Journal of Engineering Technology and Applied Physics

Stacking Ensemble Approach for Churn Prediction: Integrating CNN and Machine Learning Models with CatBoost Meta-Learner

Tan Yan Lin[1], Pang Ying Han[1,*], Ooi Shih Yin[1], Khoh Wee How[1] and Hiew Fu San[2]
[1]*Faculty of Information Science and Technology, Multimedia University, Ayer Keroh, Melaka, Malaysia.*
[2]*Infineon Technologies, Free Trade Zone Batu Berendam, Batu Berendam, Melaka, Malaysia.*
**Corresponding author:* yhpang@mmu.edu.my, *ORCiD*: 0000-0002-3781-6623
https://doi.org/10.33093/jetap.2023.5.2.12
Manuscript Received: 22 June 2023, Accepted: 1 August 2023, Published: 15 September 2023

***Abstract* — In the telecom industry, predicting customer churn is crucial for improving customer retention. In literature, the use of single classifiers is predominantly focused. Customer data is complex data due to class imbalance and contain multiple factors that exhibit nonlinear dependencies. In these complex scenarios, single classifiers may be unable to fully utilize the available information to capture the underlying interactions effectively. In contrast, ensemble learning that combines various base classifiers empowers a more thorough data analysis, leading to improved prediction performance. In this paper, a heterogeneous ensemble model is proposed for churn prediction in the telecom industry. The model involves exploratory data analysis, data pre-processing and data resampling to handle class imbalance. In this proposed model, multiple trained base classifiers with different characteristics are integrated through a stacking ensemble technique. Specifically, convolutional-based neural network, logistic regression, decision tree and Support Vector Machine (SVM) are considered as the base classifiers in this work. The proposed stacking ensemble model utilizes the unique strengths of each base classifier and leverages collective knowledge to improve prediction performance with a meta-learner. The efficacy of the proposed model is assessed on a real-world dataset, i.e., Cell2Cell. The empirical results demonstrate the superiority of the proposed model in churn prediction with 62.4 % f1-score and 60.62 % recall.**

Keywords—Customer Churn Prediction, Heterogeneous Ensemble Learning, Machine Learning, Deep Learning, Stacking Ensemble.

I. INTRODUCTION

The widespread use of mobile phones has contributed to an exponential surge in the importance of telecommunication. The telecommunications industry has evolved into an indispensable aspect of our daily lives, as evidenced by the statistics showing that the number of smartphone users in 2023 is 6.92 billion [1]. According to the World Advertising Research Center, by the end of 2025, 72.6% of all Internet users will access web pages via smartphones. Even with a large number of telecommunication service subscribers, the telecommunication industry remains highly competitive with significant churn. The intense competition among telecommunications companies, technological advancements, and competitive pricing plans are the factors that contribute to churn.

Retaining existing customers is generally less costly than acquiring new customers as it requires additional marketing expenses to capture the attention of potential customers [2]. Churn leads to potential revenue and profit losses for a company. In order to mitigate customer churn, telecom service companies should implement effective marketing strategies that focus on retaining existing customers. Hence, effective customer analytics, particularly churn prediction, play a crucial role in helping telecom service companies reduce customer churn. Churn prediction enables telecom service providers to detect and identify customers who are likely to churn. This allows them to strategize and take proactive retention measures to reduce churn rates. With churn prediction, telecom companies may examine various factors, including customer usage patterns, billing information, customer feedback, service interactions, etc., to detect early warning signs of potential churn. Providers can respond promptly to retain those at-risk customers before they switch to a competitor.

In churn prediction systems, machine learning classifiers are widely deployed for analysing and predicting customer churn. Generally, machine learning classifiers can be categorised into two spheres: single classifiers and ensemble classifiers. A single classifier is a standalone model or algorithm used for making predictions or classifying data. Examples of single classifiers include decision trees, logistic regression, decision tree, Support Vector Machine and naïve Bayes [3 - 5]. On the other hand, an ensemble classifier is a method that incorporates multiple individual classifiers to make predictions or classify data. It integrates the predictions of multiple classifiers to produce more accurate predictions compared to using a single classifier. Examples of ensemble classifiers include Extreme Gradient Boosting (XGBoost), Adaboost, CatBoost, random forest, and voting classifiers [6 - 9].

Ensemble classifiers can be further classified into two types: homogeneous and heterogeneous ensembles. Homogeneous ensembles use the same-type classifiers but adopt different sampling methods during the training phase. Employing different sampling methods provides subsets of the training data that are slightly different from each other. This variation helps in capturing various facets of the underlying patterns. On the other hand, heterogeneous ensembles combine multiple different types of classifiers or models to make predictions [10]. Each base classifier contributes to the ensemble with its unique strengths and characteristics for better prediction performance.

Considering the complexity of customer churn prediction and the multitude of factors in customer data, the utilization of heterogeneous ensembles is highly advantageous. Heterogeneous ensembles can efficiently deal with the intricacies and leverage the unique strengths of each base classifier. This enables a comprehensive analysis of customer churn, leading to a deeper comprehension of churn behaviour and improved prediction performance.

The contributions of this paper are as follows:

- a heterogeneous ensemble model is proposed for customer churn prediction. This ensemble model empowers a more thorough data analysis by leveraging the knowledge of different base classifiers.
- a series of customised preprocessing processes is meticulously performed before data modelling. Exploratory data analysis and data pre-processing are conducted to delve profoundly into the database, gaining valuable insights and enhancing data quality.
- Class imbalance is a usual challenge in binary classification tasks, including churn prediction. The Synthetic Minority Over-sampling Technique (SMOTE) is applied to tackle this problem. Hence, the proposed model can be trained on a balanced dataset, avoiding biases towards the majority class.

