2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)
Zeal Education Society, Pune, India, Feb 24-26, 2017

Forecasting of sales by using fusion of Machine Learning techniques

Mohit Gurnani , Yogesh Korke[y], Prachi Shah[z], Sandeep Udmale[x], Vijay Sambhe[{], and Sunil Bhirud[k]

Department of Computer Engineering and Information Technology, VJTI, Mumbai, India Email:
mohitgurnani1@gmail.com, [y]yogesh.korke@yahoo.in, [z]prachirshah1994@gmail.com, [x]ssudmale@vjti.org.in,
[{]vksambhe@vjti.org.in, [k]sgbhirud@vjti.org.in

Abstract—**Forecasting is an integral part of any organization for their decision-making process so that they can predict their targets and modify their strategy in order to improve their sales or productivity in the coming future. This paper evaluates and compares various machine learning models, namely, ARIMA, Auto Regressive Neural Network(ARNN), XGBoost, SVM, Hy-brid Models like Hybrid ARIMA-ARNN, Hybrid ARIMA-XGBoost, Hybrid ARIMA-SVM and STL Decomposition (using ARIMA, Snaive, XGBoost) to forecast sales of a drug store company called Rossmann. Training data set contains past sales and supplemental information about drug stores. Accuracy of these models is measured by metrics such as MAE and RMSE. Initially, linear model such as ARIMA has been applied to forecast sales. ARIMA was not able to capture nonlinear patterns precisely, hence nonlinear models such as Neural Network, XGBoost and SVM were used. Nonlinear models performed better than ARIMA and gave low RMSE. Then, to further optimize the performance, composite models were designed using hybrid technique and decomposition technique. Hybrid ARIMA-ARNN, Hybrid ARIMA-XGBoost, Hybrid ARIMA-SVM were used and all of them performed better than their respective individual models. Then, the composite model was designed using STL Decomposition where the decomposed components namely seasonal, trend and remainder components were forecasted by Snaive, ARIMA and XGBoost. STL gave better results than individual and hybrid models. This paper evaluates and analyzes why composite models give better results than an individual model and state that decomposition technique is better than the hybrid technique for this application.**

Index Terms—**ARIMA, Auto Regressive Neural Network, Sales Forecasting, STL decomposition, SVM, XGBoost**

I. Introduction

Time series forecasting has been the basis for the study of the behavior of any process for over a period of time. The forecasts are generated using the flow of demands from the past as well as by considering other known factors in future. Various machine learning models are developed for the same. The purpose of forecasting sales is mainly to help the organization predict their targets and modify their strategy to improve their productivity in the coming future.

ARIMA model has been used in [20], [21] for prediction of Infant Mortality Rate (IMR) and automobile demand prediction. ARIMA has a limitation that it can deal only with linearity. Our paper proposes hybrid techniques which over-come this drawback thereby producing more reliable forecast.

Auto Regressive Neural Network (ARNN) using Back Propagation algorithm has been used in [17], [18] to forecast number of infectious diarrhea and 24-hour electricity prices respectively. It has been shown in [18] that the performance of ARNN is better in the absence of external regressors. But in our paper, external regressors have improved the performance because they greatly influence the predictor variable and have high correlation between them.

XGBoost is used for binary classification in supervised machine learning system capable of matching internet devices to web cookies in [22]. XGBoost is used in [23] to improve the marketer's ability to identify individual users as they switch between devices and show relevant content/recommendation to users wherever they go. Our paper evaluates the performance of XGBoost for regression objective by modelling time series components and behaviour.

Support Vector Machine (SVM) has been used for time series forecasting in [13] which states that SVM gives accurate forecast as compared to ARNN for power load forecasting. However, in our case, artificial neural network outperforms SVM.

A hybrid of ARIMA and ARNN has been used in [5], [24] for prediction of short-term electricity prices, forecasting resource consumption. From [24] it is inferred that hybrid model performed best for both short and long term prediction. Results from [5] showed that the neural network predictor was able to provide much higher prediction accuracy than hybrid model. Another Hybrid model used is of ARIMA and SVM. [15] and [16] clearly state that this hybrid model gives better accuracy than the individual models. Our paper evaluates precisely on the success criteria of a hybrid model.

STL decomposition has been used as a part of [6] to detect abnormal event detection using Spatiotemporal social media data. Additive and Multiplicative model has been used in [19] to forecast electricity load. Our data reveals that seasonal variation is roughly

978-1-5090-4083-4/17/$31.00 ©2017 IEEE

constant, hence Additive model has been used for STL Decomposition in our paper.

Forecasting has various applications in the area of Sales, Weather, Stock Market, Electricity, Networking, Transportation and Meteorology.

This paper analyzes various models and their composite formed by Hybrid and Decomposition techniques.The paper is structured as follows. Section II describes various machine learning models. Section III contains their design and implementation results. All the individual and composite models are analysed and compared in section IV. Final conclusions are included in section V

II. Methodology

This section describes various prediction models and their composite models designed by hybrid and decomposition technique. [11] contains basics of forecasting and various machine learning models.

A. Forecasting Models

Initially, various linear models have been applied and it is found that ARIMA performs better than any other linear model. This is mainly because of ARIMA's dynamic regression. Then, among nonlinear models, ARNN, Xgboost and SVM will be discussed.

1) ARIMA: ARIMA is one of the prediction models that is used to forecast future values by finding the relationship between past values and past error. Its internal working is described in [1], [12].

