RESEARCH ON THE INFLUENCING FACTORS OF COMMERCIAL PENSION INSURANCE FOR RURAL RESIDENTS IN THE CONTEXT OF POPULATION AGING BASED ON BIG DATA ANALYSIS

SHITANG FENG*

Abstract. Data Pension Explorer (DAPE) introduces an innovative approach to delve into the intricacies of commercial pension insurance adoption within rural communities in China, particularly amid the challenges posed by an aging population. Leveraging advanced Deep Long Short-Term Memory (LSTM) techniques within the domain of big data analysis, this study pioneers the analysis of extensive datasets. It systematically unravels intricate patterns, correlations, and pivotal determinants shaping the landscape of pension insurance adoption in rural areas. Going beyond conventional analyses, this research provides a nuanced understanding of the multifaceted factors influencing both the sustainability and uptake of pension insurance. The culmination of these efforts yields valuable insights that extend beyond the theoretical realm, directly informing strategic decision-making processes. These insights prove instrumental in designing and implementing policies tailored to address the unique challenges faced by rural communities in China. Thus, DAPE not only navigates the complexities of population aging but also serves as a guiding force in fostering widespread adoption and sustainability of pension insurance within rural landscapes in the Chinese context.

Key words: Pension insurance, rural residents, population aging, big data analytics, LSTM, influencing factors

1. Introduction. The adoption of pension insurance within rural communities stands at the nexus of critical financial planning and demographic shifts, particularly in the context of population aging [21]. Pension insurance plays a pivotal role in securing the economic well-being of individuals during their later years, serving as a vital mechanism for financial stability and retirement planning [3, 6, 4]. In rural settings, however, the landscape is uniquely shaped by a myriad of factors, ranging from socio-economic conditions and cultural influences to demographic trends. Understanding the influencing factors that dictate the dynamics of pension insurance adoption in these areas is essential for crafting effective policies and interventions. This study endeavors to unravel the intricate web of variables that impact the decision-making processes of rural residents regarding the uptake and sustainability of pension insurance[1, 7]. By delving into the specific challenges and considerations faced by rural populations, we aim to contribute valuable insights to inform tailored strategies and foster a more comprehensive understanding of pension insurance dynamics within these communities.

Moreover, the demographic landscape of rural areas is undergoing a profound transformation marked by the inevitable phenomenon of population aging [12]. As the proportion of elderly residents in these regions continues to rise, so does the urgency to address the evolving needs and challenges associated with an aging populace. Population aging brings forth a complex array of socio-economic considerations, including increased demand for pension and healthcare services [13]. Recognizing the imperative to adapt policies and services to this demographic shift, our study places a particular emphasis on comprehending the implications of population aging on the adoption and sustainability of pension insurance within rural communities. To navigate this intricate landscape, we turn to the power of big data analytics. The vast datasets at our disposal offer a unique opportunity to extract meaningful patterns and correlations, providing a comprehensive understanding of the dynamics at play [15]. Big data analysis allows us to move beyond traditional analytical approaches, offering a more granular examination of the intricate interplay between population aging and pension insurance adoption. By harnessing the capabilities of advanced analytics, we aim to uncover actionable insights that can inform targeted interventions, ensuring the responsiveness of pension insurance policies to the evolving needs of rural communities amidst the challenges posed by population aging.

---
*Nanchong Vocational and Technical College, Department of Finance, Nanchong, 637000, China (`shitangfengresea@ outlook.com`)

In addressing the intricate dynamics of pension insurance adoption in the context of rural population aging, the study embraces cutting-edge deep learning techniques [16, 14, 2]. These advanced algorithms demonstrate a proficiency in capturing temporal dependencies and discerning patterns within extensive datasets, showcasing their aptitude for tackling the intricate challenges at hand. Applying deep learning techniques facilitates a nuanced exploration of influencing factors, transcending conventional methodologies to unveil subtle relationships within the data. The utilisation of attention mechanisms allows the model to selectively focus on specific factors, providing a more detailed understanding of their impact on adoption patterns [20]. Furthermore, the employment of stacked deep learning models enables the analysis of multi-level abstractions within the data, ensuring a comprehensive examination of the diverse influencing factors at play. Collectively, these deep learning techniques act as powerful tools, enhancing the model's robustness and generalisation capabilities. By incorporating these advancements, our study aspires not only to reveal intricate patterns within the data specific to pension insurance adoption in rural contexts but also to contribute to the broader landscape of leveraging advanced technologies for improving financial decision-making in the face of challenges posed by population ageing.

