

RESEARCH ARTICLE



OPEN ACCESS

Received: 30-04-2020

Accepted: 10-06-2020

Published: 20-06-2020

Editor: Dr. Natarajan Gajendran

Citation: Selvakumar V, Satpathi DK, Praveen Kumar PTV, Haragopal VV (2020) Modeling and prediction of third-party claim using a Machine learning approach. Indian Journal of Science and Technology 13(21): 2071-2079. <https://doi.org/10.17485/IJST/v13i21.465>

***Corresponding author.**

V Selvakumar

Research Scholar, Department of Mathematics, BITS-Pilani, Hyderabad, Telangana, India. Tel.: +91-996-611-0500

Assistant Professor, Bhavan's Vivekananda College of Science, Humanities, and Commerce, Hyderabad
p20170106@hyderabad.bits-pilani.ac.in

Funding: None

Competing Interests: None

Copyright: © 2020 Selvakumar, Satpathi, Praveen Kumar, Haragopal. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

Modeling and prediction of third-party claim using a Machine learning approach

V Selvakumar^{1,2*}, Dipak Kumar Satpathi³, P T V Praveen Kumar³, V V Haragopal³

¹ Research Scholar, Department of Mathematics, BITS-Pilani, Hyderabad, Telangana, India. Tel.: +91-996-611-0500

² Assistant Professor, Bhavan's Vivekananda College of Science, Humanities, and Commerce, Hyderabad

³ Department of Mathematics, BITS-Pilani, Hyderabad, Telangana, India

Abstract

Background/Objectives: The main objective of this research paper is to build an appropriate mathematical model that helps in forecasting overall claim amount based on the chosen characteristics of the data. **Methods/Statistical analysis:** In the field of actuarial research, forecasting the third-party claim amount for Motor vehicles is a challenging task, and only limited empirical research studies are done in predicting the claim. In this context, the annual time series historical claim data were collected for a period of 34 years to examine the predictive performance of the linear regression model, exponential smoothing model, autoregressive integrated moving average (ARIMA), artificial neural network (ANN), and hybrid ARIMA-ANN models to predict third party claim amount of motor insurance data in India. **Findings:** The data are analyzed, and the empirical evidence from the study shows that the ANN model improved the accuracy prediction when compared to Linear Regression, Exponential smoothing model, ARIMA and a hybrid model with respect to the performance criteria such as root mean squared error (RMSE) and mean absolute percentage error (MAPE). Therefore, the ANN model is more potent in forecasting TP claim amounts by considering the adequacy, suitability, and accuracy of the data modeling. **Novelty/Applications:** This data analytics approach would help motor insurance companies in India to have an idea about the expected future claim amounts. Also, this modeling approach will help the Motor Insurance companies of India to provide a better customer-centric forecasting model, which ensures better claims settlement and management.

Keywords: Claim amount; linear model; stationarity; ARIMA; neural network; TRAINLM

1 Introduction

In India, the motor insurance sector has heavily suffered based on motor own damage claim from a commercial motor vehicle concerning the viability of the insurance companies, the long-term claims intimation, and the non-cooperation of claim settlement

