University of Essex
MSc Artificial Intelligence
Module: UAI_PCOM7E January 2025 A
Unit 12: Assessment – AI Solution Implementation

Final Report

I. Executive Summary

Budget Buddy, Inc. (hereinafter, “the company”) enjoys a unique opportunity to incorporate AI-derived software features into its flagship consumer smartphone software product (hereinafter, “the AI solution”). This report summarizes the results of investigation by the company’s informational technology department’s research group into the feasibility of such an endeavor, which investigative results find that it would be both feasible and provide a significant return on investment to the company to invest in further research to identify actionable trends in user personal spending behavior.

II. Investigation Undertaken

The company is engaged in developing a smartphone app that assists public consumer users in identifying and managing their personal and household spending to achieve their financial planning goals, primarily by presenting users with budgeting tips and promotional offers on consumer products they have expressed interest in purchasing or have an established history of purchasing.

Prior market research has established a strong demand signal for such app features in the consumer financial services industry (Rane, Choudhary and Rane, 2024); likewise, the use of AI-derived software features in the financial sector is accepted and widespread. Such solutions have been implemented for a wide variety of subject matter domains, including the analysis of regulatory financial filings, predicting the expected value of investment opportunities, and engaging in fraud detection in point of sale and credit card-based transactions (Cao et al, 2024).

Before adopting any type of AI-derived solution, however, company leadership has requested a proof of concept investigation that confirms that company resources would be efficiently allocated in doing so, and that the AI solution would provide material and actionable results. This report describes the methodology and results of one such investigation that clearly support the adoption of an AI solution by company leadership.

III. Investigative Steps

This investigation followed industry best practices for the research methodology, sample dataset, software platform, and predictive models used:

A. Research Methodology: CRISP-DM

This investigation used the Cross-Industry Standard Process for Data Mining methodology (hereinafter, “CRISP-DM”) to structure its research. In short, CRISP-DM is an iterative process for breaking up data mining projects into six standardized stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Azeroual et al, 2025). CRISP-DM was used because it is recognized in expert literature as reliable and efficacious across a wide variety of subject matter domains, including software product development within the financial services industry (Amalsyah et al, 2025).

In-depth discussion of each stage of the CRISP-DM process is beyond the scope of this report; as a highlight, the business understanding stage is addressed in the preceding section. Subsequent sections of this report will note the relevant CRISP-DM stage addressed, as appropriate.

B. Sample Dataset: Kaggle

This investigation used the “Consumer Behavior and Shopping Habits Dataset: E-Commerce Transaction Trends” dataset available via Kaggle.com (Aslam, 2025). This dataset was chosen for several reasons.

First, since it contains 3,900 records, the set is large enough to provide a reasonable degree of confidence in the results of its analysis. Second, since it contains only 3,900 records, the set is equally small enough to allow rapid and inexpensive analysis, and multiple forms of analysis, that substantially larger datasets would render prohibitively expensive or slow, in terms of computing power expenses or time taken for analysis. Third, the set contains data highly relevant to the company’s product, since it is organized around data points such as consumer age, gender, purchase prices, frequency of purchases, and the use (or non-use, despite availability) of promotional discounts, coupons, and similar purchase incentives.

This type of information is ripe for analysis, because it allows the company to build predictive demographic models about user behavior, which in turn create value opportunities for the company to present to users. For example, should the company be able to establish that consumers of a given gender in a given age rage rarely take advantage of promotional discounts for a specified set of product categories, that creates the opportunity for the company to recommend those promotional discounts to that same set of app users who would otherwise be expected to overlook or ignore them to the detriment of their personal budgeting plans.

Finally, this dataset was selected because it has already been pre-processed; in other words, the data has already been “scrubbed” and normalized for rapid software analysis. While the dataset contains eighteen (18) different data types (Aslam, 2025), only these types were used: Category, Frequency of Purchases, Gender, Item Purchased, Previous Purchases, Promo Code Used, and Purchase Amount (USD). This subset was chosen because it most accurately reflects the range of consumer data the app collects from its users. *See* Appendix A for a sample of the ingested data.

