Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhancing Finance Education Through Computational Tools: Implementing GitHub Classroom for Industry-Ready Graduates

Mehul Zawar
The Institute of Chartered Accountants of India, Mumbai, India
mehul.zawar92@gmail.com https://orcid.org/0009-0001-2848-2318

ARTICLE INFO

Received: 21 Sep 2024

Accepted: 19 Dec 2024

ABSTRACT

This research paper discusses the integration of GitHub Classroom into finance education to foster industry-ready graduates equipped for the fast-paced digital landscape. Using GitHub as a central repository, students can upload assignments and provide feedback, while automated features in GitHub Classroom offer instant feedback and progress tracking. The shift towards a code-first approach in teaching econometrics and statistics aims to engage students, particularly those from non-technical backgrounds, by reducing the frictions associated with computational learning. This paper outlines the benefits of embedding computation into the curriculum, emphasizing the development of essential communication skills to articulate the importance of data-driven outputs. The findings suggest that by instilling reproducibility and workflow principles early in the learning process, students not only improve their technical abilities, but also prepare for an evolving industry landscape. Ultimately, the research aims to contribute to a more meaningful, transparent, and effective approach to data problem solving in business contexts.

Keywords: Finance, Educational Technology, Fintech, Computational Tools, Financial Analytics Education, Data-Driven Learning.

1 INTRODUCTION

The fast algorithmic change of monetary administrations has pushed the monetary innovation (FinTech) industry from the sidelines to the actual center of the worldwide monetary framework. FinTech addresses the conjunction of monetary administrations and state-of-the-art advances, reshaping how people, organizations, and establishments are accessed with monetary items and administrations. This change is especially obvious in the Unified Realm, where FinTech has been perceived as a progressive mechanical wave that is forever modifying the monetary landscape. As expressed in the Kalifa Survey of UK FinTech, FinTech has developed into a public vital need, with an unmistakable focus on encouraging development, increasing tasks, and spanning computerized abilities holes.

The multidisciplinary idea of FinTech is driving a shift from the ongoing period overwhelmed by social, portable, examination, and cloud (SMAC) innovations to a high level stage described by arising distributed record innovation, man-made brainpower, broadened reality, and quantum computing (DARQ) standards. These cutting-edge advances are ready to be the critical differentiators in the monetary administrations area. Consequently, understudies and money experts should adapt to this development by securing computational education and mechanical familiarity. Without implanting computational reasoning into finance schooling, the abilities hole might obstruct development and upset the capacity of future alumni to stay serious in a quickly evolving industry.

While finance remains in a general sense a sociology, current money progressively depends on quantitative methodologies. Mechanical advances in calculation, information examination, and AI are changing the way financial professionals handle pragmatic issues. The rise of enormous information, elective datasets and computerized reasoning (man-made intelligence) has opened new roads for tackling complex monetary issues, empowering progressions in regions like gamble the executives (Lin and Hsu, 2017), portfolio streamlining (Jaeger et al., 2021), venture banking (Speculation Banking Board, 2020) and protection examination (Society of Statisticians, 2020). As López de Prado (2019) recommends, the algorithmisation of money is both unavoidable and irreversible, setting out open doors for information-driven advancement.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The strength of thin man-made intelligence, described by rule-based AI calculations, has worked with the mechanization of monetary errands, from extortion discovery to credit scoring. However, the following boondocks of monetary innovation lie in digitizing dynamic cycles and computerizing informed decisions (López de Prado, 2018). This change raises significant difficulties for finance experts, especially in resolving trust, interpretability, overfitting, predisposition, and moral information utilization issues. The trustee obligation of money experts to act to the greatest advantage of their clients requires a harmony between utilizing man-made intelligence driven innovations and keeping up with straightforwardness, decency, and responsibility. In this specific situation, there is a developing interest in computationally educated graduates fit to coordinate monetary AI with standards of supportability, morals, security, and social obligation to convey trustworthy monetary arrangements (Mahdavi and Kazemi, 2020).

The Unified Realm has arisen as a worldwide forerunner in FinTech development and is effectively situating itself to scale its FinTech biological system in the post-Brexit scene. The Kalifa Audit frames an aggressive five-direct procedure toward reinforce the UK's situation in the FinTech area, stressing the basic requirement for labor force advancement through upskilling and reskilling programs. A central component of this procedure is cooperation with colleges and educational establishments to conduct specific instructional classes that address the developing demands of the monetary business.

