Cecily Wang/Muñoz-Pastrana

DS537

Professor Nolte

December 15, 2023

Assessing Land Value Impacts and Climate Resilience in Massachusetts Counties: A Neural Network Approach (2040-2049)

Policy Area

The goal of this analysis is to explore how urban planners and environmental policymakers can use this insight to make informed decisions about where to invest in infrastructure improvements and climate resilience initiatives. With climate change affecting coastal cities such as Boston with rising sea levels, the community must be warned and protected from such hazards that are a side effect of our changing climate (Rosen 2021). By concentrating on areas and factors identified as high-risk, the state can prioritize funding and resources for measures like coastal barriers, building elevation, and reforestation to mitigate damage from sea level rise and storms (City of Boston 2016). Being able to model the climate change elements impacting land value estimates can also empower individuals to make informed decisions about where they choose to live in the future (Palter 2022). To convince policymakers and future homeowners of these changing conditions and their potential impact on future residences, this study examines the ability of neural networks to predict land value estimates and assesses how climate factors affect them.

Abstract

This study employs a neural network estimator—specifically sci-kit learn's MLP Regressor which can handle non-linear models—and feature engineering techniques using principal component analysis (PCA) for a broad analysis of land value impacts in each of Massachusetts's counties (Sci-Kit Learn Developers 2023). I focused on integrating climate risk data from the Center for International Earth Science Information Network (CIESIN) into the evaluation. This research examines the predictive accuracy of neural networks in combination with PCA and investigates the consequences of alternative sets of predictors facilitated by advanced feature engineering, incorporating key climate risk metrics. My goal is to showcase the potential of neural networks as a tool in real estate valuation as well as identify counties that may need additional infrastructure support from the risks of climate change (Coomes 2023).

Introduction

The valuation of land is a vital factor in urban planning, real estate, and policymaking. Traditional methods like regression forests have been widely used to determine land valuations, but emerging technologies such as neural networks offer a promising alternative (Vincenza 2023). This study looks at the performance of a supervised neural network in predicting land values across Massachusetts counties. Additionally, it explores the consequences of employing alternative predictors through advanced feature engineering, with a focus on the strategic incorporation of key climate risk metrics. To provide context, defining "climate risk data" for a clear understanding is crucial. The climate risk data that we will be using from CIESIN encompasses three key metrics: the potential impact of climate and weather **hazards**, socioeconomic **vulnerability** in relation to climate risks, and **exposure** to the risks that come with climate change. This study addresses a research gap by not only evaluating the capabilities of neural networks but also incorporating new climate risk projection data into the predictive land value modeling framework (Nolte Lab 3 2023).

The research aligns with global initiatives emphasizing the link between environmental actions and increased land values. Drawing on international examples, such as the analysis conducted in Zhengzhou, China, which directly correlates environmental improvements with escalating land value, reinforces the importance of environmental considerations in real estate valuation and preventative infrastructure that could strongly hold against the changing climate (Lord, Alexander, Dong, 2022).

Land value capture, another tool used globally to fund public initiatives, also gains significance as trillions are allocated toward environmental transition (GFDRR 2023). This study contributes to the global discourse by exploring the transformative potential of neural networks in predicting land values via climate risks, aligning with the argument that environmental benefits and health can enhance the economy (Coomes 2023). Notably, it considers the impact of climate change on land values, linking environmental actions with economic incentives.

In summary, this research contributes to the global conversation about aligning financial incentives with environmental action. Incorporating neural networks and emphasizing climate risk data, the study aims to showcase their transformative potential in predicting land values and understanding the future impact of climate change.

1. Objective

In this analysis, I examine the methodology's usefulness (neural network estimators) in combining future climate data with land cover data to minimize the mean squared error (MSE) between predicted and actual land values using sci-kit's MLPRegressor, RandomizedSearchCV, and PCA. MSE is calculated as the average of the squared differences between predicted and actual values.

