Journal of Engineering Design and Analysis
Volume 4, Issue 1 - 2021, Pg. No. 34-38
Peer Reviewed Journal

Article

Time Series Forecasting Models for Prediction of Conjunctivitis Disease Cases

Shobhit Verma[1], Nonita Sharma[2]

[1,2]Computer Science and Engineering, NIT Jalandhar, Jalandhar, India.

I N F O

Corresponding Author:
Shobhit Verma, Computer Science and Engineering, NIT Jalandhar, Jalandhar, India.
E-mail Id:
shobhit7800@gmail.com
How to cite this article:
Verma S, Sharma N. Time Series Forecasting Models for Prediction of Conjunctivitis Disease Cases. *J Engr Desg Anal* 2021; 4(1): 34-38.

Date of Submission: 2021-6-02
Date of Acceptance: 2021-06-18

A B S T R A C T

In Recent times, Machine learning is a powerful technique for the data analysis and for making a future prediction. There are many existing forecasting models which are useful for the prediction in different areas. Acute conjunctivitis, commonly known as "pink eye", is one of the most common eye infections, particularly among schoolchildren. Because of its highly contagious nature everyone is susceptible especially those in crowded places such as kindergartens, indoor amusement parks and swimming pools. Hence as a precautionary measure, there is an imperative need to predict the future possibilities of Conjunctivitis cases. Therefore, in this manuscript, machine learning based forecasting models are used for prediction of conjunctivitis cases in Hong Kong. Analysis is conducted on the data of past years conjunctivitis cases in Hong Kong. Mean Forecast, Seasonal Naive, Auto ARIMA, Neural Network techniques are used for analysis and forecasting. The surpassing model is adopted based on the accuracy factor. Accuracy of the models are compared with respect to Root Mean Square Error and Auto Correlation Function. Result analysis reveal that the neural network model produces least error and hence is the best prediction model for our dataset in terms of accuracy.

Keywords: Conjunctivitis, Time Series, Forecasting, Seasonal Naïve, Neural Networks, Mean Forecast

Introduction

In today life machine learning model are used too much for prediction and decision making in different areas such as: Stock Market, Medical field, Banking, Sales forecasting etc. With this there are many diseases which are the reason of concern for human's health like- Heart disease, Alzheimer, Tuberculosis, Conjunctivitis, Dengue and many more. In Hong Kong each week many cases of conjunctivitis are occurring. Health organization is taking many initiatives to put a stop on occurrence of the different diseases. But we can't do anything until we have pre-information about the disease occurrence and the number of cases of particular disease. Therefore, there is a need to predict the future occurrences so that the total number of cases that might occur in near future can be predicted in advance and further necessary action can be taken to curb that.

Conjunctivitis is an inflammation of the conjunctiva, the protective membrane that lines the inner eyelids and covers the outer surface of the eyeballs. Conjunctivitis takes various forms, e.g. infectious and allergic. The most common form is acute infectious Conjunctivitis (red-eye syndrome) which is mostly caused by bacteria and viruses. Bacterial Conjunctivitis can be caused by a variety of bacteria, with Haemophilus influenzae and Streptococcus pneumoniae being the commonest.[1] Chlamydia trachomatis may also be a cause for bacterial Conjunctivitis in both neonates and adults. Viral conjunctivitis is often associated with an upper respiratory tract infection, and is often

Copyright (c) 2021: Advanced Research Publications

caused by adenoviruses and enteroviruses. Both eyes are often affected though the symptoms usually start in one eye first. The discharge is thick, whitish or yellowish in bacterial conjunctivitis, and watery in viral conjunctivitis. Other symptoms include tearing, foreign body sensation, itchiness, pain, swelling and redness of the eyes, matted eyelids after sleep, and sensitivity to light.[2] Vision is not affected unless scarring of the cornea occurs after the infection. Government is taking several actions to prevent the occurrence of this disease. Therefore, by prediction of number of cases of Conjunctivitis can help the health organization and government to take required action against the disease.