In this paper, we investigate the open doors and difficulties that finance training faces as it adjusts to the rapid pace of the advancement of monetary innovation. In particular, we look at three critical areas of change: (1) the development of econometrics in an information-rich climate, (2) the ascent of monetary AI as a center capability, and (3) the coordination of computational instruments to work with a consistent, innovation-driven way to deal with educating. We give a contextual investigation featuring how the Administration School at Sovereign's College Belfast has tended to these difficulties through the execution of big business scale distributed computing framework and endeavor grade web programming devices.

2 BACKGROUND

2.1 What is monetary AI (FML)?

AI (ML) has seen quick reception in genuine applications; however, it lingers behind in financial examination, where econometric methods rule. Driving financial analysts attribute this to social and epistemological contrasts between econometrics and ML (Athey and Imbens 2019; Efron and Hastie 2016). Monetary AI (FML) is a subfield of man-made intelligence that tries to accommodate these distinctions by tending to three key contentions (Lommers, Harzli and Kim, 2021):

- Factual surmising.
- Causality.
- Speculations and model presumptions.

ML, in contrast to econometrics, focuses on expectation by advancing calculations to learn designs in information without accepting an unequivocal information-producing process (Efron and Hastie, 2016). Econometrics depends on precharacterized speculations and asymptotic hypothesis for true measurable achieving, yet battles with out-of-test expectation (López de Prado 2019).

2.2 Statistical Inference

Factual induction spans math and observational science, however, generally confronts computational bottlenecks, driving dependence on little example asymptotic arrangements (Efron and Hastie 2016). Econometrics stresses substantial surmising under severe suspicions (Gelman, Slope, and Vehtari 2020), yet is frequently contrary with present day finance assignments like out-of-test expectation.

ML models center around expectation, advancing calculations to fit noticed information while progressively consolidating logical computer-based intelligence (XAI) and measurable consistency (Barredo Arrieta et al. 2020; Blume et al. 2019; Zuo, Stewart and Blume 2021).

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

2.3 Modelling Paradigm

The two predominant factual standards are frequentism and Bayesiannism. Conventional econometrics is based on frequentist techniques, inclined toward objectivity under asymptotic circumstances. However, Bayesian surmising, with its instinctive refreshing of convictions, is building up some forward momentum because of computational headways (Efron and Hastie, 2016).

2.4 Causality

Causal derivation, established in measurements and financial matters (Haavelmo 1943; Rubin 1974), gauges the effect of occasions on results. While ML focuses on expectation, it can upgrade causality by working on model particular and finding new monetary speculations (López de Prado 2018; Easley et al. 2020).

2.5 Hypotheses, Suppositions, and Social Clashes

Econometrics is speculation driven, zeroing in on substantial deduction, while ML is information driven, focusing on prescient execution (Breiman, 2001). Econometrics accepts consistency, ordinariness, and effectiveness, mirroring its underlying foundations close by determined models. Propels in ML, like group strategies (e.g., irregular backwoods) and hypothetical turns of events (Athey and Bet 2017; Zuo, Stewart, and Blume 2021), exhibit the potential to coordinate ML into econometrics.

2.6 Brief History of Processing in Money and the Cloud

Money and calculation have been interwoven for quite some time. Bachelier (1900) laid the basis for quantitative money, while figuring as a utility arose during the 1960s (McCarthy).

In buy-side, Markowitz's mean fluctuation improvement (1950s) changed portfolio the executives, and Thorp and Simons (1960s) utilized PC supported calculations to take advantage of exchange. On the sell-side, the Dark Scholes-Merton (BSM) model (Dark and Scholes 1973; Merton 1973) prodded dangerous development in choices markets, driving advances in monetary registering to address BSM's impediments (Reisinger and Wissmann 2018).

