Received: 28 Mar. 2025 | Revised: 05 Apr. 2025 | Accepted: 14 Apr. 2025 | Published online: 25 Apr. 2025

RESEARCH ARTICLE

Contemporary Education and Teaching Research
2025, Vol. 6 (4)122-127
DOI: 10.61360/BoniCETR252018030403

Research on Exploring and Practicing the Improvement of Financial Big Data Literacy Based on Python

Jie Su[1] & Chang Liu[*, 1]
[1]Nanning College of Technology, China

Abstract: Financial big data literacy is the core competitiveness of financial personnel in the digital era, which directly affects the enterprise's data-driven decision-making ability. Based on Python technology, this article proposes a 4- fold strategy for financial big data literacy enhancement and practice, including building a layered training system to strengthen the technical foundation, creating a lightweight tool chain to lower the barrier of application, implementing a practical case base to enhance the scenario-based capability, and establishing a cross-departmental collaboration mechanism to promote the integration of technology and business. This strategy system will effectively solve the problem of technology implementation and promote the financial transformation from "accounting-based" to "value-based". The active implementation of the above strategies will help improve the data analysis efficiency of enterprises, deepen the integration of business and finance, reduce the cost of technology use, and help enterprises build a data-driven financial decision-making system.

Keywords: financial big data literacy, layered training, case study base, cross-departmental collaboration, digitization

1. Introduction

The trend of the digital economy is reshaping the finance function of enterprises, data has become the core asset to drive decision-making, and the financial decision-making efficiency of enterprises that make good use of data analysis can be greatly improved. However, the traditional finance model is facing the transformation challenges of large data volume, low processing efficiency, and low value, Python technology, with its open-source ecosystem, flexibility, and powerful data processing capabilities, has become a critical tool for solving the financial digitization problems. From automated report generation to intelligent risk warning, from financial forecast modeling to unstructured text analysis, Python not only reconstructs the financial workflow, but also promotes the transformation of finance staff to the role of data decision supporter. Therefore, based on Python technology, it is necessary to build a financial personnel literacy system in the era of the digital economy, and through technical empowerment and scenario-based practice, it will promote the strategic leap of financial functions from after-accounting to intelligent decision-making center.

2. The Importance of Financial Big Data in The Era of Digital Economy

After entering the era of the digital economy, data has become a key production factor to promote enterprise value creation, and the importance of financial big data is becoming more and more obvious. The traditional financial model relies on structured reports and lagging indicators, making it difficult to deal with the huge, multi-dimensional, real-time data flood, while financial big data builds a comprehensive decision-making perspective by integrating internal and external information of the enterprise. On the one hand, it can break data "island", integrate and analyze heterogeneous data sources such as ERP systems, market opinion, supply chain trends, etc., so that the process of budget forecasting, cost control and so on can be shifted from empirical judgment to data-driven. On the other hand, its powerful correlation analysis capability can reveal hidden business logic, for example, through the cross-analysis of customer consumption data and

Corresponding Author: Chang Liu
Nanning College of Technology, China
Email: 121240946@qq.com

financial indicators to accurately locate high-value customer groups; or through the correlation of cash flow fluctuations and public opinion hotspots to explore, to warn of potential business risks in advance. In industry application scenarios, financial big data is releasing unprecedented business value. The manufacturing industry integrates Internet of Things (IoT) equipment data and costing systems to achieve dynamic cost monitoring and optimization of production line performance to reduce operating costs; the retail industry builds personalized pricing models based on the correlation analysis of customer consumption behavior and financial data, such as improving inventory turnover through demand forecasting algorithms; and the banking industry builds a credit assessment mechanism based on customer transaction and social network data to improve the accuracy of risk recognition and identification.

