Customer Churn Prediction on E-Commerce Data using Stacking Classifier

SUBMISSION DATE / POSTED DATE

12-07-2022 / 14-07-2022

CITATION

Awasthi, Shashwat (2022): Customer Churn Prediction on E-Commerce Data using Stacking Classifier. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.20291694.v1

DOI

10.36227/techrxiv.20291694.v1

Customer Churn Prediction on E-Commerce Data using Stacking Classifier

-by Shashwat Awasthi

Abstract

Retaining the customers for business is one of the most import insight of any business which run on customers. Nowadays, budding organizations look to have a firm hold in the market by building upon their existing customer base and fortify them. Thus, Customer Churn is a major problem for them. Companies aim at reducing this customer churn and to reduce customer churn, it's important to analyse the factors which increase customer churn and then build on various strategies in order to reduce it.

This article on Customer churn using Stacking Classifier aims at analysing and anticipating the churn in ecommerce data using various features such as Gender, Tenure etc. For this purpose, we established a Stacking Classifier, which is based on Ensemble learning technique, using meta-classifier which learned from various base learners. The model used four different base learners- KNN, SVM, RF Classifiers and Decision Trees. The results showed that the stacking-based classifier outperformed various individual ML approaches. In terms of performance metrics, the system achieves the highest accuracy of 98.2%, with Area under ROC curve of 98.1% and F-Measure score of 95.0%.

Introduction

'Churn' means whenever a customer leaves a particular business services and opt for others or either if he stops using the products of that particular business. Customer relationship management is fundamentally an investment decision for the long run. It is the business as a whole, as well as the customer base, that must be handled in order to get the most out of the client portfolio. Customer's churn is a really important metric for business professionals who use aggressive business policies in their business. However, if done early enough, anticipating consumers who are likely to quit the company could represent a sizable additional revenue source. Customer retention in the markets today is as important as growing your business, as retained customers not only contribute to easier business flow but also to the increase in spread of word which help to boost the business much more. According to data, increasing client retention by just 5% can result in a profit increase of at least 25%[2]. Thus, Customer churn prediction is necessary as it helps forecast and anticipate users which might be vulnerable to leave.

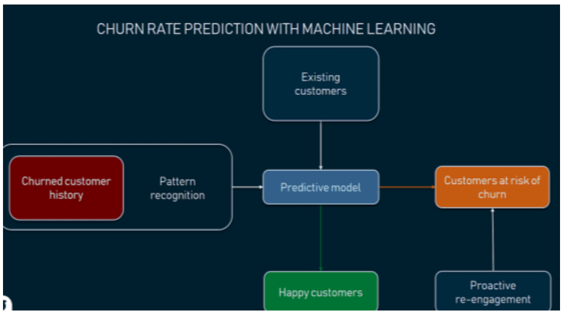

Fig 1.1 - Churn Prediction Technique[3]

The chance of a client churn is predicted by a churn forecast. It reduces the cost of gaining new customers while also assisting in the retention of existing customers. The amount of time and money spent on marketing and attracting new customers is much Keeping existing consumers is more difficult and expensive. Customers who are reluctant to buy or who are ready to transfer shopping sites owing to cash flow issues can be persuaded and captured. Customers are free to leave if they are ending their relationship owing to serious and unavoidable conditions. Despite the fact that the end outcome is excellent, we are investing in required churners.[4].

The interesting technology of data mining and analysis provides various tools and algorithms to dig deep into data and extract various summaries and predict the future course of events based on that. This method is vividly used to identify patterns of data and extract useful conclusions. These can include predicting a price decrease or fall, assessing the mindset of people and various other insights. To perform churn analysis, the data must first be collected and pre-processed in a way that allows it to be input into a Machine Learning Model. Then, using a variety of strategies, we select the best techniques to develop a model that provides the greatest prediction results. Eventually, the model which has the best accuracy is selected as the models for the Customer Churn Analysis.

This research offers a Stacking technique-based e-commerce customer churn prediction model. It is based on the premise that the accuracy of a group of classifiers, such as decision trees, KNNs, and SVMs, is usually higher than the accuracy of any individual classifier. Various researches in the field of Machine Learning have shown that there is no such one classification algorithm which gives accurate measures known as No Free Lunch theorem[5]. The method suggested in this study is based on a stacking classifier that uses several base learners whose findings are collected in a meta learner to build upon a final result.