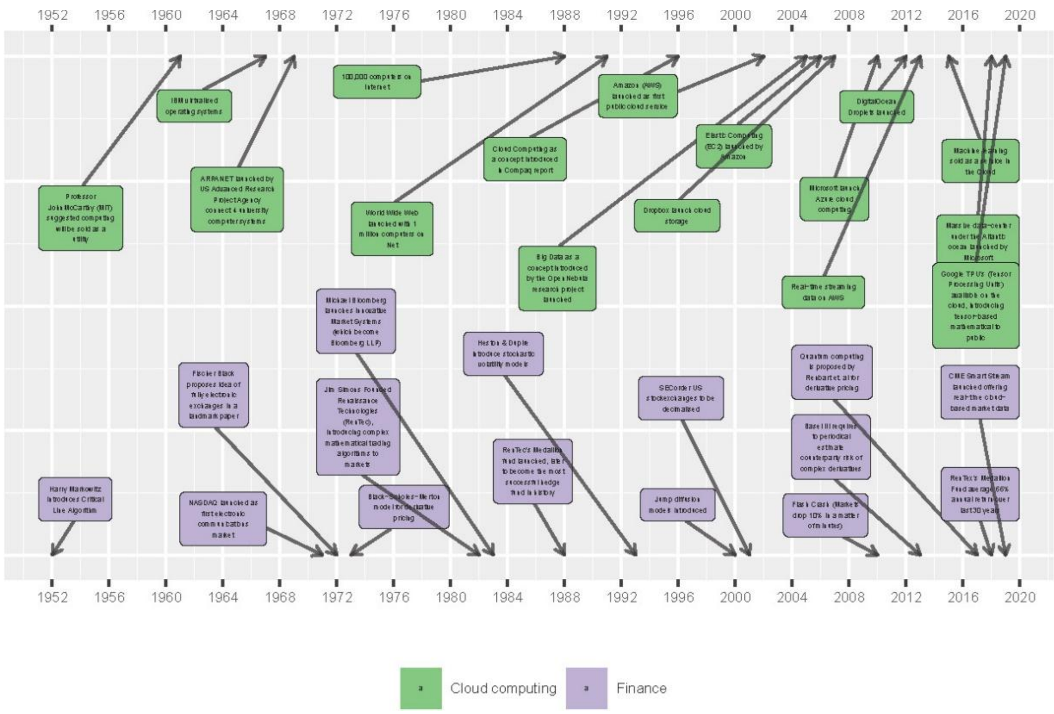

Figure 1: Computing landmarks finance and cloud computing. The data for the cloud computing timeline is sourced from Varghese et al. (2019), while the finance timeline is the authors' calculations.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3 TEACHING CLIMATE FOR COMPUTING

Performing a registration on a monetary examination course offers three key benefits:

1) Creating significant result inside the five star, empowering quick information examination.

2) Coordinating calculation as a center piece of the money educational program (Kaplan 2007; etinkaya-Rundel and Rundel 2018.

3) Demystifying the folk hypothesis of measurable computing}, where understudies characteristic result quality to the figuring climate.

Conventional registrar laboratories need authoritative control, limit content preloading, and minimize dynamic commitment. Interestingly, program-based distributed computing arrangements, for example, RStudio Groups, offer a frictionless encounter through Docker-based holders. The RStudio Workbench IDE upholds R and Python prearranging, improving learning with negligible arrangement grinding ("RStudio Workbench" 2021).

Unlike laboratories, our methodology gives:

• Dynamic talks and every minute of every day admittance to figuring assets;

• Practical answers for understudies lacking equipment;

• Worked on code, information, and climate sharing.

3.1 Why R and Python?

R and Python overwhelm information examination in industry, making them fundamental for employability. Although GUI-based apparatuses work on early learning, they thwart generalizable abilities, increment mistakes, and disengage calculation from examination (Baumer et al. 2014). Showing programming close-by measurements better prepares understudies for current work processes and perplexing, chaotic monetary datasets.

The two dialects are educated at the MSc level (e.g., MSc in Quantitative Money), with plans to venture into actuarial science programs. Understudies gain adaptable coding abilities, improving industry availability.

3.2 Why RStudio Teams?

RStudio Groups packs venture-grade devices - Workbench, Bundle Supervisor, and Associate - to scale and oversee information science work processes ("RStudio Group" 2021). Scholarly limits (saving $15, 000 - $20,000) allow centering around a nimble framework.

