15; 1–5) | 4.05 | 0.86 | 4 | 78.6% [66.2, 87.3] (very/extremely useful) |
| Importance of accurate prediction (Q13; 1–5) | 4.29 | 0.89 | 4 | 87.5% [76.4, 93.8] |
| Intention to use (Q12; 1–5) | 3.98 | 1.21 | 4 | 67.9% [54.8, 78.6] (probably/definitely) |

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

| Item | Mean | SD | Median | Top-2 / favourable share (95% CI) |
|---|---|---|---|---|
| Belief ML can predict accurately (Q11; 0–3) | 2.11 | 0.91 | 2 | 37.5% [26.0, 50.6] "Yes"; 82.1% [70.2, 90.0] "Yes" or "Maybe" |

Respondents value accuracy highly (87.5%) and rate ML as useful (78.6%), but trust is hedged: only 37.5% answered an unqualified "Yes" to ML accuracy, 44.6% answered "Maybe", and 17.9% answered "No" or "Not sure". Intent to use is nonetheless high, with 48.2% answering "Definitely" and only 12.5% "Probably not" or "Never". This pattern, in which usefulness and intent outpace confidence in accuracy, suggests that users see prediction tools as decision *aids* rather than authoritative valuations.

**Table 6.** Principal concerns about ML-based house price prediction.

| Concern | $n$ | % | Mean intent to use |
|---|---|---|---|
| Outdated property data | 16 | 29.1 | 3.81 |
| Incorrect price prediction | 14 | 25.5 | 4.36 |
| Lack of location information | 14 | 25.5 | 3.71 |
| Privacy and data security | 8 | 14.5 | 4.38 |
| No major concerns | 3 | 5.5 | 3.00 |
| *(missing)* | 1 | — | — |

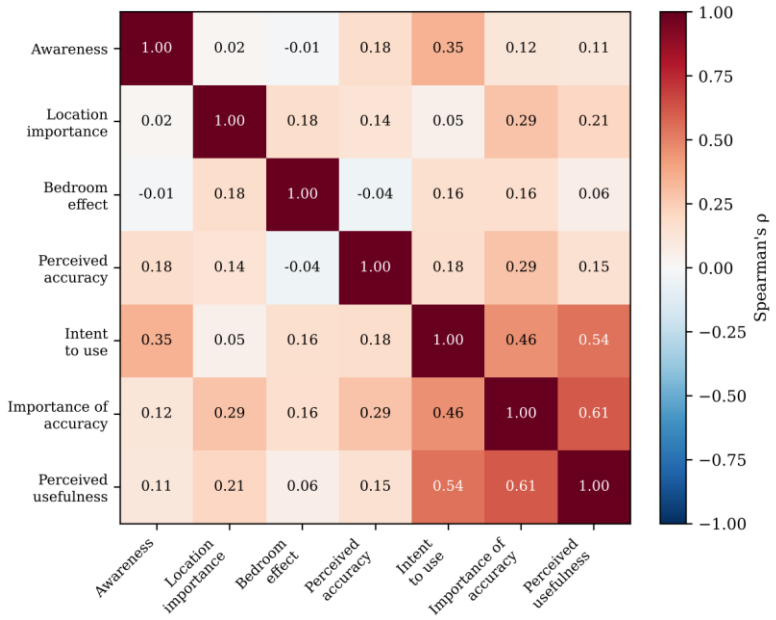
Data-related concerns (outdated data plus missing location information) account for 54.5% of valid responses, exceeding concerns about the prediction error itself (25.5%) or privacy (14.5%). This aligns with the observation that input data quality, especially spatial data, governs AVM performance [13], [14]. Intent differed by concern type (Pearson $\chi^2(16) = 27.63$, $p = .035$, $V = .35$), but expected cell counts were mostly below 5, so this result is fragile; notably, the three respondents with "no major concerns" reported the lowest intent (mean 3.00), a reminder that absence of concern is not equivalent to enthusiasm.

