Predicting Outpatient Visits using Time Series Data based on ARIMA Model

1Liv T. Onic, 2Minle Mar M. Ayalin, 3Hiezyl M. Gorantes, 4Jay C. Liza
1Researcher, 2Researcher, 3Researcher, 4Research Adviser
School of Information and Communications Technology

Abstract - **The purpose of the research is to find out who is visiting the community health center by forecasting patients visits using the ARIMA method, which is very suitable for processing time series data, The time series data of outpatient visits at the Rural Health Unit of the Municipality of Janiuay from January 1, 2020, to August 11, 2022, was gathered through secondary data collection method. Forecasting the health data of the patients using the 95% confidence interval showed that symptoms trended up per day for twenty-one (21) days, resulting in a 16.13% increase, rising by 44.00% in 4 days. The number of symptoms jumped from 25 to 36 during its steepest incline. Acute respiratory tract infection is the most common symptom, followed by urinary tract infection and gastroesophageal reflux disease. Acute respiratory tract infection accounted for 17.83% of all barangays, and adults made up 7.13% of those with acute respiratory tract infection, which is the most prevalent symptom covering almost all barangays, followed by urinary tract infection and gastrointestinal reflux. It is concluded that acute respiratory tract infection tops the list of symptoms that the patients are experiencing, as recorded on the outpatient visits, and covers almost all of the barangays affected. Most of all, it also affects adults. It is recommended that the local government unit of the Municipality of Janiuay prioritize medicines that would address minimizing respiratory tract-related concerns and other related health problems.**

keywords - **outpatient, visits, ARIMA, time series**

I. INTRODUCTION

The public's awareness of health issues is increasing, leading to an increase in patient visits, which the health center must pay extra attention to as it prepares for the completion of facilities [5]. Primary care systems provide comprehensive health care and education, making referrals to hospitals and specialists as needed, allowing for effective delivery, cost management, and access for everyone [10]. Inequities in health are a hallmark of the Filipino healthcare system. Compared to, life expectancy is more than ten years higher in wealthy provinces [5,6,7]. To solve the issue, public health centers must carefully plan their operations, one of which is forecasting efforts. In order to improve the current state of health in the local community setting, community health centers were established. Public health centers are health units that deal with health issues that exist in the community.

Forecasting the number of various outpatient visits in an accurate and reliable manner aids in the scientific allocation of vital medical resources like equipment. Hence, accurate outpatient visit forecasting is useful for the prudent planning and allocation of healthcare resources to satisfy medical needs or foresee prospective resource shortages. [6] Being relatively straightforward, the ARIMA model can also be used to forecast the number of outpatient visits without the inclusion of variables [9]. Covariates have been employed in some earlier studies, but you must first collect covariate data for that specific time period because they cannot be used to forecast the future. For instance, the incidence of brucellosis in 2015 was predicted using the ARIMA model with the covariates of atmospheric pressure, wind speed, and mean temperature in 2015 [3].

In this work, we retrospectively analyze the time series of outpatient visits at the Rural Health Unit of the Municipality of Janiuay from January 1, 2020, to August 11, 2022, using an ARIMA model. Several studies also investigate the influence of independent covariate delay effects on outpatient visits [3,8]. To investigate the connection between symptoms and outpatient visits, as well as to create an easy-to-use model that may be applied to forecast outpatient visits.

According to Oracle (2022), forecasting can be used to create operational goals, estimate quality and standard compliance, estimate return on investment, estimate growth and the strategic impact of innovations, and address tactical issues like estimating costs, inventory needs, and customer happiness.

Luo et al. (2017) say that to effectively plan and allocate healthcare resources to meet patient demand, reliable forecasting of hospital outpatient visits is helpful when it comes to the many characteristics of daily outpatient visits, such as randomness, cyclicity, and trend.

According to Bellot and Schaar (2020), discuss how to estimate (dynamically, as new information becomes available) tailored survival distributions using sparsely sampled longitudinal data, missing metrics indicative of the underlying health condition, and static data. Because there are many times very few samples available in a patient's history, and when there are, their information content is exceedingly variable. This makes using electronic medical data to build tailored risk profiles extremely difficult.

In order to create an annual local budget for any level of government, including national and municipal, time-series algorithms are also utilized in budgetary analysis. Using such an algorithm to conduct budgetary analysis makes sure that the funds provided are spent for the right purposes.

II. METHODOLOGY

Data Gathering

Selecting the data source: important considerations for choosing data include whether or not the key variables are available to appropriately define an analytic cohort and identify exposures, outcomes, covariates, and confounders. The researcher gathers the time series data of outpatients at the Rural Health Unit of the Municipality of Janiuay from January 1, 2020, to August 11, 2022.

