

# Machine Learning Models and Transfer Models for Measuring Impact of the Pandemic on Communities

Timur Berdibekov  
George Mason University  
tberdibe@gmu.edu

Research Report for Work presented by Timur Berdibekov at Information theory section in the Machine learning conference.  
Research Contribution: Timur Berdibekov

**Abstract**—This paper studies the relationship between the 2019 Novel Coronavirus (COVID-19) pandemic, its public health and economic impact, and rates of economic inclusion and access to banking services throughout the pandemic in the United States. For select U.S. counties, this paper examined COVID-19 infection and mortality rates, unemployment rates and the number of bank closures, and the rate of economic inclusion to discover any notable relationships. Lastly, select features are evaluated for the predictive capability of the county and county-equivalent rates of unbanked households to better inform policy making given that the unbanked household rates are unknown for most counties.

**Index Terms**—Pandemic, COVID-19, Transformer, Machine learning, Artificial Intelligence

## I. INTRODUCTION

The coronavirus pandemic has created fundamentally new disruptions and continues to exacerbate existing long-term issues and underlying disparities in the United States, with significant impact to both public health and the economy. The coronavirus pandemic has quickly reached community-spread level virtually in every locale in the United States and has had a significant impact on the physical and economic well-being of millions of U.S. households. Mounting evidence continues to emerge which demonstrates major disparities in COVID-19 incidence and mortality rates, as well as unemployment and general economic impact for different racial and income groups, underscoring the insidious nature of the pandemic in traveling along existing fault lines of inequality in the United States.

Direct access to the formal economy in the United States via the established banking system similarly matches patterns of structural inequality. The most recent analysis by the Federal Reserve indicates that nearly 22% of adults are either unbanked, having no checking, savings, or money market account, or underbanked, defined as having a bank account but still relying on alternative financial service products, with higher percentage of black and Hispanic adults classified unbanked as opposed to white adults (Board of Governors of the Federal Reserve System, 2020). Unbanked and underbanked individuals rely more on alternative financial products, such as such as check cashing, payday lending, and pawn-shop

loans lack consumer protections found in mainstream financial products and services, and tend to be more costly in the long-run, and stands in contrast to full participation in the banking industry which reduces vulnerability both to theft and discriminatory and predatory lending practices (Federal Deposit Insurance Corporation, 2017).

Maintaining access to a bank account not only establishes access to credit from traditional mainstream sources, but further supports consumer participation in Fintech products, including digital payment platforms and mobile banking, that provide convenience as well as seamless access to eCommerce platforms (Detrixhe, 2020). The unbanked population, numbering approximately 14.1 million adults and 6.4 million children, are further removed from federal pandemic relief efforts and will experience longer waits in accessing stimulus funds provided on paper checks and prepaid cards relative to fully banked households which can utilize near-immediate direct deposit of funds into their bank accounts [1] [2] [3] [4]

We Use transfer learning models [5] [6]–[8] [9] [10] and machine learning [11] [12] [13] models [14] [15]. [16], [17], and BERT [18]. for this research project.

[11] [19]. [12] [20] [21]. [22], [23]

### A. problem

The current general academic and practitioner consensus is that the ongoing coronavirus pandemic has had an outsized impact on the most financially-vulnerable population in the United States – the unbanked and underbanked. The goal of our project is to more clearly understand the linkages between the coronavirus pandemic and unbanked and underbanked communities by examining unemployment rates, bank closure rates, and small business closures in largely unbanked and underbanked counties.

As increasing economic and financial inclusion and increasing participation in the economy is an accepted and expressed bipartisan goal, it is critical to understand how unbanked and underbanked populations may potentially be more vulnerable to the COVID-19 pandemic, as measured by unemployment rates, bank closure rates, and small business closures. Ultimately, we seek to provide an answer to the

following: can unbanked/underbanked and pandemic rates, in addition to available demographic information be used to predict unemployment, bank branch closure, and small business closure rates for a given county in the United States?

## II. METHOD

The focus of our project would be to determine a relationship between the pandemic and current rates of financial inclusion, unemployment, small business closures, and bank branch closures rates of select counties across the United States.

In our research for the problem statement we have collected data for coronavirus rates provided by the Johns Hopkins University Center for Systems Science and Engineering, as well as the dataset collected by the FDIC and Census Bureau as part of the Current Population Survey supplement for unbanked and underbanked households. We are using local area unemployment statistics provided by the Bureau of Labor Statistics for county-level unemployment rates, and a dataset of bank branch closures compiled manually through the Office of the Comptroller and Currency Corporate Applications Search (CAS). We are currently in the process of selecting and finalizing a dataset capture the rate of small business closures by county to use a proxy for economic impact of the coronavirus pandemic, in addition to local area unemployment rates. These datasets combined give us the data about the drastic changes occurred during the COVID-19, which gave us an insight about our problem statement. [24] [25].

Data cleaning is a process of identifying unwanted or unused data and eliminating them from the dataset after integration. There are few actions we can perform in the cleaning process like add, delete, edit, replace in order to achieve a cleaned data set from analysis. There are three main steps important from cleansing our data. Standardization: It is better to use the universal and replacement format while performing data entry. Validation: To know the accuracy of our data, we need to verify it.

Deduplication: It is necessary that we remove duplicate data for an efficient and accurate data set. Data cleaning has not been started at the time of milestone 1 submission. This task will be achieved using the languages R and python and tools like Jupyter and RStudio. We will be replacing the missing values with mean or zero and search for any duplication or standardization error using filter options and replace them with the opted standard.

