

## Applying Convolutional-GRU for Term Deposit Likelihood Prediction

<sup>1</sup>Shawni Dutta and <sup>2</sup>Prof. Samir Kumar Bandyopadhyay

<sup>1</sup>Lecturer, Department of Computer Science, The Bhawanipur Education Society College,  
Kolkata, India

<sup>2</sup>Academic Advisor, The Bhawanipur Education Society College, Kolkata, India

### Abstract-

Banks are normally offered two kinds of deposit accounts. It consists of deposits like current/saving account and term deposits like fixed or recurring deposits. For enhancing the maximized profit from bank as well as customer perspective, term deposit can accelerate uplifting of finance fields. This paper focuses on likelihood of term deposit subscription taken by the customers. Bank campaign efforts and customer detail analysis can influence term deposit subscription chances. An automated system is approached in this paper that works towards prediction of term deposit investment possibilities in advance. This paper proposes deep learning based hybrid model that stacks Convolutional layers and Recurrent Neural Network (RNN) layers as predictive model. For RNN, Gated Recurrent Unit (GRU) is employed. The proposed predictive model is later compared with other benchmark classifiers such as k-Nearest Neighbor (k-NN), Decision tree classifier (DT), and Multi-layer perceptron classifier (MLP). Experimental study concludes that proposed model attains an accuracy of 89.59% and MSE of 0.1041 which outperform well other baseline models.

**Keywords:** *Term deposit subscription, Neural network, GRU, Convolutional layers, DT, MLP, k-NN*

### Introduction-

Banking sector in globe plays a significant role to boost up socio-economic structure. Actually banks are service sector and provide support to its account holders. Deposits by account holders to banks are essential key factors for sustaining financial health of the banks. The opening of new bank or banks needs not only marketing but also campaigns. Mass campaigns target at different places for general mass and direct marketing campaigns are made with the target of a specific group. In direct marketing, the response is very low [1]. Direct marketing is not always fruitful since people may incline to the established banks. Due to evolving telemarketing through Computer technology/mobile it became now easy to generate a variety of reports through marketing campaigns and also other types of information require for the organizations. Several savings schemes offered by banks include term deposit, recurring deposit, fixed deposit, and deposits in savings account, current account and many more [2]. In the paper, only term deposit investment scheme is considered since it increase health of the bank as well as health of the account holders as

one of the important investing scheme since it facilitates the bank as well as the customers. Telemarketing campaigns influence on term deposit account subscription and the impact of these campaigns is taken into consideration by this paper.

A recommender system has been proposed in this paper that automatically predicts the possibilities of term deposit investments from client side. Term deposit is beneficial from the client side as well as the bank's perspective. Fixed amount of money is locked up for definite period of time with higher interest rates than traditional saving accounts. This will assist in gaining maximized profit not only for customers but for the banking sectors' investment as well. Term deposit is often seen as an outcome of bank market campaigns. Data mining and knowledge discovery processes [3] often play interesting role while analyzing and identifying hidden patterns and/or relationship in an enormous amount of data. Bank campaign data can be analyzed using data mining techniques and term deposit possibilities of clients may be determined beforehand. If term deposit investment possibilities are known in advance, bank sectors can look into the matter to attract clients towards their term deposit schemes.

Machine Learning (ML) techniques are useful for learning and utilizing the patterns discovered from large database. ML techniques can be applied on set of information in order to recognize underlying relationship patterns from this information set. Later, the learning can be tested in terms of incoming unknown set of patterns. Deep learning (DL) [4] often regarded as subfield of ML, can process information with minimal processing due to its self-adaptive structure. Deep learning is an expansion over conventional artificial neural networks since it facilitates the construction of networks by incorporating more than two layers [4][5]. DL based framework is implemented in this paper that is dedicated for improving the efficiency in term deposit subscription prediction using bank campaign data.

This paper proposes Convolutional-RNN based model for determining term deposit probabilities agreed by customers. To address the mentioned problem, convolutional layer[4] and Gated Recurrent Unit (GRU) [4] layers are stacked under single platform for implementing recommender system. Strength of campaign results, customer loan history, job profile, marital status etc. are considered as influential factors while identifying whether customer will place term deposit or not. All these features are given as input to the recommended system implemented in this paper. This implemented classifier model is compared with other baseline models such as k-Nearest Neighbor (k-NN) [6], Decision tree classifier (DT) [7], and Multi-layer perceptron classifier (MLP) [8].