1.1 Decision Variables

In order to come to a conclusion that looks at the hyperparameters of the neural network estimator that could best fit this merged dataset to visualize the climate features that will impact Massachusetts's land values the most in the future, I will deepen my understanding of the way that the climate risk data affects the model's ability to estimate land value. I will investigate the influence of different hyperparameters on the performance of neural networks to enhance model robustness by adjusting the hyperparameters such as the number of hidden layers, learning rate, and using a randomized search technique instead of a grid search to accommodate the complexity of the data. And for the matter of feature engineering, I have integrated principal component analysis (PCA) into the process to identify the most informative climate risk features for land value prediction.

2. Methodology

2.1 Data Sources

The data used in this analysis is a merge of "PLACES" (Nolte 2023), a dataset that includes land cover and building-related variables in Massachusetts, and "US Climate Risk Projections by County" (CIESIN 2023), a dataset that contains a projection for 2040-2049 risk for the U.S. at the county level with a climate risk index that combines hazards, exposures, and vulnerabilities. The hazards are characterized as weather and climate and are grouped as the frequency of heat waves, cold spells, drought, and heavy precipitation events along with anomalies of temperature and precipitation. Exposure is characterized by projections of population, infrastructure, and built surfaces prone to multiple hazards including sea level rise and storm surges. Vulnerability is characterized by projections of demographic groups most sensitive to climate hazards. Analysis was performed at the county level for each of the datasets with the GEOID codes of the CIESIN data set to 25 to narrow the data down to only Massachusetts (US Census Bureau 2021).

2.2 Setting up the Data

Nine variables from the CIESN data were chosen for the neural network based on the availability of information and relevance towards climate change in Massachusetts and possible discrete information that could help with land valuation (Flint 2022). These variables were used in conjunction with the land value estimation variables from the PLACES data. The variables chosen from the CIESN data are the average weather variations in the 2040s such as temperature (precipitation, extreme cold, and heat waves), estimates of housing density based on "population values used to drive housing density growth as depicted by the Spatially Explicit Regional Growth Model (SERGoM v3)", houses that would be exposed to 0.6 meters of sea level rise or storm surge above the current level, hazard, vulnerability, and exposure (CIESIN 2023). The merge of the climate risk data with the existing land value dataset was done and I further ensured consistency in spatial and temporal dimensions by changing the coordinate reference system of the CIESIN and PLACES to match that of the Massachusetts county parcel (MassGIS 2022) and assigning centroid geometries to the merged dataset that filtered based on only Massachusetts GEOIDs (2 5 - - -).

---

*It is important to note that the geometry of each land parcel in the merged data corresponds to the "coordinate center" of their respective counties, which means that the points heavily overlap on a spatial map. To see more of the points, I have implemented a random jitter of standard deviation 100 to the geometries of each land parcel. Although this gets rid of some of the advantages of a spatial map, this jitter is just to show how *each county* (not specific land parcel) in Massachusetts is affected by climate risk and how land value estimates capture this added data once PCA is performed.

So as we can see in Figure 1 below, the first thing that comes to the viewers' mind should be how counties near the coast are experiencing higher land value estimates which *may* be due to enhanced climate risk in these areas as we will see later in Figure 4. *

