IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

"Modelling Insurance Premiums using Deep Learning approach"

Dr B Navatha[1] Dr Obbu Ramesh[2]
Assistant Professor[1] Assistant Professor[2]
Department of Mathematics[1,2]
Presidency Univesity,Bangalore-560064,India

Abstract

Based on total premiums from a commercial about the insurance firms' viability, the long-term claims notification, and the industry's lack of cooperation in claim settlement, the health insurance business in India has suffered greatly. Predictive modeling and forecasting of the total claims paid in the insurance industry are crucial for determining the appropriate premium level for various individuals. Verylittle empirical research has been done on modeling the quantity of insurance premiums, and few authors have looked into deep learning models for predicting future health care insurance costs associated with inpatient diagnoses, property premiums, etc. Nevertheless, using two capable models—the combination ARIMA and the Moving Average Model—in the Republic of Macedonia life insurance data, it was determined that the combination Box Jenkins model.

Keywords:
Deep learning Techniques, Premium,claimspaid,ANN,RNN,CNN,LSTM

1.1 Introduction

Modeling implied volatility is made easier by a study on accurately pricing deposit insurances and by using a machine learning approach with a regularized cost function and derived quadratic model approach. Another study that compared the standard logistic regression method to the machine learning algorithm XGBoost for predicting accident claims found that the XGBoost method had a higher predictive performance. Nonetheless,the extent of research on health insurance in India has been limited thus far. In comparisonwith the previous research, it continues to be a motivating factor for us to improve and develop various predictive models

. 1.2 Objective of the Study

This research study aims to develop deep learning models to forecast premium amount based on the historical sample data. To arrive at this, we have used monthlypremium data for two years. The major contributions of the study are as follows:

- To forecast the premium amount of insurance companies by using ANN,CNN,RNN and LSTM.

- To identify the most suitable forecasting technique among the four models mentioned above by considering the data modeling's adequacy, suitability, and accuracy.

1.3 Data Context and Use

Secondary data for the two years 2019 and 2020 in India is gathered monthly fromdifferent insurance providers. We have collected three premiums paid of different sizes. The actual data consists of 56 column variables distributed over 5,56,560 rows (i.e., approx. Five lakhs fifty thousand five sixty policy holders). After identifying the data set's outliers, it was limited to premiums paid with sizes 1446,2149, and 601.

1.4 Methodology

For the same set of data, many deep learning models, including ANN, CNN, RNN, andLSTM, are run and compared to the fitted models.

Modeling of Premiums paid

We explored various premiums data by considering ANN ,CNN,RNN and LSTM models to predict total premiums paid with accuracy and a best model fit. In this chapter modeling the Insurance premiums data with various deep learning algorithms such as ANN,CNN,RNN, and LSTM models were used to estimate future values of premium amounts with improved precision. In this study, we have studied three premium amounts datasets with sizes 1446,2149 and 601 respectively. These were evaluated by using STATGRAPHICS Version 18.1.12 and MATLAB 2019 Version.

1.4.1 Comparison of, ANN,CNN,RNN and LSTM Model for Three Premiums paid datasets:

The model with the least MSE,MAPE and RMSE was chosen as the suitable modelfor predicting for premium data. The experimental results showed that the LSTM model yields a more precise forecast than any other model in each category of various premiums paid concerning MSE, RMSE and MAPE. Hence, the Long Short Term Memory is better than ANN,CNN and RNN for forecasting the premium amounts From table 1.4(a),1.4(b) and 1.4(c) It has been noted that the MSE, RMSE, and MAPE performance criteria indexes for prediction for differentmodels, including ANN, CNN, RNN, and LSTM Model for three premiums paid data, are compared. Comparative results showed that, compared to other models, the LSTM model produced more accurate forecasts and had lower MSE, RMSE, and MAPE values.As a result, the LSTM model predicts these premiums paid better than the other models.

