**Identifying Determinants of Rural Economic Resiliency via**

**COVID-19 Firm Closures:**

**A Multi-Model Machine Learning Approach for Investigating Successful Economic Development in Rural Texas**

Reece Iriye, ririye@smu.edu

**Abstract**

This study aims to analyze determinants of rural economic resiliency by predicting the number of firm closures in rural towns during the COVID-19 pandemic. This analysis is divided up into 5 classes of population ranges (500 – 2,499; 2,500 – 9,999; 10,000 – 24,999; 25,000 – 49,999; 50,000+) to provide stronger insights for individual towns, rather than conducting the county-level study. The research incorporates demographic, economic, energy industry, other industry, and county-level variables. Two machine learning approaches, the Elastic Net and Random Forest models, are utilized to highlight key predictors of firm failure while maintaining predictive accuracy with the complex relationships among socio-economic variables. The resampling scheme will involve 10-fold cross-validation with five repeats for each population category to minimize bias and overfitting. The optimal hyperparameter combinations for each model will be calculated to maximize its performance. This study will provide insights into how policymakers can allocate resources towards the sustainment and development of rural communities. By leveraging machine learning tools, policymakers can make more informed decisions about how to best develop policies that promote entrepreneurship and sustained economies in rural communities. Research papers incorporating machine learning into their design often focus on analyzing urban areas. With over a billion dollars being invested by the U.S. government into rural America in 2022, this study aims to provide some insights into how investments into rural America should be invested by policymakers.

**Introduction**

**Firm Closure in Rural America as a Determinant of Economic Resiliency**

Urbanization has fostered an environment in America that has caused large metropolitan areas to become the focal point of economic activity. In turn, the trend towards urbanization has left rural America behind. Technological advances in agriculture, improvements in automobile manufacturing and accessibility, and the construction of modern highways all have holistically benefitted society but left harmful side-effects in rural communities as a result [1]. People who formerly lived in many rural areas are flocking to denser, more-populated locations, causing job loss and firm closures. Small businesses are the predominant employers of rural America—providing 60-80% of America’s overall net new jobs in the last decade—but “small businesses are also prone to ‘destroy’ more jobs due to layoff and closure” [2]. With large-scale job loss being a threat to several rural communities (especially those that rely on specific industrial sectors) [3], small towns should try to minimize firm failure as much as possible to ensure their survival.

Speaking with a variety of locals in a case study of Trinidad, TX, I learned that residents there were simultaneously concerned about the survival of their town and about maintaining the small-town charm that makes Trinidad special and unique. Based on their responses, I decided that researching what causes firm closures rather than firm growth would provide stronger insights into how rural communities should approach the creation of successful economic development policies. Rural entrepreneurship research has grown significantly since 2000 [4], but many of these papers home in on firm creation as their primary point of focus. Analyzing firm creation in entrepreneurship research is important for helping rural communities to some extent. For one, firm creation is an indicator of sustained economic growth, as it helps communities create jobs and mitigate trade shocks [1]. Focusing only on growth, however, undermines the priorities of many rural town residents. It ignores how these people often fear the small-charm elements of their town they know and love disappearing with the influx of outsiders. Instead, focusing on economic resiliency via firm closures addresses both the sustainability and longevity of economic activity in rural communities. The implications of urban and rural entrepreneurship are entirely different from one another, and while researchers have attempted to analyze rural firm growth in the same way that urban entrepreneurship has been studied without considering regional contexts [5], their analyses do not always translate well to the wants and needs of rural communities. By assessing what causes firm closures rather than firm growth, policymakers and local stakeholders can assess the overall health of the local economy and implement policies to enhance its resiliency.

**Rural Texas During the COVID-19 Pandemic**

During the COVID-19 pandemic, rural communities largely experienced the consequences of economic shock and the health implications of the pandemic as well. Due to the lack of data collection in these communities, rural America had been hidden as a portion of the country that actively had been hit hard by the pandemic [6]. The lack of healthcare, telemedicine, social services, and broadband made rural areas susceptible to the virus. Nevertheless, local leaders in some rural Texas counties were largely much more concerned about the pandemic’s impact on their small businesses [7].

