

Transforming Industries: Empowering Organizations through Big Data Analytics

Prof. Haipeng Shen

Dr. Dan Yang

Dr. Zhengli Wang

Innovation and Information Management

HKU Business School

2023/5

Summary of the Impact



Prof. Shen, Dr. Yang, and Dr. Wang have consistently participated in knowledge exchange (KE) activities, leading to significant advancements in the application of big data analysis and machine learning across diverse industries. Their involvement has sparked transformative changes in organizational operations within mainland China and the Greater Bay Area, spanning sectors such as healthcare, finance, real estate, banking, dairy, and manufacturing.

The impact encompasses the following key aspects

- **Research Impact:** Profound contributions are made through multiple publications and grants, with a constant drive for exciting and innovative developments on the horizon.
- **Teaching Impact:** By imparting industry-specific knowledge and nurturing a diverse skill set encompassing methodologies, algorithm implementation, and a business mindset, we empower individuals across all levels, from students to senior executives, to revolutionize their respective industries.
- **Development and Implementation of Frameworks and Robust Tools:** We create and implement frameworks and rigorous tools that enhance business outcomes in various business contexts, resulting in significant and measurable improvements. For instance, the Yili Project in Shandong Province successfully increased sales by 12.85% by implementing an algorithm-based store recommendation system. This has led to numerous privileged awards within industries.
- **Systematic and Automatic Data Gathering and Analysis:** We design streamlined procedures for systematically and automatically collecting data, extracting relevant information, and analyzing their importance within the specific business context. Such streamlined procedures can be utilized by people without training in data analytics.
- **Corporate Engagement:** We actively drive corporate engagement to promote widespread awareness, highlight the necessity, and enhance understanding of the significance of employing big data approaches at all levels within companies, including employees, managers, and senior executives.

Table of Contents



谢谢!

Explain the quality of the knowledge. An outline of what the underpinning research was.

- At the heart of our research lies machine learning, time series forecasting, and network analysis. These core components have extensive applications across various industries and scenarios, proving their significance in today's data-driven world.
- In recent years, time series data has grown exponentially, forming the backbone of big data. As the structure of time series becomes increasingly complex, modern data collection techniques are producing "tensor" data (multi-dimensional arrays) more frequently than ever before. This shift necessitates the development of novel methodologies to analyze and interpret such data accurately.
- Our cutting-edge time series analysis techniques are specifically designed to handle large and complex datasets, providing more accurate forecasting than traditional state-of-the-art methods. Furthermore, our approach generates a deeper understanding of the fundamental evolution of the processes involved, allowing for more informed decision-making.
- In addition to forecasting and comprehending complex tensors, we also place significant emphasis on the estimation and testing of their covariance. This focus is crucial for understanding the correlation structure and relationships within the data, providing valuable insights to drive further analysis.
- Network formulation plays a fundamental role in our research as well. We are particularly interested in the impact of network structures on specific outcomes and are dedicated to exploring this relationship. By examining the intricacies of network formulation, we aim to uncover valuable insights that can inform future predictions and enable better decision-making across a wide range of applications.

Explain the quality of the knowledge. More findings:

- Through the application of our newly developed methodologies, we have achieved several notable accomplishments across various domains. These concrete achievements demonstrate the potential of our innovative techniques in providing valuable insights and enhancing decision-making processes.
- One of our key accomplishments involves formulating a novel understanding of taxi traffic flow in New York City. By applying our cutting-edge methods, we have successfully uncovered previously hidden patterns and trends, leading to a more comprehensive understanding of the city's transportation dynamics.
- In the realm of global trade, we have discovered new findings regarding the patterns of import-export networks across various industries. Our innovative approach has allowed us to uncover unique relationships and dependencies, enriching our knowledge of international trade dynamics.
- When examining global macroeconomic indices, we have found that they comove in a more systematic manner, allowing for more accurate forecasting. This improved prediction capability can have significant implications for policymakers and market participants alike.
- Our analysis of international stock market data has revealed intriguing patterns and validated economic propositions through rigorous statistical testing. This accomplishment demonstrates the potential of our methods in enhancing investment strategies and understanding market behavior.
- Lastly, we have successfully predicted currency risk premiums more accurately by leveraging the global trade network. This achievement has led to the generation of lucrative trading strategies, highlighting the practical applications of our innovative techniques in the world of finance. Overall, our research has demonstrated significant potential for providing valuable insights across various industries and applications, ultimately leading to better-informed decision-making.