Workbench gives a natural IDE to R and Python, highlighting

• Incorporated Devices: R Markdown, Git, and Docker

• Intelligent sheets for information, plots, and records

• Uniform conditions, diminishing arrangement issues

Understudies access normalized cloud-based assets while logically fostering their nearby arrangements.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

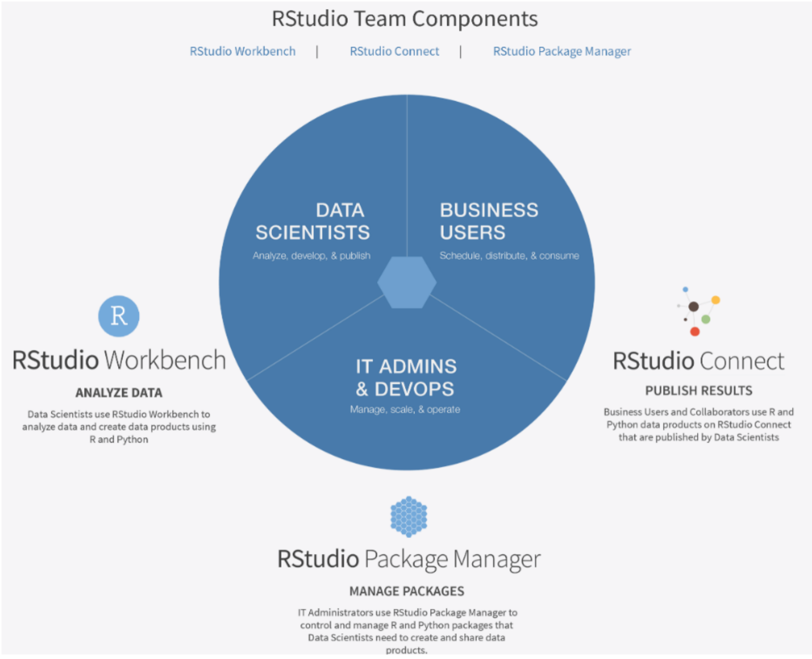

Figure 2: The three components of the RStudio Enterprise Team Bundle.

3.3 Remote RStudio Workbench Platform

We use virtualized equipment on Microsoft Purplish blue to run RStudio Workbench. The understudies interface through departmental verification, guaranteeing steady figuring conditions and settling the "but it chips away at my machine" issue.

Cloud stages enable incorporated command over bundles, asset checking, and meetings, facilitating IT loads while supporting reproducibility. Regardless of the board above, the input features of the understudy are critical employment benefits.

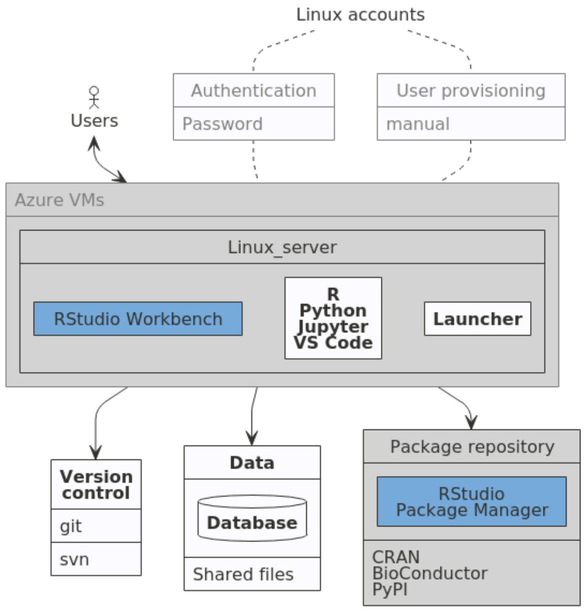

Figure 3: Current set up of RStudio workbench on Azure.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.4 Containerisation in Finance

Containerization bundles applications with their runtime surroundings, guaranteeing a lightweight and convenient arrangement. Unlike virtual machines, holders share the host operating system, decreasing the asset above (Mavridis and Karatza 2019).

The open source Docker system "build once, run anywhere" speeds up reproducible examination, allowing sandboxed RStudio occurrences (Worldwide Investor 2017). Kubernetes and the purplish blue Kubernetes Administration (AKS) can coordinate adaptable holder groups.

To improve administration, we embrace Duke College's arrangement (etinkaya-Rundel and Rundel 2018), shared by Imprint McCahill (etinkaya-Rundel and Rundel 2018), shared by Imprint McCahill. This plan empowers controlled, adaptable conditions while limiting complexity.