When recalling their perceived narratives of the pandemic, several local leaders surveyed in select rural counties in Texas in 2020 characterized the pandemic as a disruption in their communities’ lives. At this time, a severe oil and gas crisis also was affecting the industry worldwide [8], and rural communities relying on these industries were especially experiencing economic turmoil. Local leaders in these rural counties constructed a narrative with a hero/villain archetype, where foreign governments involved in price wars and state and federal leaders instating lockdown policies were frequently cast as their villains [7]. Victims in these narratives were typically local business owners and residents, showcasing how the survival of small businesses was many of these leaders’ utmost priority [7]. Their concerns showcase how firm closures spiked during this period in 2020, especially since they were not able sustain during business closures. By observing firm failure during 2020 when businesses were closing their doors during the pandemic, we are able to see if rural communities were able to sustain the economic shocks brought onto them by the pandemic. Governmental figures in these rural communities tended to label businesses shutting down as their primary concern, so analyzing how firm closures were caused during the pandemic will provide some insight into the concerns of rural leaders at the time. Moreover, it can specify exactly how these rural economies react in periods of economic shock, highlighting what characteristics tend to make rural communities more resilient when markets perform poorly. Especially with high inflation rates in 2022 and 2023 causing fears of a potential recession [9], it is essential to provide rural policymakers with a direct study showcasing why exactly some rural communities tend to perform better than others in an economic crisis.

Analyzing firm closures in 2020 showcases how rural economies act under pressure and why they act in that manner, but why am I not analyzing firm failure during the 2008 financial crisis instead? Unemployment was a major issue at this time that affected both urban and rural communities, but it especially impacted rural towns because of their tendency to lack industry diversification and large-scale competition [10]. A Canadian study of unemployment during the 2008 financial crisis showcased how the magnitude of coefficients related to diversification and competition severely increased in an OLS model when compared to the same study of a different timeframe [10]. I am analyzing firm closures during the COVID-19 pandemic instead for a multitude of reasons. For one, the pandemic is more recent to the present than 2008. More recent data can relate better to current trends facing our society today (e.g. technological advancement, increases in remote work opportunities, broadening of telemedicine accessibility). The data also highlights how even the oil and gas industry suffered during this economic shock, but that did not happen to the same extent in 2008. Rural counties with stronger oil and gas economies performed much better than counties without these strong assets [11], but the oil crisis that also occurred alongside the pandemic caused that failsafe in their economies to not active in 2020. The pandemic rendered businesses inactive for several months, and rural communities’ ability to withstand that economic shock is important to understand in the context of rural entrepreneurship.

**Classification of Urban-Rural Towns through a Continuum, An Alternative to County-Level Analysis**

Defining what constitutes a rural community versus an urban one can be difficult, and the official definition espoused by the U.S. government has changed over time. In their study of firm growth, In a study of rural economic resiliency, distinguishing urban and rural towns can be difficult, so I decided to use an intuitive way to divide up my analysis to produce the best results possible. For the sake of simplicity, I will be dividing up my analysis into 5 population brackets for each town:

* 500 – 2,499
* 2,500 – 9,999
* 10,000 – 24,999
* 25,000 – 49,999
* 50,000+

These population ranges that I defined for the context of my analysis are derived from the Frontier and Remote (FAR) code system, which was developed in the 2000’s [12]. Rural health specialists created this tool for classifying regions to conceptualize access to healthcare access, and while the model is much more complex than just dividing up areas by their population count, I will solely categorize my analysis into these population subsets for simplicity and reproducibility purposes. The U.S. Department of Agriculture acknowledges urban and rural as multidimensional concepts and identifies the threshold “used to differentiate rural and urban communities range from 2,500 up to 50,000, depending on the definition” [13].

Research on the economic development often focuses on non-metropolitan county-level analysis defined by the Office of Management and Budget (OMB) [13]. While county-level analysis can be a great tool for conceptualizing regional population and economic trends, it is flawed in telling the stories of individual towns. Hand et al., for example, found that population was their most influential variable in predicting firm growth [4], but that finding might illuminate a problem in how not separating analyses by population range and by town can mask other important predictors in a model on the rural level. I will be focusing specifically on towns divided up by the population divisions indicated above, because these ranges roughly characterize the types of towns that they are. For example, Athens is a town in Henderson County, TX, with a population of 12,890 as of 2021 [14]. Trinidad is also a town in Henderson County, and its population as of 2021 is 861 [15]. Figure 1 presents a breakdown of these two towns.

