

**Matthew Bullen**

**MSc Artificial Intelligence**

**Module: UAI\_PCOM7E January 2025 A**

**Unit 9: Individual Essay – Artificial Intelligence and Its Applications**

## **Individual Essay**

### **I. Executive Summary**

Budget Buddy, Inc. (hereinafter, “the company”) is presented with a unique opportunity to incorporate an AI-based data analysis solution into its flagship consumer smartphone software product, albeit one that carries a non-trivial degree of practical software development and product performance risks.

### **II. Industry Landscape Positioning**

The company’s mission is to develop a consumer smartphone app that allows users to manage their personal and household spending to achieve their financial planning goals. Prior market research by the company has confirmed that a strong consumer demand exists for such an app (Rane, Choudhary and Rane, 2024). The company has the competitive advantage of acting as a first mover, as it would be one of the first publicly prominent companies of its type to undertake building such an app using a machine learning-based (hereinafter, “AI”) solution.

### **III. Solution to Be Implemented**

The company’s flagship product will be a smartphone app designed to assist users in tracking and managing their household spending budgets. The company will use an AI solution as a core component of its data management and database platforms to analyze user-supplied data about their spending habits in order to automatically generate budget recommendations aligned with their spending goals.

The use of AI-based software solutions in the financial sector is not completely novel; such solutions have been successfully implemented for subject matter domains as varied as analyzing yearly regulatory filings, identifying investment opportunities, and fraud detection (Cao et al, 2024), indicating that the app’s use of AI faces no unreasonable pre-existing technological barriers to success, despite its specific scope being unique to the market.

### **IV. The Scope of the Solution**

Should the company executive management agree to incorporate an AI solution into the app, the scope and boundaries of such are straightforward:

#### **A. Datasets to Be Used**

The majority of raw data will be supplied by app users themselves, in the form of user-generated notes and logs of what products and services they purchased each month added directly into their app accounts, either by manual entry or by using the app’s camera to scan paper receipts. In addition, users will be provided

the option within their app accounts to voluntarily import retailer account and bank account transaction statements into the app. The AI solution will also draw upon publicly available and anonymized macroeconomic datasets about consumer spending patterns, localized by country, region or state, and major metropolitan areas, as relevant.

In accordance with long-established industry best practices, all user-sourced data will be processed into a secure set of cloud-hosted databases and/or e-document warehouses (Gade, 2024; Jain, Gyanchandani and Khare, 2016).

## **B. Algorithmic Approaches to Be Used**

The AI solution will employ several algorithmic approaches to processing user data. As a first line approach, both supervised and unsupervised linear and logistic regression methods will be applied in a series of first passes of the initial set of user data. These methods will be used because they are well-known, straightforward to implement in the technical sense, and amenable to quickly capturing actionable insights from large data sets (Burkov, 2019; Hewamalage, Ackermann, and Bergmeir, 2022; Montgomery, Peck and Vining, 2021).

For example, linear regression will be applied to specific categories of consumer spending to predict recurring trends in individual user spending and budget savings preferences aligned by product or service category.

Logistic regression will be applied largely for data cleaning or categorization (Burkov, 2019); for example, by sorting purchases as “recurring” or “one time only”, which results would then be used in support of the linear regression models as applied to specific spending behavior.

After these initial methods have captured the most readily available insights they are designed to capture, more complex approaches may be applied, such as a random forest approach to apply decision trees across subsets of users or across subsets of spending categories (Burkov, 2019). These more complex approaches are expected to be more costly to implement in terms of developer time and cloud platform compute costs (Mehrabi et al, 2021); as such, they will only be used after a reasonable expectation of their utility has been confirmed.

## **C. Staffing and Budgetary Requirements**

Current staffing levels are assessed to be adequate to complete the project as scoped, although a portion of the company budget should be reserved for a discretionary hiring fund in the case additional staff, including third party expert domain consultants, are needed over the course of development (Ze et al, 2024).

A significant portion of the annual budget, currently estimated to be at least twenty-five percent (25%), will be used to lease state of the art cloud-based virtual machines suitable for large scale data processing – leasing is preferable to purchasing computing hardware, as infrastructure maintenance costs and platform risk management are thereby outsourced to expert providers (Abbas, 2024).

