# Time Series Analysis of Disease Outbreaks in Eastern Samar Province Philippines

#### **Rolando Real Codilan\***

College of Computer Studies, Eastern Samar State University Main Campus, Borongan City, Philippines; roland.r.codilan@gmail.com

## Abstract

**Objective**: To analyze the 5-year occurrence of different diseases based on hospital records in Eastern Samar Provincial Hospital; to present a time series analysis model for the occurrence of disease in Eastern Samar. **Methods**: On analyzing the data this study used the data mining model to process the activities and used Orange as the tool to show the result of the analysis. The study chooses 5 common diseases which are: Acute Gastroenteritis, Urinary Tract Infection, Pneumonia, Hypertensive Cardiovascular and Pulmonary Tuberculosis from the year 2007-2011 and through Orange widget the gathered data imported and used graph visualizer. The study gets the sum of all patients every month, year and by a quarter to get the different analysis of the occurrence of the disease. **Findings:** The result showed the time series analysis of the highest number of the patient from the year 2007-2011 was the Pneumonia followed by Acute Gastroenteritis, UTI, Hypertensive Cardiovascular and Pulmonary Tuberculosis, respectively. Borongan Eastern Samar has the highest number of occurrence of all the 5 diseases. **Application:** Consider common diseases that are affected by distinctive weather condition (rainy season and dry season).

Keywords: Data Mining Tool, Disease Outbreak, Time Series Analysis

## 1. Introduction

Disease outbreaks nationwide, especially to remote areas in the Philippines have been an inevitable and a major concern to the government. The concern agencies trying their best, they have attended training and seminars in and outside of the country just to address these disease outbreaks. The main focus of this study was the analysis of the series of time of the disease outbreaks with the use of Orange data mining tool and using it for classification in the field of medical bioinformatics. The use of time-series for the monitoring of disease is useful in preparing public health authorities to be forewarned of an impending disease outbreak and gives adequate time for disease outbreak mitigation efforts to be implemented<sup>1</sup>. Forecasting the future trajectory of cases during an infectious disease outbreak can make an important contribution to public health and intervention planning<sup>2</sup>. Planning by forecasting trends in the future is one way to apply statistical knowledge to analyze data in the past that are related to the current events<sup>3</sup>. The study can contribute to the Eastern Samar Provincial Hospital Philippines to have an advanced preparation on what kind of medicine is most important or appropriate within the month. With the use of the data from the hospital records from Eastern Samar Provincial Hospital, the study conducted the Time Series Analysis of Diseases Outbreak in Eastern Samar that has been used in analyzing the data by comparing data from the year 2007-2011 and analyzed the occurrence of disease in each month.

Orange is a component-based visual programming software package for data visualization, machine learning, data mining and data analysis.

\*Author for correspondence

There are several different levels at which Orange can be used especially in data analysis or the time series analysis. Orange contains modules for data classification and accuracy to analyze diseases. It has been used in bioinformatics for diagnosis and analysis of diseases (Orange, edited 2018).

The goal of the study was to analyze a time series of disease outbreak in Eastern Samar that will show the occurrence of disease by month, year, within five years. It contributed other information to Eastern Samar Provincial Hospital on the possibility of occurrence of what disease in every month and year.

#### 1.1 Objectives

This study aimed to:

- Analyze the 5-year occurrence of different diseases based on hospital records in Eastern Samar Provincial Hospital.
- Present a time series analysis model for the occurrence of disease in Eastern Samar.

# 2. Methodology

## 2.1 Research Design

This study used the investigative type of research in which it used facts or information already available and analyzed them to make a significant assessment of the data. It involves deeply study and assessment of available information in an attempt to explain the complex occurrence.

Figure 1 shows the example of hospital records that have been used a dataset for analyzes and save it into an Excel file. The dataset was separated or grouped into 4 columns according to the month, age, gender, address and diagnosis.