#### **Related Work-**

Several researches have been carried out for term deposit subscription prediction. This section discusses some of the studies carried out in this field. Term deposit investment prediction is performed using three classification models such as Decision Trees, Naïve Bayes and Support Vector Machines[1]. The predictive results acquired from these models

are compared in with respect to ROC and Lift curve analysis. Support Vector Machine obtained the best results. An analysis was applied to extract useful knowledge from this classification model [1]. Using SPSS Modeler, both classification and clustering models are established in [2]. Boosted C5.0 model exhibited the best performance with highest accuracy in terms of classification. Next, clustering algorithms are applied to classify clients who have subscribed to a term deposit in order to discover and understand customers' behaviours and characteristics, social and economic context attributes [2].

Safia Abbas [9] focused on improving the efficiency of the marketing campaigns. Number of interfering features were reduced which will help in predicting the deposit customer retention criteria based on potential predictive rules. By applying decision tree (DT) and rough set theory (RST) classification module predictive results are obtained. This study concludes that application of feature reduction process, RST obtains a better summarization to the data set [9]. Another study [10] constructed logistic regression model by considering relationship between success and other factors. The classifier model predicts the success of bank telemarketing to identify the top consumer set. Some basic classifier model including Bayes, Support Vector Machine, Neural Network and Decision Tree are implemented and compared in this study. As a result, the prediction accuracy and the area under ROC curve prove the logistic regression model outperforms well in classifying than other models [10].

The concept of lifetime value (LTV) is used by Moro et al. in [11] to improve the return and to invest money/assets into bank marketing. Recency, frequency and monetary value are considered as parameters for this purpose. The results in [11] are useful for contact companies with an improved predictive performance. A comparative study is drawn in [12] among four models such as logistic regression, decision trees (DT), neural network (NN) and support vector machine in terms of two metrics AUC and ALIFT. Concluding study states that the NN achieved the best results with AUC of 0.8 and ALIFT of 0.7.

Hung et. al. [13] presented term deposit subscription prediction model using PySpark and its machine learning frameworks such as Decision Tree, Random Forest and Gradient Boosting techniques. Their study concluded accuracy rates of detection and classification reach 71% and 86% respectively[13].

Application of deep convolutional neural network is presented to predict whether a given customer is proper for bank telemarketing or not. Collection of 45,211 phone calls for 30 months is utilised for such prediction. Prediction results achieve an accuracy of 76.70% of accuracy which outperforms other conventional classifiers [14].

## **Background-**

Multiple layers of learning-nodes are stacked in Deep Learning (DL) system for understanding the features present in the raw input data. During the process each layer

transforms the output obtained from the previous layer into a representation at a higher and more abstract level. The depth allows the system to learn complex features and enables it to draw inferences [4]. DL technique exemplifies the use of neural network that mimics human brain like operations for inferring complex problem solving approach. It recognizes underlying relationships in a set of data with the provision of necessary adaptation of changing input. This will generate the best possible result without altering the output criteria [15]. Neural network comprised of several neurons. Each of these neurons will accept necessary parameters and apply some activation functions in order to produce outputs. Activation functions [16] are useful to perform diverse computations and produce outputs within a definite range. In other words, activation function is a step that associates input signal into output signal. Among several types of activation function, sigmoid and relu are two popular activation functions. A brief description of the functions are discussed as follows-

- Sigmoid activation function [16] transforms input data in the range of 0 to 1 and it is shown in equation (1).

$$f(x) = 1/(1 + \exp^{-x}) \quad (1)$$

- Rectified Linear Unit (ReLu) activation function [16] is the most successful and widely used faster activation function. It performs a threshold operation to each input element where values less than zero are set to zero whereas the values greater or equal to zeros kept as intact and it is shown in equation (2).

$$f(x) = \max(0, x) = \begin{cases} x_i, & \text{if } x_i \geq 0 \\ 0, & \text{if } x_i < 0 \end{cases} \quad (2)$$

Recurrent neural network (RNN) [4] is a type of neural network architecture that processes both sequential and parallel information. Similar operations like human brain can be simulated by incorporating memory cells to the neural network. The RNN differs from traditional neural network in terms of relationship observed among input and outputs. RNN introduces cycle in its structure which enables it to memorise past input data. Gated Recurrent Unit (GRU) is a variant of RNN. The RNN has disadvantage of having gradient vanishing problem which is eliminated by implementing GRU. The gating units present in GRU can control the flow of information inside the unit, without considering separate memory cells. GRU lacks of having memory cells in it and it has a lesser number of gates which are activated using current input as well as previous output. GRU controls the information flow from the previous activation while computing the new, candidate activation, but does not independently control the amount of the candidate activation being added. GRU can converge faster due to reduction of parameters [17]. GRU consists of update gate and reset gate in its structure. The update gate determines the amount of past

information to be passed to future processing. The amount of information to be remembered is determined by reset gate [4].