ARIMA is composed of 3 parts namely Auto Regressive(AR), Integrated(I) and Moving Average(MA):

1) Auto Regressive(AR) is used to predict future values by extracting the influence of past values.
2) Moving Average(MA) is used to predict future values by extracting the influence of past errors.
3) Integrated(I) means differencing the series (i.e. subtracting time series with its lagged series to extract trend) to make the time series stationary.

The appropriate order of ARIMA model i.e. p,d,q is determined with the help of ACF and PACF plots. Many candidate models are selected in this section. After determining the order, parameters are estimated using least square estimation, Yule-Walker estimation or Maximum likelihood es-timation. After parameter estimation, best candidate model is selected based on minimum AICc value [3].

2) Auto Regressive Neural Network: Neural network is a prediction model that is based on the working of the mathematical model of a brain. It is capable of finding complex nonlinear hidden relations between the response variable and predictor variable.

It consists of the set of neurons which are arranged in the form of layers. Input layer accepts input i.e. the training data, Hidden layer helps in determining the complex nonlinear relationship and Output layer gives the result.

Neurons in a layer are connected with neurons in next layer by coefficients also known as weights. These weights are self-adjusted and retrained each time a new input is fed into the network which helps in accurate capture of patterns in data. Internal working of Neural network is given in [4].

Auto Regressive Neural Network(ARNN) has been used to forecast time series because they accept past values as input and predicts the future values. ARNN uses back propagation algorithm to train the network. It allows inclusion of external regressors as well.

3) XGBoost: Xgboost is short for eXtreme Gradient Boost-ing. It is based on gradient boosting framework. Gradient boosting is the machine learning technique to deal with classification, regression and ranking problems. Xgboost has regularized model formalization to control overfitting, which boost its performance. It provides a good result for most of the datasets involving linearity and nonlinearity. It is efficient as it supports parallel computation on a single machine.

Model of Xgboost is based on the concept of Gradient boosting which believes that single trees are not strong enough to give accurate prediction. Hence, the ensembles of decision trees are used, where trees are added in such a way that they optimizes the current error. The tree ensemble model uses classification and regression trees (CART). It optimizes the result by using objective functions with certain set of parameters [9], [10].

Objective functions consist of two parts:

1) Training loss (L)
2) Regularization $^{1}!^{o}$

And it is represented as,

$$Ob\,j^{1\,o} = L^{1\,o} + {}^{1\,o} \qquad (1)$$

where is used to indicate parameters. The training loss is used to measure predictive power of model on training data. Mostly used training loss measure is mean squared error (MSE). The regularization controls the complexity of the model to prevent overfitting.

Mathematical model can be written in the form:

$$y\hat{}_{i} = \overset{K}{\underset{k=1}{\mathbf{O}}} f_{k}\,{}^{1}x_{i}{}^{o};\ f_{k}\ 2\ F \qquad (2)$$

where "K" is the number of trees, "f " is a function in the functional space "F", and "F" is the set of all possible CARTs. Thus, objective to optimize can be represented as

$$Ob\ j^{1\ o} = \overset{n}{\underset{i}{\hat{o}}}\ l^1yi;\ y\hat{\ }i^o +^1 \overset{K}{\underset{k=1}{\hat{o}}} fk\ ^o \qquad (3)$$

4) Support Vector Machine: Support Vector Machine(SVM) can be used for classification or regression problems. SVM constructs a hyperplane or a set of hyperplanes in n dimensional space. It selects the hyperplane which has the largest margin. Where, margin is the sum of the shortest distance from the separating hyperplane to the closest data point of all categories. Such a hyperplane is likely to correctly classify unknown test data points [13].

To support non-linear classification, SVM does the mapping from input space to feature space. The kernel method is used for doing this. By using this, linear classification in the new space (the feature space) becomes equivalent to nonlinear classification in the original space (the input space). Here, SVR(Support Vector Regression) has been used for time series forecasting which works on the same principle.

In SVM, the input vector x is mapped to the high dimensional feature space to enhance linear separability using the mapping function $^1x^o$. The decision function if training data is linearly separable after mapping into feature space, is:

$f\ ^1x^o = w^T\ ^1x^o + b$ (4) where w=weight factor, b=bias term, $y^1x_i^o\ f\ ^1x_i^o > 0$ for i=1,...,M

B. Hybrid Approach

A question might arise to you, what is the need of a hybrid model. Well, linear models are not able to capture nonlinear patterns accurately, hence to improve the prediction result, their residue (which contains nonlinear pattern) is forecasted by nonlinear model. This forecast result is added to the forecast obtained by the linear model in order to get the resultant forecast of time series [3], [7], [8]. Hence, hybrid approach takes into account a linear and a nonlinear model for better results.

Time series can be represented as

$$Y_t = L_t + N_t \qquad (5)$$

where Y_t is time series, L_t is Linear component of time series and N_t is its nonlinear component.