#### 2.2 Data Mining Tool

In order to carry out experimentations and implementations, Orange was used as the time series analysis tool; it is an open source data visualization and analysis tool. Orange consists of a canvas interface onto which the user places widgets and creates a data analysis workflow. Orange has also featured for different visualizations, such

Z ABC	tone bloot Page Lay	out Parmulas Data R	view		34,4200	N arais	At Excel	and Shake 0	-	
Spetting Rese	anth Thesaurus Translate Language	New Delete Prevenue Hart Comment	Show At Comment	Prob	ent Protect et Workhoo	Share	Track O	hers to Edit hanges +	Rangel	
CI	· · · /	GENDER	1.00						1.1	_
A	8 C	D		F	6	H	1	1	К.	1
1 MONTH	AGE GENDER	ADDRESS	DIAGNOSIS							
a JAN	70 M	BORONGAN	HEVD							
-	15 0	MATDOLONG	VA.							
The Land	13.5	MAXDOLONG	405							
6 14.5J	10.14	BORONGAN	100							
7 LAN	63 E	BORONGAN	LOVER EALES	S CRAW	TABLINA .					
E LAN	2 M	BORONGAN	AGE	a anare						
P LAN	20 F	BOBONGAN	UT							
10 JAN	80 F	LLORENTE	HCVD							
11 JAN	1.6	BALANGKAYAN	UTI							
12 JAN	71 M	BORONGAN	AGE							
IT JAN	68 F	ORAS	PNEUMONIA							
14 JAN	1 M	BORDNGAN	PNEUMONIJ							
15 JAN	1.5	BORONGAN	AGE							
16 JAN	74 F	TAFT	MCA							
17 JAN	1.6	BORONGAN	PNEUMONIA	h						
IS JAN	29 M	MAYDOLONG	AGE							
19 JAN	34 M	ORAS	TYPHOID FET	VER						
AAL OS	64 F	BORONGAN	PNEUMONU	4. ·						
21 JAN	545 M	DOLORES	PTB							
22 JAN	1.5	BORONGAN	AGE							
23 JAN	45 F	BORONGAN	EMPYEMA T	HORACI	5					
24 JAN	77 8	BORONGAN	PNEUMONIA	4						
25 JAN	54 F	BORONGAN	5V8							
26 JAN	71 M	BORONGAN	HCVD							
27 JAN	69 M	BORONGAN	LIVER CIRRH	OSIS						

Patient Record

Figure 1. Screenshot view of hospital database.

as scatter plots, bar charts and trees to dendrograms, networks and heat maps. By combining the various widgets the design of data analytics framework can be done.

Figure 2 represents the data analytics framework of the study in which the researchers used the five various widgets which include File, Data Table, Tree, Tree Viewer and the Distribution.

## 2.3 Research Locale

The study was conducted at the Eastern Samar Provincial Hospital Philippines Record's Office. This study limits to patients within the Province of Eastern Samar from 2007-2011.

#### 2.4 Data Analysis

The data gathered from Eastern Samar Provincial Hospital was encoded manually into Excel and the 5 common diseases Acute Gastroenteritis, UTI, Pneumonia, Hypertensive Cardiovascular Disease and Pulmonary Tuberculosis were chosen. This study used Orange as data mining tool to analyze the data gathered to show the possibility of occurrence of disease, through different chart and tables the analyzed data was shown different distribution of diagnosis for every month and year; the percentage of diseases for every year; the distribution of address by diagnosis and get the structured result of time series analysis for every month, year and by quarter of year 2007-2011.

# 3. Results and Discussion

The widget was used to input the data each type will appear and can be numeric or categorical and choose the target data, the diagnosis and apply to reveal the result.

Figure 3 represents the File widget used from Figure 2 that contains the following datasets with its respective data type such as categorical for the month, gender, address, diagnosis and numeric type for the age. The main target of the study is the Diagnosis or specific disease. The figure above represents the month of January 2007. The file contains 218 instances which represented the total number of



Orange Widgets Graphical User Interface

Figure 2. Screenshot view of Orange widgets.

patients, 4 features and 1 target. The figure represents all file widget for different months each year.