In this work, the performance of the proposed ensemble learning model is evaluated on a real-world telecommunications dataset, known as Cell2Cell. From the obtained empirical results, a promising performance in the application of churn prediction is obtained.

II. RELATED WORK

Churn is a significant concern for all businesses for business sustainability. Extensive methods have been presented for customer churn prediction in the telecommunication industry. For example, the work of [3] explored different classifiers on multiple datasets on customer segmentation and customer churn prediction. The classifiers included logistic regression, decision tree, random forest, Naïve Bayes, AdaBoost and Multi-Layer Perceptron. In the paper, the importance of data sampling for addressing class imbalance was studied. Furthermore, the authors also reported that AdaBoost and random forest produced better performance compared with the other classifiers. Nguyen & Duong [6] highlighted that the techniques for handling imbalanced data could be classified into two categories: resampling methods and cost-sensitive learning methods. The former methods balance the data before model training; the latter methods adjust the relative costs of the errors during model training. The authors compared the performance of two resampling methods, which are SMOTE and Deep Belief Network (DBN) against the two-cost sensitive learning methods (i.e. focal loss and weighted loss). The reported experimental results showed that the focal loss and weighted loss methods performed better than the SMOTE and DBN. The best-performing method was the combination of XGBoost with focal loss and weighted loss with an AUC of 66.18 % and an AUC of 65.92 % on the Cell2Cell dataset.

Furthermore, Wang *et al.* [11] proposed an ensemble learning on churn prediction. LightGBM, XGBoost, random forest, and decision tree were employed as base classifiers in the ensemble learning model. A soft voting technique was used. The proposed ensemble model exhibited a promising performance. Hammoudeh *et al.* proposed a selective ensemble model for predicting the customers who are more likely to churn in the mobile telecommunication industry [12]. The proposed selective ensemble model dynamically selected a combination of machine learning models from a pool of models to contribute to the formation of the final outcome. The empirical results showed that the proposed selective ensemble model outperformed its constituent models and the averaging ensemble model. Idris *et al.* developed a particle swarm optimization (PSO)-based undersampling method [13]. The proposed method could handle the issue of imbalanced data in collaboration with various techniques such as Principal Component Analysis, Fisher's ratio, F-score and Minimum Redundancy and Maximum Relevance (mRMR). To assess the performance on optimally selected features, random forest and K-Nearest Neighbour classifiers were used. The empirical results revealed that the proposed approach based on PSO, mRMR and random forest, coined as Chr-PmRF, performed well for churn prediction.

In recent years, the rise of deep learning has revolutionized the field of churn prediction. Ahmed *et al.* [14] proposed a model, called TL-DeepE, for churn prediction. This model employed the principle of Transfer Learning and Ensemble-based Meta-Classification. In this model, a multi-layer convolutional neural network (CNN) and Genetic Programming (GP) were integrated, whereas the AdaBoost model was adopted as the meta-classifier. In this work, the telecom data was transformed into two-dimensional data. Three pre-trained convolutional neural networks (i.e. AlexNet, Inception-ResNet-V2, and a custom 6-layer deep neural network model) and transfer learning were explored. The features extracted from the CNN models were used as the input for the GP-AdaBoost ensemble classifier. The proposed TL-DeepE achieved an encouraging churn prediction performance. In addition, a study on churn analysis using the deep learning method is conducted by Albrecht *et al.* [15]. This paper presented an analytical framework for effectively implementing deep learning techniques in customer churn prediction. The framework outlined a structured application procedure tailored for practical users, providing guidelines and recommendations.

In 2022, Wael Fujo *et al.* applied deep learning to customer churn prediction [16]. In this study, a deep back-propagation artificial neural network, known as Deep-BP-ANN, was implemented using two feature selection methods (i.e. Variance Thresholding and Lasso Regression). They worked on two popular telecom datasets which are IBM Telco and Cell2cell datasets. To address the issue of data imbalance, the Random Oversampling technique was employed to balance both datasets. The proposed model outperformed the existing deep learning techniques that used holdout or 10-fold cross-validation on the same datasets.

The telecommunication sector has witnessed significant improvements in churn prediction models due to advancements in machine learning and deep learning techniques. Overall, the incorporation of machine learning and deep learning techniques, along with ensemble learning, has led to significant advancements in churn prediction models. These advancements have paved the way for more accurate and reliable predictions, ultimately benefiting businesses in managing customer churn effectively.

III. PROPOSED METHOD DESIGN

The overview of the proposed churn prediction framework is depicted in Fig. 1. Firstly, exploratory data analysis (EDA) and data pre-processing techniques are performed on the customer data. The aim of conducting EDA is to understand the data and identify the data patterns. On the other hand, data pre-processing involves cleaning the dataset and handling those missing values to prepare the data for modelling. Furthermore, data transformation is carried out to transform the raw data into a suitable format for machine learning. This process involves data encoding and scaling to optimise model performance. Class imbalance is a prevalent challenge in churn prediction. Hence, a data sampling technique is conducted to address the imbalanced classes. Lastly, a churn prediction model is built and evaluated.

