Ensemble Approach and Enhanced Features for Precise Bank Churn Prediction Analysis

Lare Samuel Adeola
l.adeola@bolton.ac.uk

Celestine Iwendi
c.iwendi@bolton.ac.uk

School of Creative Technologies
University of Bolton
November 2023

Abstract

Numerous studies and research work has been undertaken in the area of creating predictive models for studying Bank Churn. In these studies, the end goal was to create a high accuracy predictive model; while this is commendable, this research focuses on creating an architecture for a predictive model by aggregating the power of various predictive models. The architecture and model proposed in this paper achieved an accuracy of 91% in the test data (35% of the original data set), and an AUC of 96% - confirming the generalized nature of the model. Also, various feature extrapolation techniques were introduced which provide valuable insights to the banking sector.

Background

The significance of understanding why customers of a certain business enterprise leave for another enterprise in the same domain. Various studies have been conducted and various predictive models concocted. Then the question remains: why study bank churning? One major reason is that customer retention leads to increased profits (Verbeke et al. 2012). However, several research have been carried out that analyzes bank churning from the aspect of profitability and building a model to predict churning (Keramati et al. 2016). This research would rather focus on two things:

- understanding the reason for the bank churn, studying the sociological factors that sometimes can span generations or even creating simple associations in the existing data. The purpose is to better understand the factors that lead to customer churn and to use this information in building a model to predict customer churn.
- build an architecture that combines the predictive power of various machine learning models

Research Questions

- Can one build a predictive model with a high accuracy on new data?
- What other features can be extrapolated from the current ones in the data?

Brief Overview of Relevant Literature

Previous works have looked at building predictive models for bank churning. In the work by (Keramati et al. 2016), there was a focus on using data mining methods so as to be able to group customers that would churn and otherwise. Then the Decision Tree Algorithm was used to make the classification. There are two main cornerstones of this research, one is the data set that was used, as unlike other research works, their data set was sourced primarily from a bank and secondly their methodology which was the CRISP-DM method.

The first phase has to deal with understanding the business, the second is understanding and critically looking at the data, the third involves pre-processing the data which would be fed to the model, this could be through data cleaning e.g filling missing values (which was present in their data), modeling involves building the machine learning model, in their case they used the Decision Tree model, evaluation has to deal with how well the model performed and deployment involves taking the model and putting it in action e.g. the banks recommendation engine or sales dashboard etc.Their model had an accuracy of 99.70%. Their sample size was 4,383, which was imbalanced by over 98.6%: churners - 63, non-churners - 4320. Unfortunately, in the research, the table of results didn't provide information on the class labels for both the precision and recall rates so one can't ascertain the model's inter-class performance. This may mean that the model would be very good at only identifying a certain class of customers - non-churners. And, the real insight for the bank would be in identifying the churners. However, another logic could be, if a customer doesn't fall into the non-churner class then that customer may likely be in the churner class. It would be interesting how the model performs when presented with a data sample of the churner class because despite using the k-folds cross validation of 10, this wouldn't matter much as the non-churner class makes up for most of the data, the k's or sample chunks would be dominated by the non-churner class hence the model would still be predicting a majority of customers in the non-churner class.

In the work by (Sahu n.d.), the focus was on building simple machine learning models through a heuristic approach and selecting the one with a high predictive power. It used the same dataset as this research - which is the banbk dataset. It was'nt mentioned if any attempt was made to balance the data but the best algorithm (going by accuracy) was the Random Forest algorithm which had an accuracy of 85%.

For (Baby et al. 2023), they took a different approach. Their approach was to use Artificial Neural Networks(ANN) to build a predictive model for the problem. Their dataset was also sourced from Kaggle. Their model was evaluated on more metrics like the Precision/Specificity, Recall/Sensitivity, F1-score and Accuracy. The architecture of their neural network contained all the relevant independent features, three hidden layers and one output layer with two neurons - customers that would churn and customers that would not churn. Also, they computed separate values for specific metrics like the Precision, Recall and F1-score, taking the time to know the inter-class performance of their model. Their model achieved an accuracy of 86%, Precision in the class label 0 of 88%, Recall in this same class of 97% and F1-score of 92%.