Figure 4 represents the second widget that was used which is the Tree Viewer which contains and display the tree of the analyzed data from Figure 3. The figure represents all Tree Viewer widget for different months in each year.

Figure 5 represents the dataset presented on a table. The dataset was placed according to their role. The first column represented by the diagnosis dataset filled with gray color represented the target class and the rest of the data represented as the feature. The figure represents all Data Table widget for different months each year.

Figure 6 represents the distribution of each data or variable grouped by the variable Diagnosis. The distribution represented by a graph with a different color which depends on specific diagnosis/disease. The figure below represents the distribution of month grouped by diagnosis. The figure represents all file widget for different months each year. The 5-year occurrence of different diseases is based on hospital records in Eastern Samar Provincial Hospital.

Figure 7 shows the highest percentage of diseases occurring in every month in 2007 which was Pneumonia represented with green color. The second disease that occurred frequently was the Urinary Tract Infection which is represented by the yellow color and the third disease that occurred frequently was Acute Gastroenteritis which is represented by the color blue and fourth is Hypertensive Cardiovascular Disease represented by the color red and the last disease that occurred frequently is Pulmonary Tuberculosis which was represented by the Orange color. Distribution of Different Disease is by Percentage for 2007.

Table 1 shows the percentage of occurrence of the five specific diseases in each month from the above figure. The five diseases that most occurred each month was placed into five different rows and for the month 12 different columns. The results which represented in the table, the disease that occurred mostly from the five diseases were Pneumonia from the month of January-May and October-December.

Figure 8 shows that Can-avid has the most percentage of patients from Pneumonia. The graphs represented by Diagnosis are grouped according to address.

Figure 9 represents the distribution of each month where the disease occurred. The graph shows that the highest to the lowest frequency of diseases occurred in every month for 2008 was Pneumonia, the second highest disease that occurs as the Urinary Tract Infection, the third disease was Acute Gastroenteritis, the fourth was Hypertensive Cardiovascular Disease and the disease that occurred less frequently was Pulmonary Tuberculosis. Each disease is represented with different colors. Distribution of Different Disease is by Percentage for 2008.

Table 2 shows the percentage of occurrence of the five specific diseases in each month from the table above. The five diseases that occurred most frequently each month were placed in five different rows and 12 different columns.

Figure 10 shows that Borongan has the most percentage of patients from Pneumonia. The graphs represented by Diagnosis are grouped according to address.

Figure 11 represents the monthly distribution of each of the five (5) diseases that occurred. The graph shows the highest to lowest percentage of diseases in every month in 2009. Pneumonia has the highest percentage of occurrence, followed by Urinary Tract Infection, Acute Gastroenteritis, Pulmonary Tuberculosis and Hypertensive Cardiovascular Disease respectively. Distribution of Different Disease is by Percentage for 2009.

Table 3 shows the percentage of occurrence of the five specific diseases in each month from the table above. The five diseases that most frequently occurred each month were placed in five different rows and 12 different columns. The disease that occurred mostly from the five diseases was Pneumonia.

Figure 12 shows that Borongan has the most percentage of patients from Pneumonia. The graph represented by Diagnosis was grouped according to address.

Figure 13 represents the distributions of each month for each of the five (5) diseases that occurred. The graph shows the highest to lowest percentage of occurrence of these diseases in every month for 2010. The Acute Gastroenteritis was the most highly diagnosed disease within this year, followed by Pneumonia, Urinary Tract Infection, Hypertensive Cardiovascular Disease and Pulmonary Tuberculosis respectively. Distribution of Different Disease is by Percentage for 2010.

Table 4 shows the percentage of occurrence of the five specific diseases in each month. The five diseases that occurred most frequently each month were placed into five different rows and 12 different columns.