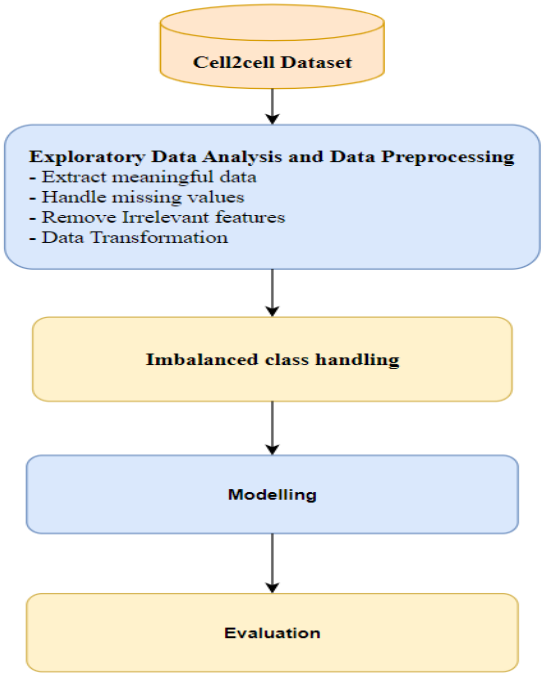

Fig. 1. The overview of the proposed churn prediction model.

A. Cell2cell Dataset

The adopted dataset in this study is the Cell2cell dataset which consists of 51,047 instances and 58 attributes (recorded in the Appendix). It is collected by the Teradata Center at Duke University and is a publicly available dataset [17]. The selection of this dataset in this study is based on its popularity in churn prediction studies [3, 6, 14 - 16].

B. Exploratory Data Analysis & Data Pre-processing

Exploratory Data Analysis and data pre-processing are crucial steps prior to modelling. EDA provides insights and a comprehensive understanding of the dataset. The process helps in identifying missing values, eliminating irrelevant variables, and data transformation to facilitate effective data pre-processing.

Handle Missing Value

Handling missing values is one of the crucial steps to ensuring reliable data analysis and averting biases. If missing values are not handled appropriately, results may be skewed or erroneous. Figure 2 visualises Cell2Cell's features with missing values. There are 14 features containing missing values, such as Monthly Revenue, Monthly Minutes, Total Recurring Charge, etc. In this study, the features with missing values are imputed with a value of 0.

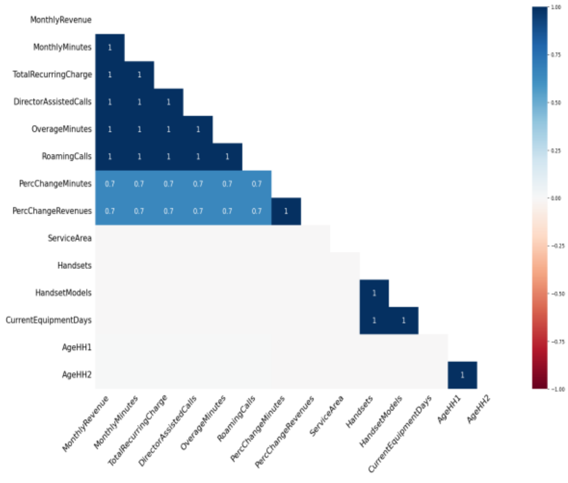

Fig. 2. Features with missing values.

Remove Irrelevant Features

During EDA, it is determined that the features of "Customer ID" and "Service Area" do not contribute meaningful insights. “Customer ID” is only an identifier and does not bring any predictive information; whilst “Service Area” is a near-unique feature: NYCBRO917 with 3 %, HOUHOU281 with 3 % and other values 94 %. The near uniqueness impacts the high cardinality, resulting in computational challenges and memory inefficiency during data modelling. Besides that, with an enormous quantity of unique values, some areas may have very few instances, resulting in sparse representation. Hence, it is challenging for a model to learn meaningful patterns. In other words, these features are determined to be irrelevant or lack significant relevance to the modelling process. Thus, they are dropped in this study to improve the data quality, reduce computational complexity and improve model performance.

Data Transformation

Data Transformation techniques can significantly enhance the overall performance of the model [18]. Categorical variable transformation is indeed crucial for the majority of machine learning models because most of them can only deal with numeric values. Therefore, the categorical features will be transformed into a numeric machine-friendly form before modelling. The label Encoding method is used to transform the categorical data in this study. For instance, the feature ‘Churn’ which contains two classes (Yes, No) is transformed into 0 and 1; the categorical feature PrizmCode’ with 4 classes (Rural, Suburban, Town, Others), is transformed into 0, 1, 2, and 3. Moreover, data scaling is also applied to scale the features’ values to a range of [0 1] with the min-max scaling method.

C. Imbalance Class Handling

Customer churn is a binary classification. Binary classification often encounters imbalance class problems in which churners belong to the minority class and non-churners belong to the majority class. Imbalance class problems can bring a huge negative impact on the model, leading the model biased towards the majority class. The Cell2cell dataset exhibits a high degree of data imbalance, with a churn rate of 71.2 % (23,238 samples) and a non-churner rate of 28.87 % (9431 samples), depicted in Fig. 3. The work of [19] utilises sampling techniques to handle the imbalance class problem with Synthetic minority oversampling technique, random oversampling (ROS), etc. The empirical results showed that the SMOTE method is the most effective approach for data balancing. Thus, we adopt this technique to solve the imbalance class problem. SMOTE involves generating synthetic instances of the minority class. From Fig. 4, an instance x_i from the minority class is selected. Then, several nearest neighbours of the same class, represented as x_{i1} to x_{i4}, are chosen using a distance measure. A randomised interpolation is then performed, and results in the generation of new instances labelled as r_1 to r_2 [6].