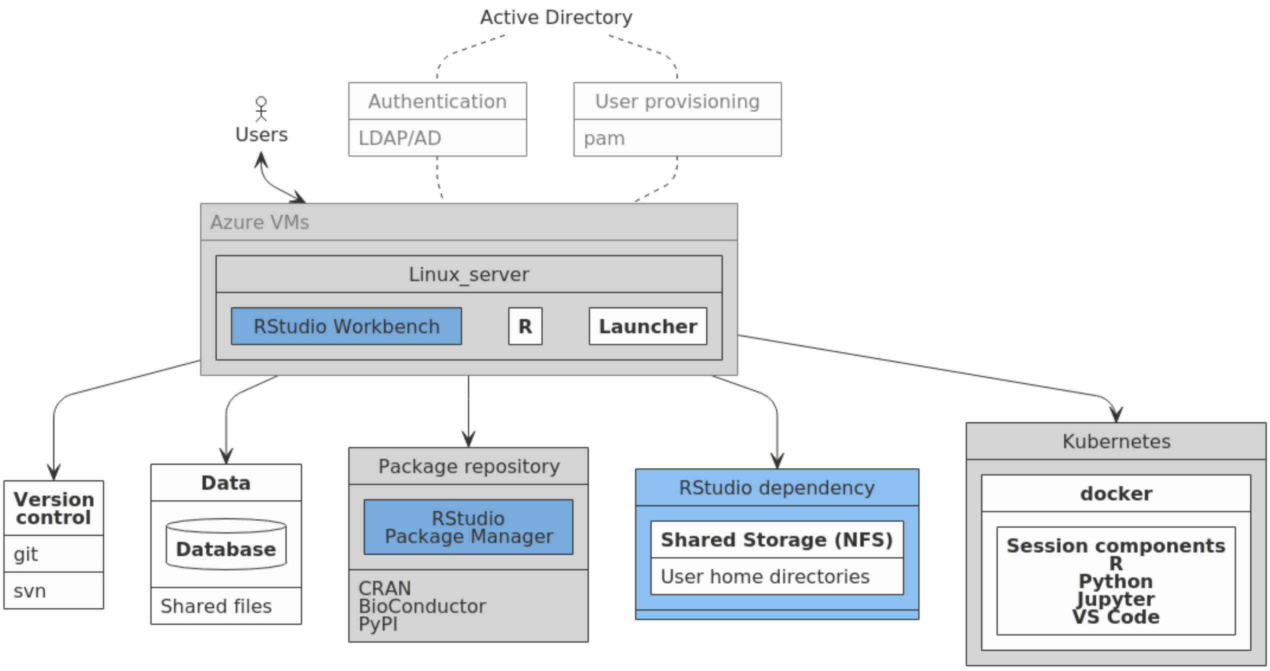

Figure 4: Dockerised set up of RStudio workbench on Azure.

4 COURSE IMPLEMENTATION

We steered our new framework, Q-RaP (Sovereign's Administration School Remote Examination Stage), in 2020-2021 for two MSc modules: Algorithmic Exchanging and Venture, and Time-Series Monetary Econometrics. The input from the study was great, especially during remote instruction. In 2021-2022, Q-RaP expanded to two more MSc courses (Exploration Strategies in Money and Computational Techniques in Money) and selected business examination modules. In past education, it reduces PC lab pressures.

4.1 Reproducibility with Computational Notebooks

Computational note pads consolidate code, story, and result, improving straightforwardness, compactness, and reproducibility. They execute the PPDAC (Issue, Plan, Information, Examination, Correspondence) work process, the expert norm for solid investigation (Spiegelhalter 2019).

Two essential journal designs are Jupyter Note pads and R Markdown. Both use Markdown, a lightweight, stage-free grammar for current information examination. Jupyter upholds various programming dialects, including Python through the IPython portion, empowering non-specialized understudies to rapidly begin coding. R Markdown, incorporated with RStudio, allows dynamic results such as reports, dashboards, and web applications utilizing apparatuses like Shiny and reticulate for Python joining.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Key educational advantages of computational note pads (Knuth, 1984) include:

1) Bound together code, result, and story streamline troubleshooting and trial and error;

2) Normalized formats ease reviewing;

3) Variant control upholds cooperation and tracks commitments;

4) Benchmark layouts progressively empower dynamic advancing as platform is taken out.

4.2 Version Control, Git, and GitHub

Git, a generally accepted rendition control instrument, and GitHub, its facilitating stage, are fundamental for cooperative monetary determination. In spite of beginning industry opposition over security concerns, acquisitions like Microsoft's GitHub (2018) and Google's interest in GitLab have established Git's status as an industry standard.

Understudies use GitHub for tasks, group work, and reproducibility. GitHub study hall works with robotized criticism, reproducibility checks, and live competitor lists for expectation projects (etinkaya-Rundel and Rundel 2018). Mix with learning the executives frameworks like Material improves on work processes.