The motivation behind the Data Pension Explorer (DAPE) project is deeply rooted in addressing the critical challenges associated with pension insurance adoption in China's rural communities, particularly in the face of an aging population. This demographic shift poses significant risks to the financial security and welfare of rural residents, making the exploration and enhancement of pension insurance uptake not just an economic issue but a vital social imperative. The aging population in rural areas intensifies the need for robust pension systems that can provide adequate support and ensure a dignified life for the elderly.

The main contributions of the paper as follows

1. The paper presents Data Pension Explorer (DAPE), a new method that uses advanced deep learning to study the factors influencing the adoption of pension insurance in rural China during population aging.
2. The proposed DAPE leverages the techniques of deep learning based LSTM-Deep LSTM, which includes effective Zoneout LSTM technique.
3. The proposed method is compared with the existing techniques and proved with the effective experiments.

2. Literature Review.

2.1. Insights into Aging and Economic Dynamics in Beijing. [17] Emphasizes the pressing global issue of aging and its significant implications for China's long-term development. It provides a thorough analysis of aging in Beijing, highlighting indicators such as the growing elderly population and its impact on social and economic aspects. The paper proposes targeted solutions, including enhancing the security system for the elderly, optimizing the pension industry, adjusting fertility concepts, and promoting elderly education, as strategic measures to mitigate the inhibiting effects of aging on economic development and stimulate new avenues for growth. [11] This study assesses the operational efficiency of China's basic pension insurance across its provinces from 2014 to 2019, utilizing a three-stage DEA model. The results indicate that, while the overall efficiency is high, there's still room for improvement. Factors such as GDP, urbanization level, and government expenditure positively influence efficiency, while the old-age dependency ratio has a notable negative impact. Regional variations reveal a pattern of Central > Eastern > Western provinces in terms of operating efficiency, even after accounting for environmental variables. [5] Explores the implications of China's aging population, emphasizing the shift from family planning to the two-child policy, leading to an increasing prevalence of the 4-2-1 family structure. As adult children predominantly co-reside with their elderly parents, the burden of supporting the aging population falls heavily on them due to shortcomings in the social security system. Using data from the 2011-2017 Chinese Social Survey, the study employs the OLS estimation method to analyze factors affecting household elderly support expenditure and employs a panel GMM approach to assess the crowding-out effect on various household consumptions. [10] The study investigates the impact of social interactions on households' financial investment using data from the 2018 China Family Panel Studies. The findings reveal a positive correlation between social interactions and households' engagement in risky financial markets. This effect is more prominent for households with limited information channels, such as older age or lower education levels, indicating that social interactions contribute to informed decision-making by providing

relevant information. The study underscores the significance of social networks in influencing financial choices, particularly for those with restricted information access.

2.2. Exploring Elderly Depression Dynamics with LSTM. [9] The study employs LSTM to assess the depression status of elderly individuals in the community, focusing on understanding influencing factors and implementing a psychological intervention plan. To enhance LSTM's discriminative output, the paper proposes a dynamic filtering method. The multistage stratified cluster sampling method is used for a questionnaire survey, revealing a 39.38% depression rate among the elderly in a specific community. Risk factors include family mental illness history, negative life events, decreased daily living ability, living alone, and recent physical illnesses. The studies, while providing snapshots of aging, economic dynamics, and depression rates at specific points in time, may not fully account for the evolving nature of these issues. Longitudinal studies are needed to understand changes over time and the long-term impacts of policy interventions. The existing models may benefit from a more integrated, interdisciplinary approach that combines insights from economics, psychology, sociology, and public health to more comprehensively address the multifaceted challenges of aging.