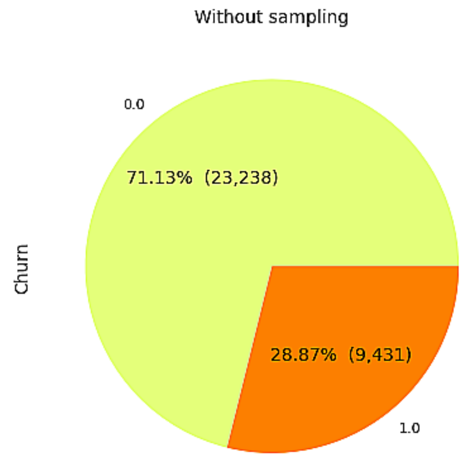

Fig. 3. Percentage of churners and non-churners without sampling.

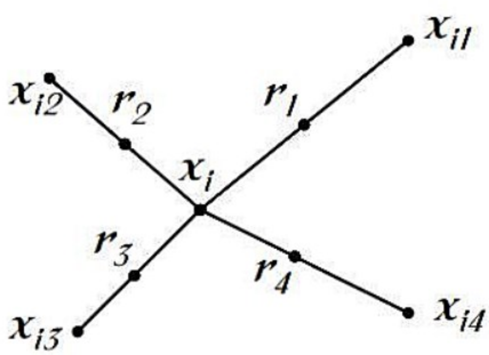

Fig. 4. SMOTE [6].

After data sampling, the distribution of churners and non-churners is balanced, which results in 50 % for each class, as shown in Fig. 5. By applying the SMOTE method, the number of churners has increased from 9,431 to 23,238.

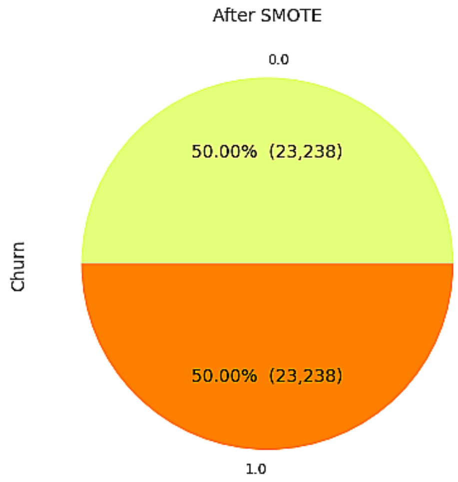

Fig. 5. Percentage of churners and non-churners after SMOTE.

D. Modelling: Ensemble Learning

Ensemble learning is a technique that combines multiple machine learning algorithms' predictive information and aims to improve the performance of the classifier [20]. Multiple models with different characteristics are integrated into a meta-classifier to eliminate the limitations of every single classifier and enhance its generalisation potential, which is known as heterogeneous ensembles. This paper presents a new approach by proposing a heterogeneous ensemble model for customer churn prediction. This proposed ensemble model incorporates multiple base classifiers and each classifier produces an independent prediction. After that, the predictions are input into a meta-learner, which learns from the combined predictions to produce the final prediction.

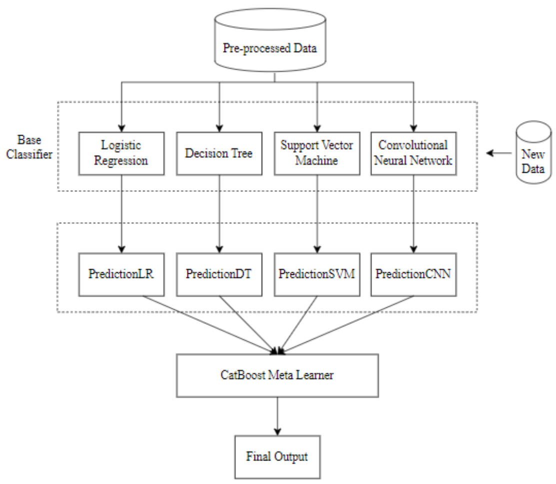

Fig. 6. The proposed Ensemble Model.

The proposed stacking ensemble model is illustrated in Fig. 6. In this stacking ensemble, there are two levels: (1) the base classifier level and (2) the meta-learner level. At the base classifier level, each adopted base classifier is trained by using the pre-processed data and individual predictions are produced. Next, the predictions of the base classifiers are mapped to the corresponding actual label via meta-learner. In this work, different types of base classifiers, including logistic regression, decision tree, support vector machine and custom-convolutional neural network models, are considered for feature diversity. Besides, the CatBoost classifier is adopted as the meta-learner because of its capability to handle diverse data. The capability of CatBoost to handle a wide range of data characteristics makes it well-suited for aggregating the predictions of the base classifiers, thereby ensuring the overall performance of the ensemble.

Base Classifier: Logistic Regression

Logistic regression is a widely adopted statistical modelling technique that falls under the category of supervised machine learning algorithms. It can be binary, multinomial, or ordinal. In this study, binary logistic regression is being adopted. In logistic regression, if the probability of prediction is more than 0.5, the output will be assigned to class 0 (non-churners); otherwise, the output is assigned to class 1 (churners). The logistic regression can be formulated as [21]:

$$P = \frac{1}{1 + e^{-z}} \tag{1}$$

, where P is the probability of churn and z represents the linear combination of customer's features weighted by their coefficient which can be written as:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \tag{2}$$

, where $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are coefficients to the independent variables $X_1, X_2, \ldots, X_n$.