Related Publications:

- Xin Chen, Dan Yang, Yan Xu, Yin Xia, Dong Wang, and Haipeng Shen (2023). Covariance matrix testing for matrix-valued observations and its application to portfolio construction. *Journal of Econometrics*, 232(2):544-564.
- Rong Chen, Dan Yang, and Cun-Hui Zhang (2022). Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association*, with discussion, 117(537): 94-116.
- Rong Chen, Han Xiao, and Dan Yang (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics*, 222(1):539-560.
- Junhui Cai, Dan Yang, Wu Zhu, Haipeng Shen, and Linda Zhao (2023+). Network regression and supervised centrality estimation. Available at [arxiv:2111.12921](https://arxiv.org/abs/2111.12921).
- Zhengli Wang, Kevin MacMillan, Mark Powell and Lawrence Wein (2020). "A Cost-Effectiveness Analysis of the No. of Samples to Collect and Test From a SAK." *Proceedings of the National Academy of Sciences* 117, no. 24: 13421-13427.

When your research was undertaken, and your role (and your team members' roles, if applicable) in the creation of such knowledge.

Any relevant key contextual information about this area of research, e.g. where it is a wider body of research in collaboration with other institutions.

- Our research and collaboration efforts, which began in 2016, have led to multiple publications and grants, and continue to thrive with numerous exciting and innovative developments on the horizon. Prof. Shen and Dr. Yang have joined forces on a series of projects, demonstrating the fruitful results of their combined expertise.
- Our collaborative approach extends beyond working with colleagues within the IIM. We have also partnered with professionals from the Finance department at HKU Business School, creating opportunities for cross-disciplinary insights and knowledge exchange.
- Further expanding our collaborative network, we have engaged with statisticians from prestigious institutions such as Wharton Business School and Rutgers University, as well as economists from both Wharton and Tsinghua University. These partnerships have allowed us to draw upon the unique strengths and perspectives of each expert, resulting in a truly multidisciplinary approach to our research.
- This diverse range of expertise not only enriches our collective understanding but also ensures that our work has a more profound and far-reaching impact across multiple fields. By fostering a collaborative environment that encourages learning and growth, we aim to continue driving innovation and pushing the boundaries of what is possible in our research endeavors.

Innovativeness of the knowledge arising from your research at HKU

The research stands out as groundbreaking and innovative due to several reasons, which are highlighted below:

1. **Pioneering Analysis:** Our study is the first to delve into the complex and extensive domain of tensor time series data by employing auto-regressive models and factor models. This innovative approach allows us to gain unprecedented insights and understanding of the underlying patterns and structures within the data.
2. **Rigorous Statistical Testing:** We have taken the lead in performing thorough statistical testing of covariance matrices of tensor data, thereby ensuring the reliability and validity of the results obtained from our analysis. This rigorous testing ensures that our findings are robust and can withstand scrutiny from the scientific community.
3. **Identification of Drawbacks:** Our research is the first to uncover the limitations of naive regression on certain summaries of network information. This discovery highlights the need for more advanced and sophisticated approaches to analyzing network data and paves the way for the development of more effective methodologies.
4. **Broad Application Scenarios:** We have demonstrated the versatility and applicability of our developed methods by applying them to a wide range of new and diverse scenarios, such as multiple trade networks across various industries, taxi traffic tensors, and unified analysis of macroeconomic indices from multiple countries. This showcases the potential of our methods to be employed across various fields, driving advancements in multiple sectors.

Significance of the key insights or findings from the research that relate to the impact achieved by the KE project

- The development of innovative methodologies tailored to address the unique requirements of emerging data structures is of paramount importance. In particular, the Yili and China Resource projects generate a plethora of data sets that necessitate the adoption of such cutting-edge approaches.
- Through the successful implementation of these research-driven techniques, forecasts can be significantly enhanced, ensuring a deeper comprehension of the sales dynamics across various scenarios. Consequently, this will facilitate more informed decision-making and improved business outcomes.
- The cold start launch is widely recognized as a particularly challenging problem in the industry. In this context, the efficient and accurate identification of target stores is crucial for the profitability and overall success of new products. Our research has the potential to provide invaluable insights and support in this regard, thanks to the multidimensional nature of tensor data, which enables the integration of information from diverse sources. This, in turn, will empower businesses to make more strategic decisions and optimize their product launches.