|  |  |  |
| --- | --- | --- |
|  | **Athens, TX** | **Trinidad, TX** |
| **Total Population** | 12,890 | 861 |
| **Population Growth (%) from 2000 – 2021** | + 11.0 % | - 21.0% |
| **Residents Below Poverty Line (%)** | 30.2% | 23.3% |
| **Residents with Completed Bachelor’s Degree (%)** | 22.2% | 10.6% |
| **Median Income ($)** | $38,149 | $31,324 |
| **Households with Broadband Internet (%)** | 73.2% | 55.3% |
| **Hospital Located in Town** | Yes | No |
| **Grocery Store Located in Town** | Yes | Nothing larger than a Dollar Store |

**Figure 1:** Breakdown of Athens, TX & Trinidad, TX in 2021 [14, 15]

By OMB standards, Henderson County is considered non-metropolitan, and thus both towns would be categorized together when performing a county-level analysis. Trinidad and Athens are so different to the point where I believe categorizing them together alongside other different towns would cause the overall regional context of these towns to heavily outweigh local context. Such a discrepancy in information using county-level analysis would ignore the needs of each individual town when it comes to analyzing firm failures. Especially for smaller towns with population ranges 500 – 2,499 and 2,500 – 9,999, import trends that could be revealed about why firms fail would be overshadowed by larger micropolitan towns in the same county. Discrepancies in population, human capital, income, broadband access, population dynamics, and existing businesses (per Figure 1) showcase why Athens and Trinidad should be separated instead of grouped together for analysis.

An issue that could emerge with the analysis of towns instead of counties is the risk of scarce or even inaccurate data in the smaller rural towns. Some reliable database services such as the U.S. Census Bureau’s QuickFacts interface only display data for towns with a population greater than 5,000 [16]. While data reporting and reliability could potentially pose issues in my analysis, I believe that conducting a study that focuses on towns as their own separate entities is essential for revealing trends that are actually important for policymakers. A study of individual towns could pose some contextual issues. While county-level analysis over-compensates for regional context, focusing on towns as entirely separate entities from their county could potentially under-compensate for regional context. Understanding a town’s surrounding region is important for understanding how they function for multiple reasons. Residents in one town, for example, may commute to work in surrounding towns or regularly purchase products in their surrounding region. To compensate for this potential error, I will include some variables that are county-dependent in the model. However, a county-level analysis may not perfectly capture a town’s context and relationship with their entire surrounding region either. I expect that providing insights into towns as separate entities from their overall county will help illustrate how local policymakers should behave to reduce firm failure in their individual communities as much as possible.

**Machine Learning in Rural Entrepreneurship and Policymaking in General**

The effects of causation variables in rural entrepreneurship are collinear and mutually enforcing and stretch among a vast set of disciplines [4]. Because of this issue, utilizing standard econometric methods to research a phenomenon as complex as rural entrepreneurship can be difficult, because identifying variables responsible for firm failure requires models that can see past these correlations to the best of their ability [4]. Hence, employing various machine learning algorithms that can identify collinear relations and/or capture unusual trends may better reveal exactly how firm failure is caused in small towns.

Machine learning was developed to maximize prediction performance best intuitively [17]. These methods utilize the data at hand to choose where a model should ideally stand with the bias-variance trade-off, and they take more flexible functional forms to uncover trends that a standard OLS regression model could not. A random forest model, for example, interprets independent variables by using the range of values for all the parameters to establish the resulting output.

Several important studies pertaining to rural economic development have incorporated machine learning models into their analysis in recent years. Celbiş conducted a study using classification and machine learning to determine “why some entrepreneurs succeed in rural areas at the very first stage of…entrepreneurial activity while others fail” [18]. In his research, Celbiş uses data from the Life in Transition Survey (LiTS) series conducted by the European Bank for Reconstruction and Development. The data showcases responses from a survey asking subjects if they ever tried starting a business, and it is split into YES and NO responses for binary classification to model the data more purposefully. Then, Celbiş conducts a training and testing split and performs 10-fold cross-validation as a resampling scheme. Following the data preprocessing, he implements bagging, random forest, and stochastic and non-stochastic gradient boosting models, because they are ensemble models that perform variable selection to compensate for collinearity. These ML models also showcase Variable Importance Plots to highlight which variables play stronger roles in predicting the y-variable [18].

