Embedding a Data-Driven Decision-Making Work Culture in a Social Housing Environment

Srinidhi Karthikeyan[a], Takao Maruyama[b] and Sankar Sivarajah[c]
Faculty of Management, Law & Social Sciences, University of Bradford, Richmond Rd, Bradford BD7 1DP, U.K.

Keywords: Social Housing, Artificial Intelligence, Machine Learning, Rent Arrears, Explainable AI.

Abstract: This paper explores the issue of delayed rent payments in social housing in the United Kingdom and its impact on tenants and housing providers. Our approach is to use machine learning algorithms to analyse payment patterns and identify tenants who may be at risk of falling behind on rent payments. By doing this, we aim to equip housing providers with the necessary tools to intervene early and maintain consistent tenancies. We have conducted research using machine learning models such as decision trees and random forests to address this issue. The paper emphasises the potential benefits of Explainable AI, which can help build trust in data-driven decision-making and AI among employees unfamiliar with AI and machine learning.

1 INTRODUCTION

Social housing is vital in the UK, providing accommodation to over 5 million households (Stone, 2003). Timely rent payments are crucial for the sustainability of social housing, as tenants' early exits can hinder social providers' objectives. The transition to Universal Credit (UC) has shifted the responsibility of rent payments to tenants, and only 8% of those in Direct Payment Demonstration Projects consistently paid their rent on time and in full (Hickman et al., 2017). By 2023, over 7 million households will receive Universal Credit (Hickman, 2021). Addressing this issue is essential to alleviate the financial pressure on social housing providers, who face an average cost of £10,000 per eviction (The Guinness Partnership and Tickell, 2015). Evictions can be traumatic for tenants and jeopardise their well-being (Bond et al., 2018). Effective rent collection strategies enable providers to function efficiently, increase housing supply, and fulfil their responsibilities to tenants. The development of artificial intelligence (AI) and machine learning (ML) has brought significant advancements in various fields, allowing for data analysis, pattern recognition, and autonomous decision-making. This has made AI and ML increasingly valuable in the current digital age. These complex algorithms enable machines to learn from data, recognise patterns, and make decisions without human intervention (Holzinger et al., 2017). Due to their immense potential, people from various application domains, including healthcare (Shaheen, 2021), finance (Cao, 2020), and marketing (Mariani et al., 2022), are increasingly interested in utilising these algorithms. As a result, AI and ML are now employed in different application domains, such as speech recognition systems (Amberkar et al., 2018) and self-driving cars (Rao and Frtunikj, 2018), to provide new solutions to previously unsolvable problems. However, AI in social housing is an area that needs more research.

The paper proposes using machine learning algorithms to predict rent arrears before they occur by analysing tenants' payment history. By identifying patterns, social housing landlords can take timely action to prevent the situation from escalating. The paper explains how a machine learning model was developed to accurately predict tenants' payment behaviour to alert Income officers of potential issues. Our team aims to prevent the situation from worsening, leading to further arrears or court cases and will offer ample support to help tenants manage their finances.

[a] https://orcid.org/0000-0001-6863-0760
[b] https://orcid.org/0000-0002-4830-7322
[c] https://orcid.org/0000-0002-6401-540X

Karthikeyan, S., Maruyama, T. and Sivarajah, S.
Embedding a Data-Driven Decision-Making Work Culture in a Social Housing Environment.
DOI: 10.5220/0012700700003690

In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024) - Volume 1*, pages 807-811
ISBN: 978-989-758-692-7; ISSN: 2184-4992

2 LITERATURE REVIEW

Despite its widespread adoption in various industries, the potential of Artificial Intelligence (AI) in social housing still needs to be explored. This review delves into the applications of AI in other domains, the causes of rent arrears in social housing, and potential areas for future AI research. Several successful applications of AI can be seen in fields such as loan default prediction (Lagasio et al., 2022; Neisen and Geraskin, 2022; Shaheen and Elfakharany, 2018; Turiel and Aste, 2019) and healthcare risk prediction (Ehlers et al., 2017; Shinde and Rajeswari, 2018; Ferdousi et al., 2021; Karthick et al., 2022; Wu et al., 2021; SK and P, 2017; Sawhney et al., 2023; Dutta et al., 2022). These examples illustrate the capability of AI to analyse data, recognise patterns, and make informed decisions. Explainable AI (XAI) has been used to minimise loan default risk (Egan, 2021) and interpret default forecasting models (Cascarino et al., 2022). By explaining AI's decisions, XAI can enhance user confidence and trust in the AI systems (Weitz et al., 2019; Druce et al., 2021).