The Working of hybrid model can be described below:

1) Initially, linear model like ARIMA is applied to time series(Y) which yields forecast. This forecast is considered as linear forecast because ARIMA captures Linear component very well as compared to nonlinear patterns. It is represented as $F^1L_{t+h}{}^o$
2) Since, ARIMA doesn't capture nonlinear patterns prop-erly, hence its residue contains nonlinear component which is given by

$$E_t = Y_t \quad F^1Y_t\ ^o \qquad (6)$$

where E_t is residue, Y_t is Actual Values of Training Set and $F^1Y_t\ ^o$ is Forecast Values of Training Set
3) Residue obtained by ARIMA is applied to nonlinear models like Neural Network, XGBoost and SVM to obtain forecast of nonlinear patterns missed by ARIMA. This forecast is represented as $F^1N_{t+h}{}^o$
4) At the end, resultant forecast is obtained by adding forecast of both linear model and a nonlinear model. It is given below

$$F^1Y_{t+h}{}^o = F^1L_{t+h}{}^o + F^1N_{t+h}{}^o \qquad (7)$$

where $F^1Y_{t+h}{}^o$ is resultant forecast of time series in Test set, $F^1L_{t+h}{}^o$ is forecast of linear component of test set and $F^1N_{t+h}{}^o$ is forecast of nonlinear component of test set.

Step 2 and 3 can be represented in the flowchart given below in Figure 1.

C. Time series decomposition using STL

Time series exhibit a variety of patterns, it is necessary to categorize them so that these categories can be studied separately and analyzed. Hence it is decomposed into several components that represents one of these categories. These components are separately studied and suitable machine learn-ing algorithms are fit into these components to produce a reliable forecast. Forecast obtained from all the individual components is combined to produce the resultant forecast of time series [6].

Some of these components obtained after decomposition are:

Trend: long-term increase or decrease in data

Cyclic: rise and fall of data that are not of fixed period

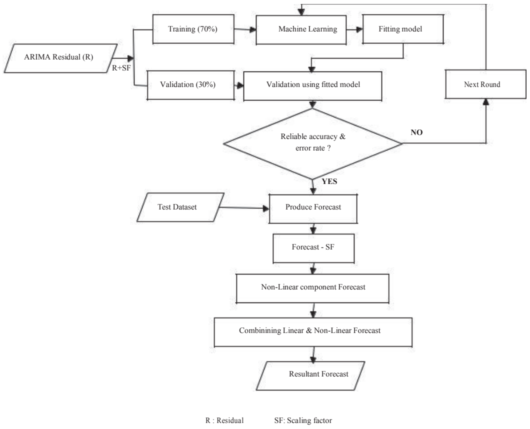

Fig. 1. Hybrid model execution flowchart to forecast Nonlinear component

Season: a specific pattern that reoccurs after fixed time period

Remainder: does not exhibit any specific behavior

Additive model of decomposition is expressed as

$$Y_t = S_t + T_t + E_t \tag{8}$$

Multiplicative model is expressed as

$$Y_t = S_t \quad T_t \quad E_t \tag{9}$$

where Y_t is time series, S_t is the seasonal component, T_t is the trend-cycle component and E_t is the remainder component. Additive models are used if the magnitude of seasonal fluctuation or variations in trend-cycle component does not change with time. If variations in trend-cycle component increase with time, then a multiplicative model is used.

There are various methods to extract these components from time series like Moving average is used to extract the trend component.

Various decomposition techniques are classical decomposition, X-12-ARIMA decomposition and STL Decomposition. STL is superior among them, hence it has been used in this paper. STL stands for "Seasonal and Trend decomposition using Loess". It can handle data of any seasonality. It allows the seasonal component to vary as per user. It allows the user to control the smoothness of trend-cycle component. It is robust to outliers. STL has some disadvantage like it does not handle calendar variation, it is suitable for additive decomposition. To get multiplicative decomposition, logarithm of data has to be taken or Box-cox transformation has to be applied.

After obtaining different components by decompositio components are fitted into different machine learning m obtain respective forecasts. These forecasts are added tc final forecast of time series which can be shown ir equation.

$$F^1Y_{t+h}^o = F^1S_{t+h}^o + F^1T_{t+h}^o + F^1R_{t+h}^o \tag{10}$$

where $F^1Y_{t+h}^o$ represents forecast of time series Y_t , $F^1S_{t+h}^o$ is forecast of seasonal component, $F^1T_{t+h}^o$ represents forecast of trend component and $F^1R_{t+h}^o$ represents forecast of remainder component.

Hence, STL decomposition will be used to forecast the sales.

III. Implementation & Results

This section describes the implementation strategy and results in graphical and tabular format.

A. Forecasting models

Results of all the individual model namely ARIMA, ARNN, XGBoost and SVM are given below.

1) ARIMA: Dataset displayed weekly seasonality, so fre-quency of 7 has been used in the prediction process. To en-hance the performance of ARIMA, external regressors namely School holiday, Promo, open, dayofweek, day, month, year have been used.

Performance measure of various ARIMA models has been computed which is shown below in Table I.

TABLE I
Performance measures for ARIMA Models

Models	MAE	RMSE
ARIMA(2,0,1)(2,0,0)[7]	562.7	827.4
ARIMA(2,0,2)(2,0,0)[7]	559.13	822.7
ARIMA(1,0,1)(1,0,0)[7]	578	822
ARIMA(1,0,1)(2,0,2)[7]	**481**	**771**
ARIMA(1,0,1)(3,0,0)[7]	560	834

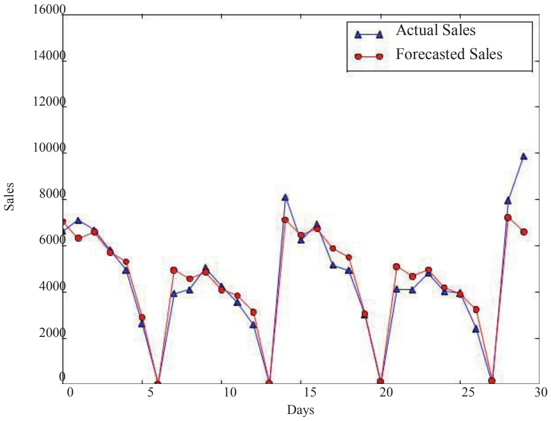

Fig. 2. Actual vs ARIMA Prediction [Test Set]

