Capstone Paper Review Week 1 (Paper 1)

YOUR PROJECT TITLE HERE

Tommy Steele

Supervisor: **Dr. Shushen Pu**

Contents

Background/Motivation

In the paper that I reviewed, the broader context of the research was to determine the overall effectiveness of the ARIMA model in predicting future Covid 19 cases in the short term and the long term. The study took place in Recife, a city in Brazil, throughout a good portion of 2020, when Covid became most prevalent. From my understanding when reading the paper, I believe that the researchers were attempting to test the overall effectiveness of the ARIMA model in predicting cases of future diseases as a whole. The authors often referenced that the ARIMA model had been used in other disease forecasts which is why I am led to believe that this would be the case. The problem that the abstract of the paper mentioned was that the ARIMA model was often overlooked in cases dealing with the forecast of any pandemics because of âperceived limitations in handling complex and dynamic scenarios,â as was mentioned in the second sentence of the abstract of the paper. Further than this, it wanted to prove the short-term ability and effectiveness of the ARIMA model in being able to determine daily cases of pandemic diseases such as COVID-19. From my account, I believe that this research question is significant because it touches upon a topic that could very well become as big of an issue as COVID-19 was in 2020. For many countries, especially those of a lower economic standing, COVID-19 became a major cause for concern because of mass shutdowns and families in need not being able to return to work and earn the weekly or biweekly paycheck that could be keeping their family afloat. I believe that studies like this form an important basis for research that could be done in the future to prevent such major pandemics from taking place again. Of course, these studies can never replace the

effectiveness of the vaccines that work to combat the diseases medically. However, these studies can serve as a depot of information for local governments and national governments to understand the measures that must be taken in the short term. An example that was mentioned in the paper, was a complete shutdown period in the city of Recife due to a fluctuation of cases in the first wave of the pandemic. Research such as this could be the measure that prevents such actions from being immediately necessary but also allows for measures that are less impactful on peopleâs everyday lives.

Methods Used

The article mentioned three major methodologies, listed under the mathematical background section of the paper, that the study used. The list of these includes Predictive Machine Learning techniques, ARIMA Models and Parameter Estimation, and, finally, Model Evaluation Metrics. To summarize first, the Predictive Machine Learning techniques were essentially defined as techniques to teach an artificial intelligence program to understand patterns and make predictions from them using the aforementioned ARIMA model that became a core asset of this study. As mentioned, the ARIMA Models and Parameter Estimation are effective formulas given to the artificial intelligence program so that it can process the data that is inputted in it through the formulas and be able to create models of the historical cases as well as predict future cases that may arise. Finally, the Model Evaluation Metrics are as they sound, they are extra estimation and error formulas that test the artificial intelligenceâs selected ARIMA model from the several possibilities to determine if the chosen was the most effective of the possibilities. I believe that the combination of these methods is the most effective for studying the research question because they can provide relatively rapid (compared to the hand calculations of a human) and concise results from the limited data set that was worked with in the set. All the while, the expected error that comes with working with artificial intelligence that is following only the information given is combatted by the error formulas and estimation metrics that ensure that test results are as accurate as possible. As mentioned earlier in this review, it would seem as if the use of the ARIMA model for this case was already particularly innovative in itself. I believe that this is the case because

it was already mentioned that these models were previously overlooked because it was believed that they had no place in determining such things. Furthermore, it can be said that applying these models while also using still early developmental stages of artificial intelligence is another innovation that the study made. To use these temperamental models along with artificial intelligence is known to have its own myriads of issues and is a risk that led to serious innovation. Given this, the risk was accounted for with the metrics they used to evaluate the produced model so that each test that was done would provide the most accurate estimate possible.

Significance of the Work

The paper mentioned that it had several key findings not only regarding the ARIMA model but also in predictive models as a whole. To start with the findings of the ARIMA model, it is duly noted that the ARIMA model performed exceptionally well in predicting cases within a short period. As mentioned in the conclusion of the paper, âThe relative error was kept below 10% until the end of the second period of analysis, which ended 42 days after the beginning of the forecast.â This shows that the model works particularly well when delving into forecasts around one to two months in the future. However, it was also noted that the long-term capability of the model was not so promising. It was further mentioned in the conclusion of the article that it âreached values greater than 50% by the end of the sixth period, around four months after the analysis started.â This is quite a significant jump from being below 10% to reaching over 50% error in what seems to be a relatively short amount of time. However, this leads to the other significant finding of the article in which the authors suggest the importance of probing the models that are used to ensure that proper models are being utilized. I believe that these results are important within the broader context of the field because the study encourages the use of the ARIMA model in the short term. However, it also encourages that in the long term, it might be beneficial to create new, better models that may be even better predictors not only in the longer term, where the ARIMA model struggled but even in the short term. These findings could be the boon necessary for researchers to decide to continue research into the practice of time series analyses and create models that are better suited for cases such as this one. In all, I believe that