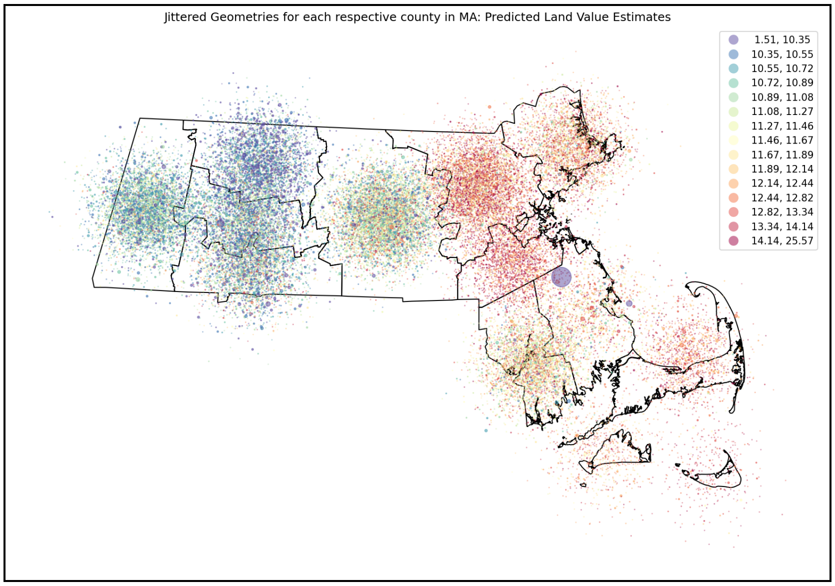

Figure 1

2.3 Neural Network Estimate

Starting from the variables reported in the merged database, the neural network has been trained to relate environmental and other considered land attributes to the land value prices. The considered neural network trains using backpropagation and uses square error as the loss function (Singh 2019). Here, I have set the initial neural network parameters to be composed of three hidden layers with eight neurons each. The input data to train consists of the nine climate risk variables selected and land attributes of the collected properties from “PLACES” (Nolte 2023). The land value estimate represents the output. The dataset (with 4424 records) has been divided into two different subsets: a training set (80% of total records) and a test set (20%). These subsets have been built to include in each of them all the price ranges of the considered land, which were then evaluated through mean-squared error metrics for the training and testing sets.

Here are more parameter specifications behind the initial integration of the neural network estimator without a tuning algorithm involved in the hyperparameter selection process:

- hidden_layer_sizes=(8, 8, 8): Three hidden layers with eight neurons each.
- activation='relu': Rectified Linear Unit (ReLU) activation function.
- solver='adam': Stochastic weight optimization method was used because we have many parameters involved (Kingma and Ba 2017).
- max_iter=500: Maximum number of iterations for optimization.

2.4 Feature Engineering

To incorporate climate variables into the land valuation model effectively, I used feature engineering, including Principal Component Analysis (PCA). PCA was utilized to condense climate risk variables, making the integration of climate data into the neural network model more efficient. PCA identifies linear combinations of climate variables with the highest variance, creating new variables known as principal components. These principal components, capturing the majority of variance, were then used as predictors in the neural network model. This approach reduces multicollinearity risk and highlights the climate variables with the most substantial impact on land values.

Furthermore, the incorporation of the nine selected variables from the CIESN data involved a manual feature engineering process, significantly enhancing the model's capacity to assess future land values. The effect achieved by merging land cover variables and prices with climate risk factors introduced a dynamic dimension to the model's predictive capabilities. Consequently, this analysis provided valuable insights into the climate risks that policymakers and urban planners should prioritize. By leveraging this feature selection technique, we gained a deeper understanding of how the integration of CIESN data into the PLACES dataset influenced the model's ability to estimate land values accurately.

3. Results

3.1 Neural Network Estimator

I aimed to decrease the MSE of calculating land value estimates so the next step was to use a randomized search to find potential parameters that could improve the MSE of the model. Although there are more benefits to using a grid search instead of a randomized search, the reason that I decided to go with a random search algorithm is due to time constraints and how big the dataset is. So although I might be getting a faster performing run by using randomized search, the results are less fine-tuned to finding the actual optimal parameters as compared to what a grid search could do.

After running the randomized search, the results showed that the best parameters for the model to return the smallest MSE were:

Best parameters found:

{'solver': 'adam', 'learning_rate': 'constant', 'hidden_layer_sizes': (50, 50), 'alpha': 0.0001, 'activation': 'tanh'}

