Personal Budget Planner Using Time series analysis with ARIMA models

Parth Dubal[1], Omkar Patil[2], Rohit Chavan[3], Gaurav Ambavle[4]

Abstract— Budget planning is very important factor to main-tain expense within the income limit and manage the daily expenses accordingly. As nowadays smartphones are easily accessible to many people, so using a smartphone app they can conveniently look over their expenses, income and budget at the reach of their fingertips. In traditional systems one has to note down their expenses in a catalogue, ledger, diary or handbook, but nowadays people dont have time to note down their expenses in a book. The applications that are currently available on play store basically focuses on creating the graphical representation of users expenditures, maintains log of daily expenses and export data that have been entered in the application. In this project, we are creating an android application Personal budget planner which can be used by the people to maintain and manage their budget easily. It will analyse the daily expenditure and provide you the predictions for the current month. Also the user will get alerts and reminders to be aware about his monthly budget. In this application we are using time series analysis algorithm for prediction purpose as they are very useful models when we have serially correlated data. Time Series Analysis, as the name suggests, involves working on time (years, days, hours, minutes) based data, to derive hidden insights to make informed decision making.

Budget planner, prediction, Time Series Analysis

I. INTRODUCTION

The applications that are currently available in market is basically focuses on creating the graphical representation of users expenditures, maintains log of daily expenses and export data that you have entered in the application. Some application also provides you the facility of payment of your bills and recharge. All the features are not available in one application each contains some pros and cons. various organizations are adopting information technology to assist their jobs by providing adequate storage units to store an up to date information and used that information to the highest benefit with various methods. 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 event. The results were used to predict future events. Time series is the order of historical data, which resembles the group or observation of the data that have been collected over time according to the continuous period of time. That collected data may already be in a daily, weekly, monthly, quarterly, or yearly format, depending on which one is appropriate to use. Time series series used forecasting techniques to identify models from the past data. With the assumption that the information will Resemble itself in the future, we can thus forecast future events from the occurred data. There are several methods of statistical forecasting such as regressing analysis, classical decomposition method, Box, and Jenkins and smoothing techniques. These techniques pro-vide forecasting models of different accuracy. The accuracy of the prediction is based on the minimum error of the forecast. The appropriate prediction methods are considered from several factors such as prediction interval, prediction period, characteristic of time series, and size of time series

[2]. There were two forecasting techniques that were used in this study; Autoregressive Integrated Moving Average (ARIMA) and Autoregressive Moving Average (ARMA). We applied these methods for detecting patterns and trends of the daily expenditure of an individual with real time series period in daily, weekly, monthly, and quarterly [1]. We used python [4], [5], for constructing the model [2], [6]. The most suitable forecasting method and the Section 2 contains the literature review. In Section 3, we present details about the computational approach for deriving time series model. Results of our system is given in Section 4. Finally, we conclude work in Section 5.

II. LITERATURE REVIEW

The current various researches have used the method of forecasting with time series data such as the electric power consumption. Saab and colleagues [] studied the forecasting method for monthly electric energy consumption in Lebanon. They used two different univariate modeling methods namely, ARIMA and AR(1) with a highpass filter. The best forecasting method for this particular energy data was AR(1) highpass filter model. In [4] Zhu, Guo, and Feng studied the issue of household energy consumption in China from the year 1980 to 2009 with construction VAR model. There were two forecasting methods that used ARIMA and BVAR. The Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models Pasapitch Chujai*, Nittaya Kerdprasop, and Kittisak Kerdprasop C Pro-ceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I, IMECS 2013, March 13 - 15, 2013, Hong Kong ISBN: 978-988-19251-8-3 ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online) IMECS 2013 results showed that both of them can predict the sustained growth of household energy consumption (HEC) trends.

Features	Daily Expense -3	My Budget Book	Andro money	Money View	Pro- posed System
Expense Log	Yes	Yes	Yes	Yes	Yes
Monthly Analysis	Yes	Yes	Yes	Yes	Yes
Cloud Storage	No	No	No	No	No
Alerts and Reminders	Yes	Yes	No	Yes	Yes
Expense Prediction	No	No	No	No	Yes
Free	No	No	No	No	Yes

Table 2.1:Comparison Table

III. OUR SYSTEM

In this research, our main objective was to find a model to efficiently predict the personal budget of a user. The suitable forecasting methods were chosen for finding the method that was suitable for short term analysis in daily, weekly, monthly, and quarterly.

a) : For prediction we have used Time Series Analysis to get more accurate results. Following are the steps that we need to perform for the time series analysis:

Step 1: Visualize the Time Series It is essential to analyze the trends prior to building any kind of time series model. The details we are interested in pertains to any kind of trend, seasonal- ity or random behaviour in the series. We have covered this part in the second part of this series.