are some of the issues faced by the industry. The predictive modeling and forecasting the time series statistics is essential in the uncertainty occurrence of the event fields like Insurance. The empirical research on modeling of the Insurance claim amount is very inadequate, and few authors have considered the ARIMA model for prediction with respect to the property damage claim⁽¹⁾, future health care insurance cost of inpatient diagnosis⁽²⁾ and fire insurance loss⁽³⁾. However, in the Republic of Macedonia life insurance data, by applying two competent models, i.e., Combined ARIMA and Moving Average Model, concluded that the combined Box Jenkins model gives the best fit to the given data series⁽⁴⁾. Based on the earlier study on various forecasting techniques proposed, modeled, and suggested by multiple authors applied in predicting inflation and stock exchange, the ARIMA model is widely used for forecasting the time series data with accuracy and prediction⁽⁵⁾. Although ARIMA models are more flexible to predict different types of problems in various fields, they have their limitations that the data should be linear^(6,7). However, the exact nature of the data series is uncertain to predict, which has a non-linear structure of the data points; we apply the ANN. Several works have been done in forecasting the time series data with the ANN architecture model in a different field by using Box Jenkins methodology as the criterion to test the efficiency of the neural network model^(8–10). In recent years, hybrid models are developed due to the fact that there is no appropriate model that exists for prediction accurately. Hence, the hybrid ARIMA-ANN model is developed to forecast the data with greater accuracy^(11–13). Also, the multiple Machine learning predictive model are developed based on the study on spine surgery patients. It can be used to improve risk adjustment when assessing the patients with varying co-morbidities with higher accuracy⁽¹⁴⁾. Another study to forecast future Health care expenditure based on insurance claims data, the Recurrent Neural Network (RNN) predictive model outperforms the linear regression, least absolute regression, and gradient boosts machine model⁽¹⁵⁾. But, another study on six health insurance societies claims data to formulate a Japan population health management, a machine learning predictive regression model LASSO developed, which may be useful to assess a population health management plan⁽¹⁶⁾. A comparative study to predict stock exchange data shows that the techniques such as stacking and blending offer higher predictions as compared with bagging and boosting using machine learning approaches such as Decision Trees, Support Vector Machines and Neural Network⁽¹⁷⁾. A study on correctly pricing deposit insurances, a machine learning approach with a regularized cost function, the derived quadratic model approach helpful for modeling implied volatility⁽¹⁸⁾. Another study based on predicting accident claims by comparing the machine learning algorithm XGBoost⁽¹⁹⁾ and the Logistics Regression approach, the classical logistic regression method showed a high predictive performance as compared with the XGBoost method⁽²⁰⁾. However, there has been no substantial amount of research work done concerning Motor Insurance in India. It remains a motivating factor for us to evolve and build predictive models as compared with other studies.

2 Research Methodology

2.1 Data used for research

The secondary data are collected from different Insurance companies of India for distinctive 34 years from 1985 to 2018. The original data consists of 108 column variables distributed over 15,15,600 rows (i.e., approximately One hundred fifty million data points). After removing outliers of the data set, it was reduced to 12,72,311. Out of this data set, we have divided randomly 8,90,618 (70%) of the sample into a training data set and 3,81,693 (30%) of the sample into a testing data set. We have studied the Third-party (TP) claim variable that is considered for fitting Linear Model, Exponential Smoothing, ARIMA, ANN, and Hybrid Model. The data analytics are analyzed by using STATGRAPHICS Version 18.1.12 and MATLAB 2019b Version. We now discuss the techniques of the model building below.

2.2 Modeling of time series

In modeling the time series data, we first applied the traditional time series models such as Linear Model, exponential smoothing, and ARIMA. Further, we have applied the ANN model and the mixed-use of both linear domain & non-linear domain, a hybrid model for predicting the future value with greater accuracy.

2.2.1 Linear model

Generalized Linear Model is a generalization of conventional linear regression models for continuous response variables (TP claim amount) given continuous predictors (time). To fit a model to TP claim data provides an idea to Insurance companies for predicting the future claim amount based on historical data.

A simple general linear regression model structure is given by

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad (1)$$

Where x_i stands for time variable in a yearly unit;
 y_i is claims amount, e_i is a residual error term

2.2.2 Exponential smoothing model

In this methodology, the weighted mean of past data is used to forecast the claim amount. It can be applied when time-series data don't have any seasonality or trend and only have a level. This model will be represented as an equation as

$$\hat{y}_{(i+1)} = \alpha y_i + \alpha(1 - \alpha)y_{i-1} + \alpha(1 - \alpha)^2 y_{i-2} + \dots \quad (2)$$

Where α is the smoothing parameter, $0 \leq \alpha \leq 1$

y_i is claim amount, \hat{y}_{i+1} is the forecast value of the claim amount

Hence, α is the smoothing parameter that defines the weighing. If the actual TP claim data appears to be relatively stable over time, then we would select a smaller value for α . On the other hand, if the TP claim data tends to fluctuate rapidly, then we would choose a larger value for α to adjust the smoothened.

2.2.3 ARIMA (p, d, q)

In order to apply an ARIMA model to the data efficiently, the model needs to be a stationary series. When the TP claim data is not stationary, to bring the data into stationary processes, then the original process can then be reconstructed by integrating the differenced series to convert it into stationarity, and then it will be modeled by ARIMA (p, d, q) by using autocorrelation function and partial autocorrelation function. These suggest the most applicable model for the data.