Best MSE from random search: 1.5164442953604933

Test Mean Squared Error for the tuned model: 1.5111221382483384

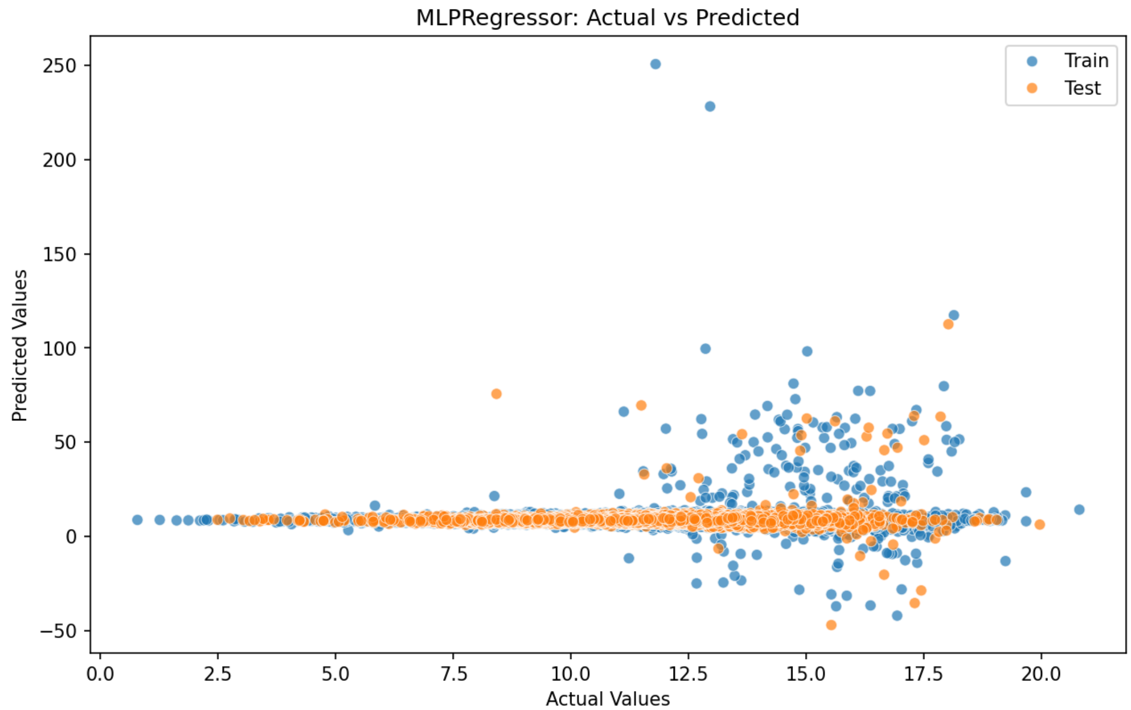

Figure 2

It was interesting to see how a two-dimensional neural network worked best for this data as compared to the initial setup that I implemented with the three dimensions of eight neurons in each. After running this multiple times, the randomized search consistently returned these values. And we can see that the MSE is significantly lower than that of the original neural network parameters output.

Practical Implications:

Resource Allocation: By having a more accurate understanding of the factors influencing land values, policymakers can allocate resources more effectively. They can target areas with higher climate risks for infrastructure improvements, prioritize climate resilience initiatives, and optimize land use planning to reduce vulnerabilities.

Resilience Planning: The identification of climate risk as a significant predictor of land values highlights the importance of resilience planning. Urban planners can use this information to create zoning regulations, building codes, and development guidelines that account for climate-related risks, such as sea-level rise and extreme weather events.

3.2 Feature Engineering Impact

In the subsequent phase of my analysis, where I explored the integration of climate risk data with land cover information, I employed PCA to discern the key features influencing land values. Following the scaling of the data, the examination revealed that various attributes from the climate risk CIESIN dataset played a crucial role in assessing land value estimates. Notably, the 'Expos(E)' feature in the CIESIN data, encompassing projections of population, infrastructure, and built surfaces vulnerable to multiple hazards like sea level rise and storm surges, exhibited a substantial feature importance of approximately 0.30 in the model. Another influential feature identified was 'HouseSLR,' representing houses that would be exposed to a 0.6-meter rise in sea level or storm surge above the current level in the future (Figure 4).

Upon implementing PCA with MLPRegressor, the Mean Squared Error (MSE) for the training data was 1.122. However, the testing data yielded a considerably different MSE of 1.610, which closely aligned with the initial parameters for the MLPRegressor (Figure 3). This discrepancy implies a potential issue of overfitting, suggesting that the regressor closely tailored itself to the training data, leading to a suboptimal performance on the testing data.

Practical Implications:

Key Climate Risk Factors: Policymakers and urban planners can use the knowledge of key climate risk factors, such as population exposure and vulnerability to hazards to develop targeted interventions. This may involve building regulations that require higher elevations for properties in vulnerable areas or investments in stormwater management infrastructure.