By incorporating ML models into his analysis, Celbiş can provide much stronger insights into the performance of independent variables and the magnitude of their impact. Celbiş incorporates several variables into his model that are collinear in nature and express complex relationships, so the use of ML is particularly suitable for addressing this research question that deals with complex variable relationships [19]. Previous research homing in on entrepreneurship establishes exactly how entrepreneurial activities can be perceived as “complex social problems,” because their non-linear relationships are extremely hard to predict [20, 21]. Moreover, the socio-economic attributes involved in predicting entrepreneurial y-variables are severely interdependent with one another [22, 23]. A simple logistic regression model could not correctly model these predictors and provide insight into their value due to the non-linear relationship in socio-economic data in its entirely for predicting entrepreneurship.

Machine learning has emerged more prominently as a tool to investigate complex problems in the entire field of public policy, especially in recent years. Some areas where ML and public policy have intertwined with one another include predicting policy outcomes in the U.S. using Random Forest models [24] and predicting EPA policy implementations based on their rules using Decision Trees [25]. Prediction is often the main priority of ML papers, but establishing leading predictors is a leading strength of ML, especially in cases where severe non-linear relationships exist [22]. Studies focusing on identifying leading predictors in why a certain phenomenon occurs can inform policymakers on how to allocate their resources towards minimizing harmful effects caused by that phenomenon. Some examples include the analysis of credit payments during COVID-19 using gradient-boosting and tree-based models and identifying factors that led to diminished willingness and ability to pay off loans and interest [26], forecasting determinants of economic recessions with the MARS model and other alternatives while identifying specifically how to deal with these predictors [27], and identifying predictors of misdemeanor recidivism and how it can be reduced through social service interventions [28].

In the context of public policy, however, both efficiency *and* equity play major roles in identifying strengths and weaknesses of how a model performs and provides accurate solutions that do not neglect parts of a community [28]. Algorithmic fairness through the collection of unbiased data plays a key role in mitigating this crisis, but such a task is much easier said than done. Crime statistics reporting, for example, is notorious for being riddled with human biases. Governments often use data from these sources for analyzing crime patterns, creating a feedback loop that causes more data to continuously be collected in regions where police officers are dispatched more often [29]. In analyzing firm closure data, identifying incorrect predictors using biased or inaccurate data could cause policymakers to focus on projects that may harm them in the long run instead of helping them. Many of these communities do not have many resources to be able to allocate towards necessary economic development and sustainment projects, expressing how the stakes for this context are high. Moreover, identifying factors that cause firm failure in rural communities in the first place, as well as in any policy context [30], are essential for ensuring that as much reliable data as possible is being employed into the model to produce productive results for all communities that may incorporate these studies into their future policy ventures. Studies that inform public policy decisions are not conducted in controlled environments, so bias is inevitable and must properly be addressed.