Exploratory Analysis

During the exploratory analysis, we perform data reduction in order to have a more compact, easily interpretable representation of the target concept. This is done by focusing on the variables most relevant to the scope of this study. We explore how important the seasonality of the data is and if there are strong relationships among the variables available.

Choosing and fitting models

Due to process of data cleaning 852 features the data set is become smaller due to and resulted in 702 features, The cleaned dataset was used to for forecasting. In practice, overfitting the test set can be accomplished by refining and selecting the models with the best test accuracy classification model.

Forecasting

A time series forecasting technique called Auto Regressive Integrated Moving Average (ARIMA) uses autocorrelation measurements to model the temporal features in the time series data and predict future values. The model's autoregression component assesses how dependent a given sample is on a small number of prior data. In order to make the data patterns stable or reduce the obvious link with previous data, these differences are measured and integrated.

A common autoregressive model is shown in the equation below. The new values of this model, as its name implies, are solely determined by a weighted linear combination of its previous values. This is known as an autoregressive model of order p or AR(p) given that there are p past values. White noise is indicated by epsilon (ε).

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1 \quad (1)$$

Next, the moving average is defined as follows:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \quad (2)$$

In this case, the future value y(t) is calculated using the mistakes t introduced by the prior model. So, each subsequent term digs further into the past to account for the errors made by that model in the current calculation. The value of q is determined by the window we are willing to see past. As a result, the aforementioned model can be indicated individually as a moving average order q or just MA (q) [13].

III. RESULTS AND DISCUSSION

Table 1 shows the number of outpatient visits at the Rural Health Unit of the Municipality of Janiuay from January 1, 2020 to August 11, 2022.

Table 1. Distribution of Patient according to their Personal Characteristics

Description	Number of Visitors	% of Visitors	Average per day
Total	701		0.7356
Sex			
Male	299	42.65	0.3137
Female	402	57.35	0.4218
Age Group			
Infant	2	0.29	0.0021
Toddler	10	1.43	0.0105
Child	85	12.13	0.0892
Teens	55	7.85	0.0577
Adult	259	36.95	0.2718
Middle Age Adult	181	25.82	0.1899
Senior Adult	109	15.55	0.1144

Table 2 shows sample data retrieved from all 852 patients visited the Rural Health Unit of the Municipality of Janiuay.

Table 2. Patient Data retrieved from the Rural Health Unit

Description	Data
Date	02-01-2020, 08-10-2020, 09-10-2021, 01-21-2022…
Birthday	23-09-1959, 28–02-1999, 30-03-2012, 30-03-2012…
Bracket	Infant, Toddler, Child, Teen, Middle Age, Adult, Senior Adult
Age	1, 9, 15, 28, 44, 58…
Sex	M, F
Symptoms	Urinary Tract Infection, Vertigo, Hypertension…
Distance	1.3km, 1.9km, 2.8km, 17.0km…
Travel Time	1min, 3min, 24min, 37min…

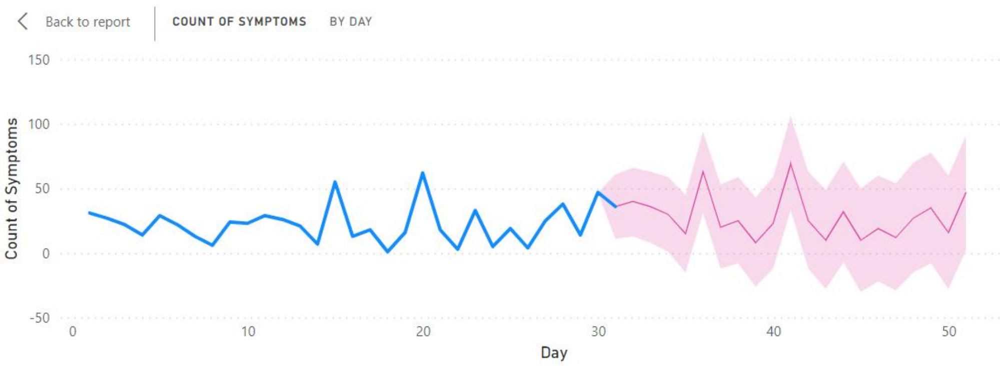

Figure 1. Forecasted Symptoms per day.