ARIMA model (1,0,1)(2,0,2)[7] gave the lowest error value (RMSE=771 and MAE=481) as compared to other candidate models. Also, it has lowest AICc (=14177) value. Hence, this model is chosen for prediction. Order of this model is d=0 that represents data is already stationary. Fig 2 depicts the forecasted sales by ARIMA model and actual sales , where the horizontal axis represents the days and the vertical axis represents sales. By analyzing the graph, it can be concluded that ARIMA has successfully captured most of the linear patterns involved in the series but nonlinear patterns are not properly captured.

3) *ARNN:* The Auto Regressive Neural Network (ARNN) has been used whose configuration is shown in Fig.3 Different ARNN models have been analyzed out of which ARNN with 36 neurons in input layer, 18 neurons in hidden layer and 1 neuron in output layer gives out the best result. 28 past

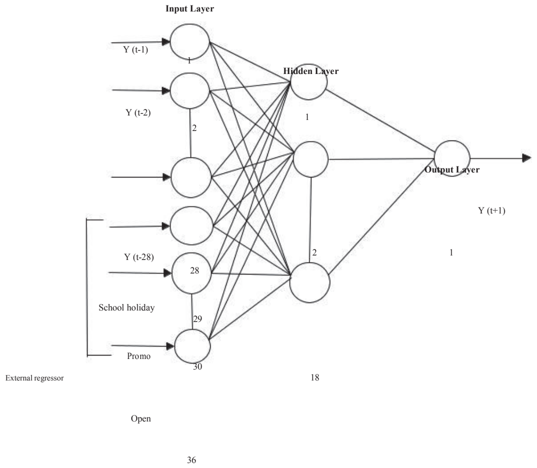

Fig. 3. Auto Regressive Neural Network

TABLE II

Performance Measure of ARNN, XGBoost and SVM

Performance Measure	MAE	RMSE
ARNN	413.47	565.58
XGBoost	346.4	670
SVM	380.6	666.7

observations and 8 external regressors have been passed as input to the network. The Result is obtained by taking average of 25 such ARNN networks with the same configuration.

Table II shows performance measure of ARNN(36,18,1). It can be seen that, an average of 25 such ARNN network has yielded prediction result of RMSE 565.58 and MAE 413.47 which is lower than ARIMA.

Fig 4 plots actual sales and ARNN prediction sales. It can be seen that ARNN has captured seasonal and trend pattern accurately along with nonlinearity involved in it except for some initial forecasts.

3) XGBoost: XGboost has produced forecast by computing 596 iterations. Xgboost uses the concept of regularization which helps to control overfitting so that performance of model remains consistent on test dataset as well. Table II shows performance measure of Xgboost. Its RMSE is 670 and MAE is 346.4. Figure 5 shows how well has xgboost captured seasonal, trend-cycle and nonlinear patterns. XGBoost is able to capture these patterns better than ARNN, still it has greater RMSE because RMSE gives more weightage to large magni-tude errors. Here, xgboost has large errors in last 2 days. It is unable to capture patterns properly at the initial and end days.

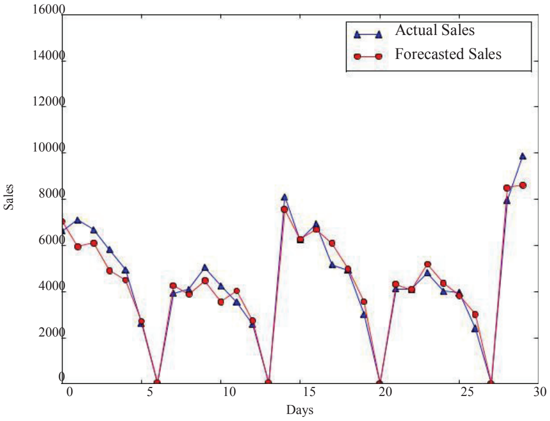

Fig. 4. Actual vs ARNN Prediction[Test Set]

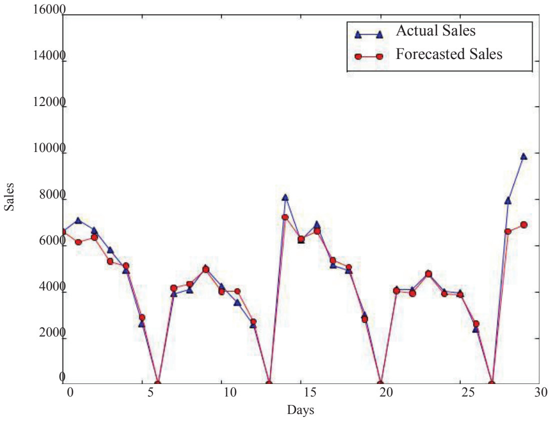

Fig. 5. Actual vs Xgboost Prediction

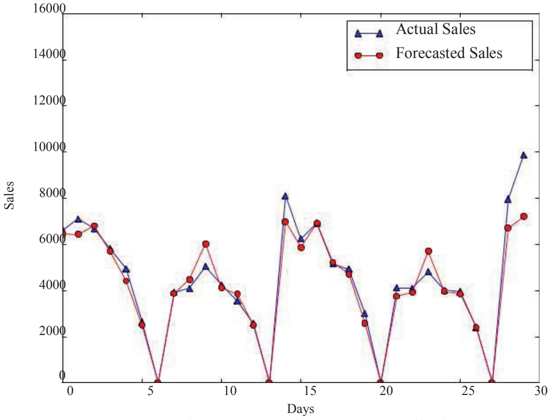

Fig. 6. Actual vs SVM Prediction

4) SVM: Table II shows Performance measure of SVM. Forecasts are computed using optimal combination of epsilon, cost and gamma parameters which yielded the MAE 380.6 and RMSE 666.7.