Convolutional Neural Networks (CNNs) are also improvement over traditional neural network. CNN can extract underlying hierarchical features by discovering the local relationship between nodes. Convolution operation is exhibited by each neighbour node in order to capture inherent relationship in adjacent nodes. Convolutional Layers are one of the components of CNN. The convolutional layer will determine the output of neurons of which are connected to local regions of the input through the calculation of the scalar product between their weights and the region connected to the input volume. The layers parameters focus around the use of learnable kernels [18]. In other words, an input data and a convolution kernel are subjected to particular mathematical operation to generate a transformed feature map. Convolution is often interpreted as a filter, where the kernel filters the feature map for information of a certain kind. The convolutional layer performs an operation called a “convolution”. It is a linear operation that involves the multiplication of a set of weights with the input. An array of input data and a two-dimensional array of weights, called a filter or a kernel, are multiplied for obtaining results. ReLu activation function [16] is popularly used in Convolutional layer and is proficient in most situations. Application of non-linearity applied after convolution assists in successful simulation [4].

Over-fitting is a problem when a network is unable to learn effectively from the dataset. In order to handle the problem, use of dropout layers are recommended. Dropout layers randomly deactivate a fraction of the units or connections in a network during each of the training iterations [19]. Incorporating the dropout layers in the deep model construction will assist in avoiding over-fitting problem.

While stacking several layers into a single framework, employing an optimizer is necessary. Adam is often regarded as one of the popular optimizers. This optimizer is computationally efficient with lower memory requirement and also easy to implement. This algorithm is appropriate for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. This algorithm is pretty well acknowledged due to its applicability on non-stationary objectives and problems with very noisy and/or sparse gradients [20].

Configuration of the neural model follows by execution of training process. An epoch is defined to be one cycle through which training process is executed where the dataset is partitioned into smaller subsections. An iterative process is executed through a couple of batch size that considers subsections of training dataset for completing epoch execution [21]. Binary cross entropy function is used as training criterion for solving binary classification problem. This function measures the distance from the true value (which is either 0 or 1) to the prediction for each of the classes. Class-wise errors are then averages to obtain the final Loss[22].

## Proposed Methodology-

The objective of this study is to determine customer term deposit subscription behaviors in advance. In this context, supervised classification algorithms assist in establishing predictive model by learning and discovering the relationship between a set of feature variables and a target variable. The feature variables include dominant reasons such as customer's age, job profession, marital status, education qualification, taken personal loan or not, taken home loan or not, has credit details or not, contact details, related details with the last contact of current campaign in terms of day, month, contact duration, related details with contact details, number of days passed, outcome of previous campaign. The above factors are acquired for identifying customer term deposit subscription, the target variable of the classification. The framework implemented in this paper proceeds through following series of steps.

### 1. Dataset Used-

In order to fulfill the objective of the study, Portugal bank marketing campaigns results are obtained from kaggle [23] as a collection of 45211 numbers of records and each of having 17 attributes. The attributes infer the related factors that affect campaign results. The target variable identifies whether a customer place term deposit or not. Hence a binary classification problem is addressed in this paper. Table1 provides details of attributes present in the dataset in terms of types of attributes and usage of them. Fig1 depicts the distribution of term deposit subscription tendency in the dataset. The attribute named as y is kept as the dependent variable during the classification procedure.