PREMIUM 1	ANN	RNN	CNN	**LSTM**
MSE	4.50E+06	9.10E+06	1.57E+07	**1.12E+06**
RMSE	2.12E+03	3.02E+03	3.97E+03	**1.06E+03**
MAPE	0.001352232	0.002031734	0.002713265	**0.000680129**

table 1.4(a) Performance Metrics Comparison of ANN,CNN,RNN andLSTM Models for Premiums paid 1

PREMIUM 2	ANN	RNN	CNN	**LSTM**
MSE	6.78E+07	1.06E+09	7.69E+08	**1.69E+07**
RMSE	8.23E+03	3.25E+04	2.77E+04	**4.12E+03**
MAPE	0.00077239	0.001984691	0.002555544	**0.000391375**

table 1.4(b) Performance Metrics Comparison of ANN,CNN,RNN andLSTM Models for Premiums paid 2

PREMIUM 3	ANN	RNN	CNN	**LSTM**
MSE	1.63E+11	5.51E+11	1.09E+12	**4.07E+10**
RMSE	4.04E+05	7.43E+05	1.05E+06	**2.02E+05**
MAPE	0.005471107	0.010781438	0.015995965	**0.0077982**

table 1.4© Performance Metrics Comparison of ANN,CNN,RNN andLSTM Models for Premiums paid 3

1.4.3 Forecast Model for three premiums Paid dataset1

After identifying the model for the data, we have predicted the premium amounts for Twelve months that is for the year 2023-24. These indicate that themodel fitted performs well in predicting the premium amount with LSTM Model. We have generated the forecasts for all the premiums paid in Table 1.4(d),1.4(e) and 1.4(f) by using all the Models. The Predicted values for twelvemonths are shown in those tables, we can observe the point-to-point comparisonbetween observed and forecasted values of all the models for three premiums paid data.

Forecasted	ANN	RNN	CNN	**LSTM**
mar-23	725858.32	737406.34	745688.56	**741547.45**
apr-23	748954.36	753970.79	761470.76	**757720.77**
may-23	758987.21	768970.73	778920.06	**773945.39**
June -23	778954.25	788869.38	797274.47	**793071.93**
July -23	798784.51	805679.57	815346.54	**810513.05**

Aug -23	812574.62	825013.52	833186.36	**829099.94**
Sep-23	837452.42	841359.20	846392.67	**843875.93**
Oct – 23	845265.97	851426.15	859854.97	**855640.56**
Nov -23	857586.32	868283.80	878527.88	**873405.84**
Dec-23	878981.28	888771.96	897135.46	**892953.71**
Jan-24	898562.64	905498.96	915593.105	**910546.0325**
Feb-24	912435.28	925687.25	935678.52	**930682.89**

table **1**.4(d) Predicted Values for premiums paid 1

1.4.4 Forecasted values of Premium dataset2:

Forecasted	ANN	RNN	CNN	LSTM
mar-23	2768612.23	2873535.40	2963579.56	**3043626.24**
apr-23	2978458.56	3053623.73	3123672.92	**3178794.38**
may-23	3128788.89	3193722.11	3233915.85	**3264883.99**
June -23	3258655.32	3274109.59	3295852.13	**3327764.51**
July -23	3289563.85	3317594.68	3359676.89	**3411237.53**
Aug -23	3345625.50	3401759.10	3462798.17	**3534736.81**
Sep-23	3457892.70	3523837.24	3606675.45	**3692594.66**
Oct – 23	3589781.78	3689513.66	3778513.86	**3865093.11**
Nov -23	3789245.54	3867514.06	3951672.35	**4041627.33**
Dec-23	3945782.58	4035830.64	4131582.31	**4209867.87**
Jan-24	4125878.70	4227333.97	4288153.43	**4331317.995**
Feb-24	4328789.24	4348972.89	4374482.56	**4458725.63**

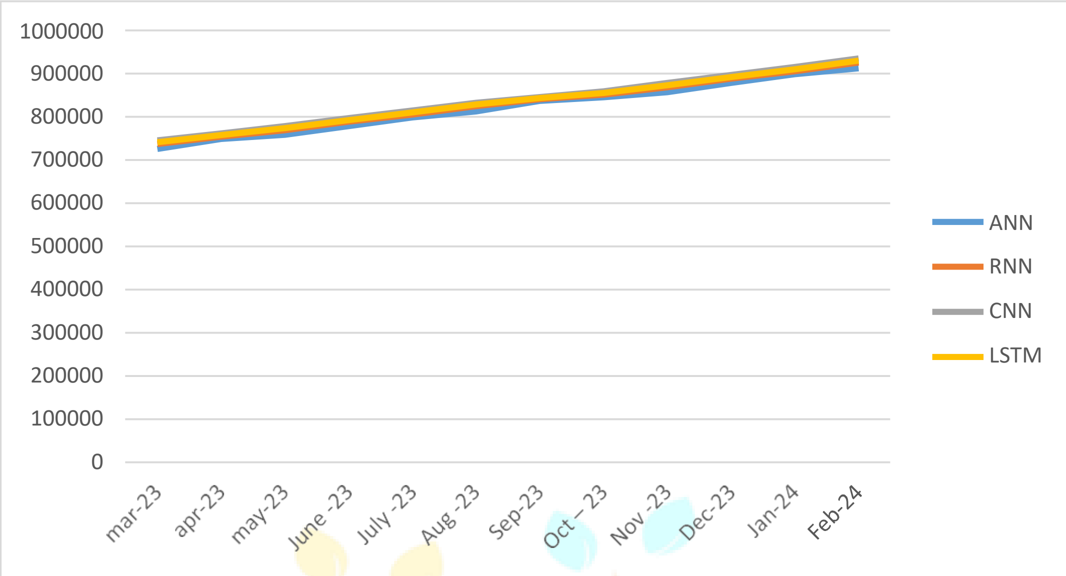
table 1.4(e) Predicted Values for premiums paid 2