Time Series Forecasting techniques can be used effectively for prediction of statistical data. Time series is a series of measurements for particular time period. This time period may be daily, weekly, monthly or yearly. Time series forecasting have many applications such as- sales forecasting, disease prediction, stock market prediction, economic forecasting, process and quality control, inventory analysis.[3] Therefore, time series forecasting models can be applied and used for predicting future cases of Conjunctivitis. In this research work, data set of previous year Conjunctivitis cases are taken for time series forecast. For forecasting, different model of forecasting are applied on the dataset and every forecasting model provides different accuracy according to data.

Further, the accuracy of the forecasting model is measured with the help of different error metrics. The error metrics that are considered here are the following- ME (Mean Error), RSME (Root Mean Squared Error), MAE (Mean Absolute Error), MPE (Mean Percentage Error, the detailed description of all is given in section IV.

Rest of the manuscript is organized as follows: Section II contains time series components, methodologies, ensemble detail information and metrics. Section III contains data collection, errors and Section IV and V contains conclusion and future work respectively.

Methodology

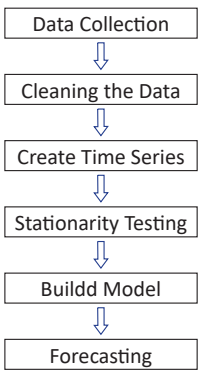

Figure 1.Time Series Forecasting Methodology

Figure 1, represents the summarized picture of Time Series Forecasting process followed. First step is Data Collection, which involves collecting the Weekly consultation rates of acute Conjunctivitis Data of Hong Kong from www.chp.gov.hk website for the time period January 2010 to January 2019. Here in our data we will consider number of Conjunctivitis cases as a single parameter. Second step is Cleaning the Data [4], which involve removing unavailable and invalid data points by replacing them with zero or not available. Third step is to create the time series, which involves converting the data in the time series format. Time series data have some essential component as following.[5]

- *Trend*: Trend is long-term increase or decrease of data. If the data contains non-stationarity, it should be removed.
- *Heteroskedasticity*: It shows that variability of data is unequal.
- *Seasonality*: It means for known and fixed period of time data repeats.
- *Stationarity*: It refers that the mean and variance of data is constant.

Fourth step is stationarity testing, here we test that our time series data is stationary or non-stationary this done by following tests:

- *ADF (Augmented Dickey-Fuller) Test:* ADF test the unit root in the time series sample. It returns *p* value which decide that time series data is stationary or not.[6]
- *Ljung-Box Test*: It is a diagnostic tool to test stationarity. Here also see the *p* value.[7]

Small values of *p* means times series as Stationary. Next step is to build model, here we apply various time series forecasting model on the statistical time series data which are following-

- *Mean Forecast:* This model based on the mean of the time series data. It is widely applicable for Statistical data forecasting.
- *Neural Network:* Artificial neural networks forecasting model is mathematical models of brain. It uses complex nonlinear relationships between the response variable and its predictors.[8] It is best suited for our data.
- *Seasonal Naive Forecast:* It is similar to Naïve Forecasting and it is best suited for seasonal data.[9]
- *ARIMA (Autoregressive Integrated Moving Average):* Combination of AR model, that does prediction on past values and MA model, that does prediction on random error terms and I stands of integration that is done to make it stationary. It can written as: ARIMA (p, d, q) where, p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order.

On applying above forecasting models on our time series data different result for various forecasting model are

obtained and the best model is selected with the help of error comparison, and the least error model get selected which shows maximum accuracy of forecasting result. The mostly use error are given below and the parameters used in formulas described in Table 1:

Table 1.Parameters used in Formula

Parameter	Description
x_i	Forecasted Value
x	Actual Value
n	Total number of observations

- *Mean Error (ME):* It is the average of errors in observation calculated by considering the positive and negative errors without sign.[10]
- $ME = \frac{1}{n}\sum_{i=1}^{n} x_i - x$ (1)
- *Mean Absolute Error (MAE):* MAE is average of absolute error values of forecasting result without direction.[11] It is absolute difference between observed and predicted values.
- $MAE = \frac{1}{n}\sum_{i=1}^{n} |\, x_i - x|$ (2)
- *Mean Squared Error (RMSE):* RMSE estimate the root of the average of squared error. This error is difference between predicted and actual values.[12]
- $RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{1,i}-x_{2,i})^2}{n}}$ (3)
- *Mean Percentage Error:* Mean percentage error is average of the percentage error. Error is variability between forecasted and actual data.[13]

$MPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - x}{x}\right)^2 * 100$ (4)

Results and Discussions

Here for Conjunctivitis incidences forecasting we have used the Tool R. R is most powerful platform used for Statistical analysis in Machine Learning. In this use various Libraries according to required functions in forecasting analysis. Now we have collected weekly Conjunctivitis incidences data in Hong Kong from website https://www.chp.gov.hk. Data is the sum of GOPC and PMP per 10000 Where, GOPC and PMP are General Out-patient Clinics rate and Private Medical Practitioner Clinics rate respectively. apply cleaning on that data. the GOPC and PMP are General Out-patient Clinics rate and Private Medical Practitioner Clinics rate respectively.

Time series diagram for the total cases of Conjunctivitis named as the object train_ts is plotted as Figure 2.

To check the trend and seasonality of the time series we decompose[14] the time series whose outputs is shown in Figure 6.

From the Figure 3, we can conclude that our time series data contains trend and seasonality.

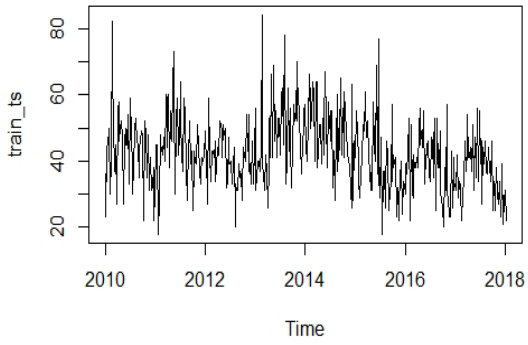

Figure 2.Time Series Plot of Cumulative Sum

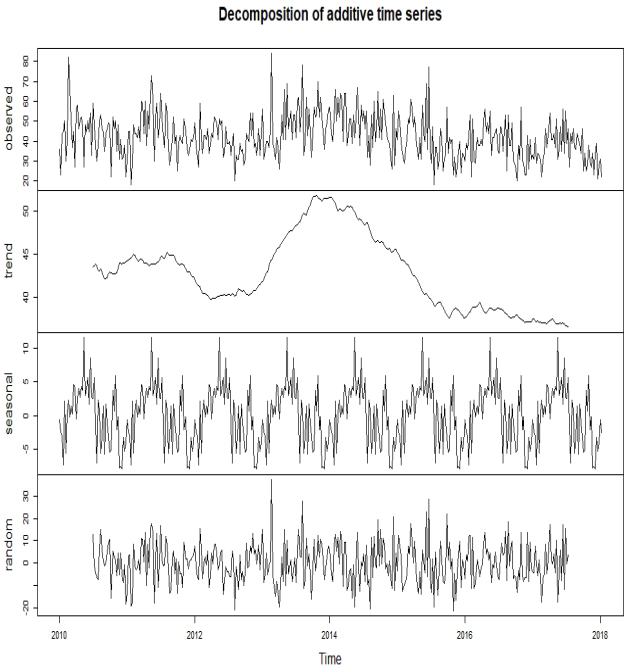

Figure 3.Decomposition Plot of Training Data

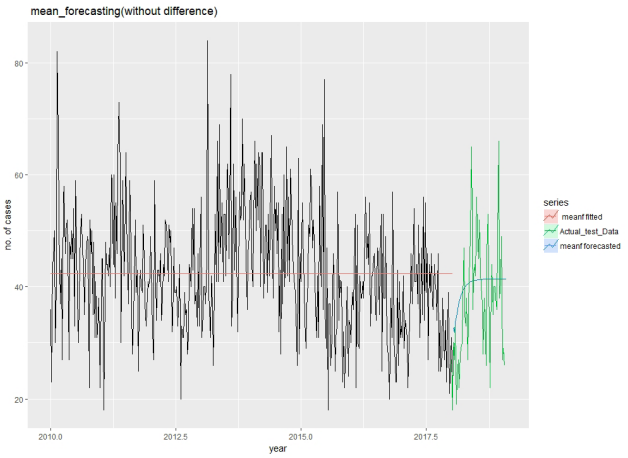

Figure 4.Mean Forecasting Graph using Without Difference Training Data

Now various forecasting model is applied on time series and resultant forecasting graph are shown in Figure 4, 5, 6 and 7.

