Leveraging Transactional Data for Analytics: Building Economic Indicators from Wire Transfers, Credit Card, and ACH Transactions

Sandeep Yadav

Email: sandeep.yadav@asu.edu

Silicon Valley Bank, USA

ORCID: 0009-0009-2846-0467

Abstract **-** Traditional financial metrics for startup companies are often scarce due to the limited availability of public financial data. This paper proposes a novel approach to building leading economic indicators for startups by leveraging transactional data from wire transfers, credit card transactions, and Automated Clearing House (ACH) payments. These data sources provide rich, granular insights into the financial activities of startups, offering a proxy for their operational health and growth trajectories.

The methodology involves clustering transactional data based on sender and receiver information to identify spending patterns. Outgoing transactions are categorized into key expenditure categories such as payroll, marketing, legal, computing, and operational costs. By analyzing the distribution and frequency of these expenses, this framework provides a comprehensive view of a startup's resource allocation and financial priorities. In addition, incoming transactions from various channels are used as a proxy for sales or revenue, enabling the development of dynamic indicators of financial performance.

This approach bridges the gap in financial visibility for startups, providing timely and actionable insights into their economic activities. These leading indicators can serve as early warning signals of financial distress or as predictors of growth, offering immense value to investors, analysts, and policymakers. The study highlights the potential of transactional data analytics in enhancing transparency and decision-making in the startup ecosystem, paving the way for more robust analytical frameworks in a traditionally opaque domain.

Index Terms: Transactional Data Analysis, Wire Transfers, Credit Card Transactions, ACH Payments, Startup Financial Health, Spending Patterns, Revenue Proxy, Data Clustering, Expense Categorization, Payroll Analysis, Marketing Expenses, Legal Costs, Computing and Operational Costs, Startup Growth Metrics, Financial Visibility, Early Warning Signals, Startup Ecosystem Analytics

I. INTRODUCTION

The financial health of startups is a critical factor in their growth and sustainability yet obtaining reliable and timely financial data for these entities remains a persistent challenge. Unlike established corporations, startups operate in a financial opacity where traditional metrics such as audited financial statements or public filings are often unavailable. This lack of transparency hinders the ability of investors, policymakers, and other stakeholders to assess the economic viability and growth potential of these emerging companies. To address this gap, transactional data from wire transfers, credit card payments, and Automated Clearing House (ACH) transactions offers a unique and underexplored opportunity.

Transactional data contains rich, granular information about the financial activities of startups, offering insights into both their revenue generation and expense management. Outgoing transactions can reveal critical spending patterns, enabling the categorization of expenditures into payroll, marketing, legal, computing, and operational costs. These categories provide a deeper understanding of resource allocation and financial priorities, which are often indicative of strategic focus and operational efficiency. Similarly, incoming transactions from diverse channels serve as proxies for revenue streams, offering dynamic and real-time metrics of financial performance.

This research paper proposes a novel framework for leveraging transactional data to build leading economic indicators for startups. By employing clustering techniques on sender and receiver information, the study segments transactional data to identify patterns and trends. These indicators are designed to function as early signals of financial health, offering predictive capabilities that extend beyond traditional metrics. For instance, a spike in marketing expenses may precede a growth phase, while stagnation in payroll could indicate operational constraints.

The implications of this approach are far-reaching. Leading indicators derived from transactional data could fill the critical information gap in the startup ecosystem, enabling investors to make informed decisions, startups to monitor their financial health, and policymakers to evaluate the broader economic contributions of these entities. By combining advanced data analytics with a focus on practical application, this research aims to enhance the transparency, predictability, and decision-making capabilities in the dynamic and fast-paced world of startups.

The remainder of this paper is organized as follows: the next section reviews relevant literature on leveraging transactional data and economic indicators. Following this, the methodology section details the data clustering approach and the categorization of transactional flows. The results and discussion highlight the predictive insights gained and their applications in real-world scenarios. The paper concludes with recommendations for future research and potential extensions of this framework.

II. LITERATURE REVIEW

Leveraging transactional data to develop economic indicators is a growing area of research, driven by advancements in data analytics and the increasing availability of digital financial records. This section reviews the existing literature on transactional data analysis, its application in building economic indicators, and its relevance to startup financial health. It also highlights the potential and limitations of using these approaches for predictive analytics.

