SECTION: TECHNICAL SCIENCES

DOI 10.70286/isu-10.09.2025.008

IDENTIFICATION OF CONSOLIDATED DATA SOURCES FOR FINANCIAL PROPOSAL CHARACTERISTICS RESEARCH

Azarov Maksym
Master of Computer Science
Lyashkevych Vasyl
Candidate of Technical Sciences, docent
Department of System Design
Ivan Franko National University of Lviv, Ukraine

Recent advances in artificial intelligence create unique opportunities for personalising financial services, so today, the concept of financial information goes far beyond transactional data. Therefore, it is understandable and relevant that describing the characteristics of financial offers is not such an easy task.

Given the diversity of financial data sources, it becomes obvious that we do not know enough about what information is most beneficial for making financial decisions. In this case, the question of the usefulness of data resources for their further consolidation is raised.

In [1], it was shown that leading financial institutions use a combination of traditional internal data, non-traditional internal data (open banking) and external data (telecom data, social networking) to improve the analytical completeness of credit scoring and decision-making, improving the modelling of financial offerings.

The details of how alternative external data complements traditional analytics for financial institutions, especially for NBO (next-best offer), were described in [2]. It was shown that banks are consolidating traditional internal data (CRM, Core banking, credit bureau) with alternative external data (income proxies, behavioural parents, geolocation, IoT, government data) to improve NBO, scoring, marketing, and risk forecasting.

The authors in [3] highlight the architecture of a platform that aggregates external economic indicators and internal banking data, providing real-time streaming processing (market tickers, social media), data validation, quality and so on. They offer a platform that consolidates internal and external sources, with a focus on streaming, quality, and governance. The proposed process follows the best practices in the IT industry, referring to data warehousing.

In [4] was highlighted that only by combining internal and external data is it possible to create a new generation of AI analytics for banks. Also, in [5], some analytical opportunities for doing analytics in meaning space were investigated.

Obviously, the consolidated data sources for studying the characteristics of the financial offer are a complex ecosystem that combines internal data of the bank or financial institution with external market and socio-economic data. They are necessary for the formation of multidimensional models of financial offers such as NBO/NBA, dynamic pricing, loan personalisation, etc. Meanwhile, while working on the problem of data sources consolidation, we should exclude internal bank data sources for some reasons related to confidentiality and security. It is not possible to get that data in an open space. Thus, we can focus on some existing accessible datasets with transactional data (payments, card operations, transfers), product data (deposits, loans, insurance, investments), scoring profiles and others. Regarding other information, we can use synthetic data if it is possible.

We can get more information from external markets and macroeconomic data. Among the data sources we have ones for: macroeconomic indicators (CPI, GDP, unemployment, interest rates, exchange rates), market data (stock prices, bond prices, ETFs, S&P 500, MSCI indices), industry data (level of competition, regulatory changes) and other expert data. The appropriate sources might be: IMF, World Bank, OECD, Eurostat, Refinitiv, Bloomberg, FactSet, etc. [6-8]

Alternative and behavioural data sources provide us with insights into customer sentiment, search query trends, ESG metrics, and other assessments. Using this type of data, we can segment customers, identifying demands for new financial offers, to get more about targeted marketing offers and so on [9]. Also, this data is accessible in different channels, like social media (Twitter/X, LinkedIn, Facebook, Reddit) [10], open banking API (PSD2, UK Open Banking), Google Trends [11], ESG Ratings, News Aggregators and others.

The main requirements [12] for financial data consolidation we would like to highlight as follows:

- Data quality (missing/duplicated, noisy data).
- Regulations (GDPR, PSD2, Basel III, local NBU/ECB requirements, etc.).
- Integration (ETL/ELT processes, format unification).
- Time synchronisation (different data update lags).

The data sources identification allowed us to make a thorough analysis of the accessible data sources for future investigation. That allows us to clarify the data we have for doing experiments and focus on the right methodology of how to correctly consolidate the required financial data, solving our problem. As a result, we understood that the effective research of the financial offering characteristics is valuable using a consolidated, multi-layered data source that leverages transactional and behavioural customer data, macroeconomic factors, market indicators, as well as alternative sources.

References

1. McKinsey & Company. (2021, June 15). Designing next-generation credit-decisioning models. McKinsey & Company. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/designing-next-generation-credit-decisioning-models.

2. PricewaterhouseCoopers (PwC). (2022). Beyond traditional data: Leveraging alternative data in banking. PwC India. https://www.pwc.in/assets/pdfs/beyond-traditional-data-leveraging-alternative-data-banking.pdf.
3. Abdullah, M., Singh, R., & Kumar, P. (2024). A big data-driven consolidated data analysis platform for the financial sector. International Journal of Data Science and Analytics, 17(2), 145–160. https://www.researchgate.net/publication/376958030_A_Big_Data-Driven_Consolidated_Data_Analysis_Platform_for_the_Financial_Sector_Authors.
4. KPMG. (2024). Put your bank's data to work. KPMG LLP. https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2024/put-your-banks-data-to-work-secured.pdf.
5. Lyashkevych L. Exploratory data analysis possibility in the meaning space using large language models / Lyashkevych V, Shuvar R. // Електроніка та інформаційні технології. - 2024. - 1, 25. - 9 (102-116). - DOI: http://dx.doi.org/10.30970/eli.25.9.
6. Anem, S. S. (2024). Predicting financial data with macroeconomic factors using machine learning [Master's thesis, West Chester University]. WCU Digital Commons. https://digitalcommons.wcupa.edu/all_theses/1465/.
7. Patsiarikas, M. (2025). Using machine learning on macroeconomic, technical, and sentiment factors to forecast the S&P 500 index. Information, 16(7), 584. https://doi.org/10.3390/info16070584.
8. Ayala, M. J., Gonzálvez-Gallego, N., & Arteaga-Sánchez, R. (2024). Google search volume index and investor attention in the stock market: A systematic review. Financial Innovation, 10, Article 70. https://doi.org/10.1186/s40854-023-00606-y.
9. Szczygielski, J. J. (2024). Google search trends and stock markets: Sentiment, predictive power, and robustness. [Unpublished manuscript]. SSRN. https://doi.org/10.2139/ssrn.4561234.
10. Elshendy, M., Fronzetti Colladon, A., & Battistoni, E. (2021). Using four different online media sources to forecast the crude oil price. arXiv. https://doi.org/10.48550/arXiv.2105.09154.
11. ResearchGate. (2024). Exploring the relationship between Google Trends and cryptocurrency metrics. [Preprint]. https://www.researchgate.net/publication/380764000_Exploring_the_Relationship_Between_Google_Trends_and_Cryptocurrency_Metrics
12. INFORMS. (2024). The use and usefulness of big data in finance. Management Science. https://doi.org/10.1287/mnsc.2022.02659.