Figure 14 shows that Borongan has the most percentage of patients from Pneumonia. The graphs represented by Diagnosis are grouped according to address. Figure 15 represents the distributions of each month where the disease occurred. The graph shows the highest percentage of occurred diseases in every month for the year 2011. Pneumonia has occurred mostly from the five diseases, the second highest disease occurred was the Acute Gastroenteritis, the third disease occurred was Urinary Tract Infection, the fourth was Hypertensive Cardiovascular Disease and the last was Pulmonary Tuberculosis. Distribution of Different Disease is by Percentage for 2011.

Table 5 shows the percentage of occurrence of the five specific diseases in each month. The five diseases that occurred most frequently each month were placed in five different rows and 12 different columns.

Figure 16 shows that Borongan has the most percentage of patients from Pneumonia. The graphs represented by Diagnosis are grouped according to address.

Figure 17 shows the occurrence of the 5 most common diseases for the month of January from 2007-2011. The time series analysis for January shows the different number of patients every year.

The sum of all patients in January shows that Pneumonia has the highest number of patients followed by Acute Gastroenteritis, UTI, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 18 shows the occurrence of the 5 most common diseases for the month of February from 2007-2011. The time series analysis for February shows a different number of patients every year. The sum of all patients in February shows that Pneumonia has the highest number of patients followed by Acute Gastroenteritis, UTI, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 19 shows the occurrence of the 5 most common diseases for the month of March from 2007-2011. The time series analysis for March shows a different number of patients every year. The sum of all patients in March shows that Acute Gastroenteritis has the highest number of patients followed by Pneumonia, UTI, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 20 shows the occurrence of the 5 most common diseases for the month of April from 2007-2011. The time series analysis for April shows a different number of patients every year. The sum of all patients in April shows that Pneumonia has the highest number of patients followed by Acute Gastroenteritis, UTI, Pulmonary Tuberculosis and Hypertensive Cardiovascular respectively.

Figure 21 shows the occurrence of the 5 most common diseases for the month of May from 2007-2011.

The time series analysis for May shows a different number of patients every year. The sum of all patients in May shows that Acute Gastroenteritis has the highest number of patients followed by Pneumonia, UTI, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 22 shows the occurrence of the 5 most common diseases for the month of June from 2007-2011. The time series analysis for June shows a different number of patients every year. The sum of all patients in June shows that Pneumonia has the highest number of a patient followed by UTI, Acute Gastroenteritis, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 23 shows the occurrence of the 5 most common diseases for the month of July from 2007-2011. The time series analysis for July shows a different number of patients every year. The sum of all patients in July shows that Pneumonia has the highest number of patients followed by UTI, Acute Gastroenteritis, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 24 shows the occurrence of the 5 most common diseases for the month of August from 2007-2011. The time series analysis for August shows a different number of patients every year.

The sum of all patients in August shows that Pneumonia has the highest number of patients followed by UTI, Acute Gastroenteritis, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 25 shows the occurrence of the 5 most common diseases for the month of September from 2007-2011. The time series analysis for September shows a different number of patients every year. The sum of all patients in September shows that Pneumonia has the highest number of patients followed by UTI, Acute Gastroenteritis, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 26 shows the occurrence of the 5 most common diseases for the month of October from 2007-2011. The time series analysis for October shows a different number of patients every year.

The sum of all patients in October shows that Pneumonia has the highest number of patients followed by, Acute Gastroenteritis, UTI, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 27 shows the occurrence of the 5 most common diseases for the month of November from 2007-2011. The time series analysis for November shows a different number of patients every year. The sum of all patients in November shows that Pneumonia has the highest number of patients followed by UTI, Acute Gastroenteritis, Pulmonary Tuberculosis and Hypertensive Cardiovascular respectively.

Figure 28 shows the occurrence of the 5 most common diseases for the month of December from 2007-2011. The time series analysis for December shows a different number of patients every year. The sum of all patients in December shows that Pneumonia has the highest number of patients followed by Acute Gastroenteritis, UTI, Pulmonary Tuberculosis and Hypertensive Cardiovascular respectively.