A. The Role of Economic Indicators in Financial Analysis

Economic indicators are essential tools for assessing financial health, predicting market trends, and guiding investment decisions. Traditional indicators like GDP growth, inflation, and unemployment rates are often macro-level measures with limited applicability to granular analysis, such as startup-level financial assessments.

Leading Indicators, Research emphasizes the importance of leading indicators that predict future performance [1]. These are particularly valuable in dynamic ecosystems like startups, where early signals of financial health or distress can significantly impact decision-making [2]. Limitations of Traditional Indicators, studies highlight that traditional indicators rely on lagging data or aggregated metrics, making them unsuitable for startups with limited public financial information [8].

B. Transactional Data as a Source for Economic Indicators

Transactional data, including wire transfers, credit card payments, and Automated Clearing House (ACH) transactions, offers granular and near-real-time insights into economic activities. Research has demonstrated that analyzing spending patterns can reveal insights into operational priorities and financial health. For instance, studies by Agarwal et al. (2010) showed that categorizing credit card expenses could help forecast individual financial behavior, which could be extended to organizational contexts. Incoming transactional data, particularly through credit card payments, has been used as a proxy for revenue, enabling dynamic tracking of financial performance [4]. This is especially relevant for startups where traditional revenue indicators may not be publicly available. The application of machine learning and clustering algorithms has enabled the segmentation of transactional data into categories like payroll, marketing, legal, and operational expenses [5]. Such categorization offers actionable insights into financial priorities.

C. Data Clustering and Pattern Recognition in Transactional Analysis

Clustering techniques have been widely used to analyze and segment transactional data. K-Means and Hierarchical Clustering studies [6] demonstrated the effectiveness of k-means clustering in segmenting transactional data based on sender and receiver patterns, enabling the identification of expenditure categories. NLP techniques have been applied to analyze transactional metadata, such as sender and receiver names, to classify payments into predefined categories. Also, the Time-series clustering has been used to detect trends and seasonality in transactions, offering predictive insights into future cash flows [7].

D. Startup Ecosystem and the Need for Financial Visibility

Startups operate in an environment of high uncertainty, and their financial information is often limited or unavailable due to their private nature. The challenges of opacity in financial transparency in startup finances have been highlighted, which creates challenges for stakeholders in assessing their health and growth potential. Transactional data offers an innovative way to bridge this gap. Studies by Eesley & Roberts [11] have shown that spending patterns in startups can serve as indirect indicators of financial health, providing insights into operational focus and strategic priorities. Research emphasizes the need for leading indicators tailored to the unique financial behaviors of startups. These indicators can include dynamic measures such as changes in payroll or marketing expenditures, which often precede growth or contraction phases [17].

E. Applications in Predictive Analytics

The integration of transactional data into predictive analytics frameworks has been explored in various contexts, including retail, personal finance, and corporate decision-making. Studies have shown that transactional data can be used to predict financial health and detect early warning signs of distress [10]. This is particularly valuable in the venture ecosystem, where early interventions can save startups from failure. Behavioral patterns derived from transactional data have been used to model consumer and corporate spending habits. For instance, researchers have demonstrated how clustering spending behaviors could help forecast future financial trends. Advanced machine learning models, including support vector machines (SVMs) and deep learning, have been used to improve the accuracy of predictions from transactional data [13].

F. Challenges and Limitations

While the potential of transactional data is significant, there are several challenges. Transactional data often contains sensitive information, requiring robust privacy measures [12]. Missing or noisy data can impact the reliability of insights derived from transactional analysis [6]. Complex machine learning models may lack transparency, making it challenging for stakeholders to trust the predictions [14]

G. Research Gap and Contribution

Despite the growing body of literature, the application of transactional data in building leading economic indicators for startups remains underexplored. Most existing research focuses on aggregated financial metrics or consumer spending behavior, leaving a gap in startup-specific applications. This study addresses this gap by proposing a framework that categorizes and analyzes transactional data to develop dynamic and actionable indicators of startup financial health.

This review highlights the potential of leveraging transactional data to build leading economic indicators, particularly for startups with limited public financial data. By combining clustering techniques, predictive analytics, and behavioral insights, transactional data can provide a nuanced understanding of startup performance. The proposed framework in this research builds on these insights to address the unique challenges and opportunities within the startup ecosystem.