The mathematical equation of Box and Jenkins⁽²¹⁾ process can be expressed as AR (p) I (d) MA (q) in terms of the backward shift operator.

$$(1 - \alpha_1 B - \alpha_p B^p)(X_t - \mu) = (1 + \beta_1 B + \dots + \beta_q B^q)e_t \quad (3)$$

Where B is the backward shift operator describing the process of differencing which is given by

$$y'_t = y_t - y_{t-1}$$

$$y'_t = y_t - B y_t \Rightarrow y'_t = (1 - B)y_t$$

Also, by comparing more than one ARIMA Model, the least Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC), RMSE and MAPE are generally considered to be best for the fitted model.

2.2.4 Artificial Neural Network (ANN Model)

The Neural network models are developed based on their interconnection of nodes. These networks are classified as multi-layer perceptron and a single layer perceptron, feedback, and feed-forward model. In feed-forward propagation, the signals are transmitted from one neuron to another in a forward direction only. In most of the time series modeling, a Multi-layer perceptron feed-forward neural network is used to model the cyclic links, which consist of the hidden layer, input, and output layer. The mathematical relationship between the input ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) and the output (y_t) for MLFFN model is

$$y_t = \alpha_0 + \sum_{i=1}^q \alpha_i f(\sum_{j=1}^p \beta_{ij} y_{t-j} + \beta_{0j}) + e_i \quad (4)$$

Here ' α_i ' is a weight from the hidden to output nodes, and ' β_{ij} ' are weights from the input to hidden nodes. 'p' and 'q' refers to input and hidden nodes of the parameter; 'f' refers the sigmoidal activation functions;

The sigmoidal function can be mathematically expressed as follows:

$$f(x) = \frac{1}{1 + e^{-ax}} \quad (5)$$

Equation (4) ANN model implements a non-linear mapping to obtain the future value from the past observations. i.e.,

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, \omega) + e_i \quad (6)$$

Here ω is a vector parameter, f is a function of weights and network structure and e_i refers error component.

2.2.5 Hybrid ARIMA-ANN model

In many real-world time series forecast modeling, the Box Jenkins model is appropriate to apply for a linear domain, whereas ANN is used for Non-linear domain. Both models have many successes in their data fields. Therefore, a hybrid model was proposed by Zhang⁽¹⁰⁾, which consists of linear as well as non-linear domains. In this approach, Zhang combined both ARIMA for modeling linear components and ANN for non-linear components separately to evaluate accurate forecasting. According to the model,

$$y_i = L_i + N_i \quad (7)$$

Where y_i stand for TP claim amount

L_i and N_i denote the linear component and non-linear component of a claim data.

According to Zhang, First, the original claim amount is fitted by the ARIMA to model the linear part, and the residuals obtained from the ARIMA contain only non-linear parts that will be fitted by the ANN model. So, the residuals of the ARIMA model is defined by $e_i = y_i - \hat{L}_i$, where \hat{L}_i denote the predicted value of ARIMA. Secondly, model the residuals using ANN that will capture the non-linear pattern of the TP claim data. Using p inputs, the ANN model for residuals will be of the form

$$e_i = f(e_{i-1}, e_{i-2}, e_{i-3}, \dots, e_{i-p}) + \varepsilon_i \quad (8)$$

Where f denotes a non-linear function of ANN and ε_i represents the random error.

The predicted value obtained from ANN denoted as \hat{N}_i , then the hybrid ANN-ARIMA forecast to the time series data is obtained as

$$\hat{y}_i = \hat{L}_i + \hat{N}_i \quad (9)$$

Now, the hybrid model is applied to third party claim data to check for better forecasting accuracies than ARIMA and ANN by comparing RMSE and MAPE.

3 Results and Discussion

The secondary sample dataset of the overall third-party claim will be taken. Further, based on other works of literature, the logarithm to base e of x of the data are used in the analysis. In this study, the Linear Regression Model and Exponential smoothing models are built using STATGRAPHICS, while ARIMA, ANN, and hybrid models are built using MATLAB 2019b. Further, the Root Mean Square (RMSE) and Mean Absolute Percentage Error (MAPE) are estimated to identify the best forecasting models. From Figure 1, we can observe that the occurrence of claim numbers and its claim amount is gradually increasing from the year 1985 to 2018. This indicates that the occurrence of claims and claim amounts are skewed.