In this paper, I have discussed abut some past researches on the same area in Section III, whereas Section IV explains the methodology and various base learners and meta learners used in this paper. Section V shows the experimental work done in writing this paper, and Section VI gives the conclusions and prospects about the future work on the area

Related Work

Customer Churn has been an area of excitement for many years. Shini Renjith, in 2015, proposed customer churn models by using Logistic Regression[6]. X Yu et al. suggested a model that employs an extended "Support Vector Machine" model and incorporates measures that indicate the influence of churners, non-churners, and nonlinearity[7]. The Combined classifier has also been adopted as a frequent measure to predict customer churn. Abinash Mishra and U. Srinivasulu Reddy conducted a comparison analysis on ensemble learning approaches, finding that the Random Forest algorithm outperformed others such as Bagging, Boosting, and Decision Trees algorithms[8]. A report published in 2020 titled Prediction of Customer Churn in e-Retailing found that ensemble strategies are the most effective in identifying churners.[9].

Sonali Agarwal, Divya Tomar and Pretam Jayswal used the bagging and boosting ensemble learning methods to predict customer churn with classifiers like Decision Trees, Random forests, Gradient Boosted trees and showed that gradient boosted trees gave the best result in customer churn analysis[10]. Also to refine their results, they used optimizations. The Combined classifier aggregates the results of two or more sub classifiers to provide with an accurate result[11]. Boosting algorithm created by Scpahire in 1990 is one example of a combined classifier[12], Breiman proposed the Bagging algorithm in 1996[13] and the Stacking algorithm which aggregates results of base learners to a meta learner and gives the result[14]. Xiaojun Wu used Boosting Algorithm along with SMOTE to predict customer Churn[15].

A three-category churn prediction model based on a retention-oriented strategy was proposed by Essam Shaaban, Yehia Helmy, and Ayman Khedr. They used Neural Networks, Support Vector Machines, and Decision Trees to demonstrate that SVM and Neural Networks outperform decision trees.[16]. Xin Hu, Lanhua Chen, Yanfei Yang, and Siru Zu developed a customer churn prediction model based on Decision Trees and Neural Networks last year, and their model had a 98.87 percent accuracy[17].

In 2019, Irfan Ullah conducted a research on churn prediction using random forest[18]. This study used customer profiling using K-Means Clustering along with RF Classifiers.

Dataset

A big online E-Commerce corporation owns the data set. There are both classified and non-categorical variables in the dataset. An online retail (E commerce) organisation needs to know which consumers are likely to churn so that they can contact them with special offers. The data was downloaded from Kaggle.[19].

Table I- Features of the dataset

Column Name	Data Type
CustomerID	Object
Churn	Object
Tenure	Float64
Preferred Login Device	Object
CityTier	Object
WarehouseToHome	Float64
PreferredPaymentMode	Object
Gender	Object
HourSpendOnApp	Float64
NumberOfDeviceRegistered	Int64
PreferredOrderCat	Object
SatisfactionScore	Object
MaritalStatus	Object
NumberOfAddress	Int64
Complain	Object
OrderAmountHikeFromLastYear	Float64
CouponUsed	Float64
OrderCount	Float64
DaysSinceLastOrder	Float64
CashbackAmount	Float64

Proposed Architecture

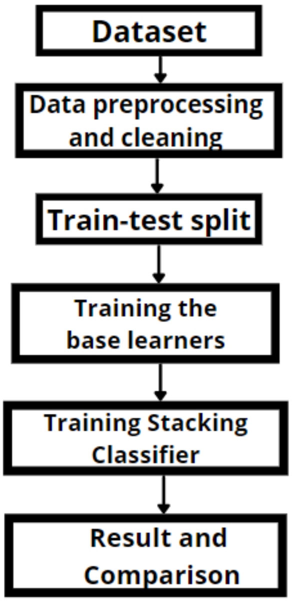

Fig- 3.1 Proposed Method Workflow

For the customer churn analysis, the dataset we used was firstly taken through the exploratory data analysis to handle the null values and find different categorical values and numerical values and see which of the columns contribute more to the customer churn in recent years. This data after exploratory data analysis was divided into train and test datasets with 90% train dataset and 10% test dataset. After this, the dataset was trained on 4 base learners namely Decision Trees, Random forests, The outputs of the Support Vector Machine and KNN classifiers were fed into the Stacking Classifier's meta classifier using logistic regression.