Our model puts together vaults per module, with groups overseeing bunch tasks through GitHub's entrance controls. Persistent combination devices allow educators to screen commitments, providing versatile and organized instruction support.

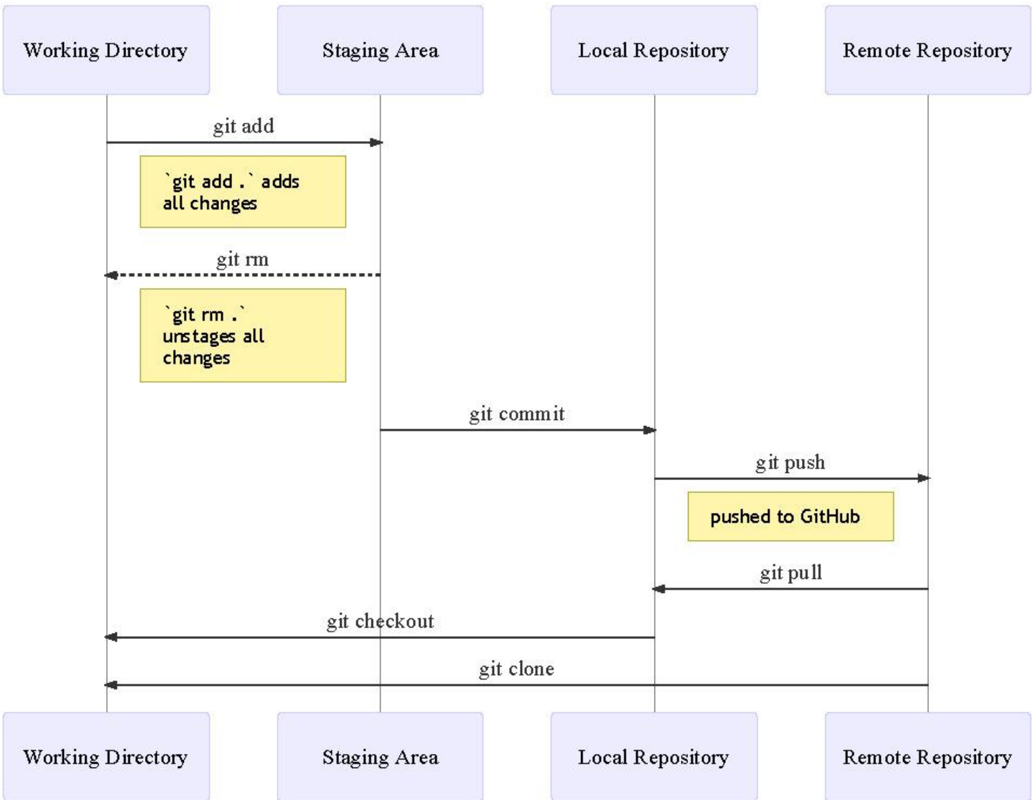

Figure 5: Seven git commands students need to learn.

5 DISCUSSION

The findings of this paper highlight the transformative potential of embedding computational tools in finance education. Using GitHub as the central repository for coursework and RStudio for statistical computing, the teaching framework reduces barriers to learning and creates an engaging, frictionless environment. This approach is aligned with industry practices, ensuring that students gain practical transferable skills.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Pedagogically, the shift to a code-first methodology has been instrumental in enhancing student engagement and understanding. By introducing computation early and focusing on practical applications, the learning curve has been significantly flattened for non-technical learners. The integration of computational notebooks has further reinforced the importance of documenting workflows, creating reproducible results, and fostering collaboration among peers. These practices mirror industry standards and ensure that students graduate with marketable skills.

Furthermore, requiring assignments in literate programming formats ensures that students internalize key principles of reproducibility and workflow management, laying the foundation for ethical and transparent data analysis. This methodology helps students move beyond rote computation to critically engage with the data, interpret the results, and communicate the findings effectively.

Despite these successes, the initiative also highlights the challenges of implementing such a model. The reliance on cloud-based tools necessitates robust IT infrastructure and technical support, which may pose obstacles for institutions with limited resources. Additionally, the initial resistance from students and educators unfamiliar with computational methods underscores the need for effective onboarding and continuous professional development. These challenges present opportunities for further refinement and collaboration.