Figure 6 shows how well has SVM captured linear as well as nonlinear patterns accurately except for some of the peaks. Structural risk minimization is taken care by SVM. Improper selection of the parameters leads to overfitting or underfitting of the training data.

B. Hybrid Approach

In Hybrid Approach, Forecasts are computed using ARIMA (1, 0, 1) (2, 0, 2) with frequency 7 and later nonlinear models like ARNN, XGBoost and SVM are applied on residue of ARIMA. Resultant forecast is obtained by adding linear and nonlinear forecast.

It can be verified from Table III that all 3 hybrid models i.e. Hybrid ARIMA-ARNN, Hybrid ARIMA-XGBoost and Hybrid ARIMA-SVM performs better than their individual models.

TABLE III
Comparison of Performance Measure

Performance Measure	**MAE**	**RMSE**
ARIMA	481	771
ARNN	413.4	565.5
XGBoost	346.4	670
SVM	380.6	666.7
Hybrid ARIMA-ARNN	372.8	530.4
Hybrid ARIMA-XGBoost	357	540.5
Hybrid ARIMA-SVM	407.4	610
STL Decomposition	328.6	426.4

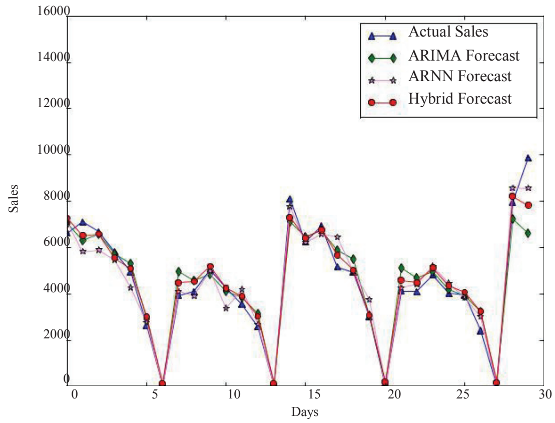

Fig. 7. Comparison of ARIMA, ARNN, Hybrid ARIMA-ARNN forecast with actual series

ARNN has very close RMSE to Hybrid ARIMA-ARNN. Thus, ARNN is capable of capturing the linear and nonlinear patterns precisely as the hybrid model.

Figure 7 shows ARIMA series, ARNN series, Hybrid series and actual data. Hybrid ARIMA-ARNN has captured seasonal and trend-cycle pattern accurately along with nonlinearity included in it and fits the time series better than its individual models.

Similarly, fig. 8 and fig. 9 shows that Hybrid ARIMA-XGBoost and Hybrid ARIMA-SVM have captured the linear and nonlinear patterns in time series better than their individ-ual models.

C. STL Decomposition

Initially, Time series is decomposed by using STL Decomposition into 3 parts i.e. Seasonal, Trend and Remainder component. Each of these component is analysed separately and are forecasted using different machine learning models.

STL has two important parameters, seasonal window and trend window. They help in adjusting smoothness of seasonal and trend component which directly influences remainder component thereby giving precise control to machine learning model so that they can fit accurately.

The seasonal component was applied to ARNN, ARIMA and Seasonal naive (Snaive) model. Among them, Snaive gave lowest RMSE of 28.3. Trend component was applied to ARIMA, ARNN and HoltWinter models. ARIMA gave lowest RMSE of 185.61. The remainder component was applied to ARNN and XGBoost models. XGBoost gave lowest RMSE of 436.31

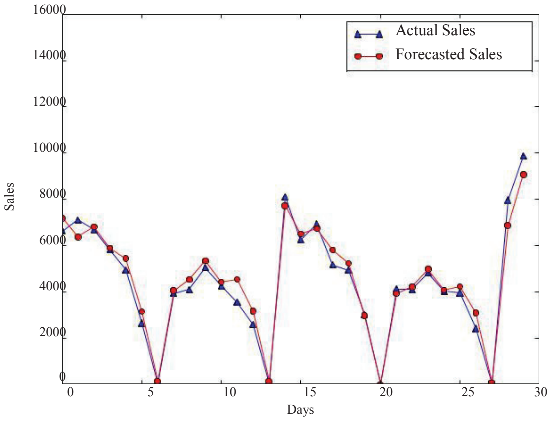

Fig. 8. Comparison of ARIMA,XGBoost,Hybrid ARIMA-XGBoost forecast with actual series