Stacking Ensemble Technique

As previously said, staking is a technique that takes into account a variety of heterogeneous base learning algorithms, trains their models in parallel on a comparable dataset, and then integrates the outcomes of these weak base learners into a meta model, which then predicts the results. In contrast to bagging, the models in stacking are usually unique (not all choice trees, for example) and fit on the same dataset (for example rather than tests (from the training data) Moreover, unlike boosting, stacking employs a single model to figure out how to integrate the predictions from the contributing models in the most effective way (for example rather than an arrangement of models that right the predictions of earlier models).

A stacking model is made up of at least two base models (sometimes referred to as level-0 models) and a meta-model (often referred to as a level-1) that connects the base models' expectations.

A stacking meta-model is used to generate a result depending on the output of certain base learners/classifiers. That is, information not used to generate the base models is input into the base models, predictions are made which are combined and aggregated along with the normal yields, offer information and yield sets for fitting the meta-model to the training dataset..

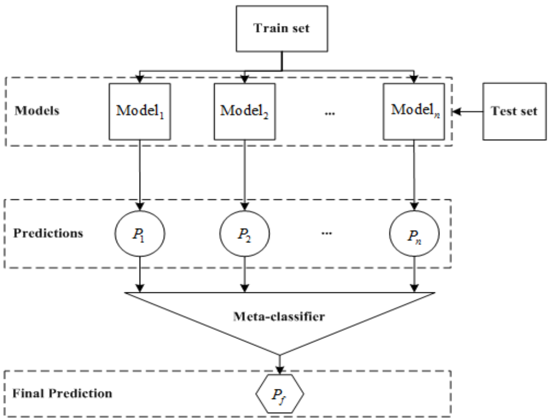

Fig 3.3- Stacking Ensemble Learning Classifier[21]

Due to relapse, the yields from the base models used as contributions to the meta-model may be real worth, likelihood esteems, likelihood like qualities, or class names, and due to arrangement, they could be likelihood esteems, likelihood like qualities, or class names.

Contributions to the base models, such as training information input components, may also be incorporated in the meta-training model's data. This can give the meta-model a new setting in terms of how to best consolidate the meta-predictions. model's The meta-model can be built in isolation on the training dataset once it is ready, whereas the base-models are built on the complete unique training dataset.

Performance Evaluation

We developed a novel Stacking Classifier using four base learners: Decision Trees, RF (Random Forest) Classifiers, KNN (K-Nearest Neighbours), and SVM (Support Vector Machines), as well as a meta learner: Logistic Regression.. Three model evaluation statistics, Accuracy, F1 score and Recall score have been used. The results of the Stacking Classifier are tabulated below-

Table II – Results of Stacking Classifier

Number/Metrics	Accuracy	F1 Score	Recall Score
1	96.64%	89.15%	88.37%
2	97.43%	91.66%	88.39%
3	98.02%	94.04%	94.18%
4	96.64%	88.50%	89.53%
5	97.43%	92,48%	93.02%
6	96.84%	91.86%	92.05%
7	97.83%	94.25%	91.86%
8	98.22%	94.04%	93.00%
9	96.83%	90.24%	84.83%
10	96.83%	89.44%	90.12%
Aggregate Results	97% ± 1%	93% ± 2%	92% ± 2%

The table below gives the results of F-Value of various Base Learners used to make the Stacking Classifier-

Table III- Results of various base learner models

Base Learner/Metrics	F1 Score	Recall Score	Accuracy Score	Precision Score	Area Under ROC curve
KNN	92.85%	95.45%	97.04%	92.30%	96.99%
SVM	61.22%	51.15%	89.87%	76.27%	74.08%
Decision Tree	92.98%	95.72%	97.02%	90.52%	97.91%
RF Classifier	56.91%	39.72%	90.58%	98.91%	69.88%

We observed through the above metrics that KNN and Decision Trees are very efficient algorithms in predicting the customer churn. But both these two algorithms slightly lag in the F1 score metrics and the precision score metrics.

The ROC curve for all the algorithms and the Stacking Classifier shows that Stacking Classifier is the most accurate when compared to the individual ML models.