In (Xie & Li 2008), the authors developed a state-of-the-art classification algorithm referred to as LD-Boosting with the ability to handle complex binary classification churn problems with imbalanced datasets. A boosting technique was applied on the algorithm to reduce error and achieve prediction results with more precision. Compared to other algorithms, the method improves prediction accuracy. The authors in (Chang et al. 2022) created a technique for identifying customers who are on the verge of leaving their service with the objective of launching efforts to retain them. This research classifies clients into clusters based on their service providers' contract. The aim of this technique is to anticipate churn and minimize mis-classification. In (Ahanger et al. 2023), the authors employed machine learning to enhance the efficiency of the Intrusion Detection Systems (IDS) but faced the challenge of redundant and irrelevant data in high-dimensional datasets. To resolve this, an advanced ensemble IDS approach was developed using Random Forest to select the initial dataset's optimal subset. Using several performance criteria, the proposed method outperformed other novel approaches with an accuracy of 99%.

The authors of (Devi et al. 2023) set out with the aim of generating a parameter elimination method in Ensemble. Their developed model achieved an 82.74% accuracy higher than novel algorithms. In (Hosseinpour & Shakibian 2023), a novel ensemble learning method was proposed using Logistic Regression and Random Forest algorithms to increase SMS spam detection and its accuracy. The primary challenge with spam detection has been the unbalanced proportion of ham and spam data, as well as the extraction of features from the short messages. The results of the proposed method show its effectiveness in spam detection. (Kaur & Kaur 2020) anticipated customer churn in the banking sector making use of an LSTM model and a SMOTE technique to pre-process the data with the aim of overcoming imbalanced information. The proposed method has an accuracy of 88% for customer churn, performing much better than the method which does not use the SMOTE technique. In (Latheef & Vineetha 2021), a machine learning model is trained using the data of a bank's customers to determine the predictability of churn. The idea is for decision makers and bank managers to be remotely aware of which and how many customers are likely to leave the bank and how to retain them by making the appropriate policies. In (Shukla 2021), a mining customer churn framework was developed using Multi-Layer Perceptron and K-means to predict customer churn in the banking sector. The data was cleaned and pre-processed, and the Silhouette technique was applied to improve the cluster quality. The evaluation results of the proposed method shows that it achieved high accuracy with low training time.

(Khine & Myo 2023) proposed a customer churn model for a private Turkish bank. Considering the challenges in creating a model for the banking industry because of the absence of contractual agreements between the bank and customer for the duration of service, the churn prediction model would be of financial use to the bank. The proposed model was evaluated on several performance criteria like sensitivity, accuracy, and specificity. In (Şen & Bayazıt 2015), the authors applied numerous churn prediction techniques including LOF and K-Means to prevent the loss of customers and the financial strain it imposes on the company. The efficiency and effectiveness of the proposed method was justified in the evaluation results. (Wang et al. 2023) explored the addition of an attention weight to three neural networks of GRU and LSTM after the data had been processed in order to obtain more accurate churn prediction results.

Methodology

Data Overview

The data set used in this research is the popular Bank Churn Data set that can be found on (kaggle.com 2018). This data-set has become the MNIST of Bank Churn predictive modelling. The image below provides a statistical overview of the data-set:

	count	mean	std	min	25%	50%	75%	max
RowNumber	10000.0	5.000500e+03	2886.895680	1.00	2500.75	5.000500e+03	7.500250e+03	10000.00
CustomerId	10000.0	1.569094e+07	71936.186123	15565701.00	15628528.25	1.569074e+07	1.575323e+07	15815690.00
CreditScore	10000.0	6.505288e+02	96.653299	350.00	584.00	6.520000e+02	7.180000e+02	850.00
Age	10000.0	3.892180e+01	10.487806	18.00	32.00	3.700000e+01	4.400000e+01	92.00
Tenure	10000.0	5.012800e+00	2.892174	0.00	3.00	5.000000e+00	7.000000e+00	10.00
Balance	10000.0	7.648589e+04	62397.405202	0.00	0.00	9.719854e+04	1.276442e+05	250898.09
NumOfProducts	10000.0	1.530200e+00	0.581654	1.00	1.00	1.000000e+00	2.000000e+00	4.00
HasCrCard	10000.0	7.055000e-01	0.455840	0.00	0.00	1.000000e+00	1.000000e+00	1.00
IsActiveMember	10000.0	5.151000e-01	0.499797	0.00	0.00	1.000000e+00	1.000000e+00	1.00
EstimatedSalary	10000.0	1.000902e+05	57510.492818	11.58	51002.11	1.001939e+05	1.493882e+05	199992.48
Exited	10000.0	2.037000e-01	0.402769	0.00	0.00	0.000000e+00	0.000000e+00	1.00