the implications of these findings lie in their use for the benefit of understanding the drawbacks and strengths of using models to predict pandemics such as Covid-19. It is clear that the field of study in using time series analyses and models of the like for pandemics is relatively new, likely due to their unnecessity. However, this article could be the key to deepening the research and giving researchers a tool to understand what methods they could apply for such research.

Connection to Other Work

Throughout the paper, it made several references to other papers that it took reference from, most of which were dealing with their studies for some facet of the Covid-19 pandemic. I believe this paper has its place amongst all of these in that it provides a look at a different model that was used to forecast the progression of the pandemic at the time. The many papers that were provided in the citations of the paper all had their unique questions and methods as well but this paper fits in with them just the same. I believe that the paper builds on the other papers in that it uses previous notions of COVID-19 to further guide the scope in which they are focusing. It noted that several other papers were using deep learning machines to assist in helping their research, which was similar to this, but the authors and researchers of this paper used a different model to help guide their research. The model is the ARIMA model that has been mentioned several times in this review. However, it also differs in a sense from the better portion of the other mentioned papers in that it focuses it sights on a specific population in Brazil. Other papers did this within other parts of the world but it is my understanding that the better portion focused on general data instead. To name one, we can reference citation 28 of the paper which looks at the paper titled âDisjoint and Functional Principal Component Analysis for Infected Cases and Deaths Due to COVID-19 in South American Countries with Sensor-Related Data.â In this paper, the focus was on the broader spectrum of infection cases and death rates throughout several South American countries. This data is great for the notion of attempting to expand research and predictions of the like to a broader scale. However, it lacks the personability of

being within certain cities such as Recife that may have adopted their own strict or loose policies that may differ from the country as a whole. That is where the paper I studied shined because it was able to include the challenges of facing strict shutdowns within its research which allowed for a more clear insight into the rapid spread that Covid would have seen in the country before and after.

Relevance to Capstone Projects

To preface this section, it should be understood that the chosen project (currently) for my group is a time series analysis of Covid-19. Our project will require us to develop a model and choose data (which will be found on the internet) to predict cases and compare those cases to the data that is available. With that being said, it is clear that the content of the paper is entirely relevant to our project as it gives a possible method that we could apply to our project, though not on as large of a scale. It also gives us a baseline for certain graphs and estimation errors that we may need to look out for when doing our modeling to ensure that we are not falling out of the confidence intervals that we may choose to use. In looking through the data, I believe that my group will be likely to include the ARIMA model in our project. The reason for this is that we do not intend to attempt to predict long-term cases that may arise from the continuing pandemic as this is not a sustainable or efficient idea for the time that is available within the class. However, like the project, we will attempt to use our own coding and learning program (likely through the use of Python or R) to help us create the models and find the data necessary to make our forecast. Unlike the paper that was read, our group is unlikely to focus on only one city or one small population in any given country or continent. It is my understanding that our groups intend to use a variety of data sets, though limited to at most four countries for the sake of time, and to use that data to try and predict a decent model of cases for them. This is certainly different as the paper I studied focused on only the city in Brazil and also had access to the site of the research itself. My group will be fully focused on using the data that is available to us online. One final point to

make is that is likely we will try to adopt the error estimation methods that the paper utilized as this will help us in ensuring that we are also receiving results that are as accurate as possible. With this, I believe that we will be able to make predictions that stay under a 10% error just as in the paper.

Bibliography

[1] Ospina, R., Gondim, J. A. M., Leiva, V., & Castro, C. An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil. *Multidisciplinary Digital Publishing Institute; MDPI*, 2023

[2] Martin-Barreiro, C., Ramirez-Figueroa, J. A., Cabezas, X., Leiva, V., & Galindo-Villardon, M. P. Disjoint and Functional Principal Component Analysis for Infected Cases and Deaths Due to COVID-19 in South American Countries with Sensor-Related Data. *Multidisciplinary Digital Publishing Institute; MDPI*, 2021