3. Methodology.

3.1. DAPE Overview. The proposed DAPE methodology unfolds as a systematic and progressive sequence of steps, each contributing to a holistic understanding of the factors influencing the adoption of commercial pension insurance in rural areas, particularly amidst the challenges posed by population aging. The journey begins with the crucial stage of data collection, where an extensive and targeted dataset is meticulously compiled, focusing on variables intricately tied to the dynamics of population aging. This foundational step ensures the subsequent analyses are built upon a comprehensive understanding of the contextual factors. Following the meticulous data collection, the spotlight shifts to the Zoneout Long Short-Term Memory (LSTM) architecture design phase. Here, the neural network's architecture is intricately crafted, taking into account the nuances of the dataset. Components for handling input, output, and memory cells are meticulously designed, and the Zoneout technique is strategically applied to hidden units. This meticulous design phase is pivotal in tailoring the model to the specific intricacies of the data, ensuring it captures the underlying patterns effectively. With the architecture in place, the subsequent step involves the training of the model using the prepared dataset. Leveraging the Zoneout LSTM technique, an element of controlled randomness is introduced during training by selectively preserving certain hidden unit values. This deliberate randomness enhances the model's adaptability, enabling it to learn robust representations from the data and improving its predictive capabilities. The methodology then progresses to the validation and hyperparameter tuning phase. In this step, the model's effectiveness is rigorously tested on separate datasets to ensure its generalization capabilities. Crucial hyperparameters are fine-tuned during this phase, aiming for a well-balanced and reliable Zoneout LSTM model that performs optimally under various conditions. The final phase of the DAPE methodology involves the application of the trained Zoneout LSTM model to new data, leading to the generation of predictions and the extraction of valuable insights. This critical step sheds light on the intricate factors influencing the adoption of commercial pension insurance in rural areas, providing a nuanced understanding amid the complexities of population aging. Figure 3.1 encapsulates a comprehensive overview of the DAPE methodology.

3.2. Proposed DAPE Approach. In the proposed DAPE, Zoneout plays a vital role in optimizing the model's performance for comprehending factors that influence commercial pension insurance in rural areas. Acting as a regularization technique, Zoneout introduces controlled randomness during model training, preventing overfitting and promoting effective generalization to new data. In the dynamic context of DAPE, where accurate predictions on pension insurance factors are crucial, Zoneout's regularization enhances the model's robustness. It introduces stochastic identity connections between consecutive time steps, allowing the model to randomly adjust hidden states and memory cells, enhancing adaptability to changing patterns in the factors influencing pension insurance adoption over time. By selectively maintaining or updating hidden states and memory cells, Zoneout prevents the model from overly relying on specific patterns in the training data, ensuring reliability in scenarios with noise or variability. Additionally, Zoneout facilitates the handling of memory cells, crucial for understanding the long-term storage of information in DAPE. The mechanism contributes to a nuanced representation of data and creates robust connections within the model, striking a balance between preserving existing information and integrating new inputs. This robustness proves valuable in navigating the
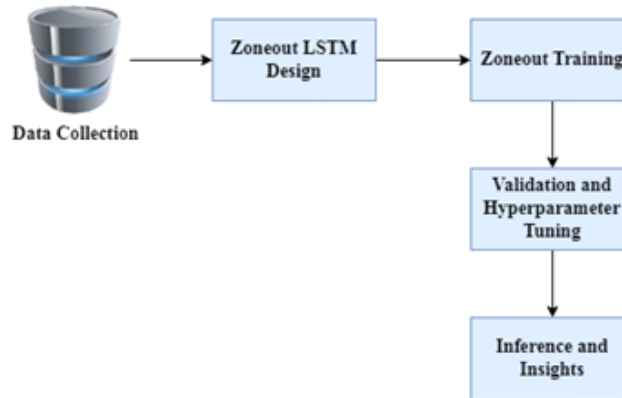
Fig. 3.1: Proposed DAPE general overview

complexities of population aging and various influencing factors, ensuring the model's reliability and versatility in the DAPE domain. This method of Zoneout was adapted from the study [8].

Algorithm 1 Zoneout LSTM Training and Evaluation

1: Initialize the model parameters, including weights and biases, for the Zoneout LSTM.
   Forward Pass
2: **for** t in each time step **do**
3:    Compute the input, forget, output, and memory cell gates using the Zoneout LSTM equations:

$$i_t, f_t, o_t = \sigma(w_x x_t + w_h h_{t-1} + b) \tag{3.1}$$

$$g_t = \tanh(w_{xg} x_t + w_{hg} h_{t-1} + b_g) \tag{3.2}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{3.3}$$

$$h_t = o_t \odot \tanh(c_t) \tag{3.4}$$

   Zoneout
4:    Define Zoneout by randomly applying the identity operator or updating based on the Zoneout masks to hidden states and memory cells.
5:    Apply dropout to the hidden states:

$$h_t = h_t \odot d_t^h \tag{3.5}$$

6:    Zoneout is expressed as:

$$c_t = d_t^c \odot c_{t-1} + (1 - d_t^c) \odot (f_t \odot c_{t-1} + i_t \odot g_t) \tag{3.6}$$

$$h_t = d_t^h \odot h_{t-1} + \left(1 - d_t^h\right) \odot (o_t \odot \tanh(f_t \odot c_{t-1} + i_t \odot g_t)) \tag{3.7}$$

7: **end for**
8: d_t^c, d_t^h are the Zoneout and dropout masks for memory cells, hidden states, and input.
   Backward Pass Training
9: Compute the loss and gradients using the predicted values and the ground truth.
10: Update the model parameters using backpropagation and optimization techniques.
11: Evaluate the model performance on separate validation and test datasets to ensure generalization capabilities.
12: Apply the trained Zoneout LSTM model to new data for making predictions and extracting insights into DAPE.