6 LIMITATIONS

Several limitations of this study should be acknowledged. First, the implementation of GitHub Classroom and RStudio was conducted in a specific institutional context, which may limit the generalizability of the findings. Different institutions may face unique challenges in adopting these tools, such as varying levels of IT support or differences in student demographics.

Second, reliance on cloud-based platforms raises concerns about data privacy and accessibility, particularly for students in regions with limited internet connectivity. Ensuring equitable access to these tools will require additional investment in infrastructure and support systems. Addressing these disparities is essential to ensure that the benefits of this approach are shared between diverse student populations.

Lastly, while the study focused on technical and pedagogical outcomes, it did not extensively assess the long-term impact on student professional success. Future research should examine how these teaching innovations influence graduates ability to apply computational skills in real-world financial settings. In addition, longitudinal studies could provide insight into how well students retain and adapt these skills over time, particularly as they face new challenges and technologies in their careers.

7 CONCLUSION

The integration of GitHub Classroom into finance education represents a significant step forward in preparing students for the demands of an industry increasingly shaped by computational tools and technologies. By embedding coding and reproducibility principles into the curriculum, the approach outlined in this document fosters both technical proficiency and critical communication skills, essential to navigate the evolving landscape of financial technology. This initiative has demonstrated that a code-first pedagogical model can bridge the gap between technical and non-technical learners, making computational education accessible and engaging.

In addition, the integration of tools such as GitHub and RStudio has reinforced a culture of reproducibility and collaborative learning. These platforms not only streamline the assignment submission and feedback process, but also instill professional habits, such as version control and data documentation, that are vital in real-world applications. The focus on using computational notebooks and adopting the PPDAC (Problem, Plan, Data, Analysis, Conclusion) framework has further enabled students to develop structured, transparent, and replicable workflows to tackle complex financial problems. These efforts align with industry trends, where proficiency in coding and data literacy is increasingly valued across sectors.

The results emphasize the importance of combining tools such as RStudio and GitHub with foundational teaching principles. These innovations enable students to produce high-quality, reproducible analyzes while cultivating habits that align with industry standards. Ultimately, this paper underscores the potential of computational tools not only to enhance learning outcomes but also to empower students to approach data-driven problem solving with transparency and rigor. By addressing the barriers traditionally associated with statistical computation, this approach contributes to the larger goal of democratizing access to financial education and ensuring that graduates are equipped for the challenges of a technology-driven economy.

8 FUTURE WORK

Although this study has shown the feasibility and benefits of integrating GitHub Classroom into finance education, several opportunities for future development remain. Expanding this model to a wider range of quantitative modules and disciplines could provide further validation and insight. Exploring additional tools, such as advanced machine learning platforms or cloud-based collaboration suites, could enhance the learning experience.

Another area of growth involves the systematic assessment of long-term outcomes, such as tracking graduates' career trajectories and their ability to apply computational skills in professional contexts. In addition, collaboration with industry partners to co-design curricula and case studies could strengthen the alignment between education and industry needs. Industry input can help refine course content, ensuring that it remains relevant and responsive to emerging trends in financial technology.

Furthermore, research into the pedagogical challenges associated with teaching non-technical students could lead to better strategies for fostering inclusivity and engagement. Tailored resources, such as interactive tutorials or adaptive learning environments, might help bridge the gap for learners with varying levels of prior experience. Finally, addressing accessibility challenges, including providing additional support for students from diverse socioeconomic backgrounds will be critical to scaling this approach globally. Future work should aim to create an adaptable, inclusive framework that ensures that every student can benefit from computational finance education.

ACKNOWLEDGEMENT

This paper builds upon the research work presented at IES MCRC's IRC-2025, incorporating further analysis, expanded discussions, and deeper insights.

REFERENCES

[1] Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(4), 1059–1086. https://doi.org/10.48550/arXiv.1612.08468]

[2] Athey, S., Bayati, M., Imbens, G., & Qu, Z. (2019, March). Ensemble methods for causal effects in panel data settings. *arXiv*. https://arxiv.org/abs/1903.10079

[3] Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*(1), 685–725. https://doi.org/10.1146/annurev-economics-080217-053433]

[4] Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics*, *47*(2), 1148–1178. https://doi.org/10.48550/arXiv.1610.01271]

[5] Athey, S., & Wager, S. (2017, February). Policy learning with observational data. *arXiv*. https://arxiv.org/abs/1702.02896

[6] Bachelier, L. (1900). Theory of speculation in the random character of stock market prices. *MIT Press, Cambridge, Mass. Blattberg*, 17–78.