Figure 1: Descriptive Statistics of the Bank Churn data set

The data set has a total of 10000 customers, with 11 total variables/features. The 'Exited' variable is the dependent feature while others are the independent features. However, as you can see from the above image, there are some variables that would be irrelevant to our analysis and model building, these are - 'RowNumber','CustomerId' and 'Surname' (which is a PII - Personal Identifiable Information (McCallister 2010)). The data set is imbalanced with the 'not churned' - when 'Exited' is 0 having 7963 entries while the 'churned' - when 'Exited' is 1 had 2037.

Data Spread

Some features with non-numerical values are present in the data set. These features include 'Surname', 'Gender', and 'Geography'. Since 'Surname' is a PII, it was removed from the analysis while the other two were left in the analysis. The image below shows the spread of the values of selected features:

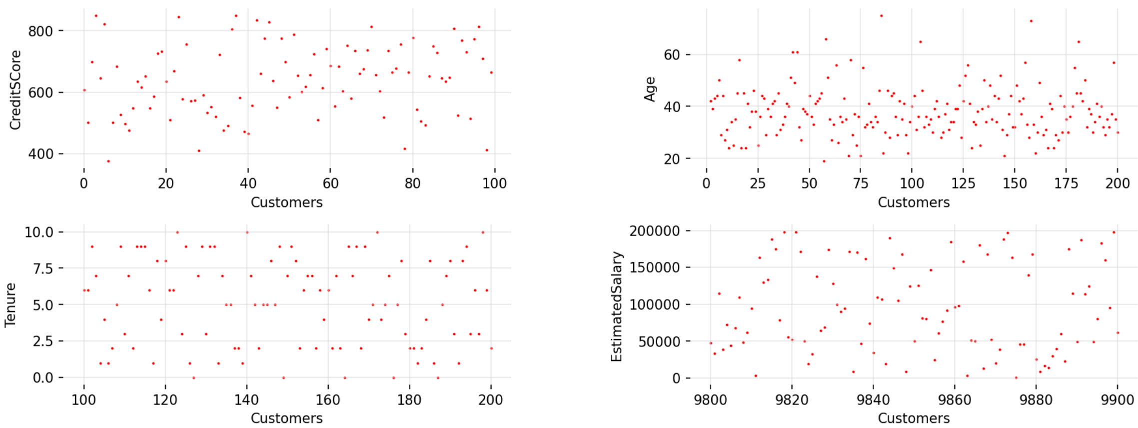

Figure 2: Image showing the non-linear spread of the values of selected features in the data set

As can be seen in the disparate spread of the features, that the model needed to both analyse and predict the customers outcome would have to be non-linear. The figure below shows the correlations among the various features and the dependent feature (Exited):

	RowNumber	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
RowNumber	1.000000	0.004202	0.005840	0.000783	-0.006495	-0.009067	0.007246	0.000599	0.012044	-0.005988	-0.016571
CustomerId	0.004202	1.000000	0.005308	0.009497	-0.014883	-0.012419	0.016972	-0.014025	0.001665	0.015271	-0.006248
CreditScore	0.005840	0.005308	1.000000	-0.003965	0.000842	0.006268	0.012238	-0.005458	0.025651	-0.001384	-0.027094
Age	0.000783	0.009497	-0.003965	1.000000	-0.009997	0.028308	-0.030680	-0.011721	0.085472	-0.007201	0.285323
Tenure	-0.006495	-0.014883	0.000842	-0.009997	1.000000	-0.012254	0.013444	0.022583	-0.028362	0.007784	-0.014001
Balance	-0.009067	-0.012419	0.006268	0.028308	-0.012254	1.000000	-0.304180	-0.014858	-0.010084	0.012797	0.118533
NumOfProducts	0.007246	0.016972	0.012238	-0.030680	0.013444	-0.304180	1.000000	0.003183	0.009612	0.014204	-0.047820
HasCrCard	0.000599	-0.014025	-0.005458	-0.011721	0.022583	-0.014858	0.003183	1.000000	-0.011866	-0.009933	-0.007138
IsActiveMember	0.012044	0.001665	0.025651	0.085472	-0.028362	-0.010084	0.009612	-0.011866	1.000000	-0.011421	-0.156128
EstimatedSalary	-0.005988	0.015271	-0.001384	-0.007201	0.007784	0.012797	0.014204	-0.009933	-0.011421	1.000000	0.012097
Exited	-0.016571	-0.006248	-0.027094	0.285323	-0.014001	0.118533	-0.047820	-0.007138	-0.156128	0.012097	1.000000