Base Classifier: Decision Tree

A decision tree is a supervised machine learning method for classification and prediction problems. It consists of a tree-like model with the internal node, branches, and leaf nodes [22]. The interval node represents a test on a feature, the branch corresponds to the possible outcome of the test, and the leaf nodes indicate the final class label. It is used for making decisions based on the input from the dataset and leads to the prediction of the target. The construction of a decision tree typically operates in a top-down manner, in which the tree is constructed and evaluated from the root node to downward. The root node indicates the initial feature that is used for classification, and the following nodes are selected based on criteria such as information gain or impurity. Entropy refers to one of the several impurity measures used in decision trees.

$$Entropy(E) = -\sum_{i=1}^{N} p_i \log_2 p_i \tag{3}$$

, where p_i is the probability of each class i within the node.

Decision tree usually involves two main phrases, which are tree building and tree pruning. In tree-building phrases, the algorithm recursively partitions the train data based on the attribute value. The data partitioning process is done until the partitions contain mostly identical values. The pruning step is to focus on selecting and removing branches that contain noisy data with the largest estimated error rate. This results in a more optimal model.

Base Classifier: Support Vector Machine

Support Vector Machine is a supervised learning algorithm for linear and nonlinear classification problems. It involves transforming the data into a higher dimensional space to identify an optimal hyperplane to separate the data [23]. The chosen hyperplane, which is known as the support vector, maximises the margin (i.e., the distance between the hyperplane and the nearest data points from each class). In binary classification, SVM seeks an optimal hyperplane for separating the data points of two different classes. The function can be defined as [23]:

$$Z = f(y) = sign\left[\sum_{i=1}^{N} y_i d_i K(x, x_i) + c\right] \quad (4)$$

where $sign$ is the sign function, y_i refers to class label for training data points, d_i and c are parameters that refer to the hyperplane. $K(x, x_i)$ denotes the kernel radial basis function (RBF).

Base Classifier: Convolutional Neural Network

Convolutional neural network is a powerful deep learning method that is widely employed in various domains. While CNNs are commonly associated with image processing tasks [24], they are also utilized for analyzing one-dimensional data, such as churn prediction in this work. CNN models typically follow a series of steps to extract crucial features from customer data while preserving the relationships between these features and the corresponding labels. Firstly, a convolutional process is performed, where filters are utilized to extract important features from the data. Non-linear activation functions, such as rectified linear units (ReLu) or sigmoids, are adopted in CNN models to create connections between input features and the hidden layers. These activation functions enable the CNN models to learn the complex correlations in the data. Lastly, the classification step takes place. The learned features from the previous layers are fed into fully connected layers to generate the predicted output. In summary, the core architecture of a CNN model comprises an input layer, hidden layers, and an output layer. The input layer receives customer data, and the hidden layers consist of calculated weights along with activation functions, to extract relevant features from the data. Finally, the output layer produces the predicted outcomes or classifications based on the learned features.

Meta-learner: CatBoost

CatBoost is a gradient-boosting on decision tree algorithm that is well-suited for handling binary classification tasks. CatBoost stores the binary features in a continuous vector B, and the values of leaf nodes in decision trees are stored in a two-dimensional float number vector. To establish the binary vector for sample y [25]:

$$\sum_{i=0}^{d-1} 2^i \cdot B(y, f(t,d)) \quad (5)$$

, where the binary vector $B(y, f)$ reads the value of binary feature f for the sample y, $f(t, d)$ retrieves the number of binary features at depth d in tree t. CatBoost constructs vectors in parallel with data, and results in faster processing through the automatic handling of categorical features.

IV. EXPERIMENT RESULT AND DISCUSSION

A. Experimental Setup and Performance Metrics

For this study, the experiments were carried out using Jupyter Notebook with Python 3.10 in Google Collaboratory. The computational resources included Intel Core i5 processors equipped with 4 GB of RAM. In churn prediction, several performance metrics were used to assess the performance of the proposed ensemble model. These include accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). Accuracy is the number of correctly classified data divided by the total number of data; precision is the measure of correctly classified data (true positive) out of the total number of data; recall measures the correctly classified data out of all actual churn data; F1-score combines the precision and recall, providing a balanced assessment for the performance of the model; and the AUC shows how well the model can distinguish between positive and negative instances.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$precision = \frac{TP}{TP + FP} \quad (7)$$

$$recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 = \frac{Precision \times Recall \times 2}{Precision + Recall} \quad (9)$$

$$AUC = \int_{-\infty}^{\infty} y(t)\, dx\, (t) \quad (10)$$

TP is true positive, where a customer is churn (positive) and classified as churn (positive);
TN is true negative, where a customer is not churn (negative) and classified as not churn (negative);
FP is false positive, where a customer is not churn (negative) and classified as churn (positive);
FN is false negative, where a customer is churn (positive) and classified as churn (negative).

B. Results and Discussion

This section will discuss the performance analysis of the base classifiers, followed by the performance of the proposed ensemble model. Besides that, the performance comparison with other existing models is also discussed.

Performance Analysis on the Base Classifiers

Figure 7 illustrates the performance of the base classifiers. The logistic regression classifier obtains better overall performance, with a 62 % AUC, 59.65 % accuracy, 66.61 % precision, 59.65 % recall, and 61.56 % F1-score. However, the CNN model has the lowest performance. One potential reason can be that CNNs are primarily designed to capture spatial dependencies in data, i.e., spatial relationships between image pixels; But customer churn prediction typically involves tabular or sequential data, where the spatial structure is less pronounced. CNNs may not fully leverage the temporal or contextual patterns present in such data, leading to suboptimal performance. On the other hand, the decision tree classifier and SVM perform reasonably well.