These cases break through the traditional financial limits and extend data analysis to the strategic decision-making level. As technology progresses, technology convergence will further develop the boundaries of financial big data. This cross-boundary capability encourages finance personnel to transform into "data strategists" who not only analyze historical data, but also anticipate future trends. The digital transformation of finance and the deepening of financial big data capacity building have become a strategic necessity for enterprises to re-establish their competitiveness and capture the commanding heights of the digital economy.

3. Python-Based Financial Big Data Literacy Dimensions

3.1 Technical capabilities in data collection and processing

Data collection and processing technology capability is an essential support for building analytical foundations. Finance employees need to master the ability to efficiently extract, cleanse and combine information from heterogeneous data sources. This capability system ensures a perfect transition from raw data to analytical models, providing high-quality inputs for credit assessment, cash flow forecasting and other scenarios, and becoming key technical support for the digital transformation of finance (Bi, 2023).

First, multi-source data crawling and API invoking capability. Skillfully used Requests library to crawl webpage data, such as financial news, regulatory announcements, etc., and obtain stock market quotes and macroeconomic indicators in real-time through financial API ports such as Yfinance, Tushare, etc.; master the Scrapy framework to cope with the scaled crawling demand of complex webpage structure, and accurately extract the key indicators in the financial report by combining with the regular expression. According to the case study of a stock brokerage firm, the annual report data of listed companies was automatically crawled by Python script, which reduced the workload that needed to be completed by labor for 3 days to 2 hours.

Second, data cleansing and preprocessing technology. Use Pandas library to achieve standardized operations such as missing value filling (e.g. median interpolation), outlier detection (Z-score or IQR method), data type conversion (e.g. converting text-type numbers to floating-point type), etc.; through functional programming to batch processing of mass millions of data, e.g., a certain group uses Pandas to convert currencies and apply account mapping to the global branch reports, so as to increase the efficiency of combined reports by 70%.

Third, data storage and structured management ability. Master SQLAlchemy to connect to MySQL/PostgreSQL databases, design star model to improve query efficiency; use HDF5 format to store high-frequency transaction data, and accelerate aggregation analysis through Parquet columnar storage. Import ERP system data into Redis in-memory database, realize real-time calculation and visualization monitoring of production cost, and improve decision-making response speed.

3.2 Algorithmic modeling ability for business situation

In Python-based financial big data literacy, the algorithmic modeling ability of business situations is the key engine to achieve the leap in data value. Finance employees need to master the technology of transforming business logic into mathematical models, and establish intelligent decision-making tools through the machine learning ecosystem of Python. This capability set enables finance employees to leapfrog traditional descriptive statistics and enter the prescriptive analysis stage. First, classification and predictive modeling capabilities. Logistic regression, random forest and other algorithms in the Scikit-learn library are used to construct credit-scoring models. The prediction accuracy can be improved by Python modeling; the use of an LSTM neural network to process time series data can reduce the sales prediction error. The key is to understand the evaluation indexes such as AUC-ROC, SHAP value, etc. to ensure that the model interpretability matches the business situation. Second, clustering and customer segmentation capabilities. Use the K-means algorithm to cluster customer relationship management (CRM) data, design differentiated service solutions, and improve customer retention; exploit anomalous transaction patterns through the DBSCAN algorithm to improve the efficiency of risk identification. It is necessary to master the evaluation methods such as the contour

coefficient to avoid the trap of "false clustering". Third, optimization and simulation modeling capabilities. Using the PuLP library to build a cash flow optimization model for supply chain finance to reduce capital costs; assessing the value at risk (VaR) of investment projects based on Monte Carlo simulation to shorten the investment decision-making cycle of enterprises. When modeling, it is necessary to balance computational efficiency and accuracy, for example, to deal with large-scale optimization problems through sparse matrix technology.