Rent underpayment is a significant problem. A study by (Irvine et al., 2007) found a strong connection between 'capability' and rent arrears accrual and identified three types of money managers: 'ordered,' 'flexible,' and 'chaotic.' The study indicated that 'flexible' and 'chaotic' money managers were more likely to experience challenges than their 'ordered' counterparts.

Rent underpayments are affected by opportunity (Johnson and O'Halloran, 2017) and mental health (Bond et al., 2018). Claimants' arrears are caused by their financial situation, including irregular or insufficient income, the five-week wait for their initial UC payment, and administrative delays in receiving the benefit (Johnson and O'Halloran, 2017). It is essential to acknowledge that many reasons tenants fall into arrears are beyond their control. If landlords allow tenants to continue accruing arrears, it can lead to a loss of profit and additional expenses such as legal processes. Despite an extensive review of the literature, the authors have not been able to find any previous research studies that focus on the application of AI in social housing, particularly in assisting social landlords in reducing their tenants' rent arrears. This indicates a significant gap in the current research.

3 METHODOLOGY

3.1 Dataset

The original dataset contained around 15 million records, which consisted of tenants' everyday payment records. It included transaction details such as the transaction date, year, amount, account type, and other relevant parameters for 20,867 out of 23,033 active tenants in the weekly payment group. The dataset was then transformed into a more usable format to provide insights into tenant behaviour. As a result of the transformation, the new dataset now includes:

- Transaction sum of WK1-WK4 (weeks 1-4), WK5-WK8 (weeks 5-8), WK9-WK12 (weeks 9-12), WK13-WK16 (weeks 13-16), WK17-WK20 (weeks 17-20), WK21-WK24 (weeks 21-24), and WK25-WK28 (weeks 25-28).
- Yearly balance of 2021-2022,2022-2023 and 2023-2024.[1]
- Arrear score of 2021-2022, 2022-2023 and 2023-2024.
- Transaction score of 2021-2022,2022-2023 and 2023-2024.
- Transaction score of all three years.
- Arrear score of all three years.

With the weekly transaction sums, tracking tenants' payment patterns and identifying inconsistencies is easy. The past three years' yearly balance helps us paint a clearer financial picture of the tenant and their payment history. The transaction and arrear scores provide valuable insights into tenant behaviour, allowing us to identify potential risks and take necessary actions to mitigate them. Overall, this transformed dataset provides a more detailed and comprehensive understanding of tenant behaviour, which can help us make informed decisions and improve our services.

During the exploratory data analysis, some interesting findings were discovered. Figure 1 shows a graph illustrating the average number of days tenants take to fall into their first arrears since the beginning of their tenancy. It appears that the period has decreased and is now the lowest in 2022 compared to 2019. One possible reason for this is inflation in the UK.

Figure 2 displays the average number of days tenants receive their first Notice of Seeking Possession (NSP) notice since the beginning of their tenancy. As with

[1]The financial year is defined from April of one year to March of the next, so years are mentioned as ranges. For the future, this will include the current financial year and the past two financial years.

the previous graph, this figure has decreased, indicating that tenants are falling behind on their rent payments more quickly than in last years. As a result, Income officers are responding more promptly by issuing NSP notices earlier to prevent the accumulation of more significant debts.