Mean forecast resultant graph shown in Figure 4, illustrate that for future prediction of years 2018-2019, the forecasted graph not follow the same trend as test data.

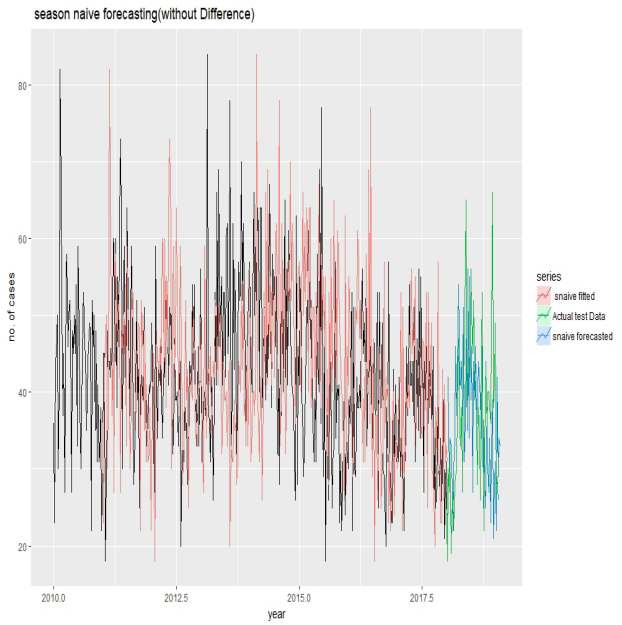

Figure 5.Seasonal Naïve Forecasting using Without Difference Training Data

Figure 5 shows output graph of seasonal naïve forecasting for prediction of years 2018-2019, the forecasted graph don't have good trend.

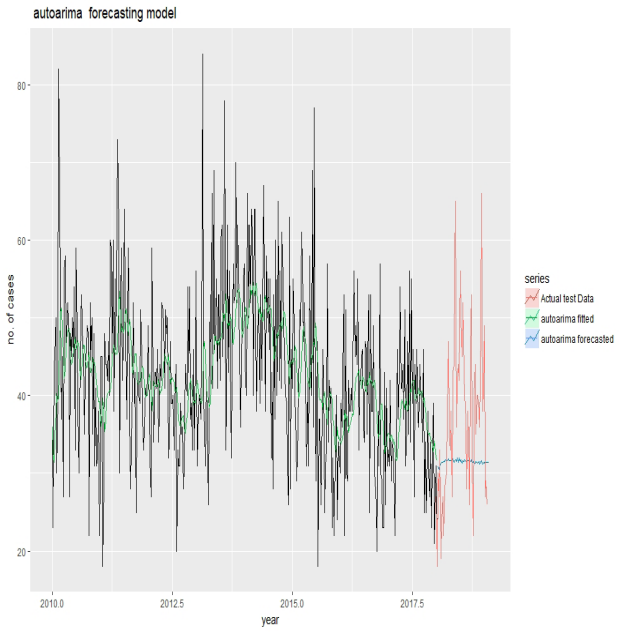

Figure 6.Auto ARIMA Forecasting Graph

Figure 6 shows output graph of auto arima forecasting for prediction of years 2018-2019 with order (2,1,3)(2,0,0), the forecasted graph is good in seen but error is too much.