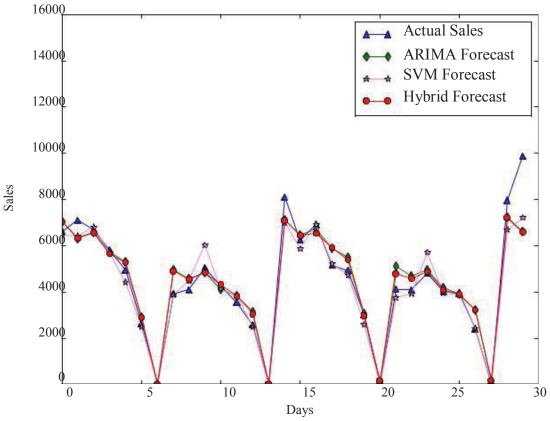

Fig. 9. Comparison of ARIMA,SVM,Hybrid ARIMA-SVM forecast with actual series

Then forecast obtained by all 3 components is added as shown in equation 10. Performance measure of all components and resultant time series can be seen in Table IV. RMSE of STL Decomposition model is 426.45 which has been the lowest as compared to all the models and their hybrids. Also it can be verified from graph 10 that STL has captured all the linear and nonlinear patterns accurately.

TABLE IV
Performance Measure of different components of STL Decomposition

Performance Measure	MAE	RMSE
Seasonal Component(by Snaive)[Test Set]	18.95	28.36
Trend Component (by ARIMA)[Test Set]	182.34	185.61
Remainder Component (by XGBoost)[Test Set]	300.79	436.31
Resultant Time Series [Test Set]	328.61	426.45

IV. Analysis of Models

This section describes the analytical comparison of various results discussed above.

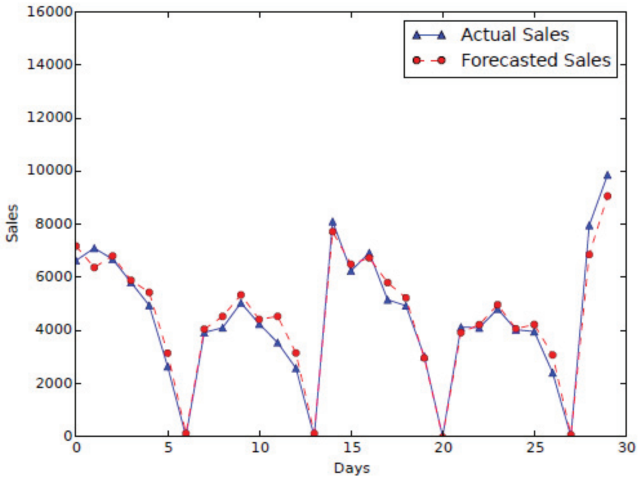

Fig. 10. Actual vs STL Decomposition Prediction

A. Forecasting Models

ARIMA is one of the best linear model that provides a good short-term forecast. However, it doesn't work well while capturing the nonlinearities. Hence nonlinear models like ARNN, XGBoost and SVM are used.

By analyzing table III, certain conclusions can be drawn

ARIMA has lowest performance measure as compared to ARNN, XGBoost and SVM because of its incapability to capture nonlinear patterns properly.

ARNN outperforms all other models because of its in-herent capability to capture complex hidden nonlinear relations.

Performance measure of Svm and XGBoost is equivalent and both of them are better than ARIMA

All the above conclusions are drawn by considering RMSE as source of truth. RMSE is preferred in this paper because large magnitude errors are undesirable. Thus, it can be concluded that among all the individual models, ARNN gives best forecasting results.

B. Hybrid Approach

Since, Hybrid approach takes advantage of both linear and nonlinear models, hence it should give better results than both the individual models. From Table III, it can be verified that performance measure of hybrid is better than individual models. In case of ARNN, its performance is similar to its hybrid ARIMA-ARNN. The reason behind this is, ARNN is itself able to capture linear and nonlinear patterns as accurately as hybrid model.

In all the hybrid models, the linear model is same. Hence, all 3 hybrid models can be compared by analysing how well they capture nonlinear patterns from residue of ARIMA.

By analyzing fig 11, certain conclusions can be drawn:

Curve of actual residue is very random and nonlinear.

SVM forecast series failed to capture random and nonlin-ear patterns. Hence, its performance is worst among all three hybrid models and should not be used.

ARNN forecast series is able to capture random walk pattern but it is not able to capture peaks properly. Its performance is better than SVM.

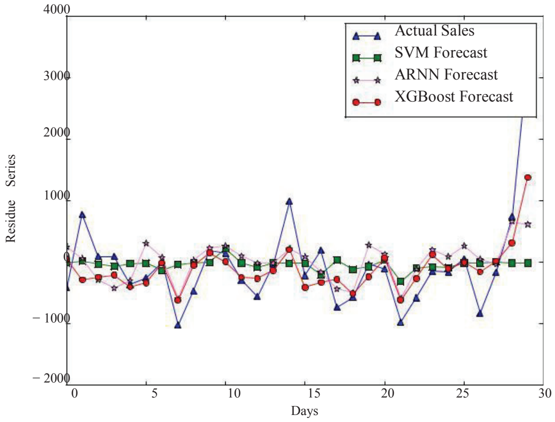

Fig. 11. Forecast over residue by SVM, ARNN and XGBoost vs Actual Residue

XGBoost forecast series is able to capture random walk pattern along with sharp peaks. It fits actual residue series better than other two hybrids.

From Table III, it can be verified that performance mea-sure of hybrid ARIMA-XGBoost and hybrid ARIMA-ARNN are better than hybrid ARIMA-SVM. Performance of hybrid XGBoost and ARNN is similar because RMSE of ARNN is less than XGBoost by 10 while MAE of XGBoost is less than ARNN by 15.