### 4.5 Correlates of Intention to Use (RQ4)

**Table 7.** Selected Spearman correlations ($n = 56$ unless noted).

| Pair | $\rho$ | $p$ |
|---|---|---|
| Importance of accurate prediction – Perceived usefulness | 0.62 | <.001 |
| Intention to use – Perceived usefulness | 0.54 | <.001 |
| Intention to use – Importance of accurate prediction | 0.46 | <.001 |
| Awareness – Intention to use | 0.35 | .008 |
| Location importance – Importance of accurate prediction ($n = 55$) | 0.29 | .033 |
| Perceived accuracy – Importance of accurate prediction | 0.29 | .033 |
| Perceived accuracy – Intention to use | 0.18 | .189 |
| Location importance – Intention to use ($n = 55$) | 0.05 | .707 |

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312




*Spearman correlation matrix of ordinal survey items.*

The strongest association is between the importance placed on accurate prediction and perceived usefulness ($\rho = .62$), followed by usefulness and intention ($\rho = .54$), which supports the TAM logic that perceived usefulness is the principal correlate of behavioural intention [15]. Awareness has a moderate positive association with intention ($\rho = .35$). By contrast, *belief in accuracy* is only weakly related to intention ($\rho = .18$, n.s.). Respondents are willing to use tools they cannot yet fully trust.

**Table 8.** Intention to use by awareness level.

| Awareness | $n$ | Mean intent (1–5) | Mean usefulness | Mean perceived accuracy (0–3) |
|---|---|---|---|---|
| Fully aware | 27 | 4.33 | 4.15 | 2.19 |
| Somewhat aware | 11 | 4.09 | 4.18 | 2.09 |
| Heard about it | 11 | 3.45 | 3.82 | 2.09 |
| Not aware at all | 7 | 3.29 | 3.86 | 1.86 |

Collapsing to "at least somewhat aware" ($n = 38$) versus "heard only / unaware" ($n = 18$), mean intent was 4.26 vs. 3.39 (Mann–Whitney $p = .013$); differences in usefulness (4.16 vs. 3.83, $p = .27$), perceived accuracy (2.16 vs. 2.00, $p = .28$), and importance of accuracy (4.39 vs. 4.06, $p = .21$) were not significant. Awareness thus appears to relate to willingness to *use* a tool more than to its perceived merit. Other group contrasts were not significant: students versus non-students on intent (4.07 vs. 3.69, $p = .36$); and by ownership, those *planning to buy* reported the highest intent (4.40, $n = 5$) and perceived accuracy (2.60) compared with owners (4.04) and non-owners (3.81), but group sizes are too small for inference. Occupation and intent showed a moderate but non-significant association ($\chi^2(16) = 24.31$, $p = .083$, $V = .33$).

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

The four-item attitude composite (Q11, Q12, Q13, Q15) had a modest internal consistency (Cronbach's $\alpha = .63$, $n = 56$), below the conventional .70 threshold, so it was not used as a single scale in subsequent tests.

### 4.6 Exploratory Predictive Modelling of Adoption Intent

**Table 9.** Repeated 5-fold cross-validation (mean ± SD across 50 folds); target: intent = "Probably/Definitely" (67.9% positive).

| Feature set | Model | Accuracy | $F_1$ | ROC-AUC |
|---|---|---|---|---|
| — | Majority baseline | 0.679 ± 0.041 | 0.808 ± 0.029 | 0.500 |
| A: profile + awareness | Logistic (L2) | 0.640 ± 0.085 | 0.763 ± 0.072 | 0.612 ± 0.155 |
| A | Decision tree | 0.575 ± 0.137 | 0.670 ± 0.157 | 0.538 ± 0.148 |
| A | Random forest | 0.651 ± 0.092 | 0.770 ± 0.078 | 0.589 ± 0.161 |
| B: full feature set | Logistic (L2) | 0.701 ± 0.093 | 0.792 ± 0.069 | 0.746 ± 0.142 |
| B | Decision tree | 0.620 ± 0.124 | 0.718 ± 0.132 | 0.577 ± 0.141 |
| B | Random forest | 0.704 ± 0.082 | 0.807 ± 0.059 | 0.712 ± 0.157 |