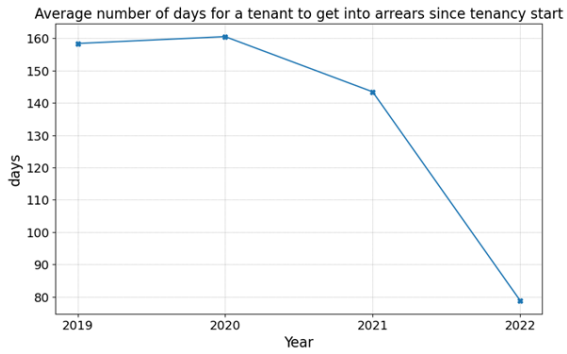

Figure 1: The Average number of days tenants take to get into their first arrears since the start of the tenancy.

Rent is collected on a weekly basis, and a cycle lasts for a maximum of 52 weeks. Each cycle has 13 periods, with each period consisting of four weeks. The payment patterns of tenants are represented using two critical features. The first feature is the transaction pattern score (ranges from 0 to 1), which indicates changes in the payment pattern of tenants over 13 periods. A higher score for the transaction pattern indicates frequent changes in payment patterns, which may be caused by financial instability or other factors. In contrast, a lower score suggests a consistent payment pattern, indicating that the tenant is financially stable and can meet payment obligations.

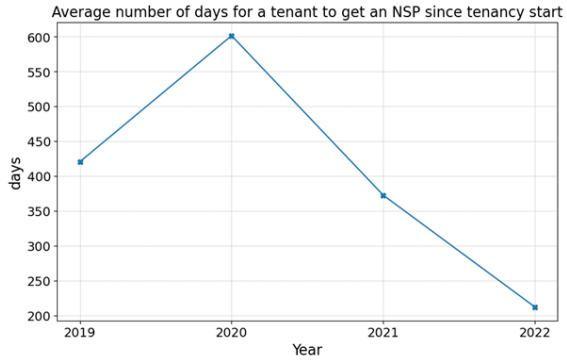

Figure 2: The Average number of days taken for tenants to get their first NSP notice since the start of tenancy.

The second feature used to represent payment patterns is the 13-period arrear scores (ranges from - ∞ to + ∞). The arrear score is calculated by summing the balances or arrears at the end of each 13-period period. A positive score indicates the tenant is in arrears, meaning they have failed to pay rent on time. In contrast, a negative score implies that the tenant has credit and is not in arrears, indicating that they have paid their rent on time or in advance. This information is crucial in determining the risk of tenants defaulting on rent payments. The target variable is whether the person will be in arrears for the next four weeks (25 to 28).

4 PRELIMINARY RESULTS

The problem at hand is a binary classification task, and two well-known machine learning algorithms, Decision Tree (Wu et al., 2008) and Random Forest (Breiman, 2001), were used to construct models. Our evaluation criteria included accuracy, precision, recall, and F1 scores. As shown in Table 1, the Random Forest model had greater accuracy than the Decision Tree model, making it the baseline model for the task.

Table 1: Evaluation results of the two models.

Metrics	DecisionTree	RandomForest
Accuracy	0.807646	0.856696
Precision	0.755488	0.839635
Recall	0.756410	0.786325
F1-Score	0.755949	0.812106

Upon further analysis of the Random Forest model's feature importance (as shown in Table 2), the top five features significantly impacting the classification were analysed. These features were Yearly balance of 2023-2024, Arrear score of 2023-2024, Transaction sum of WK17-WK20, Transaction sum of WK21-WK24, and Transaction sum of WK25-WK28. Among these five features, the Yearly balance of 2023-2024, the Arrear score of 2023-2024, and the Transaction sum of WK25-WK28 supported the findings provided by our domain experts.

However, we noticed that the Transaction sum of WK17-WK20, which corresponds to April, was a non-debit raise week due to UK holidays.

5 DISCUSSION AND FUTURE WORKS

Throughout the course of our project, we encountered a significant issue with data quality. We discovered that some of the data we were working with was less accurate, incomplete, and inconsistent, mak-

Table 2: The importance of features according to Random Forest model.