Figure 3: Image showing the Correlation Color Map of the Bank Churn Data set

In the previous section, we visually inferred that the relationship amongst the features were non-linear; figure 3 further supports this sentiment as can be seen in the correlations among the features. The most important row is the 'Exited' row which is the last one. The feature with the highest correlation is 'Age' which is 28%. In the next section, we shall expose how this feature was extrapolated for insights on the data.

Feature Extrapolation I - Salary to Bank Ratio(SBR)

The Salary to (Bank) Balance Ratio (SBR) calculates the proportion of the bank customer's salary to their bank balance. The purpose is to find out if there's a relationship between this ratio and the 'Exited' feature and also to find out the probabilities of various thresholds between the ratio and the 'Exited' variable. The formula for this ratio is expressed as:

$$\begin{aligned} SBR &= Salary/Balance \\ where: & \\ SBR &= SalarytoBalanceRatio \\ Balance &\neq 0 \end{aligned}$$

Feature Extrapolation II - Generations

Another feature extrapolated from the data was dividing the 'Age' feature column into 5 generations (sub-class labels). There are as follows:

1. Silent Generation: 1928 - 1945
2. Baby Boomers: 1946 -1964
3. Generation X: 1965 - 1980
4. Millenials: 1981 - 1996
5. Generation Z: 1997 - 2012

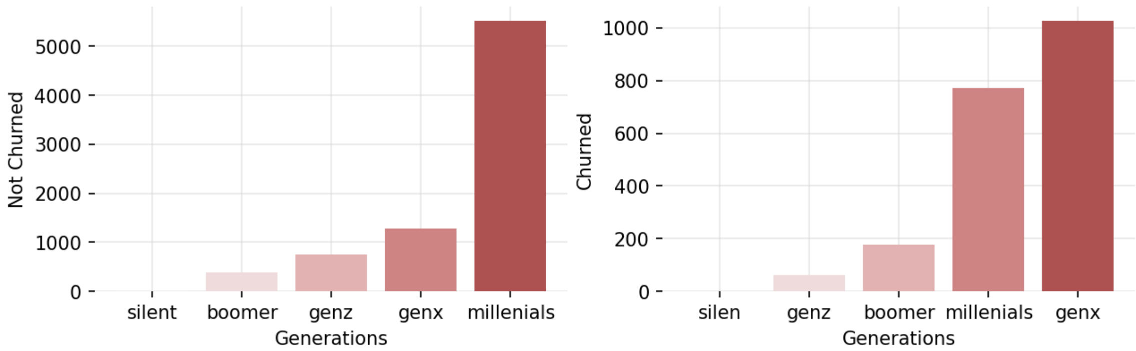

Figure 4: from left to right: image showing the generations and the amount that did not churn, image showing the generations and the amount that churned

Generation Probabilities

The probabilities for this feature was calculated as:

$$gen\_prob = \% of x in total population$$

where:
x = class label
e.g. for the silent generation
Total Class Label Population (tclp) = 24
class label (cl) = not_churned
Population of not_churned = 23
probability silent generation not_churned =

$$population of not\_churned / total class label population$$

This is how the probabilities for each of the sub-class labels in the generations were calculated for the 'not churn' class. In order to compute the probabilities of the 'churn' class:

$$1 - probability of scl$$

e.g. in the Silent Generation example shown above it would be:

$$1 - (23/24)$$

Feature Extrapolation III - Credit Worthiness

Another feature extrapolated from the data was dividing the 'CreditScore' feature column into 3 sub-segments (sub-class labels). There are as follows:

1. low_credit: CreditScore <= 720
2. fair_credit: 800 < CreditScore <= 721
3. good_credit: CreditScore > 800

Figure 5: from left to right: image showing the credit worthiness and the amount that did not churn; image showing the credit worthiness and the amount that churned

Ensemble Architecture

Stacked generalization can be defined as a method whereby various estimators are combined to reduce their biases Hastie et al. (2009). The reason for using this ensemble approach was to be able to capture the idiosyncrasies of the various predictive models. Using a Stacked Ensemble model, the figure below shows the architecture for creating the estimator:

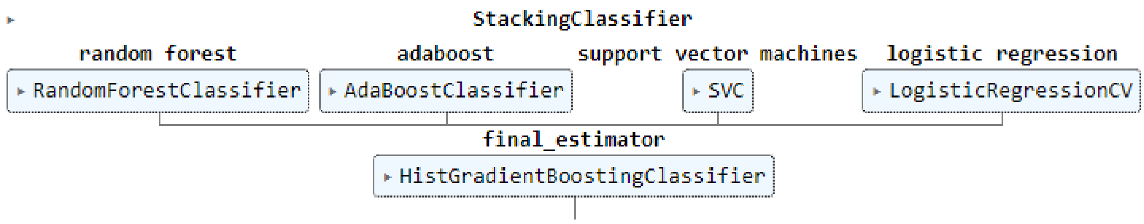

Figure 6: The Stacked Ensemble Architecture

In fig. 6, the Stacked Ensemble Classifier is made up of the random forest, adaboost, support vector machines and logistic regression in the first layer; then the second and final layer consists of the HistGradientBoostingClassifier which is a faster algorithm than the GradientBoostingClassifier of the scikit learn library for data sets that are less than or equal to 10,000 Pedregosa et al. (2011)

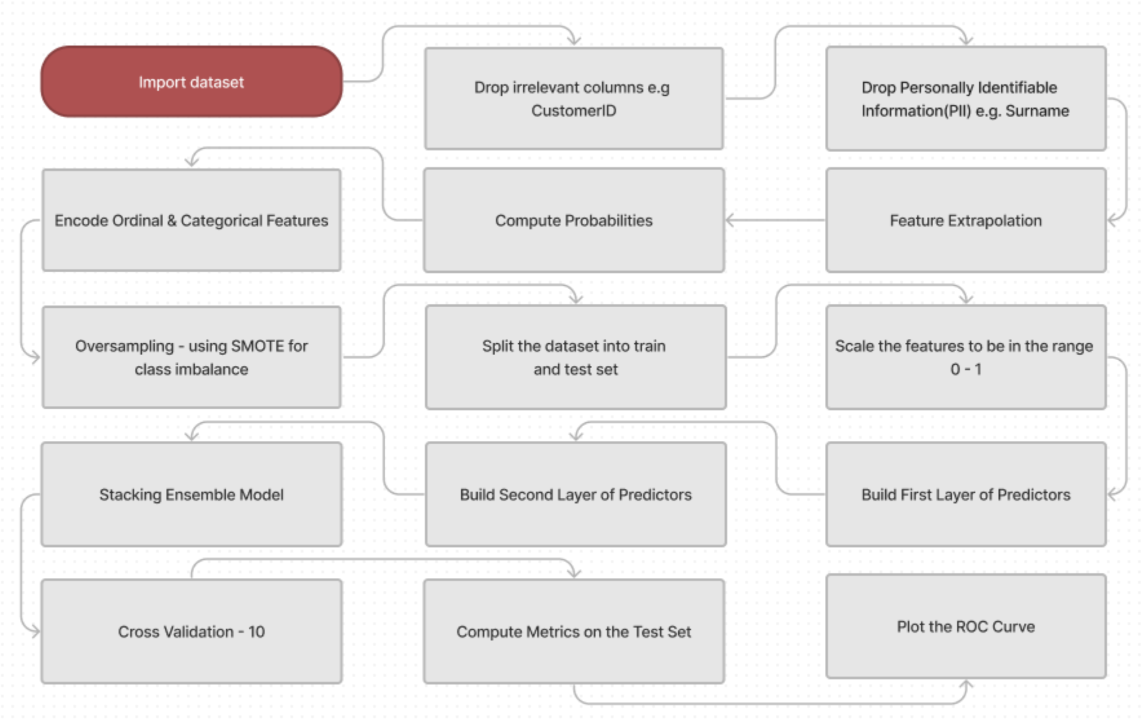

Figure 7: The Stacked Ensemble Flow Chart showing the various stages involved in building the Stacked Ensemble Model