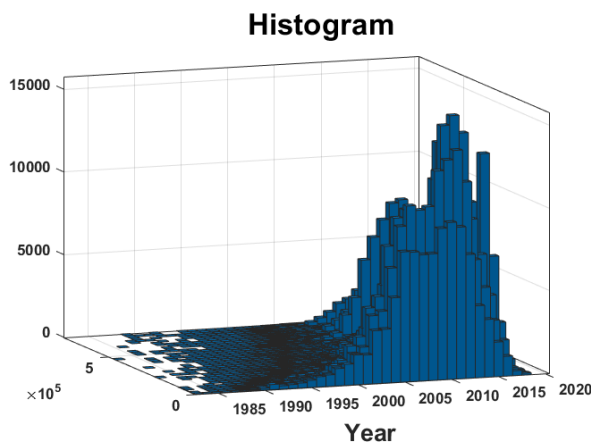


Fig 1. Histogram of Third-Party Claim Amount

3.1 Results of linear model

Firstly, we have fitted a linear regression model to the TP claim data by considering the Claim amount as a response variable and the time as an explanatory variable for the overall motor vehicles. From the observed data, the coefficients β_0 and β_1 are statistically significant.

Thus, the linear regression model equation (1) can be fitted as follows:

$$y_i = 11.0794 + 0.000001 x_i \quad (10)$$

Based on the above equation, we have predicted the claim amount for the entire data set and calculated the RMSE and MAPE accuracy measures for comparison of the model.

3.2 Results of exponential smoothing model

First, we determine the value of optimal smoothing constant α . For that, ten experimental trials are performed with different smoothing parameters from 0.1 to 0.9871. From Table 1, it is observed that the least performance accuracy measures are obtained at a higher value of the smoothing parameter (α) because the actual and forecasting TP claim amounts are fluctuating rapidly (Figure 2). The minimum measures of accuracy, such as MAPE, MSE, RMSE, occurred at an optimum smoothing parameter ($\alpha = 0.9703$). Also, it is clear that the performance measures of accuracy decrease when the smoothing constant increases.

Table 1. Performance criteria values for exponential smoothing model

Value of smoothing parameter (α)	MSD	RMSE
0.1	0.007153	0.084575
0.2	0.003992	0.063184
0.3	0.002893	0.053782
0.4	0.002342	0.048389
0.5	0.00202	0.044939
0.6	0.001817	0.042627
0.7	0.001687	0.041076
0.8	0.001607	0.040086
0.9	0.001564	0.039552
0.9871	0.001554	0.039417

An exponential smoothing plot of the actual average claim amount vs. predicted values with various measures of accuracy after fitting the model based on an equation (2) using historical data is given by

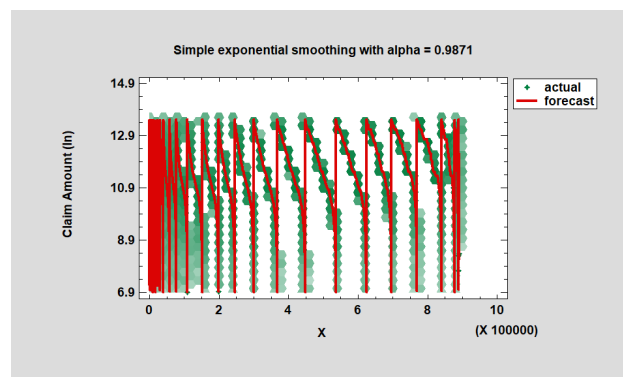


Fig 2. Predicted plots using exponential smoothing model

3.3 Results of the ARIMA (p, d, q) model

From Figure 3, it is observed that the original data series is in non-stationary form. Also, the augmented Dickey-Fuller (ADF) unit test confirmed that the data series is non-stationary and non-significant ($p=0.1714 > 0.01$). Thus, the TP claim data series can then be reconstructed by integrating the differenced series. The ADF unit root test was applied again; it is concluded that the data series evaluated is stationary ($p = 0.0000 < 0.01$). Now, the data series will be modeled by an ARIMA (p, d, q) with the parameter $d=1$.