C. Decomposition Technique

Since, Decomposition technique takes advantage of best fit of components by best model. Hence, the result obtained should be definitely better than individual models. From Table III, it can be verified that the decomposition technique outperformed all the individual models and yielded MAE 328.6 and RMSE 426.4.

Performance of Hybrid technique and Decomposition technique depends on the nature of residue and remainder com-ponent in time series respectively. Both of them are highly nonlinear which can be seen from Fig 11 and Fig 12. Hybrid may or may not capture residue pattern. But STL decom-position has seasonal and trend window which helps them adjust smoothness of seasonal and trend component which directly influences remainder component. Hence, STL gives finer control over remainder components which empowers it to capture it accurately.

By comparing fig 11 and fig 12, it can be concluded that remainder component has been captured very accurately by nonlinear model as compared to residue series. Hence, Decomposition performs better than hybrid technique. It can be verified from table III that STL Decomposition has outper-formed all the hybrid models.

D. Performance Analysis

It can be inferred from the previous section that STL decom-position using Snaive, ARIMA and XGBoost outperformed all the individual models and hybrid models.

In this section, the eff ect of the increase in weekly fore-cast against performance measure of STL Decomposition is analysed in terms of Root Mean Square Error (RMSE). Fig 13 describes performance measure of STL Decomposition for 1 week forecast, 2 week forecast, and so on till 36 weeks of forecast. X-axis represents weekly forecast and Y-axis represents performance measure of STL Decomposition in terms of RMSE. Its equation can generalized as follows:

y = 363:856 3:834:x (11)

Equation 11 indicates that with increase in weekly forecast, RMSE of STL Decomposition decreases which signifies that accuracy of STL Decomposition will increase for long termforecast as compared to short term forecast.

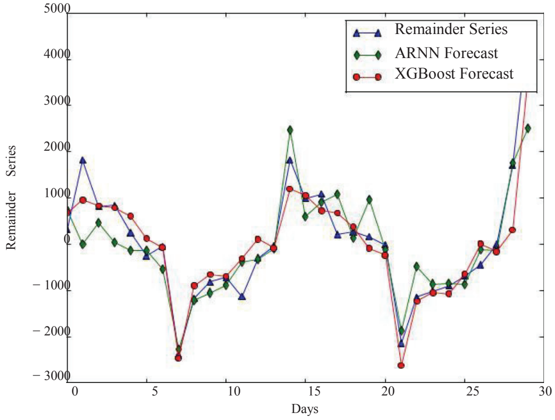

Fig. 12. Forecast of remainder component(random) in STL Decomposition by XGBoost vs Actual remainder Series

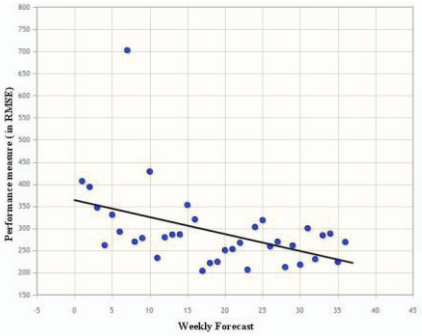

Fig. 13. Performance measure of STL Decomposition against weekly forecast

V. Conclusion

This paper presents the application of various machine learning models, hybrid models and decomposition technique to forecast the sales of Rossmann Store. Initially, a linear model such as ARIMA has been applied to our dataset. Other linear models are not taken into account because they do not allow inclusion of external regressors. Results have shown that ARIMA could not capture nonlinear patterns properly. Hence, to overcome its drawback, various nonlinear models namely ARNN, SVM and XGBoost are used.

These nonlinear models were studied separately and it was found that they captured linearity and nonlinearity better than ARIMA and ARNN gave the best result of 565 RMSE.

Now there needs to be a way out to obtain the advantage of both linear and nonlinear models. This led to the development of hybrid technique where linear component was forecasted by linear model and nonlinear component was forecasted by nonlinear model. Hybrid models gave better result than their individual model. Hybrid ARIMA-ARNN gave best result with lowest RMSE of 530.4 among all the hybrids.

Hybrid Technique can fail if nonlinear model fails to capture residue patterns as it was seen in case of Hybrid ARIMA-SVM where SVM failed to capture the residue. Also, it can fail when performance of nonlinear model outperforms the hybrid model. Performance of ARNN was as good as hybrid ARIMA-ARNN with difference of 30 in their RMSE.

Hence, a more robust technique is required. Therefore, STL Decomposition was used which decomposes the given time series into 3 components, namely Seasonal, Trend and Remainder component. These components can be modelled separately with different machine learning models that best fits them. Seasonal component was modelled accurately with snaive giving lowest RMSE of 28.36, while Trend component was modelled accurately with ARIMA giving lowest RMSE 185.61 and Remainder component was modelled using XG-Boost which gave RMSE of 436. Resultant RMSE of STL decomposition was 426.4 which has been lowest among all the individual and hybrid models.

STL Decomposition is more robust than Hybrid technique because it captures nonlinear components very precisely by adjusting its seasonal and trend window, hence nonlinearity in remainder component(STL) can be predicted better than the nonlinearity in the residue(hybrid). Also, decomposition takes advantage of best fit of components by best models, hence it outperforms all the individual and hybrid models which can be verified by result of STL Decomposition which has given lowest RMSE of 426.4. Also, it is verified in the section of ’Performance Analysis’ that accuracy of STL Decomposition improves with increase in weekly forecast which signifies that it is robust and can be used for short term and mid term forecast.