Attribute Name	Type of attribute	Description	Values Present
Age	Numeric	Customer's Age	18-95
Job	Categorical	Customer's Profession	'management', 'technician', 'entrepreneur', 'blue', 'unknown', 'retired', 'admin', 'services', 'self', 'unemployed', 'housemaid', 'student'
Marital	Categorical	Marital Status of Customer	'married', 'single', 'divorced'
education	Categorical	Education Qualification	'tertiary', 'secondary', 'unknown', 'primary'
Default	Categorical	Whether the customer has credit in default	'no', 'yes'
Balance	Numeric	Balance present in	-8019- 102127

		account	
Housing	Categorical	Whether the customer has housing loan or not	'yes', 'no'
Loan	Categorical	Whether the customer acquires housing loan or not	'no', 'yes'
Contact	Categorical	contact communication type	'unknown', 'cellular', 'telephone'
Day	Numeric	last contact day	1-31
Month	Categorical	last contact month of year	'may', 'jun', 'jul', 'aug', 'oct', 'nov', 'dec', 'jan', 'feb', 'mar', 'apr', 'sep'
Duration	Numeric	last contact duration, in seconds	0- 4918
Campaign	Numeric	number of contacts performed during this campaign and for this client	1-63
Pdays	Numeric	number of days that passed by after the client was last contacted from a previous campaign	1 - 871
Previous	Numeric	number of contacts performed before this campaign and for this client	0-275
Poutcome	Categorical	outcome of the previous marketing campaign	'unknown', 'failure', 'other', 'success'
Y	Numerical	Whether client subscribed a term deposit or not	0,1

Table1: Detailed description of attributes present in the dataset.

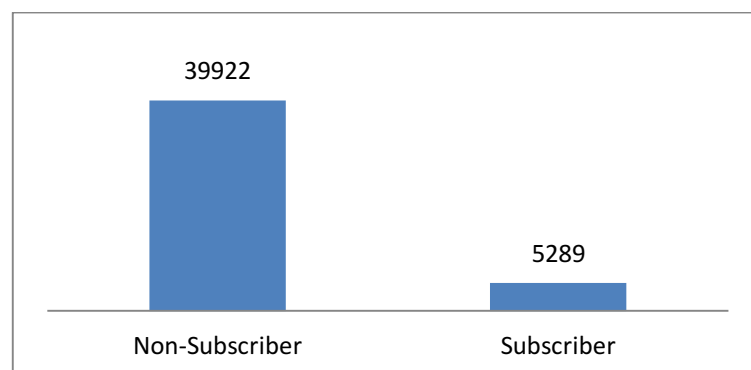


Fig.1: Subscription distribution in the dataset.



Once the dataset is collected, pre-processing techniques are applied to obtain cleaned dataset. All the categorical attributes present in the dataset is encoded into numeric data. This will be followed by scaling values of every feature with large set of data points. Feature scaling will assist the classifier to work using normalized data with an enhanced efficiency. Large set of data points are scaled down within the range of 0 to 1 using feature scaling operation. Once this feature scaling operation is performed, feature vector is fitted to classifier model as training purpose. The pre-processed dataset is bi-furcated into training and testing dataset with the ratio of 7:3. The training and testing dataset is mainly distinguished by the presence of dependent variable. The target variable is kept in training dataset whereas it is eliminated from the testing dataset. The classifier learns by extracting patterns from the training dataset during training phase. Later, prediction is acquired for the testing dataset.

## 2. Methodology-

Classification procedure is applied in this framework that is applied on the Portugal bank marketing campaigns results dataset in order to obtain term deposit subscription prediction in advance. The proposed methodology uses deep neural network which is Classification strategy is implemented by designing hybrid neural network model that assembles Convolutional and RNN layers under a single platform. While designing this model it is necessary to fine-tune hyper-parameters in order to achieve maximized efficiency. This section describes specification of the model along with its hyper-parameters.

The deep model contains two 1-dimensional Convolutional layers with filter size of 256 and 128 respectively. These layers are adjusted with kernel size of 1. Next, two GRU layers are stacked into this model with 64 and 32 nodes respectively. These four layers are followed by dropout layers with dropout rate of 0.2. Finally, four dense layers are incorporated into the deep model with 8,4,2,1 numbers of nodes respectively. In this context, either sigmoid or relu activation functions are applied in each of these specified layer. Finally these aforementioned layers are compiled using adam solver by means of 30 epochs and with a batch size of 64. Adjustment of the hyper-parameters assists the model to attain best predictive result. The deep neural network receives a total of 80,089 parameters and trains those parameters for achieving prediction. Components of the model in terms of layers, shape of output data from each layers, and number of parameters received in each layers are described in Table 2. The employed hyper-parameters while designing the proposed deep model are summarised in Table 3. The experiment has been conducted in Windows 10 Home with Intel Core i5-9300H (9th Gen), 8GB memory, and an NVIDIA GeForce GTX 1650 GPU.