Figure 29 shows the occurrence of the 5 most common diseases for 12 months in 2007. The horizontal axis represents months and the vertical axis represents the number of patients. The graph presents both the increase and decrease in the number of patients for the five most common diseases per month. The different color outlines the different diseases. Color Blue for Acute Gastroenteritis, red for Pneumonia, green for Urinary Tract Infection, violet for Hypertensive Cardiovascular Disease and blue for Pulmonary Tuberculosis.

Figure 30 shows the occurrence of the 5 most common diseases for 12 months in 2008. The horizontal axis represents months and the vertical axis represents the number of patients. The graph presents both the increase and decrease in the number of patients for the five most common diseases per month. The different color outlines the different diseases. Color Blue for Acute Gastroenteritis, red for Pneumonia, green for Urinary Tract Infection, violet for Hypertensive Cardiovascular Disease and blue for Pulmonary Tuberculosis.

Figure 31 shows the occurrence of the 5 most common diseases for 12 months in 2009. The horizontal axis represents months and the vertical axis represents the number of patients. The graph presents both the increase and decrease in the number of patients for the five most common diseases per month. The different color outlines the different diseases. Color Blue for Acute Gastroenteritis, red for Pneumonia, green for Urinary Tract Infection, violet for Hypertensive Cardiovascular Disease and blue for Pulmonary Tuberculosis.

Figure 32 shows the occurrence of the 5 most common diseases for 12 months in 2010. The horizontal axis represents months and the vertical axis represents the number of patients. The graph presents both the increase and decrease in the number of patients for the five most common diseases per month. The different color outlines the different diseases. Color Blue for Acute Gastroenteritis, red for Pneumonia, green for Urinary Tract Infection, violet for Hypertensive Cardiovascular Disease and blue for Pulmonary Tuberculosis.

Figure 33 shows the occurrence of the 5 most common diseases for 12 months in 2011. The horizontal axis represents months and the vertical axis represents the number of patients. The graph presents both the increase and decrease in the number of patients for the five most common diseases per month. The different color outlines the different diseases - Color blue for Acute Gastroenteritis, red for Pneumonia, green for Urinary Tract Infection, violet for Hypertensive Cardiovascular Disease and blue for Pulmonary Tuberculosis.

Figure 34 shows the quarterly number of patients in 2007. Q1, Q2 and Q4 Pneumonia has the highest number of patients followed by Acute Gastroenteritis which highest in Q3 UTI, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 35 shows the quarterly number of patients in 2008. Pneumonia has the highest number of patients in each quarter followed by UTI, Acute Gastroenteritis, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 36 shows the quarterly number of patients in 2009. Q1, Q2, Q3 and Q4 Pneumonia has the highest number of patients followed by UTI, Acute Gastroenteritis, Pulmonary Tuberculosis, and Hypertensive Cardiovascular respectively.

Figure 37 shows the quarterly number of patients in 2010. Q1 and Q3 Acute Gastroenteritis has the highest number of patients followed by Pneumonia which highest in Q2 and Q4, UTI, Pulmonary Tuberculosis and Hypertensive Cardiovascular respectively.

Figure 38 shows the quarterly number of patients in 2011. Q1 and Q4 Pneumonia has the highest number of patients followed by Acute Gastroenteritis which highest in Q2 and Q3, UTI, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively.