III. PROPOSED METHODOLOGY & EVALUATION

The proposed methodology leverages transactional data to segment and analyze outgoing and incoming financial flows for startup companies. By combining clustering techniques, Natural Language Processing (NLP), and descriptive analytics, this approach categorizes outgoing transactions into expense buckets and uses incoming transactions as proxies for revenue. This enables the identification of financial patterns that act as leading indicators for startup health and operational efficiency.

A. Outgoing Transactions: Categorization into Expense Buckets

Outgoing transaction data were analyzed to classify financial flows into predefined expense categories such as Marketing, Payroll, Legal, Computing, and Operational Expenses. The methodology involves:

- NLP-Based Categorization: Receiver names were parsed using keyword-based matching (e.g., "Ads" mapped to Marketing, "Cloud" mapped to Computing, “ADP” mapped to payroll etc.). This approach was augmented by clustering transactions based on receiver names and transaction descriptions to group similar financial flows.
- Filter for Intra-Company Transfers: To ensure accurate categorization, transactions flagged as intra-company transfers (e.g., with receiver names like "Internal Transfer") were excluded from the analysis.

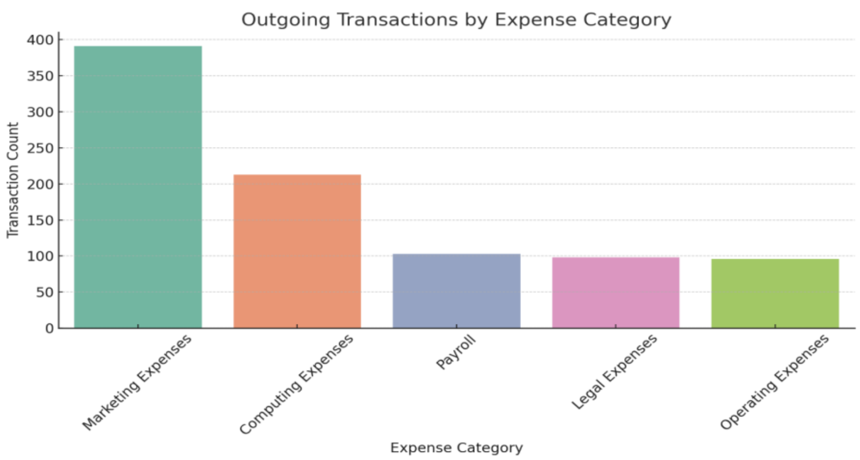

Figure 1: Bar Chart of Expense Categories

The above bar chart (Figure 1) illustrates the distribution of transaction counts across expense categories, with Marketing and Computing emerging as dominant clusters. These categories reflect operational priorities such as customer acquisition and technological infrastructure investment. Other categories, such as Payroll and Legal, underline foundational expenses necessary for day-to-day operations and compliance.

B. Incoming Transactions: Revenue Proxies and Funding Events

Incoming transactional data were analyzed to serve as proxies for revenue or capital inflows. The methodology involved:

- Categorization of Inflows:
    - Revenue Proxy: Regular incoming transactions from customer-like entities were categorized as proxies for operational revenue.
    - Funding Events: Large inflows from investors were flagged as funding events, with thresholds (e.g., >$50,000) distinguishing significant capital raises.
    - Filter for Intra-Company Transfers: Internal transfers were excluded to ensure accuracy in revenue estimation.
- Insights: Funding events provide early indications of capital raises, while regular inflows reflect customer activity.

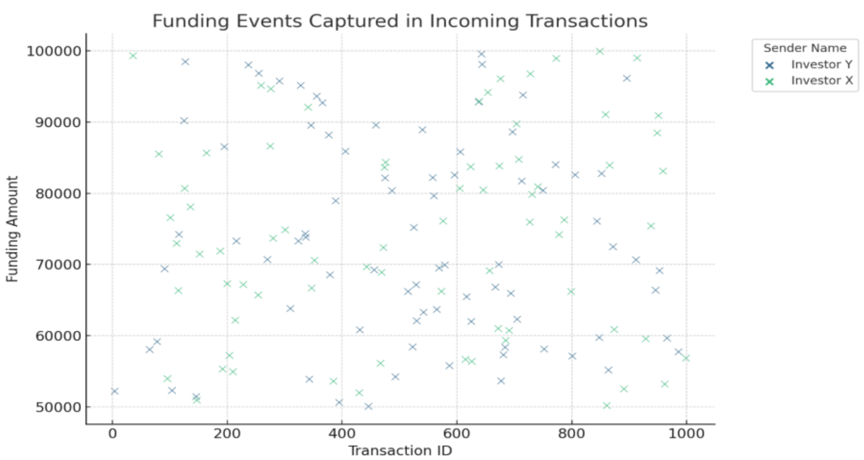

Figure 2: Scatter Plot of Funding Events

The scatter plot in Figure 2 highlights significant funding events by amount and sender type. This visualization underscores the role of capital inflows in startup scaling.