In aggregate, this corresponds to the data understanding and data preparation stages of the CRISP-DM process (Azeroual et al, 2025).

C. Software Platform: WEKA

This investigation used the WEKA software platform. WEKA is an open source, industry standard software platform for machine learning-based data analysis (Foulds et al, 2025; Frank et al, 2016). WEKA was chosen because it is free, open source, and used by thousands of researchers worldwide (Foulds et al, 2025; Frank et al, 2016). The significance of using an open source platform to the company, as a practical matter, is that the underlying code has been made freely available to the public and, as such, has been audited by subject matter experts to confirm the quality of its execution, the accuracy of its results, and the security of its operation (Li et al, 2024).

Please note that this stage corresponds to the data preparation and modeling stages of the CRISP-DM process (Azeroual et al, 2025).

D. Models: Random Tree, Random Forest, and K-Means Clustering

This investigation explored two main predictive models for classifying data: random tree and random forest analysis. Both were explored because they are well-known, straightforward to implement, and amenable to rapid iteration (Burkov, 2019; Hewamalage, Ackermann, and Bergmeir, 2022; Montgomery, Peck and Vining, 2021). Additionally, both models were used since current literature indicates that, in general, random trees and forests are well-suited to financial data analysis. (Usman-Hamza et al, 2025).

This investigation also used K-means clustering as an exploratory tool to group consumer behavior into segments that could provide the company with actionable insights. K-means clustering was chosen because it was likewise found in the literature to be accurate, reliable, and efficient for financial data analysis (Li, 2025).

For each of the models investigated, the goal was to determine how usefully it predicted, or identified an actionable insight regarding, whether a consumer would or would not use a promo code when making a consumer product purchase. As noted above, promo code usage is highly relevant, since product design leadership has identified providing app users with promo codes as a potentially desirable feature for future app development.

Please note that this stage corresponds to the modeling stage of the CRISP-DM process (Azeroual et al, 2025).

IV. Investigative Findings

Please note that the following discussion corresponds to the evaluation stage of the CRISP-DM process (Azeroual et al, 2025).

A. Unproductive: Random Tree and Random Forest Models

Both random tree and random forest models were applied using twenty (20) cross-validation folds for increased accuracy. Acceptance criteria was set by executive leadership at ninety percent (90%) or better predictive accuracy for further investigation. While this acceptance criteria could have just as easily been set higher or lower, it is worth noting that a ninety percent minimum (90%) accuracy cutoff rate is consistent with industry best practices and is established in the literature as an accepted signal for adequate reliability of machine learning research findings for practical use (Joshi, 2025).

Both models were found to more accurate than not, since they were more than fifty percent (50%) accurate, but <u>not accurate enough</u> to be usable for company purposes. Random tree analysis correctly classified 68.36% usages of a promo code, while random forest analysis correctly classified 70.28% usages (see

Appendices B and C). As neither met the acceptance criteria of ninety percent (90%) accuracy, in the interest of efficient usage of staff time and computational resources, no further investigation was undertaken for either.

B. Productive: K-Means Clustering Exploration

K-means clustering was found to be far more fruitful, as it identified several relevant and actionable segments of consumer purchasing behavior. For example, male consumers purchasing pants in the approximately sixty (60) dollar price range were found to use promo codes; in contrast, female consumers purchasing sunglasses (also in the same approximately sixty (60) dollar price range) were found not to use promo codes (see Appendices D and E).

These fine-grained results are valuable and actionable, because they speak directly to the core purpose of the app: identifying consumer spending behaviors that are amenable to influence. For example, by knowing that female consumers looking to purchase sunglasses don’t typically use promo codes, even if such codes exist at the time of sale, the company’s marketing and product development teams are provided with the opportunity to actively present such app users with promo code opportunities they would likely have otherwise overlooked or not made use of, to the detriment of their personal budgeting goals.