Policy Formulation: The insights gained from feature engineering can inform policy formulation aimed at mitigating the impact of climate change on land values. For example, policies could incentivize green infrastructure projects or encourage sustainable land use practices in areas prone to climate risks.

3.2.1 Evaluation of Neural Network Performance with Different Predictors

The evaluation of neural network performance with distinct sets of predictors underscored the dynamic nature of the model. While initial parameters provided valuable insights, the randomized search revealed a more optimized set of hyperparameters, significantly reducing the MSE. This highlights the importance of hyperparameter tuning in enhancing the model's predictive accuracy.

	Initial MLPRegressor	Random Search with MLP Regressor	PCA with Random Search
MSE Train	1.847539952	1.573648812	1.122804315
MSE Test	1.876893331	1.532269183	1.610086131
iterations	500	75	500
Parameters:			
Solver	adam	adam	adam
Learning rate		constant	
Hidden_layer	(8,8,8)	(50,50)	(50,50)
Alpha		0.0001	
Activation	relu	tanh	tanh

Figure 3

3.2.2 Identifying Key Features Influencing Land Values (PCA)

The identification of key features through PCA demonstrated that certain climate risk variables, particularly **'Expos(E)'-** composite “index covering projections of population, infrastructure, and built surfaces prone to multiple hazards including sea level rise and storm surges”– and **'HouseSLR,'-** “houses that would be exposed to 0.6 meters of sea level rise or storm surge above the current level”–exerted a substantial impact on land value estimates (CIESIN 2023). Understanding the significance of these features allows for targeted interventions and policy decisions to address infrastructural vulnerabilities associated with climate change, such as rising sea levels and increased storm surges.

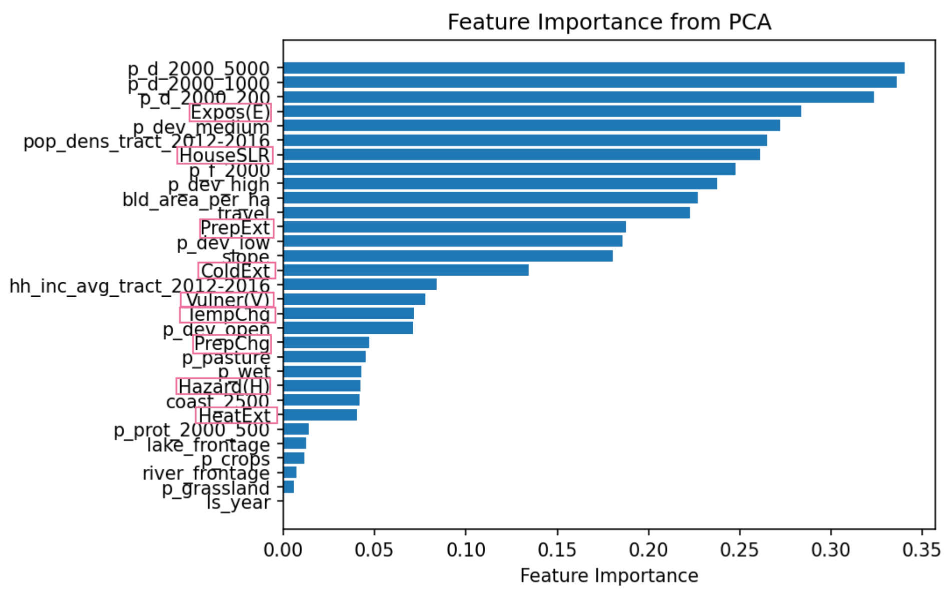

Figure 4 *(Pink Outline refers to the nine variables selected from the CIESIN climate risk dataset)*

3.2.3 Implications of Feature Engineering on Model Interpretability and Policy Relevance:

Feature engineering through PCA not only improved the model's ability to capture essential information but also raised questions about its generalizability. The observed overfitting phenomenon emphasizes the delicate balance required in feature selection, cautioning against overly complex models that may struggle to adapt to new data. This was experienced when the neural network randomized search was implemented alongside PCA (Figure 3), where we see how much worse the testing MSE was in comparison to that of the trained dataset. Despite these challenges, the insights gained from feature engineering contribute to a more nuanced understanding of the factors influencing land values, aiding policymakers in formulating informed decisions.