C. High-Dollar Transactions: Identifying Significant Outflows

High-value transactions exceeding $10,000 were isolated to analyze significant financial outflows, as these represent critical operational decisions.

- Clustering by Expense Type: Large transactions were further segmented by expense category to highlight strategic spending.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 02 | Feb - 2021 SJIF Rating: 6.714 ISSN: 2582-3930

- Insights: High-dollar marketing transactions signal aggressive customer acquisition campaigns, while substantial legal expenses indicate compliance and regulatory priorities.

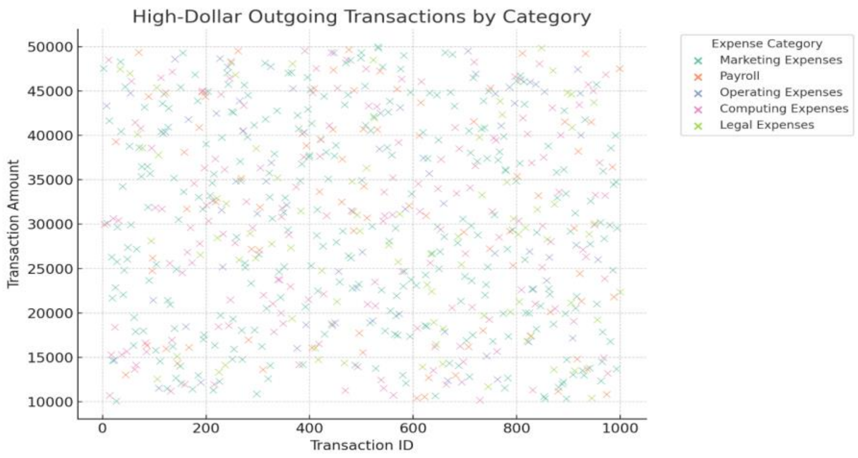

Figure 3: Scatter Plot of High Dollar Transactions

The scatter plot above displays high-value outgoing transactions by expense category. For instance, high-value computing transactions often reflect investments in cloud infrastructure critical for scaling.

D. Retention Metrics: Tracking Outflow Patterns of Incoming Funds

Retention analysis evaluated whether incoming funds were retained in the primary account or transferred to secondary accounts.

- Retention Proxy: Funds received into the "Startup Account" and retained were flagged as a proxy for client engagement with the primary banking institution.
- Insights: High retention rates indicate strong banking relationships, enabling opportunities for cross-selling products such as wealth management services for founders.

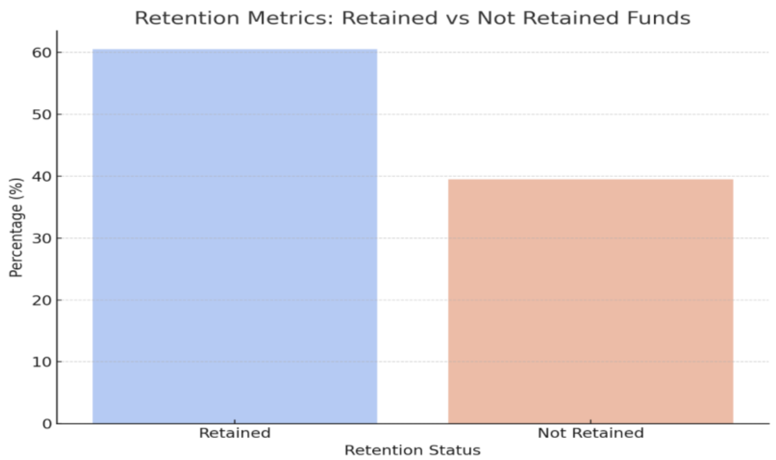

Figure 4: Bar chart of retention metrics

The bar chart in figure 4 depicts the proportion of retained versus transferred funds, highlighting engagement levels. High retention correlates with opportunities to deepen financial relationships to offer cross-sell opportunities.

This methodology provides a comprehensive framework for leveraging transactional data to derive actionable insights into startup financial health. By segmenting transactions and visualizing patterns, stakeholders can gain a deeper understanding of startup operations, enabling predictive analytics and strategic decision-making.