1.4.5 Forecasted values of Premium Dataset3:

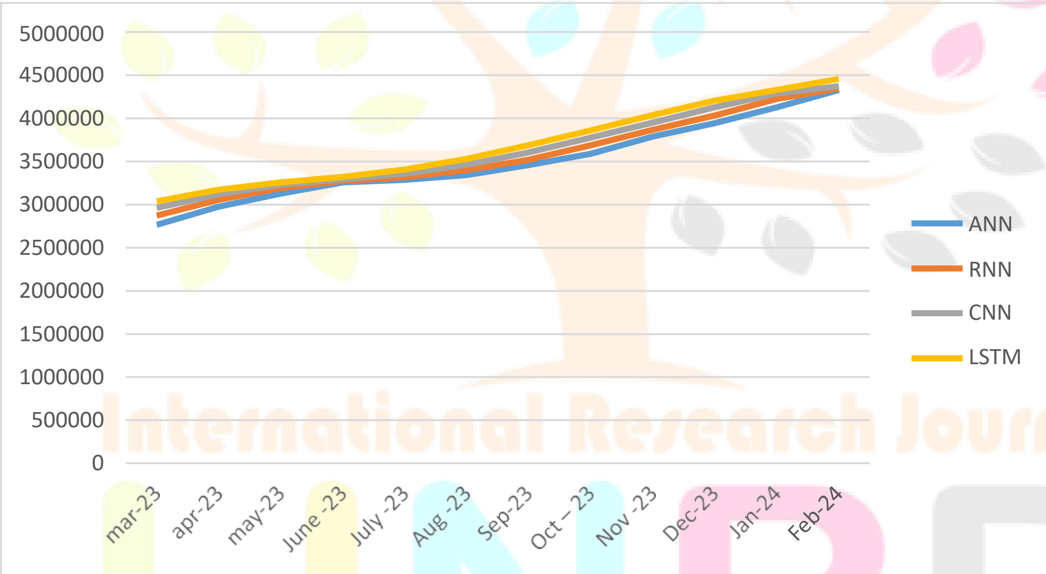
Forecasted	ANN	RNN	CNN	**LSTM**
mar-23	68987898.25	68987898.06	68987898.09	**68987898.08**
apr-23	69879874.14	68987898.16	68987898.11	**68987898.13**
may-23	69899989.38	69433886.15	69210892.15	**69322389.15**
June -23	70125898.28	69666937.76	69550411.96	**69608674.86**
July -23	72859586.35	69896418.02	69781677.89	**69839047.96**
Aug -23	73898925.25	71378002.19	70637210.10	**71007606.15**
Sep-23	75898565.35	72638463.72	72008232.95	**72323348.34**
Oct – 23	76986859.30	74268514.53	73453489.13	**73861001.83**
Nov -23	79895969.37	75627686.92	74948100.73	**75287893.82**
Dec-23	81256575.88	77761828.14	76694757.53	**77228292.84**
Jan-24	83258525.44	79509202.01	78635515.08	**79072358.54**
Feb-24	85789586.31	81383863.73	80446532.87	**80915198.30**