Figure 39 shows the occurrence of the 5 most common diseases for 2007-2011. The horizontal axis represents the year and the vertical axis represents the number of patients. The graph presents both the increase and decrease of the number of patients for the five most common diseases per year. The different colors outline the different diseases are - color blue for Acute Gastroenteritis, red for Pneumonia, green for Urinary Tract Infection, violet for Hypertensive Cardiovascular Disease and blue for Pulmonary Tuberculosis. The highest number of patient

## File Widget Menu

0 F	le: JAN2007.c	sv		👻 🚺 🕌 📄 🕻 🌌 Reload
2	JRL:			
In 218 Dat	fo 3 instance(s), 4 fe ta has no target v	ature(s), 1 meta att ariable.	ribute(s)	
Co	lumns (Double clic	k to edit)		
	Name	Туре	Role	Values
1	MONTH	Categorica	l feature	JAN
2	AGE	N numeric	feature	
3	GENDER	G categorica	feature	F, M
4	ADDRESS	G categorica	l feature	ARTECHE, BALANGKAYAN, BORONGAN, CAN-AVID, DOLORES, HERNANI, JIPAPAD, L
5	DIAGNOSIS	G categorica	l target	
iros	vse documentatio	n data sets		Apply

Figure 3. Screenshot view for file widget.



Tree Viewer Widget Menu

Figure 4. Screenshot view of Tree Viewer.

2010		DIAGNOSIS	ADDRESS	MONTH	AGE	GENDER	
100 instances features (0, 1% existing values)	1	ACUTE GASTR	BORONGAN	JAN	1	F	
iscrete class with 5 values (no	2	URINARY TRAC	BORONGAN	JAN	12	F	
issing values)	3	URINARY TRAC	BORONGAN	JAN	3	F	
meta attribute (0.1% missing	4	PNEUMONIA	BORONGAN	JAN	6	F	
sues)	5	ACUTE GASTR	BORONGAN	JAN	7	M	
	6	PNEUMONIA	BORONGAN	JAN	1	F	
ariables	7	PNEUMONIA	BORONGAN	JAN	2	M	
Show variable labels (if present)	8	PNEUMONIA	BORONGAN	JAN	2	M	
Visualize numeric values	9	ACUTE GASTR	BALANGKAYAN	JAN	38	F	
Color by instance classes	10	PNEUMONIA	BORONGAN	JAN	1	F	
	11	ACUTE GASTR	MAYDOLONG	JAN	1	F	
election	12	PNEUMONIA	BORONGAN	JAN	86	F	
Select full rows	13	PNEUMONIA	MAYDOLONG	JAN	1	F	
	14	ACUTE GASTR	MAYDOLONG	JAN	53	F	
	15	PNEUMONIA	SAN JULIAN	JAN	8	M	
	16	ACUTE GASTR	BALANGKAYAN	JAN	1	M	
	17	ACUTE GASTR	SULAT	JAN	1	F	
	18	PNEUMONIA	BORONGAN	JAN	62	F	
	19	URINARY TRAC	BORONGAN	JAN	35	M	
Bastan Original Order	20	ACUTE GASTR	BALANGKAYAN	JAN	62	F	
Restore Original Order	21	HYPERTENSIVE	BORONGAN	JAN	70	M	
	22	URINARY TRAC	BORONGAN	JAN	1	M	
Send Automatically	23	ACUTE GASTR	MAYDOLONG	JAN	1	M	

Orange Widget Data Table

**Figure 5.** Screen of Orange data table.

-



#### Sample Graph for Diagnosis

Figure 6. Screenshot of Orange distribution.

Disease Occurrence on Year 2007



**Figure 7.** Data distribution between diagnosis and month in 2007.



PNEUMONIA

PULMONARY TUBERCULOSIS URINARY TRACT INFECTION



Figure 8. Data distribution between address and diagnosis in 2007.



Figure 9. Data distribution between diagnosis and month in 2008.





Figure 10. Data distribution between address and diagnosis in 2008.

2009 Time Series for Disease



Figure 11. Data distribution between diagnosis in month in 2009.



Distribution of Address by Diagnosis

**Figure 12.** Data distribution between address and diagnosis in 2009.

2010 Time Series for Diseases



Figure 13. Data distribution between diagnosis and month in 2010.





Figure 14. Data distribution between address and diagnosis in 2010.

2011 Time Series for Disease



Figure 15. Data distribution between diagnosis and month in 2011.