Analysis

Results with the Ensemble Architecture

The Stacked Ensemble Architecture achieved an accuracy of 91% on the test set. It had a Recall rate of 96% on class label 0 i.e 'customers that did not churn' and a Precision rate of 87% in this same class. In the other class label 1 i.e. 'customers that did churn', it had a Recall rate of 87% and a Precision rate of 96%. Overall, the SEC Model outperformed other models in both classes as can be seen in figures 8 and 9. The Gradient Boosting model achieved similar performance and this was the reason in the first place to use it as the final layer in the SEC architecture model.

	Precision	Recall	f1-score	Accuracy
SVM	0.820000	0.920000	0.870000	0.860000
Logistic Regression	0.700000	0.740000	0.720000	0.720000
Random Forest	0.870000	0.950000	0.910000	0.910000
Decision Trees	0.860000	0.860000	0.860000	0.860000
Naive Bayes	0.780000	0.630000	0.700000	0.740000
Gradient Boosting	0.880000	0.950000	0.910000	0.910000
SEC	0.870000	0.960000	0.910000	0.910000

Figure 8: Table showing the various metrics (for the class label 0 or customer would not churn) of the different Machine Learning models used in the experiments. The intersections with red background are the highest values in the column.

	Precision	Recall	f1-score	Accuracy
SVM	0.920000	0.810000	0.860000	0.860000
Logistic Regression	0.740000	0.700000	0.720000	0.720000
Random Forest	0.950000	0.870000	0.910000	0.910000
Decision Trees	0.870000	0.870000	0.870000	0.860000
Naive Bayes	0.710000	0.830000	0.760000	0.740000
Gradient Boosting	0.950000	0.880000	0.910000	0.910000
SEC	0.960000	0.870000	0.910000	0.910000

Figure 9: Table showing the various metrics (for the class label 1 or customer would churn) of the different Machine Learning models used in the experiments. The intersections with red background are the highest values in the column.

The AUC Curve

In the AUC curve, we can see the average performance of the various models across different classification thresholds. Hence, it is a good measure to know how generalized the model is. Figure 10 shows that the SEC model still outperforms other models.

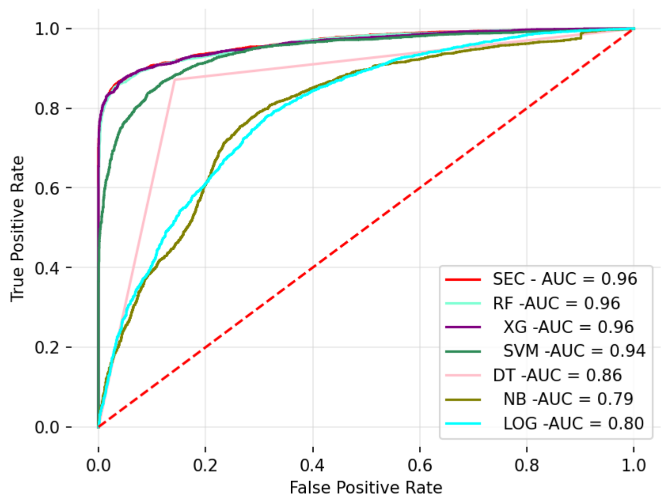

Figure 10: The ROC Curve for the different machine learning models and the SEC model

Probability Analysis of Feature Extrapolations

Salary to Bank Balance Ratio (SBR) - Customer that would not churn

- if SBR is greater than 0.59 or 0 then the customer has a 70% chance of not churning
- if SBR is greater than 0.59 i.e its average then there is a 31% chance that the customer does not churn else the customer churns.
- if SBR is greater than 1.025 i.e. high liquidity, then the customer has an 18% chance of not churning
- if SBR is equal to zero i.e. no liquid or account not in use, then the customer has a 31% chance of not churning