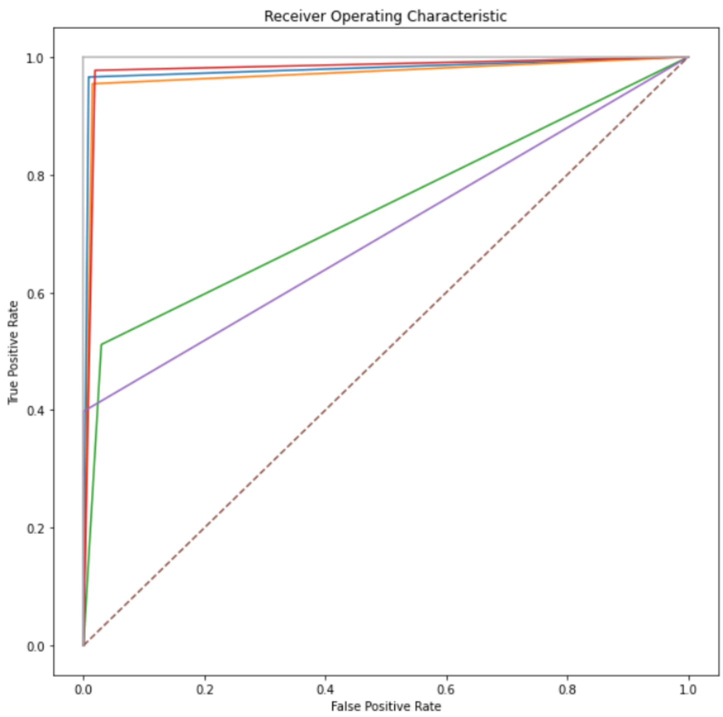

Fig-4.1 ROC curve for the Base Learners and Stacking Classifier

Comparison of Stacking Classifier with XG boost

From [25], we took the results of putting the XG boost classifier on the dataset. It was established that out Stacking classifier performed better than XG boost classifier. Below are the results tabulated-

Table V- Comparison of Result between Stacking and XG Boost Classifier

Model/Metrics	Precision	F1-Score	Accuracy	Recall Score
Stacking Classifier	**91%**	**93%**	**98%**	**92%**
XG Boost	**91%**	**90%**	**91%**	**90%**

Data Pre-processing

The E-Commerce dataset used for this particular customer churn classification consisted of 5630 samples with 20 attributes. The presence of null values in the dataset would have made it difficult for us to process the data, so they were dealt by adding median values of the data frame of other columns to avoid null values in the dataset. Also, various categorical variables which were set as numerical variables were converted to process the data. On further processing the data and plotting the pair plot, it was observed that churn was affected the most with quantities like Days Since Last Order, Tenure etc. The insights were-

- People who churn have lower tenure.
- People who churn use a smaller number of coupons
- People who have lower cashback churn more
- Hours spend on app is inversely proportional to churn

4.3.1) Outlier Treatment

Following this, we needed to remove the outliers which would have brought the accuracy of the model down. To remove outliers, I defined the lower range and upper range which is going to be 1.5 times on both sides/ends of the inter-quartile range from respective whiskers.

The algorithm used for Outlier Treatment is **IQR (Inter-Quartile Range Method)-**

- Step 1- Find the first Quartile Q1
- Step 2- Q3 is the third Quartile, find it.
- Step 3- Calculate the Inter Quartile Range as the difference between values calculated in first two steps.
- Step 4- Here, Measure the data range using Q1–1.5* (Inter Quartile Range) as the lower limit and Q3+1.5* as the upper limit (Inter Quartile Range)
- Step 5- Any data point which are beyond these data limits shall be removed as they are not required in the analysis.