[7] Barredo Arrieta, A., Diaz-Rodriguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., & García, S. et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.48550/arXiv.1910.10045]

[8] Baumer, B., Çetinkaya-Rundel, M., Bray, A., Loi, L., & Horton, N. (2014). R Markdown: Integrating a reproducible analysis tool into introductory statistics. *[Undefined Journal]*.

[9] Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637–654.

[10] Blume, J. D., Greevy, R. A., Welty, V. F., Smith, J. R., & Dupont, W. D. (2019). An introduction to second-generation p-values. *The American Statistician*, *73*(S1), 157–167.

[11] Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231.
[12] Cesa, M. (2017). A brief history of quantitative finance. *Probability, Uncertainty and Quantitative Risk*, *2*(1), 1–16.
[13] Chamberlain, G. (2000). Econometrics and decision theory. *Journal of Econometrics*, *95*(2), 255–283.
[14] Chamberlain, G. (2020). Robust decision theory and econometrics. *Annual Review of Economics*, *12*(1), 239–271.
[15] Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.
[16] Çetinkaya-Rundel, M., & Rundel, C. (2018). Infrastructure and tools for teaching computing throughout the statistical curriculum. *The American Statistician*, *72*(1), 58–65.
[17] Dixon, M. F., & Polson, N. G. (2020, March). Deep fundamental factor models. *arXiv*. https://arxiv.org/abs/1903.07677
[18] Dreze, J. H. (1972). Econometrics and decision theory. *Econometrica*.
[19] Easley, D., López de Prado, M., O'Hara, M., & Zhang, Z. (2020, July). Microstructure in the machine age. *Review of Financial Studies*.
[20] Efron, B., & Hastie, T. (2016). *Computer age statistical inference*. Cambridge University Press.
[21] Fisher, R. A. (1936). Design of experiments. *British Medical Journal*, *1*(3923), 554.
[22] Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.
[23] Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, *11*(1), 1–12.
[24] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction (2nd ed.)*. Springer Science & Business Media.
[25] International Banker. (2017). The benefits of leveraging containers in the financial services industry. *International Banker*. https://internationalbanker.com/technology/benefits-leveraging-containers-financial-services-industry/
[26] Investment Banking Council. (2020). AI in investment banking - the new frontier. *Investment Banking Council Blog*. https://www.investmentbankingcouncil.org/blog/ai-in-investment-banking-the-new-frontier
[27] Jaeger, M., Krügel, S., Marinelli, D., Papenbrock, J., & Schwendner, P. (2021, June). Interpretable machine learning for diversified portfolio construction. *The Journal of Financial Data Science*.
[28] Kaplan, D. (2007). Computing and introductory statistics. *Technology Innovations in Statistics Education*, *1*(1).
[29] Knuth, D. E. (1984). Literate programming. *The Computer Journal*, *27*(2), 97–111.
[30] Lin, S.-J., & Hsu, M.-F. (2017). Incorporated risk metrics and hybrid AI techniques for risk management. *Neural Computing and Applications*, *28*(11), 3477–3489.
[31] Lommers, K., El Harzli, O., & Kim, J. (2021, June). Confronting machine learning with financial research. *The Journal of Financial Data Science*.
[32] López de Prado, M. (2018). *Advances in financial machine learning*. John Wiley & Sons.
[33] López de Prado, M. (2019). A data science solution to the multiple-testing crisis in financial research. *The Journal of Financial Data Science*, *1*(1), 99–110.
[34] Mahdavi, M., & Kazemi, H. (2020). It's all about data: How to make good decisions in a world awash with information. *The Journal of Financial Data Science*, *2*(2), 8–18.
[35] Perrot, S. (2016). On the use of machine learning techniques in portfolio management. *Quantitative Finance*, *16*(1), 1–10.
[36] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.
[37] Tibshirani, R. (2015). The lasso: A tool for variable selection in high-dimensional data. *Statistical Science*, *28*(3), 267–275.
[38] Wager, S., Hastie, T., & Hinton, G. E. (2018). Deep learning for econometrics. *Journal of Econometrics*, *202*(2), 107–127.
[39] Wager, S., & Athey, S. (2017). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *112*(519), 1–15.

[40] Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82.
[41] Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, *5*(3), 161–215.
[42] Zuo, Y., Stewart, T. G., & Blume, J. D. (2021). Variable selection with second-generation p-values. *The American Statistician*, *75*(2), 1–21.