Step 2: Stationarize the Series Once we know the patterns, trends, cycles and seasonality , we can check if the series is stationary or not. Dickey Fuller is one of the popular test to check the same. We have covered this test in the first part of this article series. This doesnt ends here! What if the series is found to be non-stationary? There are three commonly used technique to make a time series stationary: 1. Detrend- ing : Here, we simply remove the trend component from the time series. For instance, the equation of my time series is: x(t) = (mean + trend * t) + error Well simply remove the part in the parentheses and build model for the rest.

2. Differencing : This is the commonly used technique to remove non-stationarity. Here we try to model the differences of the terms and not the actual term. For instance, x(t) x(t-1) = ARMA (p , q) This differencing is called as the Inte-gration part in AR(I)MA. Now, we have three parameters p : AR d : I q : MA 3. Seasonality : Seasonality can easily be incorporated in the ARIMA model directly. More on this has been discussed in the applications part below.

Step 3: Find Optimal Parameters The parameters p,d,q can be found using ACF and PACF plots. An addition to this approach is can be, if both ACF and PACF decreases gradually, it indicates Department of Information Technology, DBIT, Mumbai 5 LIST OF TABLES that we need to make the time series stationary and introduce a value to d.

Step 4: Build ARIMA Model With the parameters in hand, we can now try to build ARIMA model. The value found in the previous section might be an approximate estimate and we need to explore more (p,d,q) combinations. The one with the lowest BIC and AIC should be our choice. We can also try some models with a seasonal com- ponent. Just in case, we notice any seasonality in ACF/PACF plots.

Step 5: Make Predictions Once we have the final ARIMA model, we are now ready to make predictions on the future time points. We can also visualize the trends to cross validate if the model works fine. Stationary Series There are three basic criterion for a series to be classified as stationary series : 1. The mean of the series should not be a function of time rather should be a constant.. 2. The variance of the series should not a be a function of time. This property is known as homoscedasticity. 3. The covariance of the i th term and the (i + m) th term should not be a func- tion of time. Why do I care about stationarity of a time series? In cases where the stationary criterion are violated, the first requisite becomes to stationarize the time series and then try stochastic models to predict this time series.

For prediction we have used python programming for the implementation, there is two main model we have used

1.Pandas: Pandas has dedicated libraries for handling TS objects, particularly the data- time64[ns] class which stores time information and allows us to perform some operations really fast
import pandas as pd

2.STATSMODEL: Stats model is a Python module that provides classes and functions for the esti- mation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration. An extensive list of result statistics are available for each estimator. The results are tested against existing statistical packages to ensure that they are correct. statsmodels.tsa.arima m odelimportARIMA

b) : For further processing we need to eliminate the trend and seasonality

Differencing

One of the most common methods of dealing with both trend and seasonality is differencing. In this technique, we take the difference of the observation at a particular instant with that at the previous instant. This mostly works well in improving stationarity.

x(t) x(t-1) = ARMA (p , q)

This differencing is called as the Integration part in AR(I)MA. Now, we have three parameters

p : AR; d : I ; q : MA

Detrending

Here, we simply remove the trend component from the time series. If the trend is deterministic (e.g. a linear trend) you could run a regression of the data on the deterministic trend (e.g. a constant plus time index) to estimate the trend and remove it from the data. If the trend is stochastic you should detrend the series by taking first differences on it.

1.After testing the data of student the accuracy that we achieved was : 75.87 percent

2.After testing the data of house wife the accuracy that we achieved was: 73.20 percent