Figure 1 shows the forecast of predicted symptoms per day for twenty-one (21) days using the 95% confidence interval. While lower and upper bounds are the limits of such a range. Symptoms trended up, resulting in a 16.13% increase between Wednesday, January 1, 2020, and Friday, January 31, 2020, while trending up on Monday, January 27, 2020, rising by 44.00% (11) in 4 days. The number of symptoms jumped from 25 to 36 during its steepest incline between Monday, January 27, 2020, and Friday, January 31, 2020.

Frequency Distribution of Symptoms per Barangay

Figure 2 shows the frequency of symptoms per barangay, including a few cases from the municipalities of Badiangan and Lambunao, as observed in all 701-outpatient data gathered from January 1, 2020, to August 11, 2022.Acute Respiratory Tract Infection at 125 is the most common symptom of the patients, followed by Urinary Tract Infection and Gastroesophageal Reflux Disease. Acute respiratory tract infection accounted for 17.83% of all barangays. Across all 61 symptoms, which ranged from 1 to 125.

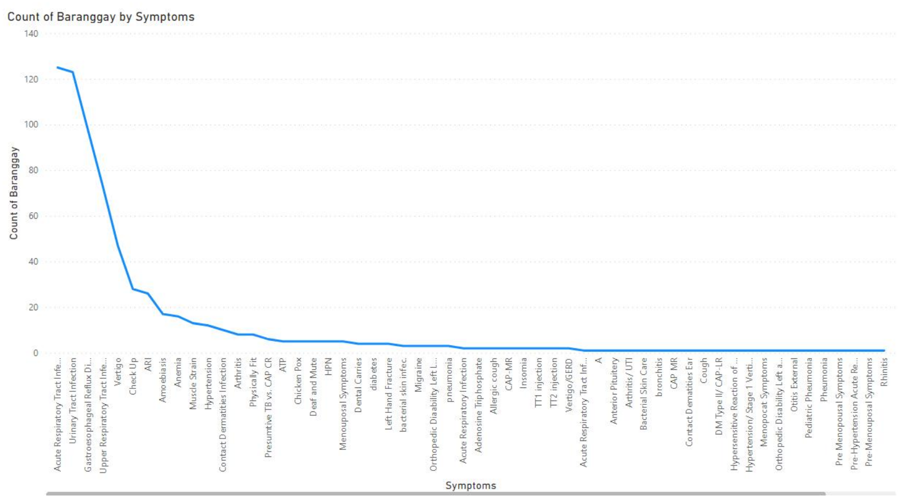

Figure 2. Count of symptoms by barangay

It is observed that Acute Respiratory Tract Infection is the prevalent symptom among adults, middle age adults and senior adults. On the other hand, amoebiasis is the prevalent symptoms among toddlers, children and teens. With just one case classified as an infant, bronchitis is the main concern.

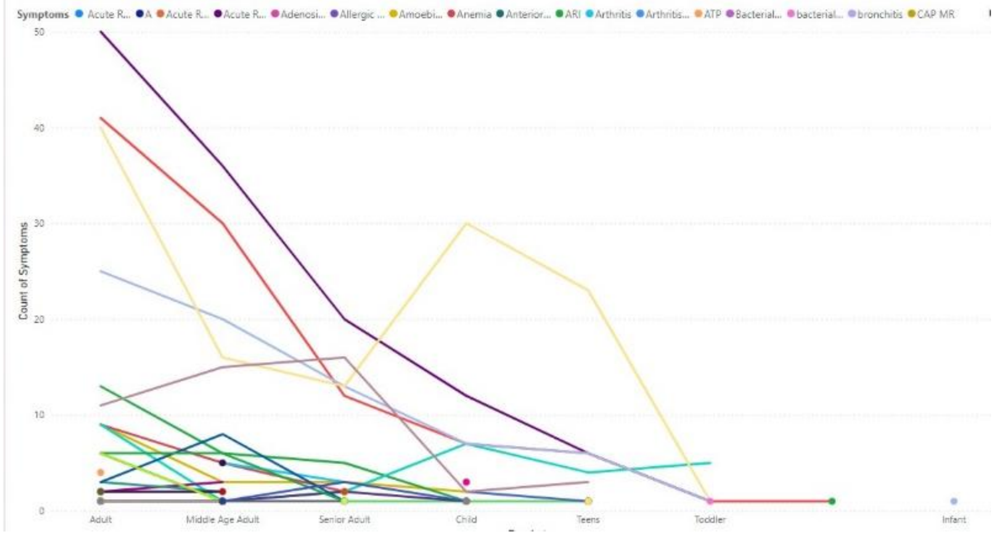

Figure 3. Distribution of Symptoms as to age brackets

Figure 3 shows the frequency of symptoms per age brackets as observed in all 701 out-patient data gathered from January 1, 2020 to August 11, 2022. It was observed that Adults made up 7.13% has acute respiratory tract infection which is most prevalent symptoms covering almost all barangays of origin of the patients, followed by urinary tract infection, and gastrointestinal reflux.