Distribution of Address by Diagnosis



Figure 16. Data distribution between address and diagnosis in 2011.



Figure 17. Time series analysis for the month of January.



Figure 18. Time series analysis for the month of February.



Figure 19. Time series analysis for the month of March.



Figure 20. Time series analysis for the month of April.



Figure 21. Time series analysis for the month of May.



Figure 22. Time series analysis for the month of June.



Figure 23. Time series analysis for the month of July.



Figure 24. Time series analysis for the month of August.







Figure 26. Time series analysis for the month of October.



Figure 27. Time series analysis for the month of November.



Figure 28. Time series analysis for the month of December.



Figure 29. Time series analysis for the Year 2007.



Figure 30. Time series analysis for the Year 2008.



Figure 31. Time series analysis for the Year 2009.



Figure 32. Time series analysis for the Year 2010.



Figure 33. Time series analysis for the Year 2011







Figure 35. Number of patient quarterly for the Year 2008.



Figure 36. Number of patient quarterly for the Year 2009.



Figure 37. Number of patient quarterly for the Year 2010.







Figure 39. Time series analysis for 5 Years



Figure 40. Data distribution for the overall address where the most common diseases occur.

Table 1. Analysis of disease by month in 2007								
Month	Acute Gastroenteritis	Hypertensive Cardiovascular Disease	Pneumonia	Pulmonary Tuberculosis	Urinary Tract Infection			
January	22.3%	19.2%	40.8%	2.3%	15.4%			
February	27.1%	7.6%	36.2%	4.8%	24.3%			
March	15.9%	8.6%	45.9%	3.2%	26.4%			
April	20.8%	2.1%	42.7%	5.2%	29.2%			
May	21.7%	8.3%	33.3%	4.2%	32.5%			
June	11.9%	9%	26.9%	13.4%	38.8%			
July	38.2%	7.9%	28.1%	8.4%	17.4%			
August	38.2%	6.5%	32.3%	21.5%	1.6%			
September	31.3%	8.1%	30.8%	9%	20.9%			
October	32%	6.3%	36.9%	4.5%	20.3%			
November	31.4%	2.9%	35.2%	3.8%	26.7%			
December	26.8%	1.2%	50%	7.1%	14.9%			

#### **Table 2.**Analysis of disease by month in 2008

Month	Acute Gastroenteritis	Hypertensive Cardiovascular Disease	Pneumonia	Urinary Tract Infection	Pulmonary Tuberculosis
January	13.90%	12.10%	35.1%	29.90%	9.10%
February	9.20%	5.60%	44.3%	32.30%	7.70%
March	31.7%	4.80%	29.60%	29.70%	4.20%
April	22.9%	10.70%	35%	26.40%	26.40%
May	20.50%	5.40%	40.40%	29.90%	25.90%
June	17.80%	7.00%	39.80%	31.90%	31.90%
July	20.20%	12.00%	37.50%	25.50%	5.90%
August	13.60%	5.70%	50.00%	28.70%	1.90%
September	14.90%	5.70%	44.40%	24.70%	5.50%
October	21.10%	8.00%	30.90%	31.70%	30.90%
November	22.40%	4.60%	39.70%	8%	8.00%
December	21.70%	1.30%	32.80%	26.00%	18.30%

## **Table 3.**Analysis of disease by month in 2009

Month	Acute Gastroenteritis	Hypertensive Cardiovascular Disease	Pneumonia	Pulmonary Tuberculosis	Urinary Tract Infection
January	27%	7.90%	30.20%	4.80%	27.80%
February	23.90%	5.50%	42.20%	3.70%	24.80%
March	17.30%	5.30%	42.70%	9.30%	25.30%
April	25.60%	7.30%	30.50%	7.30%	28.00%
May	15.70%	2.80%	35.20%	8.30%	31.50%