Underpinning Teaching & Learning



- Teaching and learning activities in the Taught Postgraduate (TPg) and Executive Education (EE) programs at HKU have significantly contributed to the exchange of knowledge and the overall impact of the learning process. These programs have equipped students with the necessary skills and understanding to excel in their respective fields.
- Prof. Shen, for instance, teaches the Master of Science in Business Analytics (MSBA) Business Statistics course, which delves into the fundamental principles and methodologies of prediction. Furthermore, Prof. Shen instructs the capstone course of the MSBA program, acting as the cornerstone of this vital knowledge exchange. Through hands-on experience, MSBA students gain a deeper understanding of what it means to analyze data within various industries.
- Dr. Wang's MSBA Machine Learning course offers students the opportunity to explore and implement numerous advanced machine learning techniques within their projects. This exposure to cutting-edge technology enables students to stay ahead in the rapidly evolving field of data analysis.
- Dr. Yang, on the other hand, teaches three distinct MSBA courses: Forecasting, Big Data, and Deep Learning. These courses equip students with a more comprehensive and state-of-the-art toolkit for data analysis, ensuring they are well-prepared for the challenges they may encounter in their careers.
- Both Prof. Shen and Dr. Yang also contribute their expertise to multiple Executive Education courses, including the EE capstone. Through these courses, students such as Mr. Qichao Zhao have acquired the knowledge and skills needed to become successful data scientists. They have also developed hands-on experience in data analysis, allowing them to tackle real business situations with tangible implications and impacts.
- Overall, the teaching and learning activities within the TPg and EE programs at HKU have fostered a dynamic environment for knowledge exchange and practical application. This has ultimately led to the development of well-rounded professionals who are prepared to make a significant impact in their chosen industries.

What specific knowledge is being exchanged in your project?

In this project, a wealth of specific knowledge and expertise is being exchanged to ensure the successful implementation and integration of advanced technologies in the realm of business operations. This includes:

1. The comprehensive understanding and application of technical tools such as machine learning algorithms, advanced forecasting methods, and tensor time series analysis, which are crucial for processing and interpreting complex data sets efficiently and accurately.
2. The skill to effectively translate real-world problems into data analysis problems, ensuring that the appropriate data sources, methodologies, and tools are utilized to address the specific challenges and objectives at hand.
3. The capacity to communicate the results and insights derived from data analysis models in a clear and concise manner, enabling stakeholders to make informed decisions based on the evidence presented.
4. The expertise to generate actionable business recommendations by drawing meaningful conclusions from the outputs of data analysis models, ensuring that the organization is well-equipped to respond to market trends, optimize operations, and drive growth.
5. The innovative mindset to embrace the potential of digitalization and artificial intelligence, transforming traditional business operations through the adoption of cutting-edge technologies and data-driven strategies. This includes exploring new ways to enhance efficiency, improve customer experiences, and create a competitive advantage in an increasingly data-driven world.

How is the knowledge that is exchanged in your project generated, curated, and disseminated?
And how it is used to drive innovation, economic development, and/or social change?
Any particular insights or ideas being shared from your project, and how are they being applied in practice?

- Utilize business language to articulate the intricacies of the machine learning model, ensuring a comprehensive understanding of its capabilities and potential applications in the organization.
- Implement A/B testing to validate the results derived from the machine learning model, providing empirical evidence of its effectiveness and accuracy in real-world scenarios.
- Execute targeted surveys in relevant stores to corroborate the model's predictions and business recommendations, ensuring that the proposed strategies are grounded in tangible consumer insights and behaviors.
- Engage in knowledge exchange with domain experts, leveraging their expertise to refine the model, enhance its performance, and ensure its alignment with industry best practices and trends.
- Foster the development of a robust, data-driven foundation within the organization, enabling it to capitalize on the opportunities presented by the growing volume and complexity of available data.
- Embrace the power of advanced technologies, such as machine learning and artificial intelligence, to optimize operations, streamline decision-making, and drive innovation across the business.
- Propel the organization towards sustained success and growth by making data-driven decisions, navigating the competitive landscape with agility and foresight, and continuously adapting to the evolving market dynamics.