Layer Type	Number of Nodes/ filter size/ Rate	Output Shape	Number of received parameters	Activation function Used
conv1d_1	256	None, 16, 256	512	ReLu



(Conv1D)				
dropout_1 (Dropout)	Dropout Rate-20%	None, 16, 256	0	None
conv1d_2 (Conv1D)	128	None, 16, 128	32896	ReLu
dropout_2 (Dropout)	Dropout Rate-20%	None, 16, 128	0	None
gru_1 (GRU)	64	None, 16, 64	37056	ReLu
dropout_3 (Dropout)	Dropout Rate-20%	None, 16, 64	0	None
gru_2 (GRU)	32	None, 32	9312	ReLu
dropout_4 (Dropout)	Dropout Rate-20%	None, 32	0	None
dense_1 (Dense)	8	None,8	264	None
dense_2 (Dense)	4	None,4	36	None
dense_3 (Dense)	2	None,2	10	None
dense_4 (Dense)	1	None,1	3	Sigmoid

Table2: Summary of the Stacked Convolutional-GRU model

Number of Epochs	30
Batch Size	64
Loss function	Binary Cross-entropy
Optimizer Used	Adam

Table 3: Specification of Parameters

### Other baseline classifier models-

Classification is a supervised machine learning technique that analyses specified set of features and identifies data as belonging to a particular class. Different classification algorithms such as decision trees, K-nearest neighbour classifier are used to predict the target class.

*K nearest neighbour* (K-NN) [6] is often considered as lazy learner which considers instances during classification process. It is known as lazy learners because during training phase it just stores training samples. This identifies objects based on closest proximity of training examples in the feature space. The classifier considers k number of objects as the nearest object while determining the class. The main challenge of this classification technique relies on picking up the suitable value of k [6].

A *Decision Tree* (DT) [7] is a classifier that exemplifies the use of tree-like structure. Top-down learning approach is exhibited by this model. Several smaller datasets are acquired from the source dataset using a statistical measure, often in the form of the Gini index or information gain via Shannon entropy. It gains knowledge on classification. Each target class is denoted as a leaf node of DT and non-leaf nodes of DT are used as a decision node that indicates certain test. The results of those tests are identified by either of the branches of

that decision node. Traversing from the beginning at the root this tree and going through it until a leaf node is reached- is the process of retrieving classification results from DT[7].

*Multi-layer perceptron* (MLP) [8] can be used as supervised classification tool by incorporating optimized training parameters. This classification algorithm is inclined to having neural network structure. MLP is often considered to be a feed-forward artificial neural network model that associates sets of input data onto a set of appropriate outputs. For a given problem, the number of hidden layers in a multilayer perceptron and the number of nodes in each layer can differ. Deciding the correct parameters depends on the training data and the network architecture[8].

These aforementioned classifiers are implemented in this framework with necessary parameter tuning. The decision tree classifier implemented in this paper uses Gini index while choosing objects from dataset. The nodes of the decision tree are expanded until all leaves are pure or until all leaves contain less than minimum number of samples. In this case, minimum number of samples is assigned a value as 2. The K-NN classifier gives a promising result for the value k=4 considering all the evaluating metric. The MLP classifier is implemented by incorporating hidden layers sizes 128, 64, 32,16,8 respectively.

The proposed hybrid deep neural network as well as baseline classifiers including decision tree, k-NN classifiers, and MLP classifier are implemented and evaluated in terms of some pre-defined metrics. These metrics will support the comparison platform while inferring the best problem-solving approach.

### **Performance Measure Metrics-**

In order to justify performance skill of a model, it is necessary to employ metrics to estimate the evaluation. The predictions obtained for the testing data is verified with actual labelled data. The verification process requires considering some pre-defined metrics. For this purpose, following metrics are used to identify the best relevant problem-solving approach.

1. Accuracy [24] is a metric that detects the ratio of true predictions over the total number of instances considered.
2. Mean Squared Error (MSE) [24] is another evaluating metric that measures absolute differences between the prediction and actual observation of the test samples. Non-negative floating value is produced as MSE. A value near to 0.0 signifies best predictive model. It can be defined as follows-

$MSE = (\sum_{i=1}^N (X_i - X_i')^2 / N)$  where  $X_i$  is the actual value and  $X_i'$  is the predicted value.