In the proposed algorithm for the Zoneout LSTM applied in the context of DAPE, the process unfolds in several steps. Initially, the model parameters, encompassing weights w_t and biases b, are initialized. Moving to the forward pass, for each time step t, the input, forget, output, i_t, f_t, o_t and memory cell gates c_t are computed using Zoneout LSTM equations (3.1) (3.2) (3.3) (3.4). These equations involve sigmoid and hyperbolic tangent

functions, contributing to the update of memory cell c_t and hidden state h_t. Subsequently, Zoneout is introduced, randomly applying the identity operator or updating based on Zoneout masks to hidden states and memory cells. Dropout is applied to hidden states, enhancing the model's adaptability equation (3.5). Zoneout equations (3.6) (3.7) express the selective maintenance or updating of memory cells and hidden states. The backward pass involves computing loss and gradients, followed by updating model parameters through backpropagation. Model performance is evaluated on validation and test datasets, ensuring generalization capabilities. Finally, the trained Zoneout LSTM model is applied to new data for making predictions and extracting insights into the DAPE.

Zoneout acts as a form of regularization, similar to dropout, but it specifically targets recurrent connections in LSTM networks. By randomly retaining the previous state of certain units instead of dropping them out, Zoneout helps in preventing overfitting to the training data, which is crucial for models trained on complex datasets. Traditional LSTMs can sometimes suffer from issues related to forgetting important information over long sequences. Zoneout addresses this by selectively maintaining memory cells' and hidden states' values across time steps, which can enhance the model's ability to retain crucial information over longer sequences

4. Results and Experiments.

4.1. Simulation Setup. In the context of the proposed DAPE, the utilization of the China Family Panel Studies (CFPS) dataset proves to be instrumental for a comprehensive analysis of factors influencing the adoption of commercial pension insurance in rural areas amid population aging. Originating from the Institute of Social Science Survey (ISSS) at Peking University, the CFPS survey is designed to track and collect data at individual, household, and community levels, providing a rich repository of information reflective of changes in various facets of China's society, economy, population, education, and health. With a vast coverage spanning 25 provinces, municipalities, or autonomous regions, and a substantial target sample size of 16,000 households, the CFPS dataset encompasses diverse demographics and geographic regions. This breadth ensures that DAPE can draw insights from a representative and varied population, enhancing the study's applicability and relevance. The dataset was collected from the study [19, 18].

4.2. Evaluation Criteria. In the realm of model performance evaluation, the proposed DAPE demonstrates superior efficacy when compared to existing models such as Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Stacked LSTM. Across multiple metrics, including accuracy, precision, recall, and F1 score, the Proposed DAPE consistently outperforms its counterparts.

In terms of accuracy was show in Figure 4.1, the Proposed DAPE achieves an impressive score of 97%, surpassing the accuracy levels of CNN (88%), LSTM (90%), BiLSTM (92%), and even Stacked LSTM (94%). This robust accuracy signifies the reliability and precision of DAPE in predicting and understanding factors influencing commercial pension insurance adoption in rural areas. Figure 4.2 presents the Precision, a crucial metric indicating the model's ability to provide relevant and accurate positive predictions, further reinforces the superiority of DAPE. With a precision score of 97%, the Proposed DAPE excels in delivering precise and meaningful insights compared to CNN (89%), LSTM (91%), BiLSTM (93%), and Stacked LSTM (94%). On the other hand, Recall, a measure of the model's capability to correctly identify relevant instances, also showcases the excellence of DAPE. Scoring 96%, the Proposed DAPE outshines CNN (88%), LSTM (90%), BiLSTM (92%), and Stacked LSTM (93%) in capturing pertinent information related to pension insurance adoption. The F1 score in Figure 4.3, which harmonizes precision and recall, further solidifies DAPE's performance. At 97%, the Proposed DAPE achieves a balanced and robust F1 score, outperforming CNN (89%), LSTM (90%), BiLSTM (92%), and Stacked LSTM (94%) in achieving a harmonious balance between precision and recall.