Using profile and awareness variables alone (set A), no model beat the majority baseline in accuracy or $F_1$. With perception items added (set B), logistic regression and random forest reached ROC-AUC of 0.75 and 0.71 respectively, indicating some ranking ability, but accuracy (0.70) was essentially at baseline (0.68) and fold-to-fold variability was large (SD ≈ 0.08–0.16). The random forest's impurity-based importances ranked the importance of accurate prediction (0.12), perceived usefulness (0.09), and awareness (0.08) highest, in agreement with the correlation results. With only 56 observations, these models are **not** suitable for deployment; the exercise demonstrates that adoption intent cannot be reliably predicted from this questionnaire alone, and it illustrates the small-sample overfitting risk in tree-based learners.

## 5. Discussion

### 5.1 Interpretation

Three themes emerge. **First**, lay beliefs match the hedonic/ML consensus: location, direct or via amenities, is perceived to dominate price formation (RQ1), supporting the practice of investing in rich spatial features when engineering AVMs [13]. **Second**, there is an *acceptance–trust gap* (RQ3): high perceived usefulness (78.6%) and intent (67.9%) coexist with limited unconditional belief in accuracy (37.5%). Intent is predicted by perceived usefulness and awareness rather than by accuracy belief, a pattern consistent with TAM [15] and with evidence that people use decision aids as *advisory* inputs [17]. **Third**, the main concerns are about data (outdated or lacking location detail; 54.5%) rather than algorithmic error or privacy, which points to data pipelines, update frequency, and spatial coverage as the priority for trust-building [14].

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Volume 02 Issue-06 November- December 2026 ISSN NO: 3108-1312

### 5.2 Practical Implications

1. **Surface data freshness.** Display the date of the underlying market data and refresh cadence, directly addressing the most frequent concern (29.1%).
2. **Expose location inputs.** Show which neighbourhood and amenity variables entered the estimate (schools, transport), since respondents prioritise them.
3. **Communicate uncertainty.** Given that 44.6% answer "maybe" on accuracy, report prediction intervals rather than point estimates, e.g. $\hat{y} \pm t_{\alpha/2}\,\hat{\sigma}$, and provide explanations (e.g., SHAP-style attributions [18]).
4. **Raise awareness.** Awareness was the only profile-type variable significantly related to intent; short explanatory onboarding may help convert hesitant users.

### 5.3 Limitations

- **Sample size and power.** $N = 56$ yields wide confidence intervals (typically ±12–13 percentage points) and low power for subgroup analyses; several cells have expected counts below 5, invalidating asymptotic $\chi^2$ approximations.
- **Sampling bias.** The sample is a convenience sample, 76.8% students aged 18–25, collected over about 20 hours; findings cannot be generalised to the home-buying population, and the 41.8% ownership rate is likely household-level rather than personal.
- **Measurement.** All variables are self-reported, closed-ended, and single-item; Q10 (algorithm recognised) is single-choice, understating multi-algorithm familiarity; ordinal codes were treated as numeric for means, an approximation; the attitude composite had weak reliability ($\alpha = .63$).
- **No price data.** The dataset has no property prices or attributes, so ML predictive accuracy for house prices cannot be assessed. The classification exercise predicts *survey responses*, not prices.
- **Multiple comparisons.** More than 40 tests were conducted without correction (Bonferroni threshold ≈ .001); only the three strongest correlations among attitude items ($p < .001$) would survive it. The awareness–intent ($p = .008$) and group-difference ($p = .013$) findings should be regarded as exploratory.
- **Social desirability and awareness bias.** Self-rated "full awareness" was not verified by knowledge questions.