## **D. Expected Development Timeline**

The internal development time to completion for a usable proof of concept version of the app will be eighteen (18) months. In alignment with software development best practices, the initial development period will be followed by

approximately six (6) months of iterative user acceptance testing and functionality refinements (Ze et al, 2024). The app will launch to the consumer public shortly after that approximately twenty-four (24) month span, once acceptance review by executive management has been completed.

## **V. Expected Return on Investment**

Three potential advantages to an AI solution justify its development cost and confirm its expected returns to the company:

### **A. More Efficient Iterative Development**

One of the primary advantages of an AI solution will be that it will allow iterative and automated refinement of the budget recommendations the app will present to end users. Instead of human analysts engaging in time-consuming manual work to define and generate financial advice (Müller and Schmidt, 2024), such advice will emerge as the end state of AI analysis of user data.

### **B. Greater Accuracy and Elimination of Human Error**

Another advantages of an AI solution will be that it mediates the potential for human error when analyzing data sets expected to contain several hundreds of millions of data points (Müller and Schmidt, 2024). The size and scope of the data collected makes it infeasible, if not impossible, for human analysts to process such a large quantity of data in an accurate, or even meaningfully actionable, manner (Müller and Schmidt, 2024), making an AI-driven, automated process far more efficient given the company's expected rapid development cycle to complete the app.

### **C. Automated and Adaptive Data Management**

Finally, because the AI solution will be automated in an iterative manner, it will be able to discover and adapt to 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 work by human employees. In other words, the AI solution will automate the process of ensuring that the app stays current and relevant to its users.

## **VI. Potential Risks and Risk Remediation**

A number of pitfalls may be encountered while implementing the AI solution:

### **A. Cost and Development Time Overruns**

As with any greenfield software development project, initial estimates of the time, cost, and labor required to meet development milestones may be incorrect; the project may take longer to complete or cost more than budgeted (Jørgensen, 2024). In particular, the subject matter domain expertise of current staff may prove to have gaps that must be remediated by drawing on the expertise of outside consultants (Abbas, 2024).

### **B. Inability to Source Necessary Data**

As with any data-driven endeavor, there exists a non-trivial risk that even after collecting a substantial quantity of user data, that data may not be broad or relevant enough to allow actionable insights to be drawn from it (Burkov, 2019; Gade, 2024).

### **C. Misleading Interpretations of Available Data**

The plain act of using an AI solution by itself is not a guarantee of the accuracy or relevance of its results. There is a non-trivial risk that despite best efforts, the AI solution provides incomplete, erroneous, or reasonable sounding, but misleading, interpretations of user data.

For example, using linear regression models runs the risk of overfitting data, or making invalid correlative predictions between input and output values (Burkov, 2019; Hewamalage, Ackermann, and Bergmeir, 2022; Montgomery, Peck and Vining, 2021). For the company's purposes, that could result in a model that misidentifies user spending trends (e.g., model output values) based on past purchases (e.g., model input values). Logistic regression models can be prone to categorizing data points based on irrelevant data features (Hewamalage, Ackermann, and Bergmeir, 2022); again, for app purposes, that could mean creating a model that fails to reliably distinguish between relevant versus irrelevant data points, rendering it unusable.

### **D. Regulatory Compliance and Legal Exposure**

Since the app will involve analysis of consumer data subject to a variety of consumer privacy protections across a spectrum of legal jurisdictions, due care must be observed that any application of AI to such data is mindful of, and compliant with, any laws and regulatory schemas that govern its use (Adanyin, 2024).

## **VII. Conclusion**

This report recommends incorporation of an AI-based data analysis solution into 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>.
- 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>.
- 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>.
- 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=tClqEAAQBAJ&oi=fnd&pg=PR13&dq=Montgomery>.
- Müller, J. and Schmidt, A. (2024). *Robotic Process Automation (RPA) in Business: Optimizing Efficiency and Reducing Human Error*. [online] Available at: <https://mzresearch.com/index.php/JET/article/view/294> [Accessed 30 Mar. 2025].
- 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>.

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>.