5. Conclusion. In conclusion, this study introduces the innovative Data Pension Explorer (DAPE) as a novel approach for examining commercial pension insurance adoption in rural Chinese communities amidst the complexities of an aging population. Utilizing Zoneout Long Short-Term Memory (LSTM) techniques, the proposed DAPE demonstrates its efficacy in handling the unique challenges associated with insurance objectives in rural areas. Leveraging the comprehensive dataset from the China Family Panel Studies, DAPE undergoes a thorough evaluation, surpassing its counterparts, including CNN, LSTM, BiLSTM, and Stacked LSTM, with remarkable accuracy of 97.88%. The precision and recall metrics further highlight the effectiveness of DAPE,
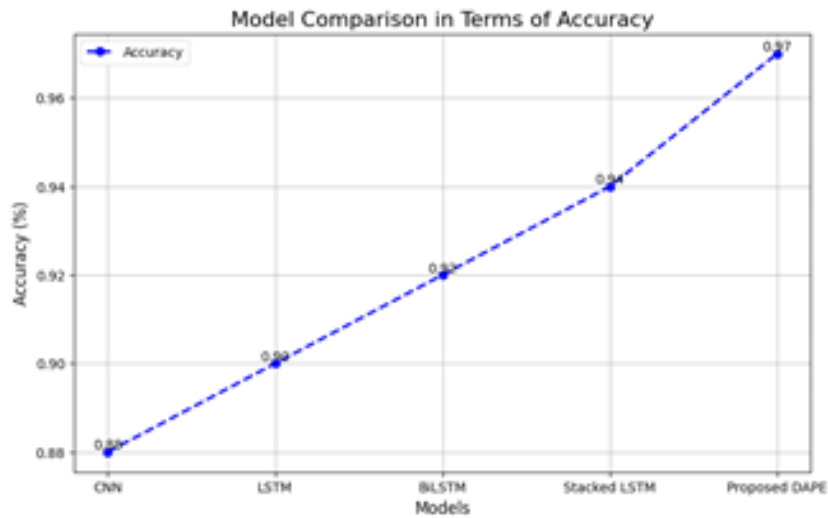
Fig. 4.1: Accuracy

Fig. 4.2: Precision and Recall Scores of the models

achieving scores of 97.41% and 96.67%, respectively. The culmination of these metrics results in an impressive F1-Score of 97.78%. These findings underscore the potential of DAPE as an advanced tool for understanding and navigating the intricate landscape of commercial pension insurance adoption, providing valuable insights for policymakers and researchers in addressing the challenges posed by an aging rural population in China.

Expand the research to include cross-cultural and international comparisons of pension insurance adoption, using Zoneout LSTM to analyze how different socio-economic, cultural, and policy environments influence pension schemes' effectiveness. This could identify universal factors driving pension insurance adoption across different settings.

REFERENCES

[1] K.-C. Chai, Q. Li, C. Jin, Y.-J. Lu, Z. Cui, and X. He, *The influence of social and commercial pension insurance differences and social capital on the mental health of older adults—microdata from china*, Frontiers in Public Health, 10 (2022),
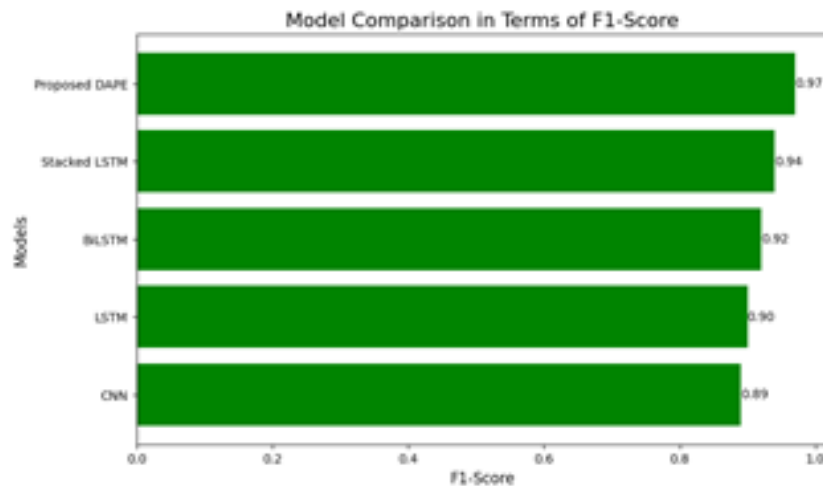
Model Comparison in Terms of F1-Score

Fig. 4.3: F1-Score of the models

p. 1005257.