Prof. Shen, Dr. Yang and Dr. Wang's engagement in KE activities has promoted the utilization of big-data analysis and machine learning in various real-world scenarios. These engagements have transformed the way organizations and its associated industry practice operated.

Impact covers the following critical elements:

- **Development and Implementation of Frameworks and Robust Tools:** We create and implement frameworks and rigorous tools that enhance business outcomes in various business contexts, resulting in significant and measurable improvements. For instance, the Yili Project in Shandong Province successfully increased sales by 12.85% by implementing an algorithm-based store recommendation system.
- **Systematic and Automatic Data Gathering and Analysis:** We design streamlined procedures for systematically and automatically collecting data, extracting relevant information, and analyzing their importance within the specific business context.
- **Corporate Engagement:** We actively drive corporate engagement to promote widespread awareness, highlight the necessity, and enhance understanding of the significance of employing big data approaches at all levels within companies, including employees, managers, and senior executives.

Innovativeness of the engagement approach

Prof. Shen, Dr. Yang, and Dr. Wang's engagement in knowledge exchange (KE) activities was characterized by an innovative and holistic approach that harnessed multiple communication channels:

- **Comprehensive Engagement Strategies:** Their approach encompassed both top-down and bottom-up engagement strategies, ensuring involvement and participation from all levels of the organizations they worked with.
- **Integration of Cutting-Edge Methods:** They successfully incorporated and integrated various cutting-edge big data and machine learning methods into the algorithm platforms of different organizations, enhancing their capabilities and efficiency.
- **Exchange and Transfer of Tools:** Through their engagement, they facilitated the exchange, development, and transfer of tools for automated data gathering and extraction. This raised awareness among organizations about the necessity and benefits of adopting a data-driven approach.
- **In-Depth Discussions with Organizational Leaders:** Prof. Shen, Dr. Yang, and Dr. Wang engaged in in-depth and comprehensive discussions with organizational leaders. This allowed for the exchange of ideas and exploration of the expansion of proposed business strategies and the implementation of big data-driven approaches.
- **Oral and Written Communications:** They effectively communicated their insights through in-depth presentations delivered to organizational leaders, middle-level managers, and individuals responsible for algorithm implementation and sales. Additionally, comprehensive project reports were widely disseminated among corporate executives and shared on internal intranets for employees.
- **Teaching and Education:** They played an instrumental role in teaching aspiring leaders through higher education and executive teaching programs, equipping them with the knowledge and skills necessary to embrace data-driven approaches.
- **Global Dissemination:** A business case outlining their systematic big data and machine learning approach is expected to be disseminated to a global audience of executive leaders in large corporations, fostering awareness and understanding on a broader scale.



Sequential Engagements

Example: Prof. Shen, Dr. Wang, and Dr. Yang have had a series of engagements with Mr. Zhao, the current head of marketing at Yili, spanning multiple phases for the past three years and are still ongoing:

- Phase I: Mr. Zhao attended classes in HKU's Data Scientist Program by EE, enhancing his skills and knowledge in the field.
- Phase II: Prof. Shen and Dr. Yang provided supervision to Mr. Zhao on a capstone project in the Data Scientist Program related to China Resources Chongqing MixC Mall Management. This collaboration resulted in successful outcomes.
- Phase III: The capstone project from Phase II was implemented in a real-life scenario, further validating its effectiveness, yielding positive results, and winning multiple awards.
- Phase IV: Mr. Zhao returned to HKU for an MSBA capstone project aimed at improving Yili's sales strategy. Under the supervision of Dr. Wang, Mr. Zhao worked alongside HKU MSBA students on this Yili project.
- Phase V: The proposed sales strategy from Phase IV was validated, leading to a remarkable 12.8% increase in sales in Shandong Province.
- Phase VI: Impressed by the outcomes, Yili's senior executives expressed their desire to expand the project to other provinces of higher strategic importance. Dr. Wang is currently supervising Mr. Zhao and Mr. Tan Limin, the current head of data analytics at Yili, on a project in Guangdong Province. This new project introduces a unique feature selection framework and a scoring mechanism for rating stores.
- Future: Yili has expressed a keen interest in further engaging with Prof. Shen, Dr. Wang, and Dr. Yang to tackle a range of critical challenges faced by the company. These challenges encompass logistics optimization, quality control and shelf life management, market competition strategies, and comprehensive consumer behavior analysis. By continuing this partnership, Yili aims to drive innovation and address these important areas within their operations.