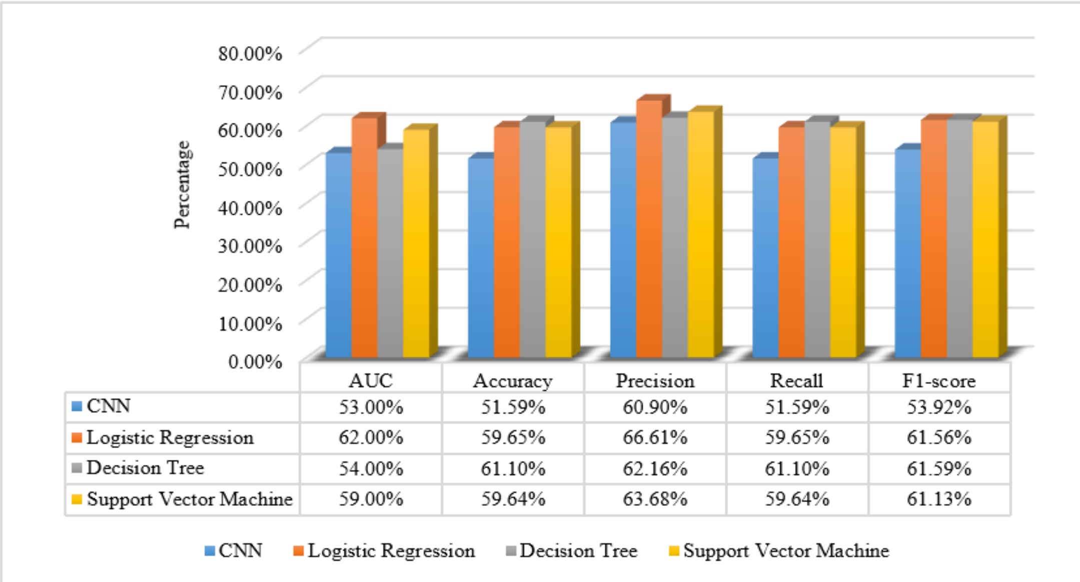

Fig. 7. Performance analysis of base classifier.

Performance Analysis on the Ensemble Model

The proposed ensemble model incorporates three traditional machine learning models, i.e., logistic regression, decision tree, and SVM, along with a deep learning model, i.e., CNN. The stacking ensemble technique is employed in this model, which involves training these base classifiers and then combining their predictions using a meta-learner, i.e., CatBoost. The prediction performance of the proposed ensemble model is illustrated in Fig. 8. From the results, we can notice that the proposed model achieves higher F1 and AUC scores. This indicates that the proposed ensemble approach has effectively improved the overall predictive performance. From the good F1 score, we can deduce that the proposed model performs better at balancing precision and recall. This denotes that the proposed ensemble is effective in detecting true positives and minimizing false positives and false negatives compared to the individual base classifiers.

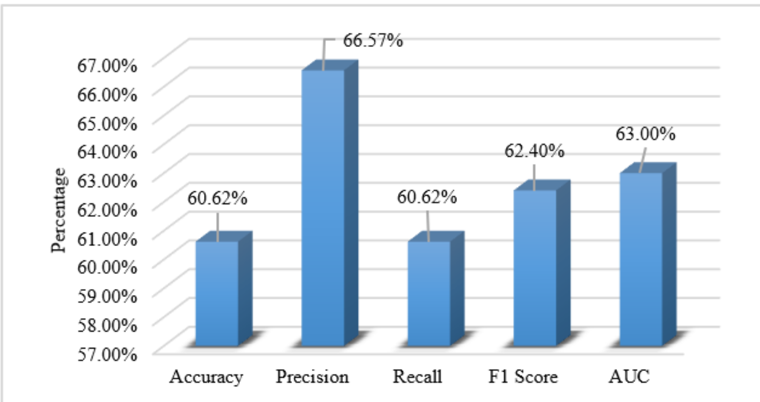

Fig. 8. The performance of the proposed ensemble model.

Performance Comparison

A performance comparison between the proposed model and the other existing models is addressed in this section. The performance of the models, in terms of accuracy, precision, recall, F1-score and AUC, are recorded in Table I. From the results, it is noticed that, generally, the proposed model achieves better precision, recall and F1 score. This achievement indicates the proposed model's effectiveness in accurately identifying positive instances, that is churn in this work, while minimizing false positives and false negatives.

Table I: Performance Comparison between the proposed ensemble model and other classifiers.

Classifier	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression [3]	0.576	0.3516	0.5367	0.4109	0.5866
AdaBoost [3]	0.5863	0.3653	0.4932	0.4052	0.5723
XGBoost [6]	-	-	-	-	0.6618
TL-DeepE [14]	0.6816	-	-	-	0.74
Deep Neural Network [15]	-	0.401	0.484	-	0.645
Proposed ensemble model	0.6062	0.6657	0.6062	0.624	0.63

On the other hand, we can notice that the TL-DeepE model achieves better accuracy and AUC. This may be because the TL-DeepE model that incorporates multiple deep learning models (pre-trained AlexNet, Inception-ResNet-V2 and CNN) is able to capture more customer behaviours, extracting more comprehensive representations of churn-related patterns. However, integrating multiple deep learning models into an ensemble may introduce additional complexities.