	A	B	C	D
1	PREDICTED	EXPECTED	DIFFERENCE	Accuracy
2	446.903248	400.01	46.893248	88.27698108
3	389.07641	500.01	110.93359	77.81372573
4	434.127574	400.01	34.117574	91.47081973
5	466.26055	670.01	203.74945	69.59008821
6	405.376537	480.01	74.633463	84.4516858
7	593.402163	580.01	13.392163	97.69104619
8	511.701724	650.01	138.308276	78.72213104
9	596.633646	520.01	76.623646	85.26496683
10	674.036042	250.01	424.026042	69.60363265
11	604.825714	150.01	454.815714	203.1902633
12	132.575192	56.01	76.565192	36.69914658
13	102.015496	110.01	7.994504	92.73292973
14	185.20973	80.01	105.19973	31.4832271
15	205.613106	230.01	24.396894	89.39311595
16	258.484366	170.01	88.474366	47.95931651
17	268.718424	320.01	51.291576	83.97188338
18	144.051564	210.01	65.958436	68.59271654
19	126.382605	160.01	33.627395	78.98419161
20	175.329295	140.01	35.319295	74.77373402
21	123.323955	200.01	76.686045	61.65889456
22	107.996727	50.01	57.986727	15.95026395
23	412.86706	170.01	242.85706	42.84869125
24	370.246982	500.01	129.763018	74.04791544
25			OVERALL	
26			ACCURACY	75.87701596
27				

Above table shows the accuracy obtained for sample Data1, which contains the data of the housewife. As a housewife tends to spend more in the beginning of than the ending of the month, Hence the pattern of expenditure is different.

	A	B	C	D	E
31	20.01	45	24.99	124.8875562	
32	15.01	25	9.99	66.55562958	
33	90.01	110	19.99	22.20864348	
34	110.01	130	19.99	18.17107536	
35	150.01	145	5.01	3.339777348	
36	30.01	40	9.99	33.2889037	
37	25.01	35	9.99	39.94402239	
38	15.01	20	4.99	33.24450366	
39	10.01	35	24.99	249.6503497	
40	50.01	61	10.99	21.97560488	
41	17.01	35	17.99	105.7613169	
42	16.01	25	8.99	56.15240475	
43	80.01	110	29.99	37.48281465	
44	50.01	90	39.99	79.9640072	
45	30.01	40	9.99	33.2889037	
46	20.01	40	19.99	99.90004998	
47	20.01	40	19.99	99.90004998	
48	15.01	40	24.99	166.4890073	
49	50.01	90	39.99	79.9640072	
50	25.01	40	14.99	59.93602559	
51	10.01	40	29.99	299.6003996	
52	15.01	25	9.99	66.55562958	
53	120.01	180	59.99	49.98750104	
54	150.01	165	14.99	9.992667156	
55	180.01	110	70.01	38.89228376	
56	25.01	50	24.99	99.92003199	
57	10.01	30	19.99	199.7002997	
58	80.01	95	14.99	18.73515811	
59	110.01	170	59.99	54.53140624	
60					
61			Overall accuracy	73.20670492	
62					

Above table shows the accuracy obtained for sample Data2,which contains the data of a student. As the expenditure pattern differ from individual to individual, a student spends specific amount on daily basis for travelling etc. that has different pattern.

IV. CONCLUSIONS

This paper proposed a method to analyze and predict the personal budget of the user. To approach this method, process model-enhanced expense and expense prediction technique were presented. Process model-enhanced executed the detail expense of each user activity in a process.thus we have implemented the basic structure of the proposed system which can store your personal expenses, show you the graphical representation and predict user's expense upto some extent.

V. CONCLUSIONS

In the future work we are aiming to add SMS retrieval module where user's online transaction will be reflected in the application this will reduce manual input work. We will provide the feature of Bill Splitting among the group. We will be adding a user's wallet functionality that will allow user to send and receive money from other user through application. We will also provide feature of saving goals where user will have his saving goal log and he/she plan future savings accordingly.

REFERENCES

[1] https://machinelearningmastery.com/time-series-forecast-study-python-monthly-sales-frenc h-chapagne/

[2] https://stackoverflow.com Belmont, CA: Wadsworth, 1993, pp.123135.

[3] https://developer.android.com

[4] Pasapitch Chujai*, Nittaya Kerdprasop, and Kittisak Kerdprasop,Time Series Analysis of Household Electric Consumption with ARIMA and ARMA ModelsInternational MultiConference of Engineers and Computer Scientists 2013 Vol I,IMECS 2013, March 13 - 15, 2013, Hong Kong

[5] K. V. Sujatha S. Meenakshi Sundaram,Stock Index Prediction Using Regression and Neural Network Models under Non Normal Con- ditions,Research Scholar Sathaybama University and Department of Mathematics Sathyabama University Tamilnadu, India

[6] Thi Bich Hong Tu,Minseok Song,Analysis and Prediction Cost of Manufacturing Process Based on Process Mining,School of Business Administration Ulsan National Institute of Science and Technol-ogy Ulsan and Department of Industrial Management Engineering POSTECH Pohang Univ. of Sci. Tech) Pohang, South Korea