Salary to Bank Balance Ratio (SBR) - Customer that would churn

- if SBR is greater than 0.59 i.e its average, then there is a 9% chance that the customer would churn

- if SBR is greater than 1.025 i.e. high liquidity, then the customer has a 6% chance of churning
- if SBR is equal to zero i.e. no liquid or account not in use, then the customer has a 5% chance of churning

Generation - Customer that would not churn

- if the age bracket falls within the silent generation, there's a 96% chance that they would not churn
- if the age bracket falls within the boomer generation, there's a 69% chance that they would not churn
- if the age bracket falls within the generation x, there's a 55% chance that they would not churn
- if the age bracket falls within the millenial, there's a 87% chance that they would not churn
- if the age bracket falls within genz, there's a 93% chance that they would not churn

Generation - Customer that would churn

- if the age bracket falls within the silent generation, there's a 4% chance that they would churn
- if the age bracket falls within the boomer generation, there's a 31% chance that they would churn
- if the age bracket falls within the generation x, there's a 45% chance that they would churn
- if the age bracket falls within the millenial, there's a 12% chance that they would churn
- if the age bracket falls within genz, there's a 7% chance that they would churn

Credit Worthiness - Customer that would not churn

- if the credit worthiness is low, there's a 79.5% chance that they would not churn
- if the credit worthiness is fair, there's a 79.5% chance that they would not churn
- if the credit worthiness is good, there's a 80% chance that they would not churn

Credit Worthiness - Customer that would churn

- if the credit worthiness is low, there's a 20.5% chance that they would churn
- if the credit worthiness is fair, there's a 20.5% chance that they would churn
- if the credit worthiness is good, there's a 20% chance that they would churn

Table of Comparison with Related Research

	Dataset Size	Test Data	class 0 Precision	class 1 Precision	class 0 Recall	class 1 Recall	F1-score	K-folds	Accuracy	AUC
keramati et al	4383	1314.900000	91.810000	91.810000	91.000000	0.000000	90.960000	10.000000	99.700000	92.900000
pinaki	10000	nan	nan	nan	nan	nan	nan	nan	85.370000	85.800000
praveen et al	0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	81.710000	84.000000
baby et al	10000	2000.000000	88.000000	79.000000	97.000000	48.000000	75.500000	0.000000	86.000000	0.000000
singh et al	10000	0.000000	90.300000	90.300000	60.100000	60.100000	61.300000	0.000000	83.900000	84.700000
kumara et al	10000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	91.950000	89.000000
SEC	10000	3500.000000	87.000000	96.000000	96.000000	87.000000	91.000000	10.000000	91.000000	96.000000

Figure 11: Table of Comparison of various selected past research ((Keramati et al. 2016), (Sahu n.d.), (Lalwani et al. 2022), (Baby et al. 2023), (Singh et al. 2023)) done in this area. The table shows the different metrics and their values gotten in the research. The intersection with the red lines shows the maximum value in the columns.

In the table above, 10 indicators were used as a comparison between SEC and other related research work in this same domain and in lots of instances using the same dataset freely available in Kaggle. As can be seen in the table, the SEC Architecture outperforms other research works in various indicators (the indicators with a red background are the maximum values in the columns) including dataset size, test size (a good indicator for the model's performance in unseen data or how generalized the model is), class 1 precision, class 1 recall, f1-score, k-folds, and AUC. For accuracy, SEC was beaten by the research work done by (Keramati et al. 2016) but with a the work done by karemani as elaborated in the literature review has a small sample size.

Discussion

The SEC Architecture has achieved an accuracy of 91% on the test data(35% of the original data), had a high AUC of 96%, performed well in both classes - customers that did not churn(class label 0) and customers that did churn (class label 1) with Precision/Specificity rates of 87% and 96% while Recall/Sensitivity rates of 96% and 87% respectively.

Three other features were extrapolated from the current relevant features. These features are: Salary to Bank Balance Ratio(SBR), Generations, and Credit Worthiness. These features further provided more insights about the behaviours of the bank customers in this data set. At this point of the research, the probabilities computed have not been tested on another data set to confirm their veracity but with this data set, the probabilities and their respective outcomes hold.