Impacts Achieved – Beneficiaries and Evidence



• Testimonials

- “本次项目我们收获很多。王教授的专业性，还有我们各位同学的思想活力，让我们收到非常大的启发。这次数据的处理、分析和建模的整个流程，我们也得到了一个结构化、体系化的学习。这次做的是金典超滤，其它产品我们也可以完全按照这个流程，复制到其它产品方面或者其它业务领域方面。这是一次非常难得的学习机会。” ——陈涛, 行销经理
- "We have gained a lot from this project. Prof. Wang's expertise, combined with the intellectual vitality of our fellow classmates, has been extremely inspiring. Throughout the entire process of data processing, analysis, and modeling, we have also acquired a structured and systematic learning experience. This project focused on gold standard ultrafiltration, and we can apply the same process to other products or different business domains. It has been a valuable learning opportunity." - Chen Tao, Marketing Manager
- “港大与伊利在山东地区的产品销售预测项目非常成功，我们接下来会继续和港大合作，把成功经验推广到广东、海南乃至全国其它地区，以及覆盖更多种类的产品，这会对伊利的销售模式产生很大的影响。” ——谭莉敏, 数据分析主管
- "The sales forecasting project between the University of Hong Kong and Yili in Shandong Province has been highly successful. Moving forward, we will continue to collaborate with the University of Hong Kong and extend the successful experience to Guangdong, Hainan, and other regions nationwide, covering a wider range of products. This will have a significant impact on Yili's sales model." - Tan Limin, Head of Data Analytics
- “数字化转型对于我们伊利来说，是一次颠覆性的尝试。通常而言，这个过程困难重重，领导层对其重要性缺乏认识，员工缺乏相关技术，而且整个过程缺乏透明度。然而，在港大团队的帮助下，我们成功走出了这条道路。我们从对机器学习做数字化转型觉得像黑匣子，到对整个流程了如指掌，到能信心满满地做商业决策。接下来，我们将继续深入与香港大学的合作。这次合作为我们带来了巨大的成果，我们相信在未来的合作中，将有更多令人激动的成就等待我们。 ——赵起超, 行销主管
- "Digital transformation has been a disruptive endeavor for Yili. Typically, this process is fraught with difficulties, including a lack of awareness among leadership about its importance, employees' limited technical skills, and a lack of transparency throughout the process. However, with the assistance of the University of Hong Kong team, we have successfully navigated this path. We have gone from viewing machine learning in digital transformation as a black box to mastering the entire process, enabling us to make confident business decisions. Moving forward, we will continue to deepen our collaboration with the University of Hong Kong. This partnership has yielded tremendous results, and we believe there are even more exciting achievements awaiting us in future collaborations." - Zhao Qichao, Head of Marketing



- Development and Implementation of Frameworks and Robust Tools: We create and implement frameworks and rigorous tools that enhance business outcomes in various business contexts, resulting in significant and measurable improvements.
- **Example: Yili Project in Shandong Province**
 - The implementation of the algorithm for store recommendations resulted in a validated increase in sales by 12.85% for the new product.
 - The project effectively addressed the challenge of launching new products with a cold-start, and it was successfully implemented and promoted at Yili's digital center and corporate headquarters.
 - A business case documenting the project's success is nearing completion and will be featured in the prestigious Asia Case Research Center at HKU.
 - Building on the successful trial in Shandong, the project received invitations to expand its scope to other products and regions, with the potential for nationwide implementation.
 - The project's impressive outcomes caught the attention of other senior executives at Yili, who expressed their interest in extending the project to provinces of higher strategic importance.
 - The project received the esteemed Best Innovation Award, a highly selective recognition within Yili. Furthermore, the knowledge and insights gained from the project equipped employees to achieve numerous other awards, which are listed on the final page.