June	17.10%	4.50%	49.50%	7.20%	20.70%
July	7.40%	8.50%	47.90%	9.60%	26.60%
August	4.10%	9.20%	54.10%	9.20%	22.40%
September	12.00%	4.30%	47.80%	3.30%	29.30%
October	19.40%	4.60%	47.20%	5.60%	21.30%
November	20.00%	5.90%	41.20%	4.70%	27.10%
December	20.50%	3.60%	34.80%	4.50%	33.00%

**Table 4.**Analysis of disease by month in 2010

Month	Acute Gastroenteritis	Hypertensive Cardiovascular Disease	Pneumonia	Pulmonary Tuberculosis	Urinary Tract Infection
JAN	38%	2.90%	35.60%	1.10%	17.50%
FEB	32.20%	7.60%	34.50%	4.30%	17.40%
MARCH	38.50%	9.30%	27.70%	6.40%	16.30%
APRIL	33.00%	4.60%	35.00%	10.30%	17.10%
MAY	32.60%	9.20%	30.60%	5.60%	17.50%
JUNE	32.80%	5.40%	35.20%	6.30%	20.20%
JULY	40.50%	9.00%	28.00%	3.00%	17.00%
AUGUST	33.80%	5.30%	19.80%	4.80%	32.40%
SEPTEMBER	30.00%	8.50%	36.80%	7.70%	17.00%
OCTOBER	36.70%	10.70%	29.60%	16.80%	6.10%
NOVEMBER	29.60%	9.20%	33.60%	7.90%	19.70%
DECEMBER	32.10%	6.50%	34.00%	18.70%	3.40%

#### **Table 5.**Analysis of disease by month in 2011

Month	Acute Gastroenteritis	Hypertensive Cardiovascular Disease	Pneumonia	Pulmonary Tuberculosis	Urinary Tract Infection
January	28%	8.6%	38.4%	2.6%	22.4%
February	18.1%	11.1%	39.7%	9.5%	21.6%
March	27.8%	13.2%	38.7%	5.2%	15.1%
April	24.7%	9.6%	34.2%	12.4%	19.2%
May	35.7%	7.1%	18.2%	9.1%	29.9%
June	39.3%	10.4%	26.8%	6%	17.5%
July	22.7%	13%	39%	7.1%	18.2%
August	44.3%	12%	33.8%	3.8%	6.1%
September	15.2%	2%	27.3%	5.1%	50.5%
October	21.9%	10.5%	43.9%	11.4%	12.3%
November	17.6%	4.5%	54.9%	7.8%	15.2%
December	19.9%	7.3%	48.1%	9.4%	15.3%

is Pneumonia, Acute Gastroenteritis, UTI, Hypertensive Cardiovascular and Pulmonary Tuberculosis respectively mostly occurred in Borongan Eastern Samar.

Figure 40 represents the overall result of each disease grouped by address. A different color represents each place where the disease occurred.

# 4. Conclusion

The study revealed the possible result of the diseases such as Urinary Tract Infection, Acute Gastroenteritis, Pneumonia, Hypertensive Cardiovascular Disease and Pulmonary Tuberculosis that occurs mostly in each month from the year 2007-2011 by the use of Orange data mining tool. The study was successfully conducted a 5-year time series analysis model of data. It was revealed that as the year changes the number of patients also changes depending on what kind of disease they are diagnosed with. The number of patients that were diagnosed among the five different diseases from 2007, 2009 and 2011 was lower than the number of patients in the year 2008 and 2010.

# 5. References

- Time series analysis and forecasting of dengue using open data. 2015. https://www.academia.edu/18569050/ Time\_Series\_Analysis\_and\_Forecasting\_of\_Dengue\_ Using\_Open\_Data
- Assessing the performance of real-time epidemic forecast: A case study of Ebola in the Western Area Region of Sierra Leone; 2014-15. 2018. https://www.biorxiv.org/ content/10.1101/177451v3
- 3. Time series analysis of household electric consumption with ARIMA and ARMA models. 2013. http://www.iaeng. org/publication/IMECS2013/IMECS2013\_pp295-300.pdf