Features	Importance score
Transaction sum of WK1-WK4	0.0668
Transaction sum of WK5-WK8	0.0556
Transaction sum of WK9-WK12	0.0721
Transaction sum of WK13-WK16	0.0734
Transaction sum of WK17-WK20	0.0904
Transaction sum of WK21-WK24	0.079
Transaction sum of WK25-WK28	0.0763
Yearly balance of 2021-2022	0.03
Yearly balance of 2022-2023	0.0416
Yearly balance of 2023-2024	0.1229
Arrear score of 2021-2022	0.0291
Arrear score of 2022-2023	0.0376
Arrear score of 2023-2024	0.1181
Transaction score of 2021-2022	0.0162
Transaction score of 2022-2023	0.0177
Transaction score of 2023-2024	0.0176
Transaction score of all three years	0.0242
Arrear score of all three years	0.0313

ing it challenging to consider other important factors, such as tenant arrangements and the number of household members. High-quality data is essential for any business to thrive, while low-quality data can lead to missed opportunities and financial losses. A study by Duvier et al. (Duvier et al., 2018) revealed that data quality issues can arise due to various organisational, cultural, or computational challenges. Therefore, we understood the importance of addressing these challenges to establish an effective data quality program.

We also faced another challenge of building trust among employees new to working with AI. We understood that trust is crucial when working with AI, and we wanted our employees to feel confident about AI's decision-making process. Therefore, we decided to use explainable AI to help employees understand and trust AI's decisions.

Based on the preliminary findings, the authors aim to broaden the dataset by adding more transaction data from the past three years. Additionally, the plan is to incorporate data from previous tenancy transactions, as the current study only utilises current tenancy data. Furthermore, the authors intend to introduce more relevant features, such as tenancy duration, property demographics, and payment methods, as recommended by domain experts, to ensure a comprehensive and robust dataset. We are preparing a follow-up paper that will present a comprehensive analysis of our findings, detailing strategies employed to overcome the identified challenges and the impact of our work on the business.

6 CONCLUSIONS

Late rent payments in social housing in the UK have been a persistent problem for many years. This paper delves into the issue and explores potential solutions and challenges when implementing them. One possible solution that the paper proposes is using Explainable AI to build trust in data-driven decision-making processes, which can improve the accuracy of rent payment predictions and reduce the number of late payments. However, more research is needed to understand the full potential of AI in addressing the problem of late payments in social housing effectively. Despite the challenges of implementing AI in social housing, there is a growing belief that it could help tackle the issue more efficiently and fairly. It is crucial to fix the issue of late payments in social housing because it affects not only the housing providers but also the tenants, who may face eviction and financial difficulties. By addressing the problem, we can ensure that tenants can maintain stable and secure housing while also achieving positive outcomes for housing providers.

ACKNOWLEDGEMENTS

The authors thank Innovate UK and Incommunities for part funding this research.

REFERENCES

Amberkar, A., Awasarmol, P., Deshmukh, G., and Dave, P. (2018). Speech recognition using recurrent neural networks. In *2018 international conference on current trends towards converging technologies (ICCTCT)*, pages 1–4. IEEE.

Bond, N., Evans, K., and Holkar, M. (2018). Where the heart is-social housing, rent arrears and mental health.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Cao, L. (2020). Ai in finance: A review. *Available at SSRN 3647625*.

Cascarino, G., Moscatelli, M., and Parlapiano, F. (2022). Explainable artificial intelligence: interpreting default forecasting models based on machine learning. *Bank of Italy Occasional Paper*, (674).

Druce, J., Harradon, M., and Tittle, J. (2021). Explainable artificial intelligence (xai) for increasing user trust in deep reinforcement learning driven autonomous systems. *arXiv preprint arXiv:2106.03775*.

Dutta, A., Hasan, M. K., Ahmad, M., Awal, M. A., Islam, M. A., Masud, M., and Meshref, H. (2022). Early prediction of diabetes using an ensemble of machine

learning models. *International Journal of Environmental Research and Public Health*, 19(19):12378.

Duvier, C., Neagu, D., Oltean-Dumbrava, C., and Dickens, D. (2018). Data quality challenges in the uk social housing sector. *International Journal of information management*, 38(1):196–200.

Egan, C. (2021). *Improving Credit Default Prediction Using Explainable AI*. PhD thesis, Dublin, National College of Ireland.