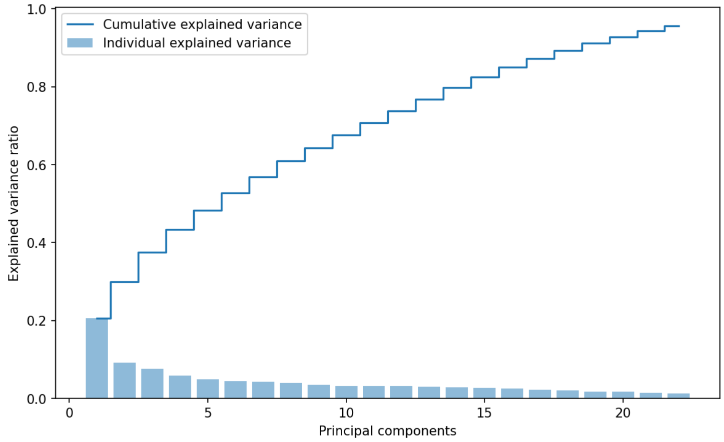

Figure 5

The implications extend beyond model interpretability, impacting policy relevance by highlighting specific climate risk factors that can guide targeted interventions. This underscores the importance of incorporating domain knowledge and interpretable model features when designing policies aimed at mitigating the impact of climate change on land values.

4. Discussion

Neural networks excel in capturing complex (multi-dimensional), non-linear relationships due to their structure, which consists of multiple interconnected layers of neurons. This configuration allows them to model intricate patterns that simpler regression models might miss. By stacking multiple layers and using non-linear activation functions, neural networks can approximate a wide range of functions, which is particularly useful for predicting land values that are influenced by a vast array of factors, such as geographic location, proximity to amenities, and environmental characteristics.

4.1 Limitations

However, one of the trade-offs when using neural networks is lack of interpretability. Unlike decision trees or linear regression where the influence of input features is more transparent, neural networks act as a "black box" where it becomes challenging to understand how the input variables are being transformed within the hidden layers to influence the output (Sci-kit Learn 2023). This can make it difficult to explain predictions to stakeholders who are not familiar with machine learning complexities.

In terms of generalizability, neural networks with a sufficient amount of relevant data and proper regularization can generalize well to unseen data. However, they can be prone to overfitting, especially if their architecture is too complex relative to the amount of training data. This is where model tuning, cross-validation, and selection of the right hyperparameters come into play to balance model complexity and generalization. This is something that was experienced when feature engineering using PCA in conjunction with randomized search hyper-tuning.

In the end our study focuses on climate risk projections for the 2040-2049 period. Climate change is a dynamic process, and longer-term projections beyond this period may reveal different patterns. Moreover, the study does not consider potential policy interventions that may impact land values and climate resilience.

