Food Predictive Anatomization using Time Series

N. Kumaran, P. Surya Sampath, P. Saran Satya Kumar

Abstract - Time series survey and forecasting upcoming values has been a research focus past years ago. Time series analysis and predict The time-series data finds its importance in various roles of implementation such as business, stock market exchange, weather forecasting, electricity demand, cost and usage of products such as fuels, etc. In this project, a detailed survey of the various techniques applied for forecasting different method of time series datasets are provided. Moving average model and Auto-Regressive Integrated Moving Average model with a case study on food predictive analysis time series data with R software.

Key Words: Moving average model, ARIMA Model, Augmented Dickey-Fuller Test, Kwiatkowski-Phillips Schmidt-Shin test.

I. INTRODUCTION

Time Series data is a special case of time-stamped data. It is alike to a number line. The events are uniformly separated in time in a variety of domains like engineering, research, medicine, and finance. A time-series denoted Y=a + b X. The important steps involved in time series analysis are

a) Identify and model the structure of the time series.
b) Estimating the Parameters
c) Diagnostic Checking with Tests and
d) Forecast the future values in the time series.

A. Noise

All-time series data will have noise or randomness in the data points that are not associated with any explain trends. Noise is unmethodical and is short term.

B. Seasonality

If there are uniform and prophetic various in the series that are correspond with the calendar it could be quarterly, weekly, or even days of the week, then the series includes a seasonality component. It's important to note that seasonality is domain specific, Also, not all time series have a seasonal method, as suitable for audio or video data.

C. Trend

When trend in time series data, it is the data has a durable .which can either be trending in the positive or negative direction. For example of a trend will be a stable growth in a company network usage.

D. Cycle

. This insert business cycle such as profitable turns or expansions, or even audio files which have cycles, but they are not connected to the calendar in the weekly, monthly, or yearly

Revised Manuscript Received on June 25, 2020.

Mr. N. Kumaran, Associate Professor, Dept. of Computer Science Engineering, SCSVMV (Deemed University), Kanchipuram, India. E-mail: kmshre19@gmail.com.

P. Surya Sampath, UG Student, Dept. of Computer Science Engineering, SCSVMV (Deemed University), Kanchipuram, India. E-mail:suryasampath0007@gmail.com

P. Saran satya kumar, UG Student, Dept. of Computer Science Engineering, SCSVMV (Deemed University), Kanchipuram, India. E-mail: saransatyakumar@gmail.com

Time series is an form methods for inspection time series data in order to draw out meaning statistics and other attribute of the data. Time series is the use of a model to predict future values based on previously observed values.. We will reveal different approaches for predicting time series.

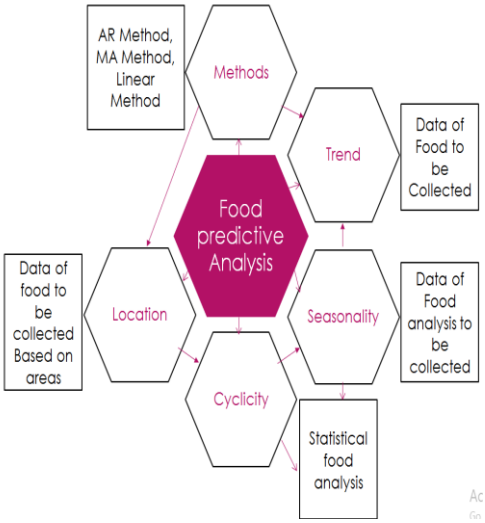

Fig.1. Block Diagram

II. IMPLEMENTATION

The moving average of a period n is a series of successive averages of n terms at a time.

In Multiplicative model

t =St * It * Tt.

To calculate seasonal & irregular components by

St * It = Y t / CMA

Where

St= Seasonality.

It= Irregularity,

CMA = centered moving average,

Y t= Sales,

Tt= Trend.

The ARIMA method is also called a Box-Jenkins methodology. The Box-Jenkins methodology is discuss with fitting a mixed ARIMA model to a given set of data. The main fitting is the ARIMA model is to identify the stochastic process of the time series and predict future values accurately. The main need to implement this are packages named Forecast, for, ex smooth, l most mt, zoo, t series need to be loaded into the R studio.