Table 1.4(f) Predicted Values premiums paid 3

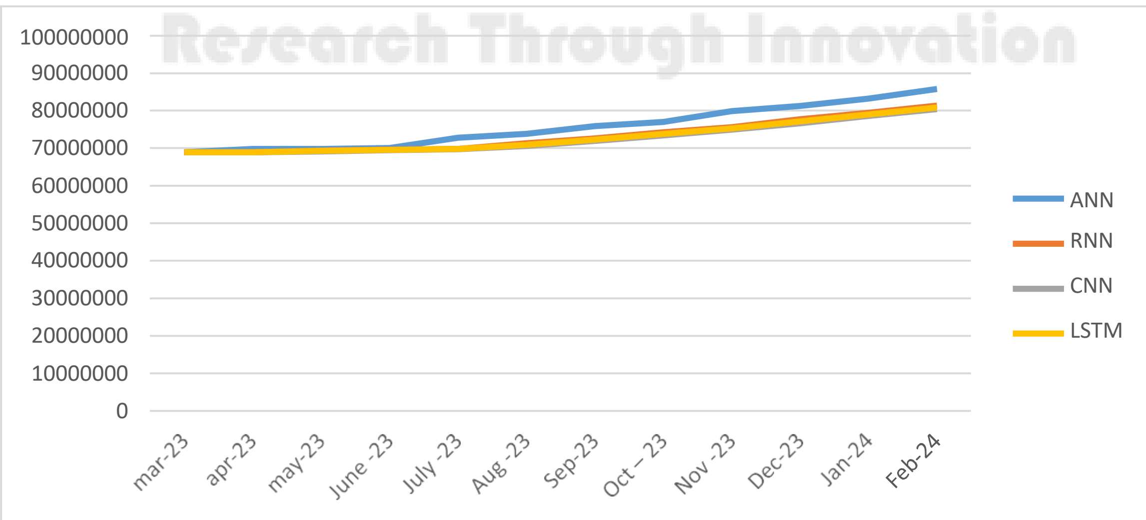
Premium paid 1



Premium paid 2

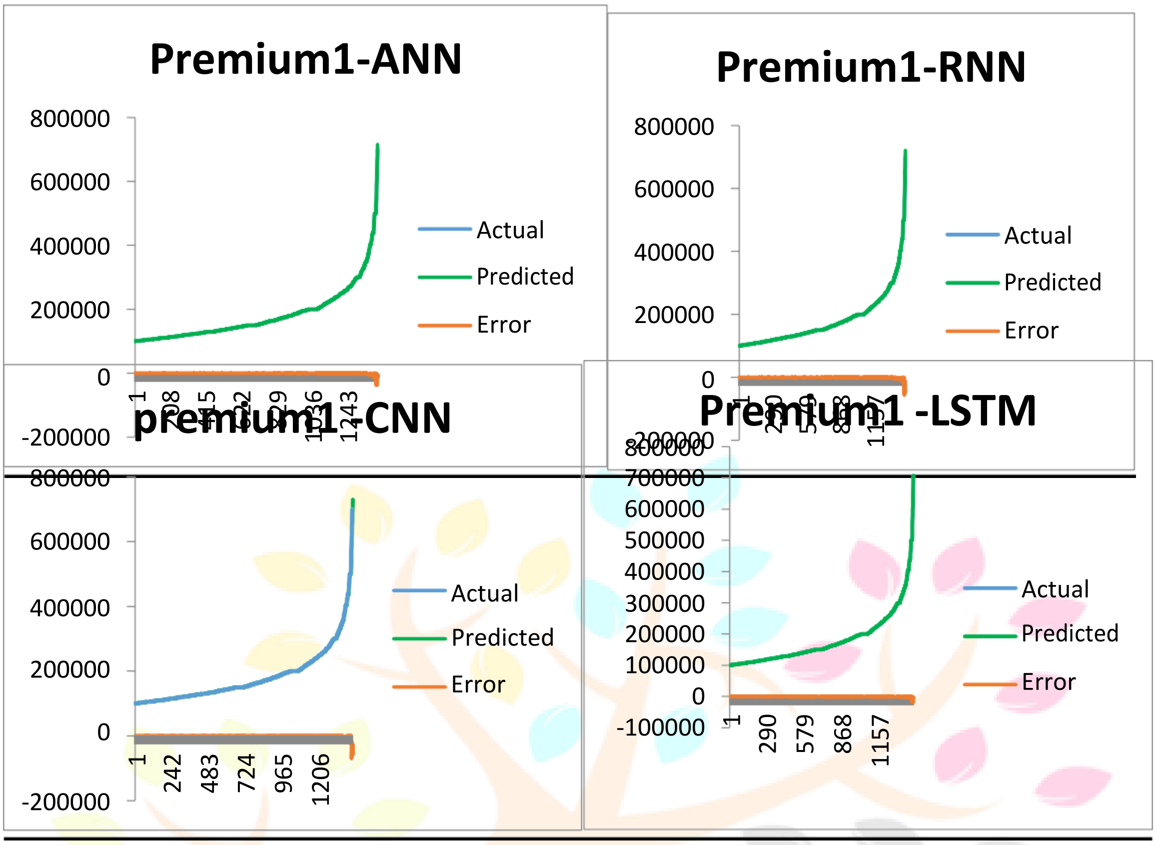


Premium paid 3

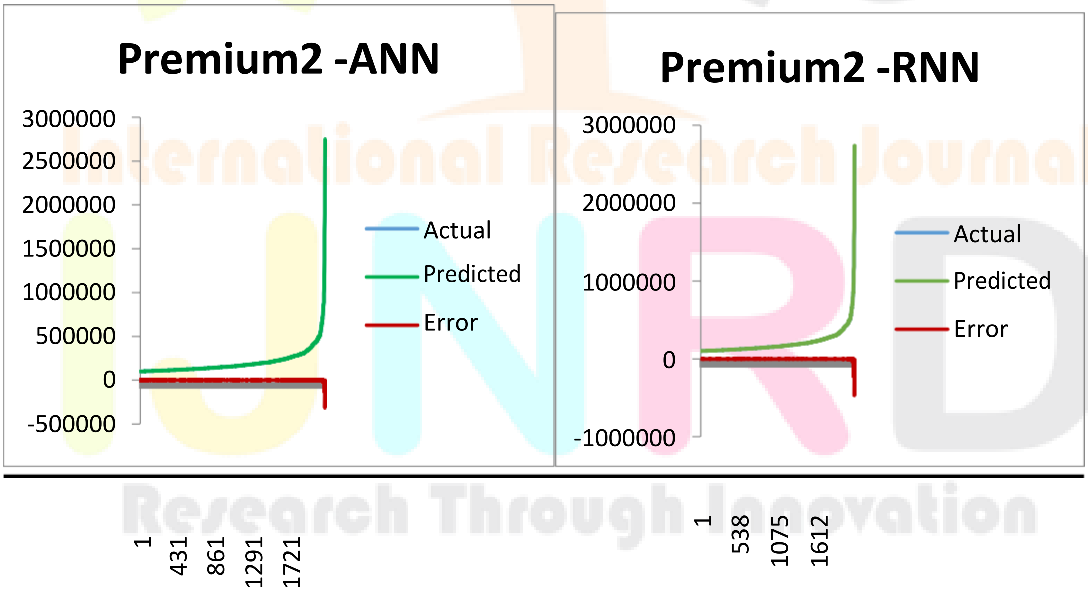


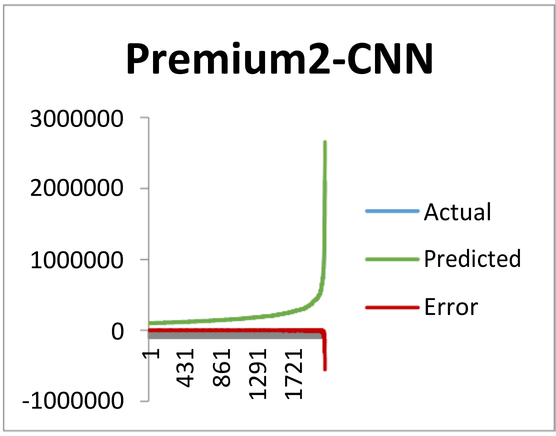
Comparison of ANN,CNN,RNN and LSTM Model Predicted Plot for all thePremiums paid datasets

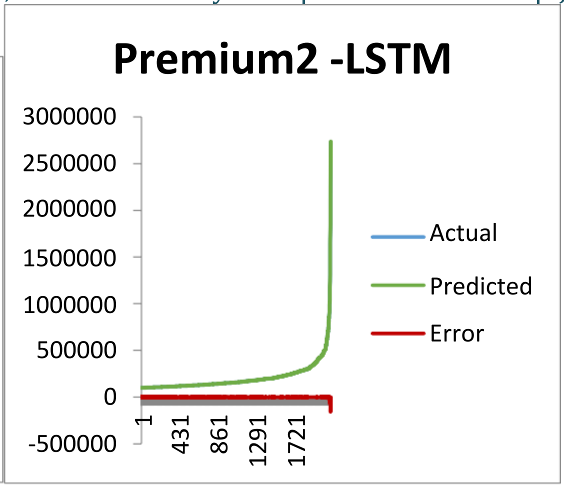
Every projected value (green line) in the figure below is quite close to the actual values (blue line) with the least amount of error (red line). Plotting the actual, predicted, and errorvalues for the premiums paid1 time series data demonstrates that, even with dynamic forecasting, the overall forecasts are accurate. The actual value (blue line) and the anticipated values (green line) correspond quite well.