V. CONCLUSION

In recent years, churn has become a significant concern for telecom companies. In this work, a customer churn prediction model is proposed. The proposed model employs a stacking ensemble approach by incorporating both machine learning and deep learning models. The proposed ensemble consists of four base classifiers: logistic regression, decision tree, SVM, and custom-CNN. The proposed stacking ensemble model utilizes the strengths of each base classifier and leverages collective knowledge to improve prediction performance with the CatBoost meta-learner. The empirical results obtained from evaluating the proposed model demonstrate a promising performance, as indicated by relatively good scores of precision, recall, and F1-score. We can deduce that the encouraging performance of the proposed model suggests its potential applicability in churn prediction tasks.

In future work, an end-to-end learning model will be explored. In other words, a single unified framework will be developed to integrate the whole process of feature extraction, data representation analysis and prediction.

APPENDIX

Table A1: Description of dataset.

#	Numerical Features	Data Format	Type
1	CustomerID	*int*	*Numerical*
2	Churn	*(Yes/No)*	*Categorical*
3	MonthlyRevenue	*float*	*Numerical*
4	MonthlyMinutes	*float*	*Numerical*
5	TotalRecurringCharge	*float*	*Numerical*
6	DirectorAssistedCalls	*float*	*Numerical*
7	OverageMinutes	*float*	*Numerical*
8	RoamingCalls	*float*	*Numerical*
9	PercChangeMinutes	*float*	*Numerical*
10	PercChangeRevenues	*float*	*Numerical*
11	DroppedCalls	*float*	*Numerical*
12	BlockedCalls	*float*	*Numerical*
13	UnansweredCalls	*float*	*Numerical*
14	CustomerCareCalls	*float*	*Numerical*
15	ThreewayCalls	*float*	*Numerical*
16	ReceivedCalls	*float*	*Numerical*
17	OutboundCalls	*float*	*Numerical*
18	InboundCalls	*float*	*Numerical*
19	PeakCallsInOut	*float*	*Numerical*
20	OffPeakCallsInOut	*float*	*Numerical*
21	DroppedBlockedCalls	*float*	*Numerical*
22	CallForwardingCalls	*float*	*Numerical*
23	CallWaitingCalls	*float*	*Numerical*
24	MonthsInService	*int*	*Numerical*
25	UniqueSubs	*int*	*Numerical*
26	ActiveSubs	*int*	*Numerical*
27	ServiceArea	*string*	*Categorical*
28	Handsets	*float*	*Numerical*
29	HandsetModels	*float*	*Numerical*
30	CurrentEquipmentDays	*float*	*Numerical*
31	AgeHH1	*float*	*Numerical*
32	AgeHH2	*float*	*Numerical*
33	ChildrenInHH	*(Yes/No)*	*Categorical*
34	HandsetRefurbished	*(Yes/No)*	*Categorical*
35	HandsetWebCapable	*(Yes/No)*	*Categorical*
36	TruckOwner	*(Yes/No)*	*Categorical*
37	RVOwner	*(Yes/No)*	*Categorical*
38	Homeownership	*(Known/ Unknown)*	*Categorical*
39	BuysViaMailOrder	*(Yes/No)*	*Categorical*
40	RespondesToMailOffers	*(Yes/No)*	*Categorical*
41	OptOutMailings	*(Yes/No)*	*Categorical*
42	NonUSTravel	*(Yes/No)*	*Categorical*
43	OwnsComputer	*(Yes/No)*	*Categorical*
44	HasCreditCard	*(Yes/No)*	*Categorical*
45	RetentionCalls	*int*	*Numerical*
46	RetentionOffersAccepted	*int*	*Numerical*
47	NewCellphoneUser	*(Yes/No)*	*Categorical*
48	NotNewCellphoneUser	*(Yes/No)*	*Categorical*
49	ReferralsMadeBySubscriber	*int*	*Numerical*
50	IncomeGroup	*int*	*Numerical*
51	OwnsMotorcycle	*(Yes/No)*	*Categorical*
52	AdjustmentsToCreditRating	*int*	*Numerical*
53	HandsetPrice	*string*	*Categorical*
54	MadeCallToRetentionTeam	*(Yes/No)*	*Categorical*
55	CreditRating	*String*	*Categorical*
56	PrizmCode	*(Rural/Suburban/Town/Other)*	*Categorical*
57	Occupation	*(Clerical/Crafts/ Homemaker/Professional/Self/Student/Retired/ Other)*	*Categorical*
58	MaritalStatus	*(Yes/No/Unknown)*	*Categorical*

REFERENCES

[1] "*How Many People Have Smartphones Worldwide* (Jun 2023)," https://www.bankmycell.com/blog/how-many-phones-are-in-the-world (accessed Jun. 20, 2023).

[2] K. W. De Bock and D. Van Den Poel, "An Empirical Evaluation of Rotation-based Ensemble Classifiers for Customer Churn Prediction," *Expert Sys. Appl.*, vol. 38, no. 10, pp. 12293–12301, 2011.

[3] S. Wu, W. C. Yau, T. S. Ong and S. C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," *IEEE Access*, vol. 9, pp. 62118–62136, 2021.

[4] N. Y. Nhu, T. Van Ly and D. V. Truong Son, "Churn Prediction in Telecommunication Industry Using Kernel Support Vector Machines," *PLoS One*, vol. 17, no. 5, pp. e0267935, 2022.