3.3 Flexibility in the application of the tool-chain

The flexible application ability of the toolchain is the key pivot to connect the data value chain. Finance employees need to establish a tool thinking, the modular components in the Python ecosystem organic combination of scenario-based solutions, the formation of the collection, processing, modeling, and presentation of the technology chain. The core lies in mastering three technical dimensions. First, cross-library synchronization and process orchestration capabilities. We need to skillfully use Pandas for data cleansing, Matplotlib for visualization, Scikit-learn for model building, and achieve code reuse through function encapsulation and class design to form configurable automation scripts; second, interactive development and deployment capabilities. Master the dynamic debugging skills of Jupyter Notebook, use Streamlit/Flask to package the analytical model as a web application, and realize the real-time connections of analysis, decision, and action; Next, the ability to integrate cloud-native tools. To familiar with AWS Lambda/Azure Functions and other serverless architectures, combined with Docker containership deployment, to build an elastic and scalable financial big data analytics platform to support real-time computation of multi-million level data. The capability emphasizes dynamic configuration of the toolchain according to business scenarios, such as stringing SQL database queries with machine learning models to build an intelligent reconciliation system; or integrating tax data and financial models through API calls to automatically generate compliance reports. Technicians need to understand the underlying logic of the tool and can call ready-made library functions as well as secondary development based on business needs. This toolbox thinking enables the finance team to quickly respond to regulatory changes, market fluctuations and other dynamic needs and transforms technical potential into business value (Hu, 2025).

3.4 Complex thought for cross-domain collaboration

Cross-disciplinary collaboration of composite thinking is the key to releasing the value of technology. Finance employees need to establish a T-shaped ability model, that is, vertical deep-plowing financial expertise, horizontal integration of technical knowledge and business insights, and the formation of three-dimensional decision-making thinking. First, the ability to deconstruct business situations. We need to break down financial goals into quantifiable data indicators, for example, transforming "optimize supply chain costs" into "establish a regression model of logistics costs and inventory turnover", and designing corresponding data collection solutions. Second, the ability to translate technical language. We need to translate algorithmic principles into an understandable decision-making basis for business departments, for example, using a "customer life cycle value prediction model" to promote data results in marketing, risk control and other situations. Third, the catalytic ability of organizational synergy. It is necessary to build a bridge between finance, IT and business departments. For example, in data governance projects, it is necessary to define data standards from the perspective of financial compliance, but also from the perspective of technical feasibility to balance the implementation program. The thought emphasizes breaking through the traditional back-end positioning of finance and taking the lead to the front end of the business chain. By understanding the data generation logic of the whole chain of procurement, production and sales, finance staff can design a more penetrating analysis framework; by anticipating the impact of regulatory policy changes on the financial model, the data capture strategy can be adjusted in advance (Yuan, 2025). This composite thinking is reshaping the role of the finance team, evolving from a data user to a data ecosystem builder, and promoting the transformation of the organization from empirical decision-making to data-driven through technological empowerment.

4. Path to Improve Financial Big Data Literacy

4.1 Build a layered python technical training mechanism

Building a layered Python technical training system is a crucial support for the digital transformation of the finance team, which needs to follow a ladder-style ability training path, layered into basic, application and advanced. The basic layer focuses on the core of Python syntax and data manipulation tool-chain, through immersive programming training to master variable control, function encapsulation and other basic logic, combined with Pandas to achieve mass data cleansing and reshaping, the use of Matplotlib to build a dynamic visualization of the board, laying the technical foundation for complex analysis (Wang et al., 2023). The application layer deepens the practical ability around financial scenarios, develops the ability to transform financial indicators into Python code through cases such as ROI sensitivity analysis

model and multi-cycle cash flow forecasting algorithm, and strengthens the SQL database interaction technology in parallel to realize the seamless connection between ERP system and local data. The high-level breaks through the traditional analysis boundary, introduces Scikit-learn to build a credit scorecard model, applies the LSTM neural network to capture the law of market fluctuation, and combines with NLP technology to parse the financial terms in the text of non-standard contracts, forming the ability of the whole chain from data collection, model training to business deployment. The system is designed to avoid knowledge faults through hierarchical progression, for example, mastering the group by aggregation of Pandas before advancing to the SQL window function, which not only ensures a smooth learning curve for beginners but also reserves technical interfaces for higher-order applications, which helps financial staff grow from script developers to data scientists gradually and realizes the value leap of technological empowerment.