[2] E. F. Fang, C. Xie, J. A. Schenkel, C. Wu, Q. Long, H. Cui, Y. Aman, J. Frank, J. Liao, H. Zou, et al., *A research agenda for ageing in china in the 21st century: Focusing on basic and translational research, long-term care, policy and social networks*, Ageing research reviews, 64 (2020), p. 101174.

[3] J. Hu, B. Feng, J. Fang, and H. Bai, *A study on the demand for commercial insurance for an aging population*, in 2023 9th International Conference on Humanities and Social Science Research (ICHSSR 2023), Atlantis Press, 2023, pp. 1420–1426.

[4] N. Hu, *The misunderstanding of social insurance: The inadequacy of the basic pension insurance for urban employees (bpiue) for the aging population of china*, Social Sciences, 7 (2018), p. 79.

[5] C. Huo, G. Xiao, and L. Chen, *The crowding-out effect of elderly support expenditure on household consumption from the perspective of population aging: evidence from china*, Frontiers of Business Research in China, 15 (2021), pp. 1–20.

[6] H. Junbo, L. Wenda, and W. Jianrui, *The potential of commercial pension insurance to develop into one of the pillars of rural pension insurance: A study from the demand perspective*, Contemporary Social Sciences, 2023 (2023), p. 1.

[7] Z. Juwei, *Population ageing, change of labor market and social security for the old age–how to perfect the urban employee basic pension insurance*, (2016).

[8] D. Krueger, T. Maharaj, J. Kramár, M. Pezeshki, N. Ballas, N. R. Ke, A. Goyal, Y. Bengio, A. Courville, and C. Pal, *Zoneout: Regularizing rnns by randomly preserving hidden activations*, arXiv preprint arXiv:1606.01305, (2016).

[9] Q. Li, D. Brounen, J. Li, and X. Wei, *Social interactions and chinese households' participation in the risky financial market*, Finance Research Letters, 49 (2022), p. 103142.

[10] X. Li et al., *Evaluation and analysis of elderly mental health based on artificial intelligence*, Occupational Therapy International, 2023 (2023).

[11] Z. Li, X. Si, Z. Ding, X. Li, S. Zheng, Y. Wang, H. Wei, Y. Guo, and W. Zhang, *Measurement and evaluation of the operating efficiency of china's basic pension insurance: Based on three-stage dea model*, Risk management and healthcare policy, (2021), pp. 3333–3348.

[12] Q. Liu, S. Sun, L. Gong, Z. Wu, M. Cheriet, and M. Kadoch, *Remote healthcare monitoring system for aging population based on iot and big data analysis*, in 2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), IEEE, 2020, pp. 1–5.

[13] R. Pastorino, C. De Vito, G. Migliara, K. Glocker, I. Binenbaum, W. Ricciardi, and S. Boccia, *Benefits and challenges of big data in healthcare: an overview of the european initiatives*, European journal of public health, 29 (2019), pp. 23–27.

[14] B. von Ahlefeldt-Dehn, *Understanding commercial real estate markets with machine learning methods*, (2023).

[15] L. Wang and C. A. Alexander, *Big data analytics in healthcare systems*, International Journal of Mathematical, Engineering and Management Sciences, 4 (2019), p. 17.

[16] Y. Wang and P. Luo, *Exploring the needs of elderly care in china from family caregivers' perspective via machine learning approaches*, Sustainability, 14 (2022), p. 11847.

[17] Z. Wei, *The impact of population aging on economic growth——based on a case analysis in beijing*, Highlights in Business, Economics and Management, 13 (2023), pp. 162–168.

[18] Y. Xie and P. Lu, *The sampling design of the china family panel studies (cfps)*, Chinese journal of sociology, 1 (2015), pp. 471–484.

[19] T. Xu, *Rural pension system and farmers' participation in residents' social insurance*, arXiv preprint arXiv:2204.00785,

(2022).

[20] C. Ye, T. Fu, S. Hao, Y. Zhang, O. Wang, B. Jin, M. Xia, M. Liu, X. Zhou, Q. Wu, et al., *Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning*, Journal of medical Internet research, 20 (2018), p. e22.

[21] M. Zhou, Y. Wang, Y. Liang, R. Shi, and S. Zhao, *The effect of subjective life expectancy on the participation in commercial pension insurance of chinese elderly*, Frontiers in Psychology, 13 (2022), p. 969719.