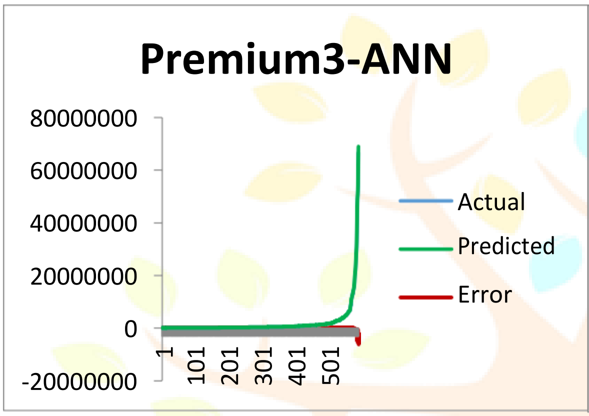
Premiums paid 1 – Actual, predicted and residual plot using ANN,RNN,CNN andLSTM

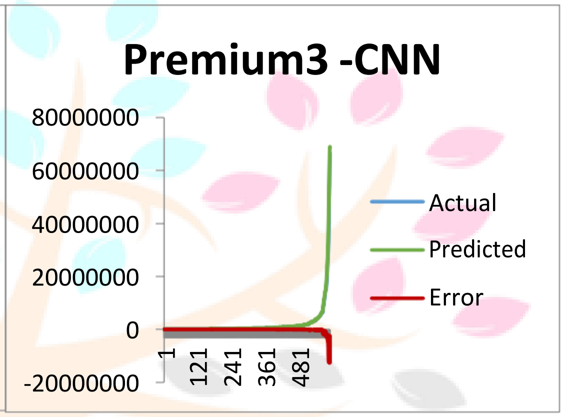


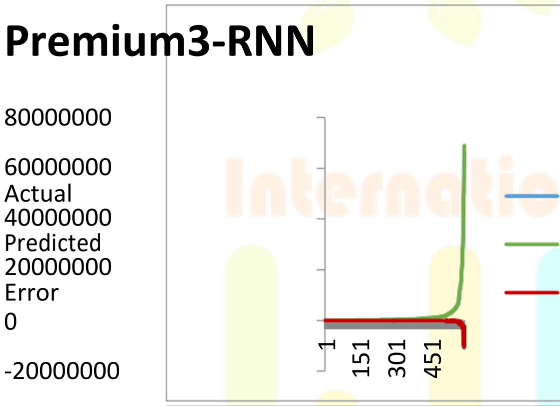


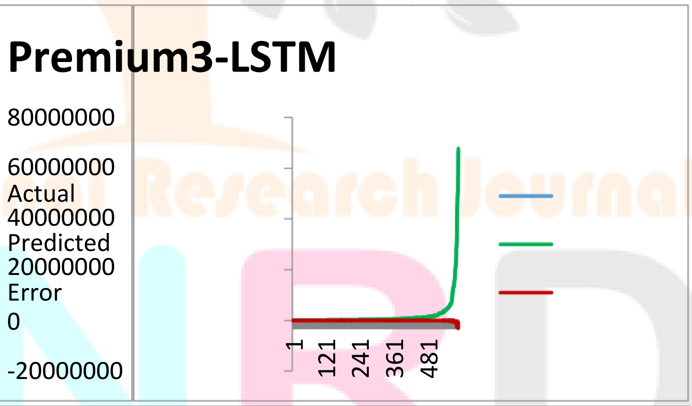


Premiums paid2 – Actual, predicted and residual plot using ANN,RNN,CNN andLSTM









Premiums paid 3 – Actual, predicted and residual plot usingANN,RNN,CNN and LSTM

1.5 Conclusion

Many financial risk problems in Health insurance companies have arisen in recentyears, and the accuracy of premium amounts forecasting has proven to be critical in solving them. Predictive modeling and forecasting are critical in dealing with theuncertainty associated with the occurrence of a premiums and also claims. In this context, the deep learning techniques were applied to predict the Total premiums paid amount and claims amount, To estimate the claim amount, premium amount, and compare the model to the non-linear capability model, we utilized a variety ofmodels.