3. Accuracy does not capture the effect of false predictions during evaluation. So, to eliminate this problem, two more metrics known as, Recall and Precision are taken into consideration. Precision [24] identifies the ratio of correct positive results over the number of positive results predicted by the classifier. Recall [24] denotes the

number of correct positive results divided by the number of all relevant samples. *F1-Score* or *F-measure* [24] is a parameter that is concerned for both recall and precision and it is calculated as the harmonic mean of precision and recall.

A model that exhibits lower MSE value and higher accuracy and f1-score result turns out to be the best problem-solving approach. Mathematically, accuracy, f1-score can be defined as follows with given True Positive, True Negative, False Positive, False Negative as TP,TN,FP,FN respectively.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+TP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{F1-Measure or F1-Score} = \frac{2 * \text{Recall} * \text{Precision}}{(\text{Recall} + \text{Precision})}$$

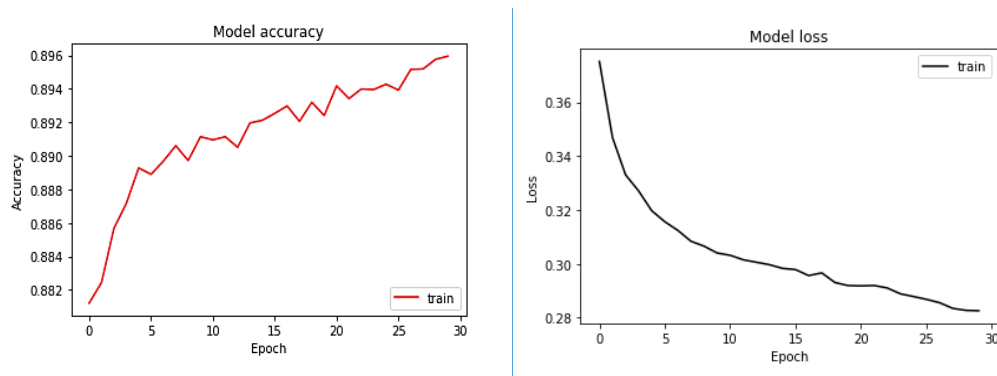
### Experimental Results-

During training, while fitting the training data into the stacked Convolutional-GRU classifier, the training process is evaluated in terms of accuracy as well as loss. For each epoch, data loss and accuracy is calculated. The best performing model will show accuracy to be increased as the number of epochs is increased. Similarly, the best model will show loss to be decreased when the number of epochs is increased. Accuracy and loss obtained for each epoch during training process of the classifier is shown in Fig 3.

After completion of training process, testing data is used for acquiring predictions. The prediction result is evaluated in terms of accuracy, f1-score and MSE. The evaluated results are shown in Table 4. Baseline classifiers such as K-NN, Decision Trees, and MLP are also evaluated with respect to aforementioned metrics. Table 4 draws comparative study between the proposed deep model and traditional ML based baseline classifiers. The comparative study indicates that the proposed model performs quite well in terms of term deposit investment possibilities.

Models	Performance Measure Metrics	Accuracy	MSE	F1-Score
Proposed Model	Stacked Convolutional-GRU Model	89.59%	0.1041	0.896
Baseline Classifier	Decision Tree Classifier	83.01%	0.17	0.83
	MLP Classifier	86.13%	0.14	0.861

	K-NN Classifier	88.83%	0.11	0.883
--	-----------------	--------	------	-------

**Table4: Performance Summarization of specified classifiers****Fig.3. Accuracy and loss acquired for each epoch of Stacked Convolutional-GRU model**

### Conclusion-

This study applies data mining techniques to forecast customers' term deposit subscription behaviours and comprehend customers' features to improve the effectiveness and accuracy of bank marketing. A hybrid deep neural model is proposed in this paper that assembles Convolutional layers and RNN based GRU layers under single platform for determining subscription behaviours at an early stage. The proposed model exemplifies fine-tuning of necessary hyper-parameters in order to maximise efficiency of prediction results. However, this method is compared with baseline classifiers such as MLP, k-NN and DT. The comparative study concludes that implemented method indicates superior result with an accuracy of 89.59%, f1-score of 0.896 and MSE 0.1168. Predictive results shown by the proposed method assist bank financial sectors to take informed decision in customer attraction process towards term deposit subscription.