5. Conclusion

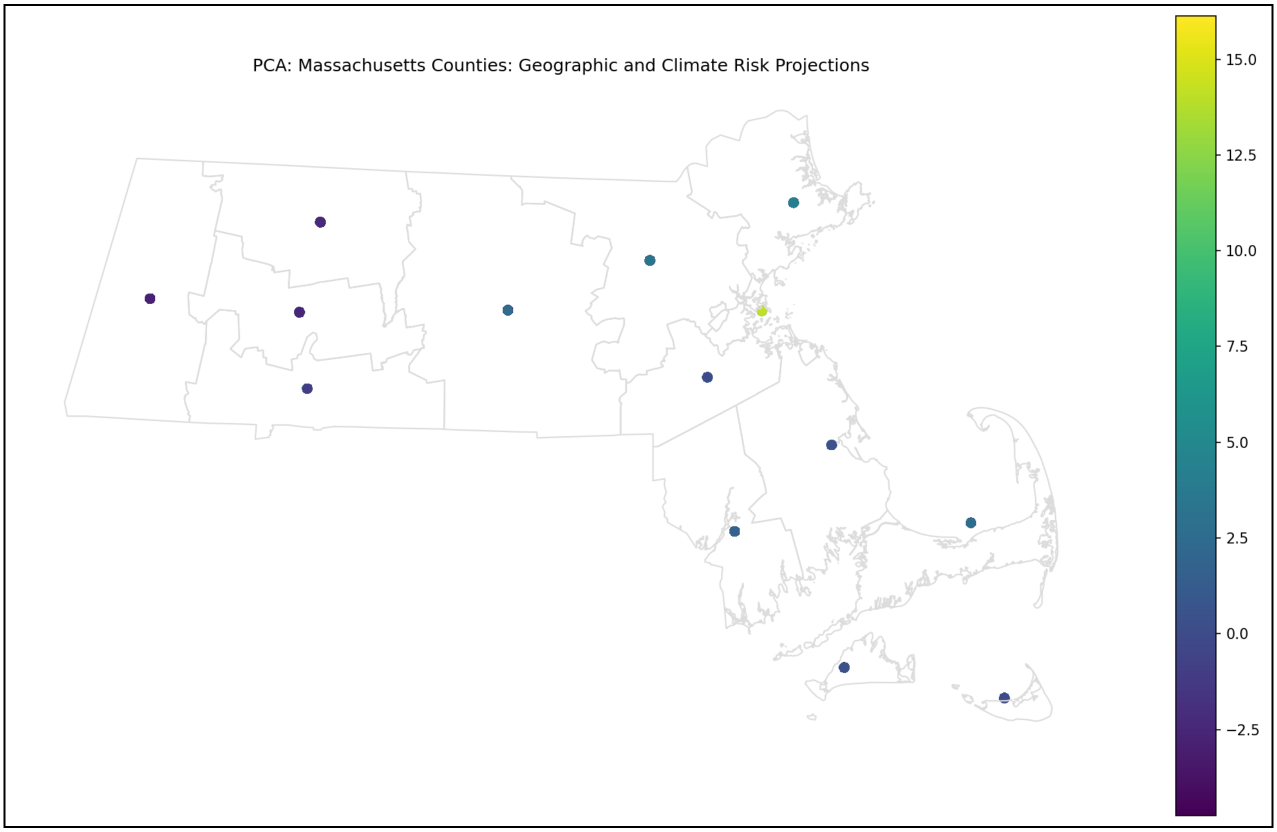

Figure 6

Rising sea levels in Massachusetts have led to a growing concern regarding sustainable infrastructure against climate change and the concurrent rise in land value (Rosen 2021). The decision problem centers on finding strategies to address the sustainability gap, ensuring that residents have access to safe housing/infrastructure systems without significant financial burdens in the present and future. Using a neural network analysis can reveal the most influential factors impacting land values, including but not limited to location-based features, economic indicators, and environmental risk factors. Additionally, feature engineering can help refine predictors, highlighting specific aspects of the environment, infrastructure, or socioeconomic factors that significantly contribute to land value fluctuations.

While this study has shed light on the effectiveness of neural networks and feature engineering in predicting land values with the use of climate risk data, there are avenues for future research and improvements. Firstly, further exploration of hyperparameter tuning methods, including grid search, could be undertaken to fine-tune the neural network and enhance its predictive performance. Additionally, investigating the robustness of the model across different, or more regions and time periods could provide valuable insights into the generalizability of the findings, something that my county-level data could not provide. Also incorporating datasets such as demographic information and economic indicators, may contribute to a more comprehensive understanding of the factors influencing land values.

Furthermore, this analysis could support efforts to improve the allocation of resources for emergency services and preparedness. Knowing which areas are most susceptible to climate risks helps target the deployment of services and training for disaster response, potentially reducing the time and cost of recovery after an event.

In conclusion, this research contributes to the evolving field of predictive modeling in land value and urban planning, emphasizing the significance of integrating climate risk data for more informed decision-making (International Renewable Energy Agency 2022). The findings presented here can serve as a foundation for policymakers, urban planners, and real estate professionals to develop strategies that address the dynamic challenges posed by climate change, particularly sea levels in Massachusetts, and escalating land values. The state could reassess land use and development policies to discourage high-density construction in the most vulnerable zones. By controlling development in these areas, the state could minimize future financial impacts related to climate risks and potentially maintain more stable land values over time.