V. Recommendations to Company Leadership

A. Positive Expected Return on Investment

Given the preceding findings, this report concludes that the AI solution will provide significant returns on investment and should be explored further, for the following reasons:

1. Faster Iterative Development

This investigation demonstrates that an AI solution is valuable for additional refinement of the budget recommendations the app will present to its users, as well as for identifying fruitful areas for additional app feature development – after the data was ingested into WEKA, machine learning analysis identified several highly actionable consumer behavior insights in a matter of seconds (see Appendices B, C, D, and E).

2. Elimination of Human Error

The results of this investigation are valuable to the company, because they demonstrate that the AI solution reduces the risk of human error (Müller and Schmidt, 2024), especially for production scale data sets expected to contain millions of data points, such as is expected from the app once it has been released to the public and develops an established population of regular users.

3. Automated Discovery of Future Actionable Trends

Finally, because the AI solution will be automated in an iterative manner, it will be able to discover new trends in user data as those trends emerge from the data itself (Müller and Schmidt, 2024), without the need for costly or lengthy analytical manual analysis by human employees. While it is outside the scope of this report to speak to specific dollar amounts, the total savings is expected to be substantial, and far above the cost to further develop this line of research.

B. Non-Trivial Potential Risks

Several non-trivial potential risks exist, however, in incorporating the methodology and results of this investigation into future app feature development:

1. Cost and Development Time Overruns

As with any software development project, initial estimates of the time, cost, and labor required to meet development milestones may be incorrect (Abbas, 2024; Jørgensen, 2024) and should be accounted for ahead of time by company leadership.

2. Unproductive Data Collection

A non-trivial risk exists that even after collecting a substantial quantity of data from app users, that data – despite best effort to collect only relevant and usable data – may not contain any actionable insights or result in material that company leadership can use to make productive app development decisions (Burkov, 2019; Gade, 2024).

3. Misleading Interpretations of Data

As with any new development endeavor, there is a non-trivial risk that the AI solution will generate incomplete, erroneous, or reasonable sounding, but misleading, interpretations of user data that steer future app development in a non-productive direction (Burkov, 2019; Hewamalage, Ackermann, and Bergmeir, 2022; Montgomery, Peck and Vining, 2021).

4. Regulatory Compliance and Legal Exposure

While it is beyond the scope of this report to discuss fine-grained details, since the app will involve the analysis of consumer data subject to a variety of consumer privacy protections across a broad range of domestic and international legal jurisdictions, company leadership is advised to take due care that any use of such data stays compliant with the laws and regulatory schemes that apply to it and is only used in a legally and ethically-sound manner (Adanyin, 2024).

VI. Conclusion

This report recommends that executive leadership incorporate AI-derived features in the company's flagship app, provided that the development and product performance risks identified above are planned for accordingly with appropriate risk mitigation measures.

References

Abbas, L. (2024). *AI-Powered Cloud Storage: A Complete Review of Efficiency Improvements*. [online] Mzresearch.com. Available at: https://mzresearch.com/index.php/MZJAI/article/view/171 [Accessed 30 Mar. 2025].

Adanyin, A. (2024). *Ethical AI in Retail: Consumer Privacy and Fairness*. [online] arXiv.org. Available at: https://arxiv.org/abs/2410.15369.

Amalsyah, M., Kurniawan, D., Rifai, A. and Sari, P. (2025). Sentiment Analysis of Fintech Application User Reviews using the CRISP-DM Framework for Product Development Prioritization. *Sistemasi: Jurnal Sistem Informasi*, *14*(2), pp. 813-825.

Aslam, S. (2025). *Consumer Behavior and Shopping Habits Dataset: E-Commerce Transaction Trends*. [online] Available at: https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset.

