

Unstructured Data Modelling for Mortgage Risk Prediction using News-media and LiDAR Images

Abstract

Predicting mortgage defaults is a key aspect of financial and risk management, enabling banks and other lenders to make smarter decisions and contribute to housing market stability. It is also important for the general public, particularly those who do not have access to structured financial data like credit scores, to understand the risks involved with mortgage defaults.

Traditionally, default predictions have primarily depended on structured data, critical for evaluating credit risk and making informed financial choices. However, this dependence on structured data can exclude small businesses, independent financial advisors, investors, and the general public who might not have the necessary resources or tools to assess credit risk effectively.

This study investigates the potential of publicly accessible unstructured data to reveal hidden insights that could be valuable for predicting defaults. This approach involves analyzing news articles across different categories and utilizing LiDAR images to assess demographic and geographical layouts. The methodology incorporates a multimodal and fusion analysis strategy, using text analysis models like RoBERTa, BERT, and ALBERT, alongside image analysis models such as VGG19, EfficientNet, and ResNet. The superior model combination from these categories integrates another layer for analyzing sentiment scores from the texts. Our final model—a composite of MLP (for sentiment analysis), VGG19 (for image processing), and BERT-CNN (for text processing)—surpasses other combinations in performance. Moreover, we benchmark this against an LLM-based model employing a "Browse-and-Concentrate" technique, which improves multimodal understanding by first surveying inputs to extract key insights and then focusing in-depth on these details, integrating LLMs with pre-trained vision models to manage multimodal inputs.

Authors & Affiliations

Mahsa Tavakoli¹, Rohitash Chandra², Cristian Bravo¹

¹University of Western Ontario, London, Canada. ²UNSW, Sydney, Australia