### A. Visualizations

Visualizations are the last part of our research which will help us to demonstrate our output efficiently. It will help us to show the summary of our research using platforms like Tableau, Python and R studio. Tableau is a tool that will simplify raw data into an understandable format. The visualizations created using tableau are generally in the form of worksheets and dashboards. Additionally, we will also utilize python and R language to build some visualizations like bar chart, heat maps etc. We will choose the best visualization

chart to portray our result of the effect of corona virus in employment/unemployment rates at the end of our project.

### B. Result Evaluation and Report

After data interpretation and visualization, bringing evaluation results together will be our final step. Data interpretation is done in order to evaluate the project from a comprehensive viewpoint of the five criteria and draw a conclusion. This is the value judgment process. Like we initially submitted the project milestone 1 report on how we will accomplish the specific tasks and steps, we will be submitting an evaluation report at last as a final submission about our results, findings, approaches, analysis, visualizations and all other steps involved in our phases of the project.

## III. CONCLUSION

In this project, our goal is to examine how the coronavirus pandemic has impacted financially vulnerable households in the United States by focusing on the connection between the rates of coronavirus incidence and underbanked and unbanked households with respect to unemployment rates, small business closure rates, and bank closure rates. We will attempt to accurately be able to predict the unemployment, bank closing, and small business closure rates of a given county based on the COVID-19 county rates and number of unbanked and underbanked households.

## REFERENCES

- [1] Centers for Disease Control and Prevention, "Coronavirus disease 2019 (covid-19) - community, work & school," <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html>, 24 July 2020.
- [2] P. Cihan, "Fuzzy rule-based system for predicting daily case in covid-19 outbreak," in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1–4, 2020.
- [3] Center on Budget and Policy Priorities, "Tracking the covid-19 recession's effects on food, housing, and employment hardships," <https://www.cbpp.org/research/poverty-and-inequality/tracking-the-covid-19-recessions-effects-on-food-housing-and>, 13 November 2020. Covid Hardship Watch.
- [4] V. Z. Marmarelis, "Predictive modeling of covid-19 data in the us: Adaptive phase-space approach," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 207–213, 2020.
- [5] M. Heidari and S. Rafatirad, "Semantic convolutional neural network model for safe business investment by using bert," in *IEEE 2020 Seventh International Conference on Social Networks Analysis, Management and Security, SNAMS 2020*, 2020.
- [6] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in *IEEE 2020 International Symposium on Technology and Society (ISTAS20)*, ISTAS20 2020, 2020.
- [7] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network model to recommend airline tickets by using online reviews," in *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA)*, pp. 1–6, 2020.
- [8] M. Heidari, S. Zad, B. Berlin, and S. Rafatirad, "Ontology creation model based on attention mechanism for a specific business domain," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

- [9] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)* (J. Burstein, C. Doran, and T. Solorio, eds.), pp. 4171–4186, Association for Computational Linguistics, 2019.
- [10] M. Heidari and J. H. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pp. 0542–0547, 2020.
- [11] R. Kochhar, "Unemployment rose higher in three months of covid-19 than it did in two years of the great recession." <https://www.pewresearch.org/fact-tank/2020/06/11/unemployment-rose-higher-in-three-months-of-covid-19-than-it-did-in-two-years-of-the-great-recession/>, 11 June 2020. Pew Research Center.
- [12] A. Klein, "A big problem for the coronavirus economy: The internet doesn't take cash." <https://www.brookings.edu/opinions/a-big-problem-for-the-coronavirus-economy-the-internet-doesnt-take-cash/>, 25 March 2020. Pew Research Center.
- [13] M. Heidari, J. H. J. Jones, and O. Uzuner, "An empirical study of machine learning algorithms for social media bot detection," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.
- [14] spatial.ai, "unemployment-sentiment-data-covid-19: 2010-2019." <https://www.spatial.ai/post/unemployment-sentiment-data-covid-19,> 2019.
- [15] M. Heidari, J. H. J. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *IEEE 2020 International Conference on Data Mining Workshops (ICDMW), ICDMW 2020*, 2020.
- [16] <https://textblob.readthedocs.io/en/dev/>, 2018.
- [17] <https://pypi.org/project/vaderSentiment/>, 2019.
- [18] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network to recommend airline tickets by using online reviews," in *IEEE 2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2020*, 2020.
- [19] C. STOLOJESCU-CRISAN, B. P. BUTUNOI, and C. CRISAN, "Iot based intelligent building applications in the context of covid-19 pandemic," in *2020 International Symposium on Electronics and Telecommunications (ISETC)*, pp. 1–4, 2020.
- [20] O. E. Ergungor, "Bank branch presence and access to credit in low-to moderate-income neighborhoods," *Journal of Money, Credit and Banking*, vol. 42, no. 7, pp. 1321–1349, 2010.
- [21] A. D. Indriyanti, I. G. L. E. Putra, D. R. Prehanto, I. K. D. Nuryana, and A. Wiyono, "Development of mapping area software for dismissal people affected by covid-19," in *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, pp. 1–4, 2020.
- [22] M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.
- [23] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *IEEE 2021 World AI IoT Congress*, 2021.
- [24] Office of the Comptroller of the Currency, "Corporate applications search (cas) - branch closings and branch establishments - 1-1- 2019 to 12-6-2020." [https://apps.occ.gov/CAAS\\_CATS/](https://apps.occ.gov/CAAS_CATS/), 2020.
- [25] U.S. Census Bureau, "Nation, states, and counties population: County population totals: 2010-2019." <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html>, 2019.