**Data**

**Variables**

Figure 2 presents a list of variables that will be employed into my model. Most of the variables are the same as the ones utilized by Hand et al. [4] in their analysis of firm growth in rural counties, but some adjustments have been made to make the model more proscriptive rather than retrospective, because I attempt to capture previous context that led to the number of firms that failed in 2020 during the pandemic and oil and gas crisis.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Type** | **Variable** | **Source** | **Timeframe** |
| **Dependent Variable** | Failed Firms | Texas Comptroller’s Office | 2020 |
| **Demographic Variables** | Population | U.S. Census Bureau | 2020 |
|  | Population (%) Aged Between 25–44 Years Old | American Community Survey | 2020 |
|  | Population (%) Aged 65+ | American Community Survey | 2020 |
|  | Ratio of Population (%) Aged 25–44 and Population (%) Aged 65+ | Feature engineered from previous two variables | 2020 |
|  | Population (%) Aged 25+ with a Bachelor’s Degree | American Community Survey | 2020 |
|  | Non-White Population (%) to Evaluate Overall Diversity | American Community Survey | 2020 |
|  | Population County In-Migration (%) | American Community Survey | 2020 |
|  | Population County Out-Migration (%) | American Community Survey | 2020 |
|  | Ratio of Protestant to Catholic Residents | The Association of Religion Data Archives | 2010 |
| **Economic Variables** | Unemployment Rate | Bureau of Labor Statistics | 2019 |
|  | Number of Banks per Town | Federal Deposit Insurance Corporation | 2019 |
|  | Total Deposits Held in Local Banks, $1000’s | Federal Deposit Insurance Corporation | 2019 |
|  | Effective Federal Funds Rate | Federal Reserve Economic Data | 2019 |
|  | Gini Index of Income Inequality | American Community Survey | 2019 |
|  | Population (%) in Poverty | American Community Survey | 2019 |
|  | Number of Newly Established Firms | Texas Comptroller’s Office | 2019 |
|  | Number of Patents Issued in Each Town | U.S. Patent Office | 2019 |
|  | Population (%) with Health Insurance | American Community Survey | 2019 |
|  | Population (%) with Access to Broadband Internet | Federal Communications Commission | 2019 |
|  | Median Household Income | U.S. Census Bureau | 2019 |
|  | Median Property Value Cost | U.S. Census Bureau | 2019 |
|  | Population Growth (%) from 2000 – 2019 | U.S. Census Bureau | 2000 – 2019 |
|  | Population Growth (%) from 2010 – 2019 | U.S. Census Bureau | 2010 – 2019 |
| **Energy Industry Variables** | Total Gas Production (barrels) | Railroad Commission of Texas (curated by Texas 2036) | 2019 |
|  | Total Gas Production (barrel of oil equivalents), Including Casinghead Gas Production | Railroad Commission of Texas (curated by Texas 2036) | 2019 |
|  | Total Solar Power Installed Capacity (in MW) | University of California Berkeley Lab (Tracking the Sun database) | 2019 |
|  | Total Wind Turbine Installations | U.S. Geological Survey (Wind Turbine database) | 2019 |
| **Other Industry Variables** | Percent of Employment in  Farming - NAICS 11 (%) | Unsure | 2019 |
|  | Percent of employment in extraction - NAICS 21 (%) | Unsure | 2019 |
|  | Percent of employment in arts, entertainment, and recreation - NAICS 71 (%) | Unsure | 2019 |
|  | Percent of employment in oil and gas extraction (%) | Unsure | 2019 |
|  | Percent of employment in Elementary and Secondary Schools - NAICS 6111 (%) | Unsure | 2019 |
|  | Percent of employment in manufacturing - NAICS 31 (%) | Unsure | 2019 |
|  | Percent of employment in Community college - NAICS 611210 (%) | Unsure | 2019 |
|  | Percent of employment in offices of physicians, dentists, and other health practitioners - NAICS 6211-3 (%) | Unsure | 2019 |
|  | Percent of employment in coal mining - NAICS 2121 (%) | Unsure | 2019 |
|  | Percent of employment in child day and elderly care services - NAICS 62441, 61412 (%) | Unsure | 2019 |
| **County-Level Variables** | Distance (miles) to a county with more than 250,000 population | ARC 2018 (Author's calculation) | 2005 |
|  | Natural Amenities | U.S. Department of Agriculture | 2005 |
|  | Difference between two major party presidential candidates | MIT Election Lab | 2020 |
|  | Social Capital Index | Northeast Regional Center for Rural Development | 2014 |

**Figure 2:** Variables that will be employed in multiple models to predict firm closure in rural towns.

Feature Engineered variables are highlighted.

Blue sources need to be verified.

Red sources have known country sources but need to be checked for individual town sources.

**Potential Setbacks to the Data**

Some issues may emerge in the analysis of firm closures in individual towns during 2020. For one, comparability in these towns is limited and severely uncontrolled. Variation exists in regional pandemic responses, lockdown measures, and COVID-19 infection rates in each town. While the narratives of county leaders were similar, variation still existed on how they believed the pandemic should be addressed in their communities [7]. Due to this issue in comparability, it may be difficult to draw extremely meaningful conclusions from the results that a perfect model using this data could describe. Also, the data is largely cross-sectional. A model using this data strong in providing an overall snapshot of economic resiliency in these rural communities, but further research would be required for telling an encompassing, meaningful story about firm closures during the pandemic [31].

**Research Methods**

**Exploring the Bias-Variance Trade-Off**

In selecting a model that will best fit a large range of parameters with complex relationships, minimizing bias and variance as much as possible would be ideal [19]. Because of the large range of parameters, I will be incorporating models that can highlight key predictors of firm failure simultaneously while maintaining predictive accuracy with the complex relationships amongst variables that likely exist. In a complete version of this paper with the data at hand, I would showcase scatterplot matrices to identify existing correlations between predictive variables.

**Penalized Regression via the Elastic Net**

One of the models I will be testing is an Elastic Net model, which maintains the same shape as a standard OLS regression model. However, it combines the best features of both the ridge and LASSO models and accurately identifies predictors that contribute to changes in the dependent variable of interest. The Elastic Net model simultaneously performs variable selection and continuous shrinkage, and it addresses multi-collinearity by grouping predictors that are highly correlated with one another [32]. I believe this model will be strong in the context of predicting firm failure and economic resiliency of rural towns during 2020, because there are such a substantial number of predictors to the point where multi-collinearity will shine as a prominent issue. The Elastic Net will also still have interpretable parameters that can easily distinguish factors that contribute to an increase or decrease in firm failures more than others.