A. Examining and Changing the Continuous Data Into Time Series Data MOVING AVERAGE METHOD

The data can be understood by using descriptive statistical functions like mean, median, min, max, standard deviation, and summary data. Data in the dataset is in the normal numerical form.

The data we have taken in an excel sheet and done some predicted formulas given above. A Graph and an example are given for an easier understanding of the topic

Table -I: Example of Moving Average Method

				Yt			Yt/CMA		de-seasor	intercept+slope*timecoc	Yk=St*Tt
t		Branch	Quarter	Sales of VEG Curries	MA	CMA	St*It	St	Yt/St	Trend	Predicted(Yk)
1	1	Hotel 1	1	2.8				0.9	3.11	3.21	2.9
2	2		2	2.1				0.77	2.74	3.36	2.58
3	3		3	4	3.4	3.5	1.15	1.14	3.51	3.51	4
4	4		4	4.5	3.6	3.7	1.2	1.21	3.72	3.66	4.43
5	5	Hotel 2	1	3.8	3.9	4	0.96	0.9	4.21	3.81	3.44
6	6		2	3.2	4.1	4.2	0.76	0.77	4.17	3.96	3.04
7	7		3	4.8	4.3	4.3	1.11	1.14	4.22	4.11	4.68
8	8		4	5.4	4.4	4.4	1.23	1.21	4.4	4.26	5.15
9	9	Hotel 3	1	4	4.5	4.5	0.88	0.9	4.43	4.4	3.98
10	10		2	3.6	4.6	4.6	0.77	0.77	4.69	4.55	3.49
11	11		3	5.5	4.7	4.7	1.15	1.14	4.83	4.7	5.35
12	12		4	5.8	4.8	4.8	1.2	1.21	4.79	4.85	5.87
13	13	hotel 4	1	4.3	4.9	4.9	0.87	0.9	4.76	5	4.51
14	14		2	3.9	5	5.1	0.77	0.77	5.08	5.15	3.95
15	15		3	6	5.2			1.14	5.27	5.3	6.03
16	16		4	6.4				1.21	5.29	5.45	6.59
17	17	predicted data	1	?				0.9		5.59	5.05
18	18		2	?				0.77		5.74	4.41
19	19		3	?				1.14		5.89	6.71
20	20		4	?				1.21		6.04	7.31

Table -III: Summary Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.748051175
R Square	0.55958056
Adjusted R Square	0.528122029
Standard Error	0.827451071
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	12.17892116	12.17892	17.78787932	0.000860653
Residual	14	9.585453843	0.684675		
Total	15	21.764375			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.064767	1.261758937	-0.68816	0.50259297	-3.574501827	1.83790571	-3.574501827	1.837905714
t	0.148796	0.287705534	4.217568	0.000860653	0.596350638	1.83048464	0.596350638	1.830484639

Scatter Plot of the MOVING AVERAGE METHOD

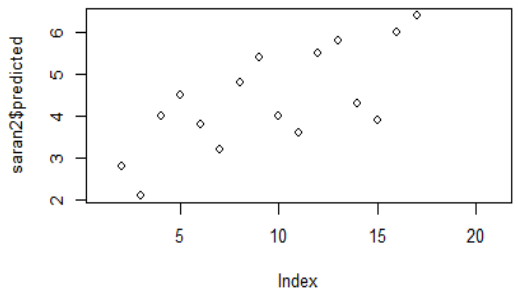

Fig.2. Scatter plot

B. Examining and Converting the Continuous Data into Time Series Data ARIMA MODEL.

So, now data has to be converted into time-series data. ts() function in R can be used for the conversion of data from continuous to time series and stored as an object.