Ehlers, A. P., Roy, S. B., Khor, S., Mandagani, P., Maria, M., Alfonso-Cristancho, R., and Flum, D. R. (2017). Improved risk prediction following surgery using machine learning algorithms. *eGEMs*, 5(2).

Ferdousi, R., Hossain, M. A., and El Saddik, A. (2021). Early-stage risk prediction of non-communicable disease using machine learning in health cps. *IEEE Access*, 9:96823–96837.

Hickman, P. (2021). Understanding social housing tenants' rent payment behaviour: evidence from great britain. *Housing Studies*, 36(2):235–257.

Hickman, P., Kemp, P. A., Reeve, K., and Wilson, I. (2017). The impact of the direct payment of housing benefit: evidence from great britain. *Housing Studies*, 32(8):1105–1126.

Holzinger, A., Biemann, C., Pattichis, C. S., and Kell, D. B. (2017). What do we need to build explainable ai systems for the medical domain? *arXiv preprint arXiv:1712.09923*.

Irvine, A., Kemp, P. A., and Nice, K. (2007). *Direct Payment of Housing Benefit: What Do Claimants Think?* Chartered Institute of Housing.

Johnson, S. and O'Halloran, A. (2017). Nudging your way to reduced rent arrears.

Karthick, K., Aruna, S., Samikannu, R., Kuppusamy, R., Teekaraman, Y., Thelkar, A. R., et al. (2022). Implementation of a heart disease risk prediction model using machine learning. *Computational and Mathematical Methods in Medicine*, 2022.

Lagasio, V., Pampurini, F., Pezzola, A., and Quaranta, A. G. (2022). Assessing bank default determinants via machine learning. *Information Sciences*, 618:87–97.

Mariani, M. M., Perez-Vega, R., and Wirtz, J. (2022). Ai in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychology & Marketing*, 39(4):755–776.

Neisen, M. and Geraskin, P. (2022). Improved credit default prediction using machine learning and its impact on risk-weighted assets of banks. *Journal of AI, Robotics & Workplace Automation*, 2(1):49–62.

Rao, Q. and Frtunikj, J. (2018). Deep learning for self-driving cars: Chances and challenges. In *Proceedings of the 1st international workshop on software engineering for AI in autonomous systems*, pages 35–38.

Sawhney, R., Malik, A., Sharma, S., and Narayan, V. (2023). A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease. *Decision Analytics Journal*, 6:100169.

Shaheen, M. Y. (2021). Applications of artificial intelligence (ai) in healthcare: A review. *ScienceOpen Preprints*.

Shaheen, S. K. and Elfakharany, E. (2018). Predictive analytics for loan default in banking sector using machine learning techniques. In *2018 28th International Conference on Computer Theory and Applications (ICCTA)*, pages 66–71. IEEE.

Shinde, S. A. and Rajeswari, P. R. (2018). Intelligent health risk prediction systems using machine learning: a review. *International Journal of Engineering & Technology*, 7(3):1019–1023.

SK, S. and P, A. (2017). A machine learning ensemble classifier for early prediction of diabetic retinopathy. *Journal of Medical Systems*, 41:1–12.

Stone, M. E. (2003). *Social housing in the UK and US: Evolution, issues and prospects*. Goldsmiths College, Centre for Urban Community Research London.

The Guinness Partnership, C. o. A. i. t. S. E. and Tickell, C. (2015). Tenancy sustainment: Summary report may 2015.

Turiel, J. D. and Aste, T. (2019). P2p loan acceptance and default prediction with artificial intelligence. *arXiv preprint arXiv:1907.01800*.

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., and André, E. (2019). " do you trust me?" increasing user-trust by integrating virtual agents in explainable ai interaction design. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 7–9.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., et al. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, 14:1–37.

Wu, Y.-T., Zhang, C.-J., Mol, B. W., Kawai, A., Li, C., Chen, L., Wang, Y., Sheng, J.-Z., Fan, J.-X., Shi, Y., et al. (2021). Early prediction of gestational diabetes mellitus in the chinese population via advanced machine learning. *The Journal of Clinical Endocrinology & Metabolism*, 106(3):e1191–e1205.