**Random Forest**

I will also be incorporating a Random Forest model into my analysis in determining what method best fits to the data. The Random Forest algorithm is an ML ensemble method that incorporates decision trees to make decisions. It addresses limitations to its predecessor (the Decision Tree algorithm) by compensating for overfitting and instability. Its ability to handle high-dimensional data and its resiliency to noisy and irrelevant predictors makes it a great choice for this task [33]. A total of 38 predictors encompassing multiple subject ranges are involved in predicting firm failures, so a model that can address this high dimensionality simultaneously while producing interpretable results is required. The Random Forest also takes a different shape from the Elastic Net, but ensemble methods like the Random Forest are strong in providing variable importance measures [18].

**Resampling Scheme**

Evaluating towns instead of counties will provide me with more data to conduct my analyses. Texas has more than 1,200 incorporated cities per the Texas Comptroller’s Office, with 400 of these towns having a population of less than 1,000. However, since I am splitting my analysis into 5 population subsets as indicated earlier, the resulting inadequate amount of data will lead me to not conduct a train-test split. Within my resampling scheme, I will perform 10-fold cross validation with 5 repeats for each separate population class to minimize bias and overfitting as much as possible.

**Establishing Tuning Parameters**

Both the Elastic Net and Random Forest models have tuning parameters embedded into their algorithms that can only be optimized through continuous iterations through the models with random combinations of these parameters [32, 33]. Calculating the optimal hyperparameter combination are computationally expensive, especially when dealing with a large dataset, but it is important because the models will perform substantially better when that ideal combination is found [34]. I will randomly incorporate 10 hyperparameter combinations into my model to tune it in a way where it calculates firm failure to the best of its ability.

**Results**

In a complete paper, include the following:

* Reported MSE & MAE for each model
* Variable importance plots highlighting which factors contributed the most to firm failures
* A visualization of the inner-workings of the tree-based method
* A description as to what model I choose for each population class as the best performing model and why
* A comparison of variable importance rankings for each population category
* A report on ideal hyperparameters, and possibly describe their implications

**Discussion**

**An Evaluation of the Model’s Findings**

After running the model, I expect to see separate results for each population combination. I likely will have a hard time finding data for some of the predictors with towns that have a population less than 5,000, so the difference in existing predictors will be something that I acknowledge throughout my discussion of findings if that ends up becoming a concern.

Hand et al. uncovered how total population, prior firm growth, oil production, banks, and immigration all played the strongest roles in predicting firm growth [4]. Prior firm growth being so important highlights how the establishment of an entrepreneurial culture leads to the maintenance of one and an incentive to increase risk and establish more firms, which verifies a claim made by Gimenez-Nadal et al. [35]. I expect to see this claim about entrepreneurial culture bleed into my analysis as well, with high firm growth in 2019 resembling that culture which would cause 2020 firm closures to not be as relatively high. Risk-taking behavior, however, may actually cause several firms to fail because they may have only been recently established, but that is a trend to look out for in my actual analysis.

I believe median household income, the Gini Index, and population growth trends will all play a major role in predicting firm failure. Median household income is a measure of existing wealth in a community, and while it is a flawed measure in that it does not evaluate the overall spread of wealth, it captures the performance of the average person or family in the region. Gini Index calculations, if available in a town-level analysis, would inform how the state of inequality affects firm failure. An increasing population for such a long time-range is a measure of sustained growth, which may lead to some resiliency towards failing markets in 2020.

**Conclusion**

With policymakers on the federal and state level investing more money recently into rural communities than before, it is essential to evaluate the state of entrepreneurship in these communities and how policymakers can best approach allocating resources towards their sustainment and development. The American Rescue Plan, for example, invested a total of $2.8 billion into coal and power plant communities in 2022, and the money was dispersed amongst multiple departments like the USDA, the EDA, and the Department of Education [36]. With these historic investments, figuring out how to adequately spend this kind of money is a topic that machine learning research in rural communities could help potentially inform.