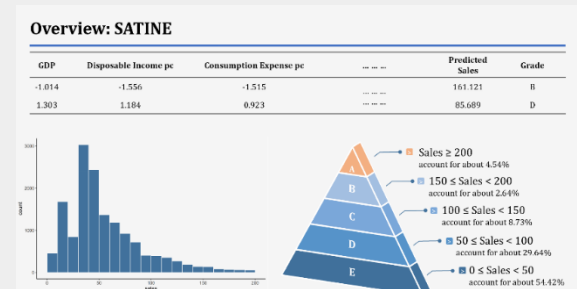
Impacts Achieved – Beneficiaries and Evidence



- Systematic and Automatic Data Gathering and Analysis: We design streamlined procedures for systematically and automatically collecting data, extracting relevant information, and analyzing their importance within the specific business context.
- **Example: Yili Project in Guangdong Province**
 - Implementation of an automated algorithm to impute missing data in store addresses and store information.
 - Creation of a semi-automated system to gather data on store surroundings, including schools, hospitals, shopping malls, bus stops, sports complexes, and more.
 - Development of an automated process to extract macro-economic data such as GDP, per capita income, population, expenditure patterns, and transportation information.
 - Introduction of a unique feature selection framework and a scoring mechanism to effectively rate stores.
 - Design of a new store classification rating system that is user-friendly for Yili sales personnel.
 - Implementation of a systematic approach to machine learning that enhances interpretability and automatically selects important features that are easily understandable.
 - These advancements in the Yili Project in Guangdong Province have enhanced data gathering and analysis capabilities, enabling Yili to make informed decisions and drive operational efficiency in a more streamlined and automated manner.

Market Segmentation: Satine Features

Class	Macro Features						Surrounding Features				
	Disposable Income pc	Transport Freight Volume	Total Retail Sales	City Passage Volume	Gender Ratio	Children %	District GDP pc	Office Building	School	Bus stop	Community
High GDP	Urban				✓	✓					✓
	Town ship							✓	✓		
	Rural							✓	✓		
Low GDP	Urban	✓	✓						✓		✓
	Town ship			✓	✓		✓		✓		
	Rural	✓							✓	✓	✓



Impacts Achieved – Beneficiaries and Evidence



- Corporate Engagement: We have actively undertaken a corporate engagement initiative aimed at promoting widespread awareness, emphasizing the necessity, and enhancing the understanding of the importance of adopting a big data approach among companies at all levels, including employees, managers, and senior executives.
- Example: China Resource MixC Mall Project (华润万象城项目) in Chongqing
 - Comprehensive approach addressing customer management, store management, and customer-store interaction problems.
 - Utilization and integration of multi-modal data, including consumption data, store information data, membership data, and others.
 - Empowering customers through various means: understanding customer behavior, acquiring new customers, retaining customers, predicting customer lifetime value, customer segmentation, predicting coupon usage, and store preference prediction.
 - Development of a highly innovative, intricate, and accurate recommendation system. The system's prediction accuracy has improved significantly, reaching 99% from the initial validation accuracy of 24%.
 - Empowering store operations through the establishment of store tags, generating store profiles, and identifying store-specific marketing opportunities.
- The successful implementation of these initiatives in the China Resource MixC Mall Project in Chongqing has showcased the immense potential and transformative impact of leveraging big data analytics.

Impacts Achieved – Beneficiaries and Evidence



Impact evidence: Other Awards

Thanks to the constant engagements of Prof. Shen, Dr. Wang, and Dr. Yang, our clients have achieved remarkable success, earning the following awards and recognitions:

- Best Innovation Award (创新达人奖): This prestigious and highly selective award within Yili recognizes their outstanding innovation and contributions.
- Talent Cup First Award & Diamond-level Data Analyst
- Data Scientist Project Excellence Prize
- Outstanding Study Case First Award
- Data Analysis Competition Outstanding Work Prize
- Digital Skill Certificate
- Excellent Lecturer Certificate

内蒙古伊利实业集团股份有限公司文件

伊利字〔2023〕9号

签发人：潘刚

关于表彰伊利集团 2022 年度“创新元”活动相关
奖项的决定

Best Innovation Award in Yili (创新达人奖)