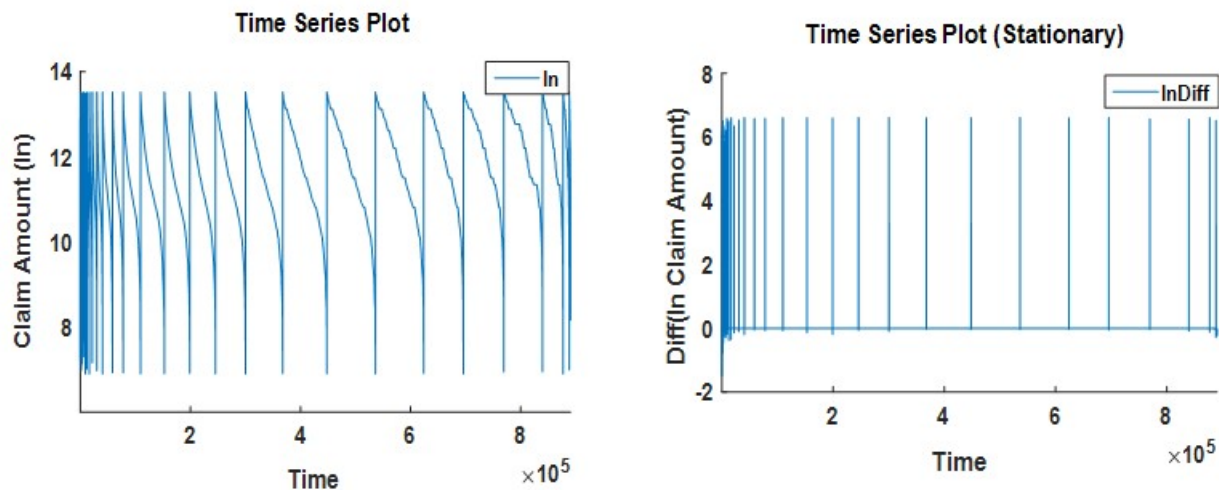


Fig 3. Overall claim data and its stationary series

From Table 2, we have computed the best possible feasible ARIMA models and their performance criteria AIC, RMSE, MAPE, and BIC values. By comparing all the fitted ARIMA models based on their performance criteria, ARIMA (2, 1, 1) model fits well with relatively smaller performance criteria values mentioned above to forecast the TP Claim Amount.

Table 2. Performance criteria values for the ARIMA models

Statistics	ARIMA (1,1,0)	ARIMA (0,1,1)	ARIMA (1,1,1)	ARIMA (2,1,0)	ARIMA (2,1,1)	ARIMA (1,1,2)	ARIMA (2,1,2)
RMSE	0.039361	0.0393606	0.0393608	0.0393595	0.0393596	0.0393595	0.0393597
MAE	0.000496	0.000497	0.0004976	0.00049931	0.0004905	0.0005008	0.0005014
MAPE	0.004348	0.0043848	0.0043522	0.00436802	0.0043475	0.004381	0.0043868
AIC	-3.63209	-3.632095	-3.632095	-2.957054	-3.632154	-3.632152	0.1518
BIC	-3.63205	-3.632056	-3.632056	-2.957014	-3.632104	-3.632102	0.3332

3.4 Results of ANN model

The modeling process was performed by using Neural Network (nntool) option in MATLAB 2019b version workspace. From the total samples, 70% of samples are selected randomly for training, 15% of them for validation and the remaining 15% of them selected for the testing network. During training the claim data, the back propagation TRAINLM is applied to predict the RMSE and MAPE. The network performance can be checked and retrain the network when performance is not satisfied. Finally, the TRAINLM method builds forecast results based on Mean Square Error (MSE) using trial and error experiments.

From Figure 4, it is observed that Regression plots of ANN-based on training, testing, validation, and overall data set to fall with a 45° line with R values in each case almost near to 0.999 with the best validation performance measure is 0.0010621. It indicates that the predicted plots represent a good fit, and also the best performance measure of the overall data set RMSE as 0.03829 and MAPE as 0.01202.