Table -3: Data for ARIMA Model

	A	B	C	D	E	F	G	H	I	J
1	year	tomata	potato	brinjal	ladies finger	soya chunks	leafy vegetables	carrot	cauliflower	
2	2015	500	320	365	220	200	480	420	490	
3	2016	323	420	400	225	235	350	413	450	
4	2017	320	425	470	250	230	300	370	360	
5	2018	335	480	425	300	320	260	320	375	
6	2019	330	488	431	321	380	220	286	392	
7	2020	331	500	438	325	453	200	241	416	
8										

The time series data into R, then store the time series thing in R

>food=ts (food $potato, start=2015)

To understand the time series data effectively we can use a graph. To draw a graph, we use a command called plot function in R studio

> plot (food)

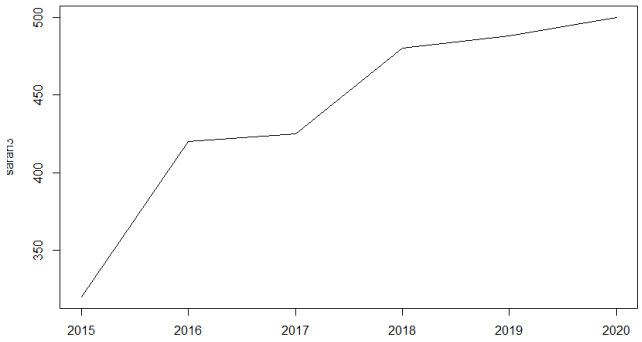

Fig.3. Graph of ARIMA Model

Model fitting for the analysis of the data ARIMA model needs the series data must be stationary. Stationary data consists of they are mean, variance, and auto covariance which are time-invariant. The ADF stands for Augmented Dickey-Fuller. This test is a formal stationary experiment for the data set.

> adf.test (food)

Augmented Dickey-Fuller Test
Data: food
Dickey-Fuller=NaN,
Lag order=1,
P-value=Na
Alternative-hypothesis: stationary

KPSS another unit root test called Kwiatkowski-Phillips-Schmidt-Shin test.

> Kpss. Test (food) # Second Test for Stationary

KPSS Test for level stationery
Data=food
KPSS level=0.4067, Truncation lag parameter=1, p-value=0.07427

We are creating an ARIMA model. ARIMA means Auto-regressive Integrated Moving Average. To create a First ARIMA model, we have a special function called ARIMA function in R studio. This function will fit the data set into ARIMA and creates a model.

>fit=ARIMA (food, order=c (0, 1,1))

This created model called fit is added as an object in the Global Environment of the R studio. ACF stands for Auto Correlation Function. This is used to find whether the data is stationary or not. To draw an ACF plot use a function called ACF () in R studio.

> Acf (diff (food))
>plot (forecast (fit)) # PACF plot of residuals

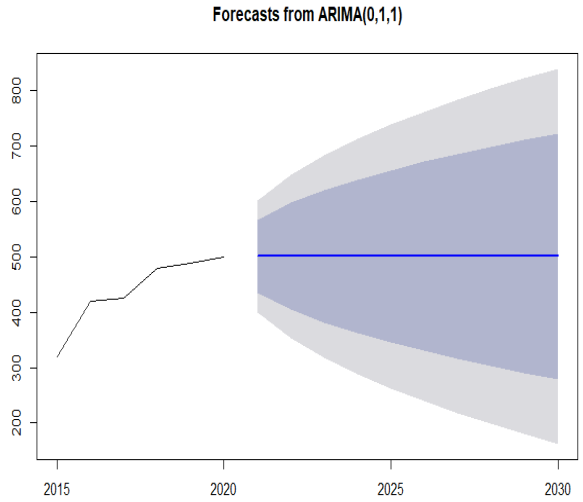

Fig.4. Forecasts from ARIMA Model

> Box. Test (residuals (fit), type="Ljung")
Box-Ljung test data: residuals (fit)
X-squared = 1.9108, df = 1, p-value = 0.1669

C. Creating New Models Using ARIMA

Auto Arima for finding the Arima(p, d, q), p, d, q parameters

> fit1=auto.arima (diff (food))

It generates a new model called fit1 and in addition to the Global Environment window of the R studio

> plot (forecast (fit1)) # Forecast plot
implementing Neural Network on Time series data
> fit2=nnetar (diff (food), maxit=1000)