Studies such as this one of firm closure where I uncovered important determinants of rural economic resiliency can be applied to policy entrepreneurship studies that seek to inform policymakers on exactly how people should address a complex policy problem. Research in rural areas is often overshadowed by research in urban areas, but with recent investments in rural communities, it is important for policymakers to have access to data-driven insights that can inform their decision-making. Machine learning approaches can be particularly useful in this context, as they can help identify patterns and trends in complex data that may be difficult to uncover through traditional methods like an OLS regression mode. By leveraging ML tools, policymakers can make more informed decisions about how to best allocate resources and develop policies that promote entrepreneurship and sustained economies in rural communities.

**References**

1. Goetz SJ, Partridge MD, Stephens HM. “The Economic Status of Rural America in the President Trump Era and beyond.” *Applied Economic Perspectives and Policy*. 2018;40: 97–118. doi:10.1093/aepp/ppx061.
2. Stephens HM, Partridge MD. “Do Entrepreneurs Enhance Economic Growth in Lagging Regions?” *Growth and Change*. 2011;42: 431–465. doi:10.1111/j.1468- 2257.2011.00563.x
3. Sherman, Jennifer. “Bend to Avoid Breaking: Job Loss, Gender Norms, and Family Stability in Rural America.” *Social Problems* 56, no. 4 (2009): 599–620. https://doi.org/10.1525/sp.2009.56.4.599.
4. Hand, M. C., Shastry, V., & Rai, V. (2023). “Predicting Firm Creation in Rural Texas: A Multi-Model Machine Learning Approach to a Complex Policy Problem.” *PLOS ONE*.
5. Karlsson, C., Dahlberg, R. Entrepreneurship, “Firm Growth and Regional Development in the New Economic Geography: Introduction.” *Small Business Economics* 21, 73–76 (2003). https://doi.org/10.1023/A:1025036125745.
6. Peters, D. J. (2020). “Community Susceptibility and Resiliency to COVID‐19 Across the Rural‐Urban Continuum in the United States.” *The Journal of Rural Health*, 36(3), 446–456. https://doi.org/10.1111/jrh.12477.
7. Hand, M. C., Morris, M. & Rai, V. (2023). “The Role of Policy Narrators During Crisis: A Micro-Level Analysis of the Sourcing, Synthesizing, and Sharing of Policy Narratives in Rural Texas.” *Policy Studies Journal*. 00, 1-21. https://doi.org/10.1111/psj.12501
8. Stevens, Pippa. “Oil Plunges 24% for Worst Day since 1991, Hits Multi-Year Low after OPEC Deal Failure Sparks Price War.” *CNBC*, CNBC, 6 Apr. 2020, https://www.cnbc.com/2020/03/08/oil-plummets-30percent-as-opec-deal-failure-sparks-price-war-fears.html.
9. Meredith, Sam. “Recession Risk and Inflation Fears Creating 'a Huge Amount of Confusion' for Investors, Strategist Says.” *CNBC*, CNBC, 27 Apr. 2023, https://www.cnbc.com/2023/04/27/recession-and-inflation-fears-stoking-investor-confusion-strategist.html.
10. Wang, C., Madsen, J. B., & Steiner, B. (2017). “Industry diversity, competition and firm relatedness: the impact on employment before and after the 2008 global financial crisis.” Regional Studies, 51(12), 1801–1814. https://doi.org/10.1080/00343404.2016.1254766
11. Abboud, A., & Betz, M. R. (2021). “The local economic impacts of the oil and gas industry: Boom, bust and resilience to shocks.” *Energy Economics*, 99, 105285–. https://doi.org/10.1016/j.eneco.2021.105285
12. National Academies of Sciences, Engineering, and Medicine. 2016. *Rationalizing Rural Area Classifications for the Economic Research Service: A Workshop Summary*. Washington, DC: The National Academies Press. https://doi.org/10.17226/21843.
13. “What Is Rural?” *USDA ERS - What Is Rural?*, https://www.ers.usda.gov/topics/rural-economy-population/rural-classifications/what-is-rural/#:~:text=This%20delineation%20of%20built%2Dup,with%20fewer%20than%202%2C500%20people.
14. “Athens, Texas Population History 1990 - 2021.” *Athens, Texas Population History | 1990 - 2022*, https://www.biggestuscities.com/city/athens-texas.
15. “Trinidad, Texas Population History 1990 - 2021.” *Trinidad, Texas Population History | 1990 - 2022*, https://www.biggestuscities.com/city/trinidad-texas.