A viable model for estimating the total claim amount and premium amount in India's insurance firms' Health Insurance Segment. Forecasting the claims paid amount and premium amount will help insurance firms for future prospects due tothe complex nature of data, the uncertainty of health insurance claims, continuing economic development, and ongoing tariff disputes. In this context, we developed and compared deep learning forecasting approaches such as artificial neural network (ANN), recurrence neural network (RNN), convolution neural network (CNN), and long short-term memory model (LSTM) using MSE,RMSE, andMAPE performance criteria. When the above performance criteria and accurate predictions are compared, the observed results with Total claims paid amount and premium amount data show that the LSTM outperforms all other models with lowest MSE, RMSE, and MAPE, it is recommended that you use the LSTM model.Furthermore, the modelling method and projections will assist Health Insurance Companies budget their future income planning for the next years for uncomplicated claim payment and premium fixing.Thus, we conclude in thischapter by stating that this comparative data analytics approach which would assistinsurance scompanies in accurately forecasting future claim amounts and premiumspaid using the LSTM model.

From the previous studies it was learnt that Multiple regression technique and ANN waspopular prediction techniques does not seems to fit well, Hence deep learning techniques have been evaluated for this claims and premiums data .Finally, In Present study found thatLSTM Model turn out to is better prediction tool for prediction in future.

we have developed and compared deep learning forecasting techniques such as Artificial Neural Network, Recurrence Neural Network, Concurrence Neural Network and Long Shirt-term Memory Model based on the performance criteria by MSE, RMSE and MAPE. After comparing the above performance criteria and accurate predictions, empirical results with Total claim amount and premium amount data suggest that the LSTM can outperform all the other models, that is with lesser values of MSE, RMSE and MAPE based various datasets. According to recent literature, the forecasts are performed by a hybrid model that fits well with the financial data series. Still, in our studies, the LSTM model is slightly better thanthe other models. Thus, the LSTM model is suggested as the best model for predicting the Total claim amount and premium amount with lesser MSE, RMSE and MAPE. Also, this modeling approach and forecasts will help the Insurance Companies budget their future revenue planning for the ensuing years for an easy disbursement of the claims. Thus, we conclude in this chapter by stating that this comparative data analytics approach which would assist insurance companies in accurately forecasting future claim amounts using the LSTM model.

References:

1. K. Abhishek, M. P. Singh, S. Ghosh, and A. Anand, "Weather Forecasting Modelusing Artificial Neural Network," *Procedia Technology*, vol. 4, 2012, doi: 10.1016/j.protcy.2012.05.047

2. S. Das, M. Devi, K. Prakash, and A. K. Barbara, "Prediction of Atmosphericvariabilities in North East India during Tropical Cyclones: Use of Numerical Weather Prediction Model and back propagation Artificial Neural Network (ANN)," *International Journal of Electronics and Applied Research*, vol. 4, no. 2, Dec. 2017,doi: 10.33665/IJEAR.2017.v04i02.002.

3. H. Assa, M. Pouralizadeh, and A. Badamchizadeh, "Sound Deposit Insurance PricingUsing a Machine Learning Approach," *Risks*, vol. 7, no. 2, Apr. 2019, doi: 10.3390/risks7020045.

4. S. S. Han, T. D. Azad, P. A. Suarez, and J. K. Ratliff, "A machine learning approach for predictive models of adverse events following spine surgery," *The Spine Journal*,vol. 19, no. 11, Nov. 2019, doi: 10.1016/j.spinee.2019.06.018.

5. T. Takeshima, S. Keino, R. Aoki, T. Matsui, and K. Iwasaki, "Development of Medical Cost Prediction Model Based on Statistical Machine Learning Using Health Insurance Claims Data," *Value in Health*, vol. 21, Sep. 2018, doi: 10.1016/j.jval.2018.07.738.

6. E. T. Lau, L. Sun, and Q. Yang, "Modelling, prediction and classification of student academic performance using artificial neural networks," *SN Applied Sciences*, vol. 1, no. 9, Sep. 2019, doi: 10.1007/s42452-019-0884-7.