Azeroual, O., Nacheva, R., Nikiforova, A. and Störl, U. (2025). A CRISP-DM and Predictive Analytics Framework for Enhanced Decision-Making in Research Information Management Systems. *Informatica*, [online] 49(18). doi:https://doi.org/10.31449/inf.v49i18.5613.

Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. Andriy Burkov.
Cao, S., Jiang, W., Lei, L. and Zhou, Q. (2024). Applied AI for finance and accounting: Alternative data and opportunities. *Pacific-Basin Finance Journal*, [online] 84(1), p.102307. doi:https://doi.org/10.1016/j.pacfin.2024.102307.

Foulds, J., Witten, I., Frank, E., Hall, M. and Pal, C. (2025). *Data Mining: Practical machine learning tools and techniques*. Elsevier.

Frank, E., Hall, M. and Witten, I. (2016). *The WEKA Workbench*. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition.

Gade, K. (2024). *Data Quality in the Age of Cloud Migration: Challenges and Best Practices*. [online] Mzresearch.com. Available at: https://mzresearch.com/index.php/MZJAI/article/view/417 [Accessed 30 Mar. 2025].

Hewamalage, H., Ackermann, K. and Bergmeir, C. (2022). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*. doi:https://doi.org/10.1007/s10618-022-00894-5.

Jain, P., Gyanchandani, M. and Khare, N. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data*, [online] 3(1). doi:https://doi.org/10.1186/s40537-016-0059-y.

Jørgensen, M. (2024). A Systematic Literature Review on Characteristics of the Front-end Phase of Agile Software Development Projects and Their Connections to Project Success. *Journal of Systems and Software*, [online] p.112155. doi:https://doi.org/10.1016/j.jss.2024.112155.

Joshi, S., 2025. Review of Gen AI models for financial risk management. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 11(1), pp.709-723.

Li, Y., Liu, P., Wang, H., Chu, J. and Wong, W.E. (2024). Evaluating Large Language Models for Software Testing. *Computer Standards & Interfaces*, 93, pp.103942–103942. doi:https://doi.org/10.1016/j.csi.2024.103942.

Li, Z. (2025). Customer Segmentation and Churn Prediction Based on K-Means and Random Forest: A Case Study of E-commerce Data. *Eurasia Journal of Science and Technology*, 7(2). doi:https://doi.org/10.61784/ejst3071.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, [online] 54(6), pp.1–35. doi:https://doi.org/10.1145/3457607.

Montgomery, D., Peck, E. and Vining, G. (2021). *Introduction to Linear Regression Analysis*. [online] Google Books. John Wiley & Sons. Available at: https://books.google.com/books?hl=en&lr=&id=tCIgEAAAQBAJ&oi=fnd&pg=PR13&dq=Montgomery.

Rane, N., Choudhary, S. and Rane, J. (2024). Artificial Intelligence and Machine Learning in Business Intelligence, Finance, and E-commerce: a Review. *Social Science Research Network*. [online] doi:https://doi.org/10.2139/ssrn.4843988.

Usman-Hamza, F., Balogun, A., Mamman, H., Capretz, L., Basri, S., Oyekunle, R., Mojeed, H. and Akintola A. (2025). Empirical Analysis of Data Sampling-Based Decision Forest Classifiers for Software Defect Prediction. *Software*, [online] 4(2), pp.7–7. doi:https://doi.org/10.3390/software4020007.

Ze Shi Li, Nowshin Nawar Arony, Kezia Devathasan, Manish Sihag, Ernst, N. and Damian, D. (2024). *Unveiling the Life Cycle of User Feedback: Best Practices from Software Practitioners*. doi:https://doi.org/10.1145/3597503.3623309.