[5] A. Sharma, P. Shukla, M. K. Gourisaria, B. Sharma and I. Ben Dhaou, "Telecom Churn Analysis Using Machine Learning in Smart Cities," in *1st Int. Conf. Advanced Innov. Smart Cities,* pp. 1-5, 2023.

[6] N. N. Nguyen and A. T. Duong, "Comparison of Two Main Approaches for Handling Imbalanced Data in Churn Prediction Problem," *J. Advances Inform. Technol.*, vol. 12, no. 1, pp. 29–35, 2021.

[7] P. Lalwani, M. K. Mishra, J. S. Chadha and P. Sethi, "Customer Churn Prediction System: A Machine Learning Approach," *Computing*, vol. 104, no. 2, pp. 271–294, 2022.

[8] J. J. R. Angelina, S. J. Subhashini, S. H. Baba, P. D. K. Reddy, P. V. S. K. Reddy, and K. S. Khan, "A Machine Learning Model for Customer Churn Prediction Using CatBoost t/Classifier," in *2023 7th Int. Conf. Intellig. Comput. and Contr. Sys.*, pp. 166–172, 2023.

[9] S. F. Bilal, A. A. Almazroi, S. Bashir, F. H. Khan and A. A. Almazroi, "An Ensemble Based Approach Using A Combination of Clustering and Classification Algorithms to Enhance Customer Churn Prediction in Telecom Industry," *PeerJ Comput. Sci.*, vol. 8, pp. e854, 2022.

[10] "*Why Use Ensemble Learning?* MachineLearningMastery.com," https://machinelearningmastery.com/why-use-ensemble-learning/ (accessed Jun. 19, 2023).

[11] X. Wang, K. Nguyen and B. P. Nguyen, "Churn Prediction Using Ensemble Learning," in *Proc. 4th Int. Conf. Machine Learning and Soft Comput.*, pp. 56–60, 2020.

[12] A. Hammoudeh, M. Fraihat and M. Almomani, "Selective Ensemble Model for Telecom Churn Prediction," in *2019 IEEE Jordan Int. Joint Conf. Electr. Eng. and Inform. Technol.,* pp. 485–487, 2019.

[13] A. Idris, M. Rizwan and A. Khan, "Churn Prediction in Telecom Using Random Forest and PSO Based Data Balancing in Combination with Various Feature Selection Strategies," *Comp. & Electr. Eng.*, vol. 38, no. 6, pp. 1808–1819, 2012.

[14] U. Ahmed, A. Khan, S. H. Khan, A. Basit, I. U. Haq and Y. S. Lee, "*Transfer Learning and Meta Classification Based Deep Churn Prediction System for Telecom Industry*," Jan. 2019, Accessed: May 28, 2023. [Online]. Available: https://arxiv.org/abs/1901.06091v2

[15] T. Albrecht and D. Baier, "Churn Analysis Using Deep Learning: Customer Classification from a Practical Point of View," *Data Science-Series A*, vol. 6, no. 2, pp. P04, 2020.

[16] S. W. Fujo, S. Subramanian and M. A. Khder, "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," *Inform. Sci. Lett.*, vol. 11, no. 1, pp. 185-198, 2022.

[17] "*telecom churn (cell2cell) | Kaggle*," https://www.kaggle.com/datasets/jpacse/datasets-for-churn-telecom (accessed May 29, 2023).

[18] N. I. Mohammad, S. A. Ismail, M. N. Kama, O. M. Yusop and A. Azmi, "Customer Churn Prediction in Telecommunication Industry Using Machine Learning Classifiers," in *Proc. 3rd Int. Conf. Vision, Image and Signal Processing*, no. 34, pp. 1–7, 2019.

[19] D. A. Kumar and V. Ravi, ''Predicting Credit Card Customer Churn in Banks Using Data Mining,'' *Int. J. Data Anal. Technol. Strategies*, vol. 1, no. 1, pp. 4–28, 2008.

[20] I. Kaur and J. Kaur, "Customer Churn Analysis and Prediction in Banking Industry Using Machine Learning," in *2020 Sixth Int. Conf. Parallel, Distributed and Grid Comput.*, pp. 434-437, 2020.

[21] Q. Yanfang and L. Chen, "Research on E-commerce User Churn Prediction Based on Logistic Regression," in *Proc. 2017 IEEE 2nd Inform. Technol., Netw., Electron. and Auto. Contr. Conf.,* vol. 2018, pp. 87–91, 2018.

[22] R. S. Shankar, J. Rajanikanth, V. V. Sivaramaraju and K. V. S. S. R. Murthy, "Prediction of Employee Attrition Using Datamining," in *2018 IEEE Int. Conf. Sys., Comput., Auto. and Netw.*, India, pp. 1-8, 2018.

[23] S. B. Jabeur, C. Gharib, S. M. Wali, and W. B. Arfi, "CatBoost Model and Artificial Intelligence Techniques for Corporate Failure Prediction," *Technol. Forecast Soc. Change*, vol. 166, pp. 120658, 2021.

[24] N. Alboukaey, A. Joukhadar and N. Ghneim, "Dynamic Behavior Based Churn Prediction in Mobile Telecom," *Expert Sys. with Appl.*, vol. 162, pp. 113779, 2020.

[25] Y. Deng, D. Li, L. Yang, J. Tang and J. Zhao, "Analysis and Prediction of Bank User Churn Based on Ensemble Learning Algorithm," in *Proc. 2021 IEEE Int. Conf. Power Electron., Comp. Appl.,* pp. 288–291, 2021.