### 5.4 Future Work

1. **Pair survey and price data.** Train AVMs (e.g. regularised regression, random forest, gradient boosting) on a transactional dataset and compare model-derived feature importances with the stated beliefs reported here.
2. **Larger, representative samples** including active buyers, sellers, agents, and lenders, with stratified sampling and a multi-item validated TAM/UTAUT instrument.
3. **Experimental designs** testing how uncertainty intervals, data-freshness labels, and explanations alter trust and willingness to rely on a prediction.
4. **Ordinal and multilevel models.** With adequate data, use proportional-odds regression of intent on usefulness, awareness, and demographics, with appropriate multiplicity control.

5. **Fairness and privacy audits** of deployed AVMs, given that 14.5% already identify privacy as the primary concern.

## 6. Conclusion

This study analysed a 56-respondent survey on ML-based house price prediction. Respondents overwhelmingly regard location as the key determinant of price (89.1% rated it very or extremely important), are broadly aware of ML prediction (87.5%), and view it as useful (78.6%), with two-thirds willing to use an online predictor. Yet only 37.5% unconditionally trust its accuracy, and their foremost concerns relate to *data currency and location coverage* rather than algorithms or privacy. Perceived usefulness ($\rho = .54$) and awareness ($\rho = .35$) were the clearest correlates of adoption intent, while accuracy belief was not significantly related. Exploratory classifiers barely exceeded a majority baseline, which cautions against over-interpreting small-sample survey data. The principal contributions are (i) an evidence-based profile of adoption readiness among a young, student-dominated sample, (ii) a documented acceptance–trust gap, and (iii) concrete design priorities (data freshness, transparent location inputs, uncertainty communication). Confirming these results requires larger, more representative samples and integration with actual transaction data.

## References

[1] J. M. Quigley, "Real estate prices and economic cycles," *International Real Estate Review*, vol. 2, no. 1, pp. 1–20, 1999.

[2] E. Čeh, M. Kilibarda, A. Lisec, and B. Bajat, "Estimating the performance of random forest versus multiple regression for predicting prices of the apartments," *ISPRS International Journal of Geo-Information*, vol. 7, no. 5, p. 168, 2018.

[3] B. Park and J. K. Bae, "Using machine learning classification techniques to predict housing prices: Evidence from Virginia," *Expert Systems with Applications*, vol. 42, no. 6, pp. 2928–2934, 2015.

[4] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[5] S. Rosen, "Hedonic prices and implicit markets: Product differentiation in pure competition," *Journal of Political Economy*, vol. 82, no. 1, pp. 34–55, 1974.

[6] D. Harrison and D. L. Rubinfeld, "Hedonic housing prices and the demand for clean air," *Journal of Environmental Economics and Management*, vol. 5, no. 1, pp. 81–102, 1978.

[7] S. Sirmans, D. Macpherson, and E. Zietz, "The composition of hedonic pricing models," *Journal of Real Estate Literature*, vol. 13, no. 1, pp. 3–43, 2005.

[8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2009.

[9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[10] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[11] D. De Cock, "Ames, Iowa: Alternative to the Boston Housing data as an end of semester regression project," *Journal of Statistics Education*, vol. 19, no. 3, 2011.

[12] V. Limsombunchai, "House price prediction: Hedonic price model vs. artificial neural network," in *Proc. NZ Agricultural and Resource Economics Society Conf.*, 2004.

[13] S. Law, B. Paige, and C. Russell, "Take a look around: Using street view and satellite images to estimate house prices," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 5, pp. 1–19, 2019.

[14] M. Kok, M. Koponen, and C. Martínez-Barbosa, "Big data in real estate? From manual appraisal to automated valuation," *Journal of Portfolio Management*, vol. 43, no. 6, pp. 202–211, 2017.

[15] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[16] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003.

[17] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err," *Journal of Experimental Psychology: General*, vol. 144, no. 1, pp. 114–126, 2015.

[18] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 4765–4774.

[19] Zillow Group, "Zillow Group reports third-quarter 2021 results and announces wind-down of Zillow Offers," shareholder letter, Nov. 2021.