16. “Census Data.” *U.S. Census Bureau*, 3 Mar. 2023, https://www.census.gov/data.html.
17. Kleinberg J, Ludwig J, Mullainathan S, Obermeyer Z. Prediction Policy Problems. American Economic Review. 2015;105: 491–495. doi:10.1257/aer.p20151023
18. Celbiş, M. G. (2021). “A machine learning approach to rural entrepreneurship.” *Papers in Regional Science*, 100(4), 1079–1104. https://doi.org/10.1111/pirs.12595Top of Form
19. Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives*, 28 (2): 3-28.DOI: 10.1257/jep.28.2.3
20. Bruyat, Chirstian & Julien, Pierre-André. (2001). “Defining the Field of Research in Entrepreneurship.” *Journal of Business Venturing*. 16. 165-180. 10.1016/S0883-9026(99)00043-9.
21. Dorado, Silvia & Ventresca, Marc. (2013). Crescive Entrepreneurship in Complex Social Problems: Institutional Conditions for Entrepreneurial Engagement. Journal of Business Venturing. 28. 69–82. 10.1016/j.jbusvent.2012.02.002.
22. Acs, Z.J., Desai, S. & Hessels, J. “Entrepreneurship, economic development and institutions.” *Small Bus Econ* 31, 219–234 (2008). https://doi.org/10.1007/s11187-008-9135-
23. de Groot, Henri, Verhoef, Erik and Nijkamp, Peter, (2001), “Energy Saving by Firms: Decision-Making, Barriers, and Policies,” *Energy Economics*, 23, Issue 6, p. 717-740.Bottom of Form
24. McGuire, S., & Delahunt, C. (2020). “Predicting United States Policy Outcomes with Random Forests.” *Institute for New Economic Thinking,* Working Paper Series No. 138. https://doi.org/10.36687/inetwp138
25. Bhattacharyya, A. M. (1999). “Policy Capturing Using Decision Trees: An Analysis Of Epa Rule-Making.” 1999 Annual meeting, August 8-11, Nashville, TN 21601. American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association). https://idead.repec.org/p/ags/aaea99/21601.html
26. Afrizal, T., & Saputra, J. “Factors that Affect Customer Credit Payments During COVID-19 Pandemic: An Application of Light Gradient Boosting Machine (LightGBM) and Classification and Regression Tree (CART).
27. Sephton, P. (2001). Forecasting recessions: can we do better on MARS. *Federal Reserve Bank of St. Louis Review*, *83(March/April 2001)*.
28. Rodolfa, K. T., Salomon, E., Haynes, L., Larson, J., & Ghani, R. (2020). “Case study: Predictive fairness to reduce misdemeanor recidivism through social service interventions.” In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 142-153). New York, NY: Association for Computing Machinery.
29. Buil-Gil, D., Moretti, A., & Langton, S. H. (2022). “The accuracy of crime statistics: assessing the impact of police data bias on geographic crime analysis.” Journal of Experimental Criminology, 18, 515-541. https://doi.org/10.1007/s11292-021-09457-y
30. Amarasinghe, K., Rodolfa, K., Lamba, H., & Ghani, R. (2023). “Explainable machine learning for public policy: Use cases, gaps, and research directions.” *Data & Policy,* *5*, E5. doi:10.1017/dap.2023.2
31. Khanal, G., & Thapa, S. (2020). Practical Guide to Conducting Cross-Sectional Studies (Quantitative) in a Community Setting. SAGE Publications Ltd.
32. Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society*. Series B (Statistical Methodology), 67(2), 301–320. http://www.jstor.org/stable/3647580
33. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. https://doi.org/10.1023/A:1010933404324
34. Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228-1242. https://doi.org/10.1080/01621459.2017.1319839
35. Gimenez-Nadal, J. I., Lafuente, M., Molina, J. A., & Velilla, J. (2019). Resampling and bootstrap algorithms to assess the relevance of variables: applications to cross section entrepreneurship data. Empirical Economics, 56, 233-267. https://doi.org/10.1007/s00181-017-1355-x=
36. The White House. (2022, March 1). Fact sheet: The Biden Administration's historic investments to create opportunity and build wealth in rural America. Retrieved from https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/01/fact-sheet-the-biden-administrations-historic-investments-to-create-opportunity-and-build-wealth-in-rural-america/#:~:text=The%20Bipartisan%20Infrastructure%20Law%20invests,safety%2C%20and%20availability%20of%20energy.