Hence, it can be concluded that, for this application, STL Decomposition using Snaive, ARIMA and XGBoost outperformed all the individual and hybrid model and has given best forecasting accuracy.

References

[1] J. Contreras, R. Espinola, F. J. Nogales and A. J. Conejo, "ARIMA models to predict next-day electricity prices," in IEEE Transactions on Power Systems, vol. 18, no. 3, pp. 1014-1020, Aug. 2003.

[2] Phatchakorn Areekul, Tomonobu Senjyu, Hirofumi Toyama, and Atsushi Yona,*A Hybrid ARIMA and Neural Network Model for Short-Term Price Forecasting in Deregulated Market*,IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 25, NO. 1, FEBRUARY 2010

[3] Ren Ye, P. N. Suganthan, N. Srikanth and S. Sarkar, "A hybrid ARIMA-DENFIS method for wind speed forecasting," Fuzzy Systems (FUZZ), 2013 IEEE International Conference on, Hyderabad, 2013, pp. 1-6.

[4] Akhter Mohiuddin Rather, "A prediction based approach for stock returns using autoregressive neural networks," Information and Communication Technologies (WICT), 2011 World Congress on, Mumbai, 2011, pp. 1271-1275.

[5] R. Skopal, "Short-term hourly price forward curve prediction using neural network and hybrid ARIMA-NN model," Information and Digital Technologies (IDT), 2015 International Conference on, Zilina, 2015, pp. 335-338.

[6] J. Chae et al., "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on, Seattle, WA, 2012, pp. 143-152

[7] R. M. Kapila Tharanga Rathnayaka, D. M. K. N. Seneviratna, W. Jianguo and H. I. Arumawadu, "A hybrid statistical approach for stock market forecasting based on Artificial Neural Network and ARIMA time series models," Behavioral, Economic and Socio-cultural Computing (BESC), 2015 International Conference on, Nanjing, 2015, pp. 54-60.

[8] Suhartono, I. Puspitasari, M. S. Akbar and M. H. Lee, "Two-level seasonal model based on hybrid ARIMA-ANFIS for forecasting short-term electricity load in Indonesia," Statistics in Science, Business, and Engineering (ICSSBE), 2012 International Conference on, Langkawi, 2012, pp. 1-5.

[9] Introduction to boosted trees [Online]. Available: https://xgboost.readthedocs.org/en/latest/model.html

[10] For XGBoost parameter tuning [Online]. Available: http://www.analyticsvidhya.com

[11] Rob J Hyndman,George Athanasopoulos[Online],*Forecasting: principles and practice*, Available: http://otexts.org/fpp

[12] Step-by-Step Graphic Guide to Forecasting through ARIMA Modeling in R[Online], Available: http://www.ucanalytics.com

[13] W. Sun, J. c. Lu and M. Meng, "Application of Time Series Based SVM Model on Next-Day Electricity Price Forecasting Under Deregulated Power Market," 2006 International Conference on Machine Learning and Cybernetics, Dalian, China, 2006, pp. 2373-2378.

[14] N. I. Sapankevych and R. Sankar, "Time Series Prediction Using Support Vector Machines: A Survey," in IEEE Computational Intelligence Magazine, vol. 4, no. 2, pp. 24-38, May 2009.

[15] Y. He, Y. Zhu and D. Duan, "Research on Hybrid ARIMA and Support Vector Machine Model in Short Term Load Forecasting," Sixth International Conference on Intelligent Systems Design and Applications, Jinan, 2006, pp. 804-809.

[16] J. H. Lo, "A study of applying ARIMA and SVM model to software reliability prediction," Uncertainty Reasoning and Knowledge Engineering (URKE), 2011 International Conference on, Bali, 2011, pp. 141-144.

[17] Y. Wang and J. Gu, "Comparative study among three different artificial neural networks to infectious diarrhea forecasting," Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on, Belfast, 2014, pp. 40-46.

[18] A. E. K., M. Eghlimi and Z. Zhang, "Forecasting the electricity price in iran power market: A comparison between neural networks and time series methods," 2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Hong Kong, 2014, pp. 1-6.

[19] H. K. Temraz, M. M. A. Salama and V. H. Quintana, "Application of the decomposition technique for forecasting the load of a large electric power network," in IEE Proceedings - Generation, Transmission and Distribution, vol. 143, no. 1, pp. 13-18, Jan 1996.

[20] D. Purwanto, C. Eswaran and R. Logeswaran, "A Comparison of ARIMA, Neural Network and Linear Regression Models for the Predic-tion of Infant Mortality Rate," 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, Kota Kinabalu, Malaysia, 2010, pp. 34-39.

[21] D. Chen, "Chinese automobile demand prediction based on ARIMA model," 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), Shanghai, 2011, pp. 2197-2201

[22] M. S. Kim, J. Liu, X. Wang and W. Yang, "Connecting Devices to Cookies via Filtering, Feature Engineering, and Boosting," 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, 2015, pp. 1690-1694.

[23] T. R. Anand and O. Renov, "Machine Learning Approach to Identify Users Across Their Digital Devices," 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, 2015, pp. 1676-1680.

[24] Y. Yan, P. Guo and L. Liu, "A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Forecasting Resource Consumption in an IIS Web Server," Software Reliability Engineering Workshops (ISSREW), 2014 IEEE International Symposium on, Naples, 2014, pp. 437-442.