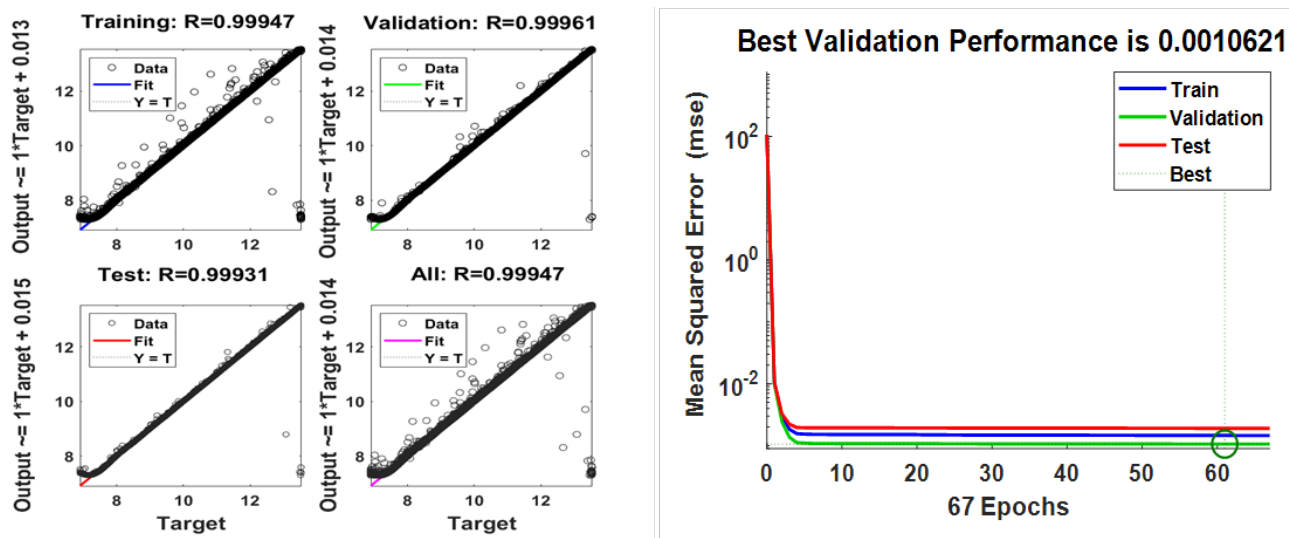


Fig 4. Regression plots and performance plot for ANN model

3.5 Results of the Hybrid ARIMA-ANN Model

In this model, First, we have fitted the Box Jenkins Methodology ARIMA Model for original data based on equation (3). The residuals obtained from the fitted ARIMA model indicates non-linearity. Then, we have applied the ANN technique to the residuals of the ARIMA by the TRAINLM algorithm. Finally, based on equation (9), the forecasted values of the hybrid model are obtained by adding the predicted values of ARIMA and ANN with the resulting best performance RMSE as 0.03921 and MAPE as 0.01341 for overall TP claim amount.

3.6 Prediction performance of forecasting models

From Table 3, it is observed that the performance criteria index RMSE and MAPE for prediction are compared for different models such as Linear Regression Model, Exponential Smoothing, ARIMA, ANN, and Hybrid. The comparative results showed that the ANN model yields a more accurate forecast compared to any other model with lesser RMSE and MAPE.

Table 3. Comparison of forecast accuracy using different Models using training and testing data set

Vehicle Category	Model	Training Data Set		Testing Data Set	
		RMSE	MAPE	RMSE	MAPE
Overall Third-Party Claim Amount	Linear Model	1.13583	8.26189	1.134610	8.247950
	Exponential Smoothing	0.03942	0.01435	0.061190	0.028091
	ARIMA	0.03935	0.01395	0.061191	0.028099
	ANN	0.03829	0.01202	0.052225	0.025334
	Hybrid	0.03921	0.01341	0.061128	0.027060

Hence from the above table, the Artificial Neural Network model is slightly better than the hybrid model for forecasting the claim amount. From Figure 5, we can observe the overall predicted data point comparison between overall claim with other predicted models. If we observe the predicted ANN plot, it is almost similar to the overall claim as compared with other models. It indicates that ANN is a better fit for the whole third-party claim amount.