4.2 Creating a lightweight financial data analysis tool-chain

Creating a lightweight financial data analysis tool-chain is the critical breakthrough in financial digital transformation, the core of which is to build a zero-code or low-code analysis environment through open-source tool combinations, and to promote technological benefits (Zhang, 2025). Tool integration layer with Jupyter Notebook for interactive analysis of the hub, support for financial staff through drag and drop controls to complete the data cleansing and visualization of the exploration; combined with the Streamlit framework will be transformed into a visualization of the Python script dashboard, the financial team only through the configuration of the parameters can be generated to dynamically monitor the development of a significant increase in efficiency. The template development layer focuses on standardized scenario packaging, and builds reusable components including financial statement automation scripts and multi-dimensional cost control model libraries, for example, encapsulating the DuPont analysis system into a function module, so that users can generate a complete analysis report inputting basic data. The eco-connection layer emphasizes cross-platform synergy, realizes direct data connection with Power BI/Tableau through PyODBC driver, supports one-key export of analysis results to visual charts; develops RESTful API interfaces to integrate with the ERP system, forming the automation capability of data collection, analysis and modeling, and decision-making push. The tool-chain is organized through Lego-style component assembly, so that non-technical personnel can also quickly build customized tools that meet business logic (Xu & Chen, 2025). For example, a manufacturing company can package a cost prediction model as a web application so that the shop floor can query the production line break-even point in real-time, via mobile. This transformational path of technological democratization is driving the evolution of financial analytics from elitist to public, so that data can truly become a productivity tool for the entire workforce.

4.3 Promote the establishment of a practice-enabled case database

The implementation of combat-driven case database construction is the key to the landing of financial big data literacy, and it is necessary to build a scenario-based, modular, evolutionary case ecosystem. Case design should follow the principles of problem orientation, data authenticity, and closed-loop decision making, and cover typical financial scenarios. For example, through the retail enterprise historical transaction data to build customer segmentation and precision marketing case, students are required to cleanse the noise data, establish an RFM model and design promotional strategies; or design cross-border merger and acquisition risk quantification case, provide exchange rate fluctuations, policy text and other sources of heterogeneous data, training integrated learning modeling and stress testing capabilities. The case database can adopt a sandwich structure, i.e., the bottom layer encapsulates standardized datasets and API interfaces, the middle layer provides benchmark code and visualization templates, and the top layer reserves space for business parameter adjustment to support trainees to explore the optimal solution under constraints (Shang, 2025). At the specific implementation level, a dual-track learning model can be implemented. On the one hand, a case study is carried out in the form of a workshop, where business backbones and technical tutors jointly teach and simulate the pressure of real decision-making; on the other hand, an online case study arena is built, and an automated scoring system and a leaderboard are set up to stimulate the internal drive for learning. After adopting this mode, the model tuning ability of students will be rapidly improved. In addition, it is also necessary to establish a dynamic update mechanism of cases, regularly inject emerging business situations, and synchronize the introduction of industry champions as a standard benchmark, so as to form a spiral upward path of "learning-practice-iteration". The establishment of a cross-departmental case co-creation team helps to promote the joint development of cases between finance and business teams based on real pain points. This real-world case system not only improves technical strength, but also fosters the mindset of communicating with data, becoming a speed booster for the digital transformation of financial organizations.