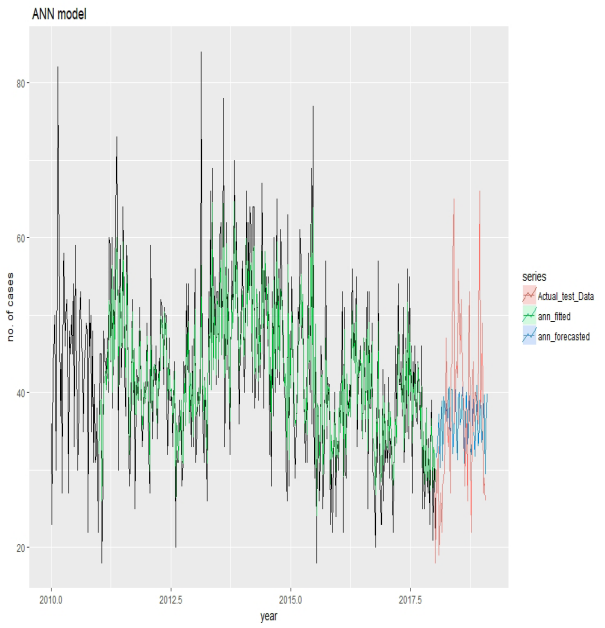

Figure 7.Neural Network forecasting graph

Neural network forecasted graph shown in Figure 7, describe that for future prediction Conjunctivitis of years 2018-2019, the forecasted graph has some similar trend as actual test dataset trend and least error generated. Errors for the above forecasting model result are quoted in Table 2.

Table 2.Error Estimation for Forecasting Models

Model	Train data type	ME	RMSE	MAE	MPE
Mean Forecast	Without diff.	-4.351	11.311	9.197	-20.182
Seasonal Naive	Without diff.	1.648	11.370	9.870	-1.196
Auto ARIMA	(2,1,3) (2,0,0)	6.647	12.305	9.654	11.000
Neural Network	Without diff.	2.055	10.276	7.819	-1.259

Conclusion

For Conjunctivitis historical data from period 2010 to 2017, the Forecasting models results are compared with respect to the trend depicted and the error values. On comparison it can be safely concluded that the neural network forecasting model has least root mean square error as 10.276, therefore neural network is best fitted for our data and other forecasting model have not shown promising results.

Future Work

In future we can enhance further work to improve the accuracy of the forecasting. Accuracy can be enhanced by combining various forecasting models, which is called ensemble technique and then apply to the dataset.

References

1. Weiss A, Brinser JH, Nazar-Stewart V. Acute conjunctivitis in childhood. *The Journal of pediatrics* 1993; 122(1): 10-14.
2. Belfort R, Marbeck P, Hsu CC et al. Epidemiological study of 134 subjects with allergic conjunctivitis. *Acta Ophthalmologica Scandinavica* 2000; 78: 38-40.
3. Mahalakshmi G, Sridevi S, Rajaram S. A survey on forecasting of time series data. InComputing Technologies and Intelligent Data Engineering (ICCTIDE), International Conference on. IEEE. 2016; 1-8.
4. Cheng X, Cook D, Hofmann H. Visually exploring missing values in multivariable data using a graphical user interface. Journal of Statistical Software. 2015; 68(1): 1-23.
5. Jebb AT, Tay L, Wang W et al. Time series analysis for psychological research: examining and forecasting change. Frontiers in psychology. 2015; 6: 727.
6. Cheung YW, Lai KS. Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business & Economic Statistics* 1995; 13(3): 277-280.
7. Chand S, Kamal S. A comparative study of portmanteau tests for univariate time series models. *Pakistan Journal of Statistics and Operation Research*. 2006; 2(2): 111-114.
8. Singh JB. Current approaches in neural network modeling of financial time series.
9. Padhan PC. Forecasting international tourists footfalls in India: an assortment of competing models. *International Journal of Business and Management* 2011; 6(5): 190.
10. Lee DK, In J, Lee S. Standard deviation and standard error of the mean. *Korean journal of anesthesiology* 2015; 68(3): 220-223.
11. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geoscientific model development. 2014; 7(3): 1247-1250.
12. Ryu K, Sanchez A. The evaluation of forecasting methods at an institutional foodservice dining facility. *The Journal of Hospitality Financial Management*. 2003; 11(1): 27-45.
13. Swanson DA, Tayman J, Bryan TM. MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecasts. *Journal of Population Research* 2011; 28(2-3): 225-243.
14. Huang NE, Shen Z, Long SR et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. *Series A: Mathematical, Physical and Engineering Sciences* 1998; 454: 903-995.