It creates a new object fit2 and adds to the Global Environment window of an R studio. Forecast plot to plot the historical data with forecasts and prediction of intervals. Drawing a Forecast plot on neural network model fit3 using plot () function.
> Plot (forecast (fit2))

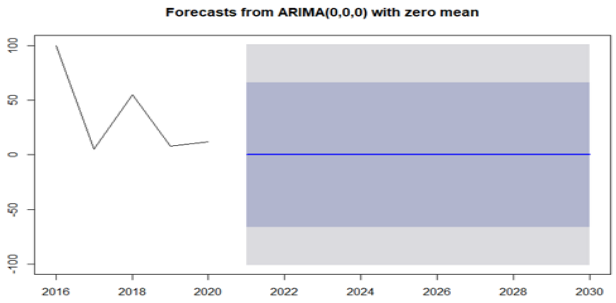

Fig.5. Graph of ARIMA Model

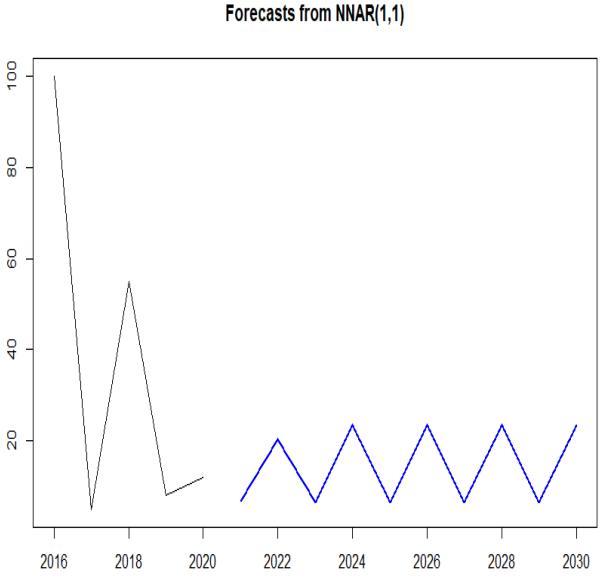

Fig.6. Graph of NNRA Model

To display the summary of the model 2 (Neural network model)

> fit2

Series: diff (food)
Model: NNAR (1, 1)
Call: nnetar (y = diff (food), maxit = 1000)
Average of 20 networks, every network of which is a 4-2-1 network with 13 weights options were - linear output units.
Sigma^2 estimated as 69.96

III. CONCLUSION

Above two models are moving average method and ARIMA method are the best method is ARIMA method .it depends on R-studio or R-Query calculation part is less in the study of Time series analysis on food predictive analysis data, the ARIMA model is used and predicted the values for the after five years. The rationality of the predicted values can be checked when the data for the foremost periods become available. However, the data need to be updated from time to time with absorb of current values.

REFERENCES

1. Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung.: Time series analysis: predicting and control.
2. John Wiley & Sons, Makridakis, Spyros, Michele Hibon, and Claus Moser.: Accuracy of forecasting: An empirical investigation. Journal of the Royal Statistical Society. Series A (General). 97145 (1979)
3. Prabakaran, K. and Sivapragasam, C.: Forecasting areas and production of rice in India using ARIMA model. International Journal of Farm Sciences. 4(1), pp.99-106 (2014)

AUTHORS PROFILE

N. Kumaran Completed his B.E(Computer Science and Engineering) from CIT, Coimbatore, Tamil Nadu, India and M.Tech (Information Technology) from Sathyabama University, Chennai, in the year 1998 and 2007 respectively. Currently pursing PhD (Video Analysis) in NIT, Trichy-15, India. His areas of interest include video processing and Computer Networks. He has got around18 years of teaching experience various Institutions.

P Surya Sampath Currently pursuing his B. E(computer science and engineering) in SCSVMV university ,Kanchipuram, Tamilnadu, India. His area of interest include in networking and data science.

P Saran Satya Kumar, Currently pursuing his B.E(computer science and engineering) in SCSVMV University Kanchipuram, Tamilnadu India. .His area of interest include data science, web development.