IV. Conclusion and Recommendations

Conclusions

As shown in the results, it is concluded that acute respiratory tract infection tops the list of symptoms that the patients are experiencing as recorded on the outpatient visits, and covers almost all of the barangays affected. Mainly, it also affects from adults to senior adults. When it comes to the daily trend per month, usually it peaks on the mid-month, usually higher in between 15th to 20th day of the month.

Recommendations

As it was concluded, as mentioned above, it is recommended that the local government unit of the Municipality of Janiuay prioritize medicines that would address minimizing respiratory tract-related concerns. But it does not mean that other chief complaints will be taken off the priority list, but rather that resources will be allocated to the rest of the possible ailments. As we are still in the post pandemic and we have lost track of other respiratory concerns like tuberculosis and asthma, it is necessary to address such concerns, especially for those patients who are classified from adulthood to senior citizens, in order to increase the possible life expectancy in the Municipality of Janiuay. In addition, since we can observe that mid-month has the highest trend, there is a need to beef up the supplies as early as the first week of the month in preparation for that surge of different complaints (symptoms). Lastly, since acute respiratory tract infections are most common in almost all areas in the Municipality of Janiuay, there should be a distribution of supplies of medicines on other Barangay Health Stations (BHS), and also it is time to deploy a doctor per barangay as an LGU initiative and allocate funds accordingly since almost all barangays in Janiuay have difficult access going into the población area.

References

[1] Bai, L., Lu, K., Dong, Y. *et al.* Predicting monthly hospital outpatient visits based on meteorological environmental factors using the ARIMA model. *Sci Rep* **13**, 2691 (2023). https://doi.org/10.1038/s41598-023-29897-y

[2] Bellot, A., & Schaar, M. V. D. (2020). Flexible modelling of longitudinal medical data: A Bayesian nonparametric approach. ACM Transactions on Computing for Healthcare, 1(1), 1-15.

[3] Cao, L. et al. Relationship of meteorological factors and human brucellosis in Hebei province, China. Sci. Total Environ. 703(2020), 135491 (2020).

[4] Capuno, J. J., Kraft, A. D., Poco, L. C., Quimbo, S., & Tan Jr., C. A. R. (2017, February 2). *Factors associated with the delay in seeking inpatient and outpatient care services in the Philippines*. Retrieved October 4, 2022, from https://ideas.repec.org/p/phs/dpaper/201703.html

[5] Department of Health. The 2013 Philippine Health Statistics [Internet]. Manila: Department of Health; 2013 [cited 2017 Jul 17]. Available from: http://www.doh.gov.ph/sites/default/files/publications /2013PHScompressed_0.pdf

[6] Department of Health (2017) Available from http://www.doh.gov.ph/sites/default/files/publications/PHS1987a.pdf3. Romualdez AJG, dela Rosa JFE, Favier JDA, et al. The Philippines health system review (Vol. 1). Geneva: WHO Press. 2011

[7] Health Intelligence Service, Department of Health 1989. Philippine Health Statistics 1987. Manila, Philippines:

[8] Huang, J. et al. Association of children wheezing diseases with meteorological and environmental factors in Suzhou, China. Sci. Rep. 12(1), 5018 (2020)

[9] How to Build ARIMA Model in Python for time series forecasting? (2023, February 2). https://www.projectpro.io/article/how-to-build-arima-model-in-python/544. Retrieved April 3, 2023, from https://www.projectpro.io/article/how-to-build-arima-model-in-python/544

[10] Luo, L., Luo, L., Zhang, X., & He, X. (2017, July 10). *Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models - BMC Health Services Research*. BioMed Central. Retrieved October 4, 2022, from https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-017-2407-9

[11] Marfori, J. R. A., Dans, A. M. L., Bastillo, M. O. C., Paterno, R. P. P., Rey, M. P., & Catabui, J. T. (2019). Equity in health benefit utilization and financial risk protection in outpatient and inpatient care: baseline survey of two socioeconomic groups of a pilot primary care benefits scheme in the catchment areas of a university-based health facility. *Acta Medica Philippina*, *53*(1).

[12] Oracle. (2022, September 27). *Time Series*. Oracle Help Center. Retrieved October 2, 2022, from https://docs.oracle.com/en/database/oracle/machine-learning/oml4sql/21/dmcon/time-series.html#GUID-C8AE0CEF-6F74-45CF-B393-63A6E1ECA168