Further Recommendations

Applying the SEC model to a new test data, preferably similar to this data set but different in size or industry to test its capability and generalized nature in the wild. Also adapting the SEC model to a new industry may be another consideration. Lastly, building the SEC Architecture with a totally new data set that is balanced without the need for using either oversampling or under-sampling algorithms and techniques is encouraged as this may further cement the architecture as one proven to predict customer churn in various scenarios.

References

Ahanger, A. S., Khan, S. M. & Masoodi, F. S. (2023), Intrusion detection system for iot environment using ensemble approaches, *in* '2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)', IEEE, pp. 935–938.

Baby, B., Dawod, Z., Sharif, S. & Elmedany, W. (2023), Customer churn prediction model using artificial neural networks (ann): A case study in banking, *in* '3ICT 2023: International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies', IEEE.

Chang, V., Gao, X., Hall, K. & Uchenna, E. (2022), Machine learning techniques for predicting customer churn in a credit card company, *in* '2022 International Conference on Industrial IoT, Big Data and Supply Chain (IIoTBDSC)', IEEE, pp. 199–207.

Devi, T. et al. (2023), Accurate parameter elimination approach in ensemble learning on nsl-kdd dataset in comparison with candidate elimination algorithm, *in* '2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)', IEEE, pp. 1–4.

Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. (2009), *The elements of statistical learning: data mining, inference, and prediction*, Vol. 2, Springer.

Hosseinpour, S. & Shakibian, H. (2023), An ensemble learning approach for sms spam detection, *in* '2023 9th International Conference on Web Research (ICWR)', IEEE, pp. 125–128.

kaggle.com (2018), 'Bank customers churn'.
URL: *https://www.kaggle.com/datasets/santoshd3/bank-customers*

Kaur, I. & Kaur, J. (2020), Customer churn analysis and prediction in banking industry using machine learning, *in* '2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)', IEEE, pp. 434–437.

Keramati, A., Ghaneei, H. & Mirmohammadi, S. M. (2016), 'Developing a prediction model for customer churn from electronic banking services using data mining', *Financial Innovation* **2**, 1–13.

Khine, S. T. & Myo, W. W. (2023), Mining customer churns for banking industry using k-means and multi-layer perceptron, *in* '2023 IEEE Conference on Computer Applications (ICCA)', IEEE, pp. 220–225.

Lalwani, P., Mishra, M. K., Chadha, J. S. & Sethi, P. (2022), 'Customer churn prediction system: a machine learning approach', *Computing* pp. 1–24.

Latheef, J. & Vineetha, S. (2021), Lstm model to predict customer churn in banking sector with smote data preprocessing, *in* '2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)', IEEE, pp. 86–90.

McCallister, E. (2010), *Guide to protecting the confidentiality of personally identifiable information*, Vol. 800, Diane Publishing.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011), 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research* **12**, 2825–2830.

Sahu, P. (n.d.), 'Unlocking the code of customer churn: Predictive strategies for banking success'.

Şen, K. & Bayazıt, N. G. (2015), Customer churn modelling in banking, *in* '2015 23nd Signal Processing and Communications Applications Conference (SIU)', IEEE, pp. 2384–2387.

Shukla, A. (2021), Application of machine learning and statistics in banking customer churn prediction, *in* '2021 8th International Conference on Smart Computing and Communications (ICSCC)', IEEE, pp. 37–41.

Singh, P. P., Anik, F. I., Senapati, R., Sinha, A., Sakib, N. & Hossain, E. (2023), 'Investigating customer churn in banking: A machine learning approach and visualization app for data science and management', *Data Science and Management* .

Verbeke, W., Dejaeger, K., Martens, D., Hur, J. & Baesens, B. (2012), 'New insights into churn prediction in the telecommunication sector: A profit driven data mining approach', *European Journal of Operational Research* **218**(1), 211–229.
URL: *https://www.sciencedirect.com/science/article/pii/S0377221711008599*

Wang, Y., Zheng, S., Liu, G. & Li, J. (2023), Research on bank customer churn model based on attention network, *in* '2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)', IEEE, pp. 346–350.

Xie, Y. & Li, X. (2008), Churn prediction with linear discriminant boosting algorithm, *in* '2008 International conference on machine learning and cybernetics', Vol. 1, IEEE, pp. 228–233.