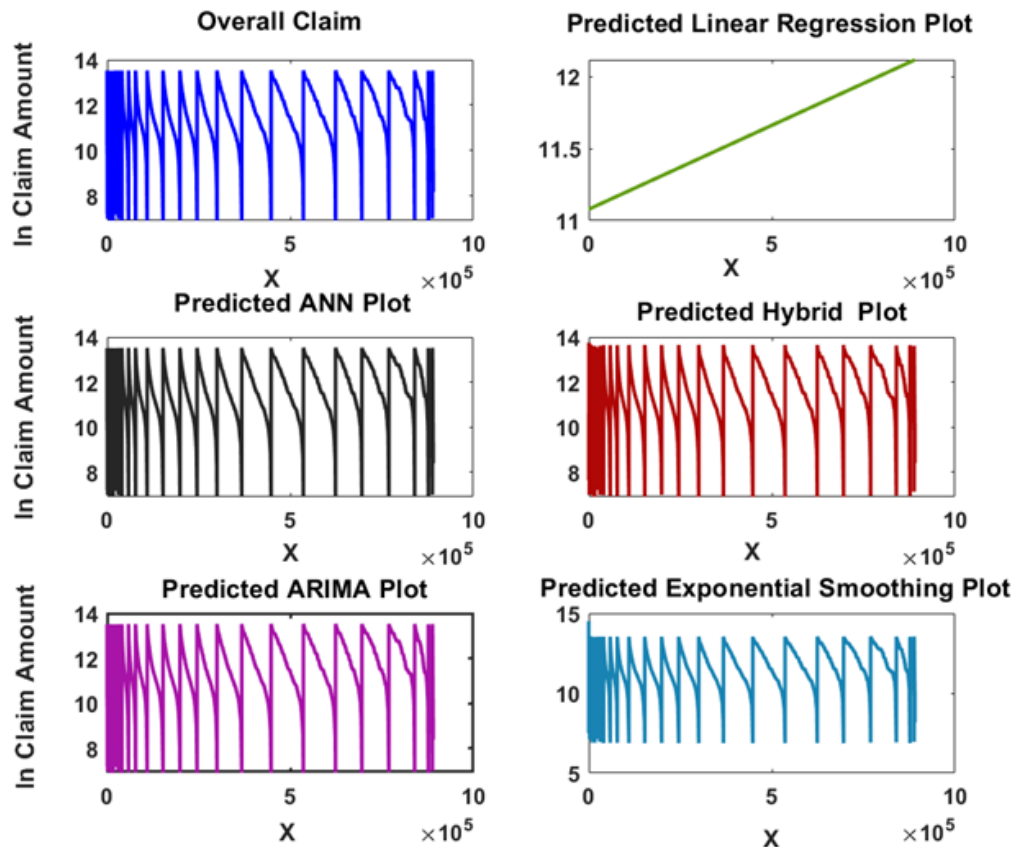


Fig 5. Predictive plot comparison of linear regression, ANN, hybrid model, ARIMA and exponential smoothing

4 Conclusion

Over the past few years, the forecasting accuracy of the claim amount is essential to many financial risk problems in Motor Insurance Companies, the predictive modeling and forecasting the claim amount is essential in the uncertainty occurrence of the claim event. With this context, the traditional time series models are applied to predict the TP claim amount. In this, the ARIMA Model outperforms the other traditional models of Linear Regression and Exponential Smoothing with lesser error measures. But, there exist a considerable variation in residuals of the ARIMA model indicating the non-linearity, which shows that the linear ARIMA model alone is not appropriate for predicting the claim amount. In this situation, we have applied a hybrid model to predict the claim amount and also compared the model to the non-linear capability model ANN. The empirical results with TP claim amount data suggest that the ANN can outperform all the other models such as Linear, exponential smoothing, ARIMA, and the hybrid model with lesser RMSE and MAPE. According to recent literature, the forecasts are performed by a hybrid model fits well with the financial data series, but in our studies, the ANN model slightly better over the hybrid model. Thus, the ANN model is suggested as the best model for predicting the third party claim amount with lesser RMSE and MAPE.

This exploratory analytics approach would help Motor Insurance companies to have an idea about the expected future claim amounts. Also, this modeling approach and forecasts will help the Insurance Companies to Budget their future revenue planning for the ensuing years for an easy disbursement of the Claims.

References

- 1) Cummins JD, Griepentrog GL. Forecasting automobile insurance paid claim costs using econometric and ARIMA models. *International Journal of Forecasting*. 1985;1(3):203–215. Available from: [https://dx.doi.org/10.1016/0169-2070\(85\)90003-2](https://dx.doi.org/10.1016/0169-2070(85)90003-2).
- 2) Mesike GC, Adeleke, Ibiwoye A. Predictive Actuarial Modeling of Health Insurance Claims Costs. *International Journal of Mathematics and Computation*.