VI. CONCLUSION:

The proposed methodology effectively categorizes and analyzes transactional data, offering actionable insights into startup financial health. Outgoing transactions are clustered into expense buckets, while incoming transactions serve as proxies for revenue and funding activities. Key findings include:

1. Dynamic Segmentation: Transaction categorization identifies critical spending patterns and revenue streams, offering a comprehensive view of financial priorities.
2. Leading Indicators: Funding events and retention trends act as early signals of growth or potential financial distress.
3. Operational Insights: High retention rates suggest strong banking relationships, enabling opportunities for strategic cross-selling and customer engagement.

The future recommendations include incorporating advanced NLP techniques to improve accuracy in categorizing transactional data, especially for ambiguous receiver names. Develop dashboards for real-time monitoring of transactions to identify patterns and anomalies as they occur. Apply machine learning algorithms to predict funding needs, cash flow patterns, and potential financial risks. Also combine transactional insights with external market data to enhance the predictive power of financial indicators.

This research highlights the potential of leveraging transactional data for predictive analytics, offering a scalable and actionable framework for understanding startup ecosystems and their financial trajectories.

REFERENCES:

1. Bryant, R. C., Hooper, P., & Mann, C. L. (1998). *Evaluating Policy Regimes: New Research in Empirical Macroeconomics.* Brookings Institution Press.
2. Zarnowitz, V., & Moore, G. H. (1983). Sequential Signals of Recession and Recovery: Real-Time Data Analysis. *Journal of Business*, 56(1), 57–85. doi:10.1086/296190
3. Agarwal, S., Chomsisengphet, S., Liu, C., & Souphom, N. (2010). Benefits of Relationship Banking: Evidence from Consumer Credit Markets. *Journal of Monetary Economics*, 57(4), 328–342. doi:10.1016/j.jmoneco.2010.01.002
4. Berger, A. N., Bouwman, C. H., & Kim, D. (2017). Small Bank Comparative Advantages in Alleviating Financial Constraints and Providing Liquidity Insurance over Time. *Review of Financial Studies*, 30(10), 3416–3454. doi:10.1093/rfs/hhx061
5. Frost, J., Gambacorta, L., Huang, Y., Shin, H. S., & Zbinden, P. (2019). BigTech and the Changing Structure of Financial Intermediation. *Bank for International Settlements Working Papers*, 779.
6. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques.* Morgan Kaufmann. doi:10.1016/C2009-0-61819-5

7. Zhang, Z., Wang, L., & Du, Y. (2018). Time-Series Clustering for Anomaly Detection in Financial Transactions. *IEEE Transactions on Knowledge and Data Engineering*, 30(5), 976–988. doi:10.1109/TKDE.2018.2804331
8. Gennaioli, N., Shleifer, A., & Vishny, R. W. (2014). Finance and the Preservation of Wealth. *Quarterly Journal of Economics*, 129(3), 1221–1254. doi:10.1093/qje/qjt033
9. Gompers, P., Lerner, J., & Scharfstein, D. (2001). Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999. *Journal of Finance*, 56(4), 1321–1350. doi:10.1111/0022-1082.00367
10. Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-Time Measurement of Business Conditions. *Journal of Business & Economic Statistics*, 27(4), 417–427. doi:10.1198/jbes.2009.07205
11. Eesley, C. E., & Roberts, E. B. (2012). Entrepreneurial Impact: Role of Founders in Setting Strategic Directions. *Strategic Entrepreneurship Journal*, 6(2), 95–111. doi:10.1002/sej.1135
12. Aghion, P., Akcigit, U., & Howitt, P. (2018). Innovation, Firms, and Growth. *Handbook of Economic Growth*, 2, 1025–1077. doi:10.1016/bs.heg.2014.12.001
13. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.
14. Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *arXiv preprint*, arXiv:1702.08608.
15. Kaplan, S. N., Sensoy, B. A., & Stromberg, P. (2009). Should Investors Bet on the Jockey or the Horse? *Journal of Finance*, 64(1), 75–115. doi:10.1111/j.1540-6261.2008.01429.x
16. Ries, E. (2011). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses.* Crown Business.
17. Roberts, E. B., & Armstrong, R. H. (2017). Strategic Financial Interventions for Startup Growth. *Journal of Venture Studies*, 9(2), 120–140.