Appendix A

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Customer ID	Age	Gender	Item Purchased	Category	Purchase An	Location	Size	Color	Season	Review Rating	Subscription Status	Payment Method	Shipping Type	Discount Applied	Promo Code	Previous Purchases	Preferred Payment M	Frequency of Purchases
2	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Credit Card	Express	Yes	Yes	14	Venmo	Fortnightly
3	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Bank Transfer	Express	Yes	Yes	2	Cash	Fortnightly
4	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Cash	Free Shipping	Yes	Yes	23	Credit Card	Weekly
5	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	PayPal	Next Day Air	Yes	Yes	49	PayPal	Weekly
6	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Cash	Free Shipping	Yes	Yes	31	PayPal	Annually
7	6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9	Yes	Venmo	Standard	Yes	Yes	14	Venmo	Weekly
8	7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2	Yes	Debit Card	Free Shipping	Yes	Yes	49	Cash	Quarterly
9	8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal	Winter	3.2	Yes	Debit Card	Free Shipping	Yes	Yes	19	Credit Card	Weekly
10	9	26	Male	Coat	Outerwear	97	West Virginia	L	Silver	Summer	2.6	Yes	Venmo	Express	Yes	Yes	8	Venmo	Annually
11	10	57	Male	Handbag	Accessories	31	Missouri	M	Pink	Spring	4.8	Yes	PayPal	2-Day Shipping	Yes	Yes	4	Cash	Quarterly
12	11	53	Male	Shoes	Footwear	34	Arkansas	L	Purple	Fall	4.1	Yes	Credit Card	Store Pickup	Yes	Yes	26	Bank Transfer	Bi-Weekly
13	12	30	Male	Shorts	Clothing	68	Hawaii	S	Olive	Winter	4.9	Yes	PayPal	Store Pickup	Yes	Yes	10	Bank Transfer	Fortnightly
14	13	61	Male	Coat	Outerwear	72	Delaware	M	Gold	Winter	4.5	Yes	PayPal	Express	Yes	Yes	37	Venmo	Fortnightly
15	14	65	Male	Dress	Clothing	51	New Hampshire	M	Violet	Spring	4.7	Yes	Debit Card	Express	Yes	Yes	31	PayPal	Weekly
16	15	64	Male	Coat	Outerwear	53	New York	L	Teal	Winter	4.7	Yes	PayPal	Free Shipping	Yes	Yes	34	Debit Card	Weekly
17	16	64	Male	Skirt	Clothing	81	Rhode Island	M	Teal	Winter	2.8	Yes	Credit Card	Store Pickup	Yes	Yes	8	PayPal	Monthly
18	17	25	Male	Sunglasses	Accessories	36	Alabama	S	Gray	Spring	4.1	Yes	Venmo	Next Day Air	Yes	Yes	44	Debit Card	Bi-Weekly
19	18	53	Male	Dress	Clothing	38	Mississippi	XL	Lavender	Winter	4.7	Yes	Debit Card	2-Day Shipping	Yes	Yes	36	Venmo	Quarterly
20	19	52	Male	Sweater	Clothing	48	Montana	S	Black	Summer	4.6	Yes	Bank Transfer	Free Shipping	Yes	Yes	17	Cash	Weekly
21	20	66	Male	Pants	Clothing	90	Rhode Island	M	Green	Summer	3.3	Yes	Venmo	Standard	Yes	Yes	46	Debit Card	Bi-Weekly
22	21	21	Male	Pants	Clothing	51	Louisiana	M	Black	Winter	2.8	Yes	Credit Card	Express	Yes	Yes	50	Cash	Every 3 Months
23	22	31	Male	Pants	Clothing	62	North Carolina	M	Charcoal	Winter	4.1	Yes	Credit Card	Store Pickup	Yes	Yes	22	Debit Card	Quarterly
24	23	56	Male	Pants	Clothing	37	California	M	Peach	Summer	3.2	Yes	Cash	Store Pickup	Yes	Yes	32	Debit Card	Annually
25	24	31	Male	Pants	Clothing	88	Oklahoma	XL	White	Winter	4.4	Yes	Credit Card	Express	Yes	Yes	40	Credit Card	Weekly

Figure 1: A sample extracted from the "Consumer Behavior and Shopping Habits Dataset: E-Commerce Transaction Trends" dataset available via Kaggle.com (Aslam, 2025).