- 2012;14:974–5718.
- 3) Taha T. Forecasting Fire Insurance Loss Ratio in Misr Insurance Company. *El- Bahith Review*. 2017;7. Available from: <https://www.asjp.cerist.dz/en/article/34449>.
- 4) Andreeski C, Milosevic B, Njegomir V. Analysis of the life insurance market in the Republic of Macedonia. *Ekonomski anali*. 2012;57(194):107–122. Available from: <https://dx.doi.org/10.2298/eka1294107a>.
- 5) Nyoni T. Modeling and Forecasting Inflation in Burundi using ARIMA Models. *Munich Personal Repec Archive MPRA*. 2019. Available from: <https://mpa.ub.uni-muenchen.de/92444/>.
- 6) Abonazel MR, Elftah AIA. Forecasting Egyptian GDP using ARIMA Models. *Reports on Economics and Finance*. 2019;5:35–47. Available from: <https://doi.org/10.12988/ref.2019.81023>.
- 7) Choden, Unhapipat S. ARIMA model to forecast international tourist visit in Bumthang, Bhutan. *Journal of Physics: Conference Series*. 2018;1039:012023–012023. Available from: <https://dx.doi.org/10.1088/1742-6596/1039/1/012023>.
- 8) SLIMANI N, SLIMANI I, SBITI N, AMGHAR M. Traffic forecasting in Morocco using artificial neural networks. *Procedia Computer Science*. 2019;151:471–476. Available from: <https://dx.doi.org/10.1016/j.procs.2019.04.064>.
- 9) Lau ET, Sun L, Yang Q. Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*. 2019;1(9). Available from: <https://dx.doi.org/10.1007/s42452-019-0884-7>.
- 10) Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 2003;50:159–175.
- 11) Khandelwal I, Adhikari R, Verma G. Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition. *Procedia Computer Science*. 2015;48:173–179. Available from: <https://dx.doi.org/10.1016/j.procs.2015.04.167>.
- 12) Omar H, Hong VH, Liu DR. A Hybrid Neural Network Model for Sales Forecasting based on ARIMA and Search Popularity of Article Titles. *Computational Intelligence and Neuroscience*. 2012;1. Available from: <https://doi.org/10.1155/2016/9656453>.
- 13) Mallikarjuna M, Rao RP. Evaluation of forecasting methods from selected stock market returns. *Financial Innovation*. 2019;5(1). Available from: <https://dx.doi.org/10.1186/s40854-019-0157-x>.
- 14) Han SS, Azad TD, Suarez PA, Ratliff JK. A machine learning approach for predictive models of adverse events following spine surgery. *The Spine Journal*. 2019;19(11):1772–1781. Available from: <https://dx.doi.org/10.1016/j.spinee.2019.06.018>.
- 15) Yang C, Delcher C, Shenkman E, Ranka S. Machine learning approaches for predicting high cost high need patient expenditures in health care. *BioMedical Engineering OnLine*. 2018;17(S1):131–131. Available from: <https://dx.doi.org/10.1186/s12938-018-0568-3>.
- 16) Takeshima T, Keino S, Aoki R, Matsui T, Iwasaki K. Development of Medical Cost Prediction Model Based on Statistical Machine Learning Using Health Insurance Claims Data. *Value in Health*. 2018;21:S97–S97. Available from: <https://dx.doi.org/10.1016/j.jval.2018.07.738>.
- 17) Nti IK, Adekoya AF, Weyori BA. A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*. 2020;7(1). Available from: <https://dx.doi.org/10.1186/s40537-020-00299-5>.
- 18) Assa H, Pouralizadeh M, Badamchizadeh A. Sound Deposit Insurance Pricing Using a Machine Learning Approach. *Risks*. 2019;7(2):45–45. Available from: <https://dx.doi.org/10.3390/risks7020045>.
- 19) Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. 2016;p. 785–794.
- 20) Pesantez-Narvaez J, Guillen M, Alcañiz M. Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression. *Risks*. 2019;7(2):70–70. Available from: <https://dx.doi.org/10.3390/risks7020070>. doi:10.3390/risks7020070.
- 21) Box EG, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. Holden-Day. 1976. Available from: <https://trove.nla.gov.au/work/10907520?q&online=true>.