### References-

- [1] S. Moro, P. Cortez, and R. M. S. Laureano, "A Data Mining Approach for Bank Telemarketing Using the rminer Package and R Tool," vol. 2014, no. September, pp. 1–23, 2013.
- [2] Q. R. Zhuang, Y. W. Yao, and O. Liu, "Application of data mining in term deposit marketing," *Lect. Notes Eng. Comput. Sci.*, vol. 2, pp. 14–17, 2018.
- [3] scar Marbn, G. Mariscal, and J. Segovi, "A Data Mining & Knowledge Discovery Process Model," *Data Min. Knowl. Discov. Real Life Appl.*, no. May 2014, 2009, doi: 10.5772/6438.
- [4] M. COŞKUN, Ö. YILDIRIM, A. UÇAR, and Y. DEMİR, "an Overview of Popular Deep Learning Methods," *Eur. J. Tech.*, vol. 7, no. 2, pp. 165–176, 2017, doi: 10.23884/ejt.2017.7.2.11.
- [5] S. H.-I. Shen Dinggang, Wu Gurrong, "Deep Learning in Medical Image Analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, no. March, pp. 221–248, 2017, doi: 10.1146/annurev-bioeng-071516.
- [6] P. Cunningham and S. J. Delany, "K -Nearest Neighbour Classifiers," *Mult. Classif. Syst.*, no. May, pp. 1–17, 2007, doi: 10.1016/S0031-3203(00)00099-6.

- [7] H. Sharma and S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954.
- [8] F. Murtagh, "Multilayer perceptrons for classification and regression," *Neurocomputing*, vol. 2, no. 5–6, pp. 183–197, 1991, doi: 10.1016/0925-2312(91)90023-5.
- [9] S. Abbas, "Deposit subscribe Prediction using Data Mining Techniques based Real Marketing Dataset," *Int. J. Comput. Appl.*, vol. 110, no. 3, pp. 1–7, 2015, doi: 10.5120/19293-0725.
- [10] Y. Jiang, "Using Logistic Regression Model to Predict the Success of Bank Telemarketing," *Int. J. Data Sci. Technol.*, vol. 4, no. 1, pp. 35–41, 2018, doi: 10.11648/j.ijdst.20180401.15.
- [11] P. R. Sérgio Moro, Paulo Cortez, "Using Customer Lifetime Value and Neural Networks to Improve the Prediction of Bank Deposit Subscription in Telemarketing Campaigns," *Neural Comput. Appl.*, vol. 26, no. 1, pp. 131–139, 2015.
- [12] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decis. Support Syst.*, vol. 62, no. February 2014, pp. 22–31, 2014, doi: 10.1016/j.dss.2014.03.001.
- [13] P. D. Hung, T. D. Hanh, and T. D. Tung, "Term deposit subscription prediction using spark MLlib and ML packages," *ACM Int. Conf. Proceeding Ser.*, pp. 88–93, 2019, doi: 10.1145/3317614.3317618.
- [14] K. H. Kim, C. S. Lee, S. M. Jo, and S. B. Cho, "Predicting the success of bank telemarketing using deep convolutional neural network," *Proc. 2015 7th Int. Conf. Soft Comput. Pattern Recognition, SoCPaR 2015*, pp. 314–317, 2016, doi: 10.1109/SOCPAR.2015.7492828.
- [15] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [16] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," pp. 1–20, 2018.
- [17] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," pp. 1–9, 2014.
- [18] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," no. December, 2015.
- [19] N. S. G. Hinton, A. K. I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting Nitish," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 15, pp. 7642–7651, 2018, doi: 10.1109/CVPR.2018.00797.
- [20] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
- [21] C. J. Shallue, J. Lee, J. Antognini, J. Sohl-Dickstein, R. Frostig, and G. E. Dahl, "Measuring the effects of data parallelism on neural network training," *J. Mach. Learn. Res.*, vol. 20, pp. 1–49, 2019.
- [22] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," *Schedae Informaticae*, vol. 25, pp. 49–59, 2016, doi: 10.4467/20838476SI.16.004.6185.
- [23] Sharan MK (2018, December). Bank Customers Survey - Marketing for Term Deposit, Version

2. Retrieved on June 1,2020 from <https://www.kaggle.com/sharanmk/bank-marketing-term-deposit>

- [24] H. M and S. M.N, "A Review on Evaluation Metrics for Data Classification Evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015, doi: 10.5121/ijdkp.2015.5201.