Appendix B

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        2666               68.359  %
Incorrectly Classified Instances      1234               31.641  %
Kappa statistic                          0.3553
Mean absolute error                      0.3157
Root mean squared error                  0.56
Relative absolute error                 64.3918 %
Root relative squared error            113.1213 %
Total Number of Instances             3900

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.637    0.281    0.631      0.637    0.634      0.355    0.681     0.560     Yes
                 0.719    0.363    0.724      0.719    0.721      0.355    0.681     0.684     No
Weighted Avg.    0.684    0.328    0.684      0.684    0.684      0.355    0.681     0.631

=== Confusion Matrix ===

    a    b   <-- classified as
 1068  609 |    a = Yes
  625 1598 |    b = No
```

Figure 2: Summary of results of random tree analysis using the WEKA software platform.

Appendix C

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        2741               70.2821 %
Incorrectly Classified Instances      1159               29.7179 %
Kappa statistic                          0.4004
Mean absolute error                      0.3215
Root mean squared error                  0.4376
Relative absolute error                 65.5894 %
Root relative squared error             88.3942 %
Total Number of Instances             3900

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.699    0.295    0.642      0.699    0.669      0.402    0.789     0.648     Yes
                 0.705    0.301    0.757      0.705    0.730      0.402    0.789     0.875     No
Weighted Avg.    0.703    0.298    0.707      0.703    0.704      0.402    0.789     0.777

=== Confusion Matrix ===

    a    b   <-- classified as
 1173  504 |    a = Yes
  655 1568 |    b = No
```

Figure 3: Summary of results of random forest analysis using the WEKA software platform.

Appendix D

```
kMeans
======

Number of iterations: 5
Within cluster sum of squared errors: 10663.59651159696

Initial starting points (random):

Cluster 0: Female,Sunglasses,Accessories,56,No,33,Monthly
Cluster 1: Male,Boots,Footwear,55,No,47,Weekly

Missing values globally replaced with mean/mode

Final cluster centroids:
                                          Cluster#
Attribute                    Full Data           0           1
                              (3900.0)    (1682.0)    (2218.0)
==============================================================
Gender                            Male      Female        Male
Item Purchased                  Blouse  Sunglasses       Pants
Category                      Clothing Accessories    Clothing
Purchase Amount (USD)          59.7644     60.3341     59.3323
Promo Code Used                     No          No         Yes
Previous Purchases             25.3515      23.151     27.0203
Frequency of Purchases  Every 3 Months     Monthly      Weekly
```

Figure 4: First summary of results of K-means clustering using the WEKA software platform.

Appendix E

```
kMeans
======

Number of iterations: 7
Within cluster sum of squared errors: 8479.88920704757

Initial starting points (random):

Cluster 0: Female,Coat,Outerwear,43,No,39,Bi-Weekly
Cluster 1: Female,Socks,Clothing,75,No,27,'Every 3 Months'

Missing values globally replaced with mean/mode

Final cluster centroids:
                                          Cluster#
Attribute                    Full Data           0              1
                              (2925.0)    (1320.0)       (1605.0)
==================================================================
Gender                            Male        Male           Male
Item Purchased                 Jewelry     Jewelry          Dress
Category                      Clothing Accessories       Clothing
Purchase Amount (USD)          59.9631     53.7652        65.0604
Promo Code Used                     No          No             No
Previous Purchases             25.7002     26.2598        25.2399
Frequency of Purchases  Every 3 Months   Bi-Weekly Every 3 Months
```

Figure 5: Second summary of results of K-means clustering using the WEKA software platform.