4.4 Establish cross-departmental technical and business collaboration mechanisms

The establishment of cross-departmental technology and business collaboration mechanisms is a critical accelerator for the landing of financial big data literacy, the core of which lies in the establishment of a trinity of technology, business and strategy integration ecology. Four major mechanisms are needed to break down departmental barriers. First, the joint team operation mechanism, the formation of an agile team including finance, IT, and business backbone, and the use of dual-responsible systems to ensure the need for accurate communication (Kang et al., 2025). For example, by pairing financial BPs (business partners) with technical experts, the demand response time is shortened. Secondly, a two-way knowledge immersion mechanism, establishing a mapping dictionary of business terms and technical logic, regularly organizing finance salons, where technical staff explain the application of LSTM neural networks in cash flow forecasting and business staff share the business logic of supply chain finance to form complementary thinking. Third, project-based collaboration mechanism, cross-functional collaboration oriented to data products, such as the development of an intelligent reconciliation system, the finance team defines the reconciliation rules, the technical team builds the NLP model, and the operation team designs the user interaction interface, forming a closed-loop from demand, development to validation. Fourth, the talent rotation training mechanism, the implementation of the two-way rotation program of the technical backbone into the business, the business backbone to learn the technology, to enhance the proportion of the financial team to independently develop reporting tools. Through the establishment of a cross-departmental KPI linkage assessment mechanism, the business transformation effect of data analysis results is incorporated into the performance of both parties, forming a community of interest. This collaboration model not only improves the efficiency of technology implementation, but also cultivates composite talents who understand business pain points and have data thinking, providing sustainable motivation for the digital transformation of finance.

Conclusion

Financial big data literacy is the core ability of financial employees in the digital era, which needs to be systematically improved in four aspects. First, build a tiered training system, from basic syntax to advanced modeling progressive training, so that financial personnel can handle data cleansing, but also master machine learning; second, build a lightweight toolchain, integrate tools such as Jupyter, Streamlit, and develop a templated code base to reduce the technical threshold; third, build a practical case database, and train the model construction and decision-making through real business situations. Finally, establish a cross-departmental collaboration mechanism to promote the in-depth integration of technology and business to form a data-driven work model. The enhancement path needs to focus on the combination of learning and use, improve efficiency through the integration of tools, refine skills through case studies, and open up the data value chain through the collaboration mechanism, so as to ultimately realize the transformation of finance from a recorder to a value creator, and provide accurate data support for enterprise decision-making.

Conflict of Interest

The authors declare that they have no conflicts of interest to this work.

References

Bi, Y. J. (2023). Research on the transformation of the "Financial Big Data Analysis" course and textbook construction. *China Training, 2023*(06), 78–80.

Hu, Y. X. (2025). Research on the application of big data analysis in corporate financial decision-making. *Modernization of the Market, 2025*(05), 171–173.

Yuan, X. W. (2025). Exploration and practice of improving financial big data literacy based on Python. *Modern Commercial and Industrial, 2025*(06), 173–176.

Wang, H. H., Wang, R., Xiao, X., et al. (2023). Exploration of teaching practice of the "Financial Big Data Application" course for improving digital literacy. *Commercial Accounting, 2023*(18), 121–123.

Zhang, W. W. (2025). Construction of the content of the Financial Big Data Analysis course from the perspective of supply and demand coupling—Based on data analysis of online job information. *Modern Commercial and Industrial, 2025*(07), 165–168.

Xu, Y., & Chen, X. W. (2025). Application of big data and artificial intelligence in financial management. *China Exhibition (China Meeting), 2025*(04), 149–151.

Shang, N. (2025). Innovation and transformation of financial management in state-owned enterprises in the big data era. *China Market, 2025*(06), 179–182.

Kang, P., Dang, Y. X., & Hu, M. (2025). Reform and innovation practice of teaching in the major of big data and accounting in agricultural

vocational colleges—Taking the "Financial Statement Analysis" course as an example. *Modern Animal Husbandry Science and Technology, 2025*(02), 177–180.

How to Cite: Su, J. & Liu, C. (2025).Research on Exploring and Practicing the Improvement of Financial Big Data Literacy Based on Python. *Contemporary Education and Teaching Research, 06*(4), 122-127. https://doi.org/ 10.61360/BoniCETR252018030403