Works Cited

Chiarazzo, Vincenza. "A Neural Network based Model for Real Estate Price Estimation Considering Environmental Quality of Property Location." *Science Direct*, 5 November 2023, https://www.sciencedirect.com/science/article/pii/S2352146514002300. Accessed 13 December 2023.

Coomes, Oliver T., et al. "Geospatial Land Price Data: A Public Good for Global Change Science and Policy." *Oxford Academic*, 5 November 2023, https://academic.oup.com/bioscience/article/68/7/481/5025644. Accessed 13 December 2023.

"Downloads » U.S. Climate Risk Projections by County, v1: Climate Risk and Vulnerability." *Socioeconomic Data and Applications Center*, 13 March 2023, https://sedac.ciesin.columbia.edu/data/set/crv-us-climate-risk-proj-county-2040-2049/data-download. Accessed 13 December 2023.

Flint, Anthony. "Return on Investment: Research Links Climate Action with Land and Property Value Increases." *Lincoln Institute of Land Policy*, 21 July 2022, https://www.lincolninst.edu/publications/articles/2022-07-research-links-climate-action-land-property-value-increases. Accessed 13 December 2023.

guide, step. "Understanding Geographic Identifiers (GEOIDs)." *Census Bureau*, 8 October 2021, https://www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html. Accessed 13 December 2023.

Kingma, Diederik P., and Jimmy Ba. "[1412.6980] Adam: A Method for Stochastic Optimization." *arXiv*, 22 December 2014, https://arxiv.org/abs/1412.6980. Accessed 13 December 2023.

Kozak, Daniel, et al. "Implementación de Infraestructura Azul y Verde (IAV) a través de mecanismos de captación de plusvalía en la Región Metropolitana de Buenos Aires." *Lincoln Institute of Land Policy*, February 2022, https://www.lincolninst.edu/publications/working-papers/implementacion-infraestructura-azul-verde-iav-traves-mecanismos. Accessed 13 December 2023.

Lord, Alexander, and Guanpeng Dong. "Building the Breathable City." *Lincoln Institute of Land Policy*, June 2022, https://www.lincolninst.edu/publications/working-papers/building-breathable-city. Accessed 13 December 2023.

"MassGIS Data: Counties." *Mass.gov*, 26 April 2022, https://www.mass.gov/info-details/massgis-data-counties. Accessed 13 December 2023.

Nolte, Christoph. *Lab 3: Predictive Modeling, Policy Instrument Design*. 2023. Document.

"1.17. Neural network models (supervised) — scikit-learn 1.3.2 documentation." *Scikit-learn*, https://scikit-learn.org/stable/modules/neural_networks_supervised.html. Accessed 13 December 2023.

Palter, Rob. "Corporate real estate strategy in the COVID-19 era." *McKinsey*, 4 February 2022, https://www.mckinsey.com/industries/real-estate/our-insights/climate-risk-and-the-opportunity-for-real-estate. Accessed 13 December 2023.

Rosen, Andy. "Boston is asking developers to help fund a sea wall as the city worries about rising water." *The Boston Globe*, 25 August 2021, https://www.bostonglobe.com/2021/08/25/business/an-experiment-edge-seaport-city-launches-climate-defense-fund/. Accessed 13 December 2023.

Rosen, Andy. "Boston is asking developers to help fund a sea wall as the city worries about rising water." *The Boston Globe*, 25 August 2021, https://www.bostonglobe.com/2021/08/25/business/an-experiment-edge-seaport-city-launches-climate-defense-fund/. Accessed 13 December 2023.

Singh, Deepika. "Machine Learning with Neural Networks Using scikit-learn." *Pluralsight*, 6 June 2019, https://www.pluralsight.com/guides/machine-learning-neural-networks-scikit-learn. Accessed 13 December 2023.

"statsmodels.multivariate.pca.PCA - statsmodels 0.14.0." *Statsmodels*, May 2023, https://www.statsmodels.org/stable/generated/statsmodels.multivariate.pca.PCA.html#statsmodels.multivariate.pca.PCA. Accessed 13 December 2023.

"World Energy Transitions Outlook 2022." *International Renewable Energy Agency*, 2022, https://www.irena.org/Digital-Report/World-Energy-Transitions-Outlook-2022. Accessed 13 December 2023.