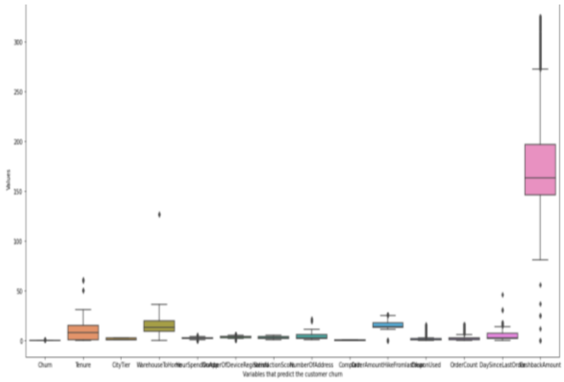

Fig 4.2- Dataset before outliers were removed

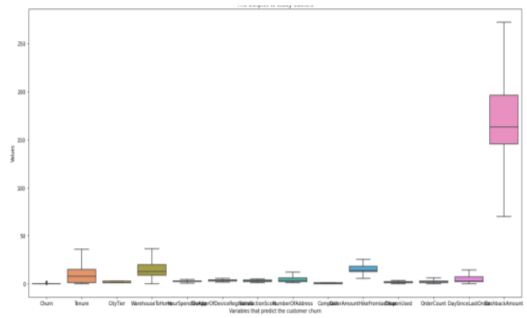

Fig 4.3- Dataset after outliers were removed

Conclusion

The model in the above empirical analysis is evaluated by the using ensemble learning Stacking Technique. The results showed that Staking Technique is better compared to Individual Models and gives better performance metrics. A 10-fold cross validation was conducted on the model and in all the cases, the Stacking Classifier had a better performance metric compared to other Individual ML models. Thus, compared to the other methods of customer churn prediction, this model can predict user customer churn more accurately. Thus, the construction of this model using ensemble learning technique proves very beneficial for ecommerce retailers in predicting the customer churn. This helps them improving their strategy, and retain upon their customers to have a firm hold on the market and to increase their business prospects. Some of the strategies which they can opt for to retain their customers based on the above data can be-

- Provide more cashback offers to the customers which they will cherish and be a loyal customer.
- Improve upon the Application UI of the store, which will help attract more customers.
- Provide the customers with various discount coupons throughout the year.
- Do small advertisement campaigns to motivate the customers to visit the store more frequently and this will help them reduce more churn.

With these measures based on the above data mining, can help improve the customer churn rate in the store of the company, thereby increasing their prospects.

References

[1] "ShashwatAwasthi04/Customer-Churn-Predicition-with-Stacking-Classifiier: This is a Customer Churn Prediction project in E-Commerce Retail using Stacking based Ensemble Learning Method using Logistic Regression as Meta Learner and KNN, SVM, Decision Trees and Random forests as base learners." https://github.com/ShashwatAwasthi04/Customer-Churn-Predicition-with-Stacking-Classifiier (accessed Nov. 16, 2021).

[2] J. A. Rozum, "Defining and Understanding Software Measurement Data 1 Why a Process for Measurement".

[3] "Customer Churn Prediction Using Machine Learning: Main Approaches and Models - KDnuggets." https://www.kdnuggets.com/2019/05/c

hurn-prediction-machine-learning.html (accessed Nov. 03, 2021).

[4] N. Forhad, M. S. Hussain, and R. M. Rahman, "Churn analysis: Predicting churners," *2014 9th International Conference on Digital Information Management, ICDIM 2014*, pp. 237–241, Dec. 2014, doi: 10.1109/ICDIM.2014.6991433.

[5] O. Mazhelis and S. Puuronen, "COMBINING ONE-CLASS CLASSIFIERS FOR MOBILE-USER SUBSTITUTION DETECTION," *ICEIS 2004 - Proceedings of the Sixth International Conference on Enterprise Information Systems*, vol. 5, pp. 130–137, Apr. 2004, doi: 10.5220/0002639901300137.

[6] S. Renjith, "An Integrated Framework to Recommend Personalized Retention Actions to Control B2C E-Commerce Customer Churn," *International Journal of Engineering Trends and Technology*, vol. 27, no. 3, pp. 152–157, Nov. 2015, doi: 10.14445/22315381/ijett-v27p227.

[7] X. Yu, S. Guo, J. Guo, and X. Huang, "An extended support vector machine forecasting framework for customer churn in e-commerce," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1425–1430, Mar. 2011, doi: 10.1016/J.ESWA.2010.07.049.

[8] A. Mishra and U. S. Reddy, "A comparative study of customer churn prediction in telecom industry using ensemble based classifiers," *Proceedings of the International Conference on Inventive Computing and Informatics, ICICI 2017*, pp. 721–725, May 2018, doi: 10.1109/ICICI.2017.8365230.

[9] M. Jaeyalakshmi, S. Gnanavel, K. S. Guhapriya, H. Phriyaa, and K. Sree, "Prediction of Customer Churn on e-Retailing," *International Journal of Recent Technology and Engineering*, no. 6, pp. 2277–3878, 2020, doi: 10.35940/ijrte.F9550.038620.

[10] P. Jayaswal, B. R. Prasad, and S. Agarwal, "An Ensemble Approach for Efficient Churn Prediction in Telecom Industry," *International Journal of Database Theory and Application*, vol. 9, no. 8, pp. 211–232, 2016, doi: 10.14257/ijdta.2016.9.8.21.

[11] J. Xiao, C. He, X. Jiang, and D. Liu, "A dynamic classifier ensemble selection approach for noise data," *Information Sciences*, vol. 180, no. 18, pp. 3402–3421, Sep. 2010, doi: 10.1016/J.INS.2010.05.021.

[12] R. E. Schapire, "The Strength of Weak Learnability," vol. 5, pp. 197–227, 1990.

[13] L. Breiman, "Bagging predictors," *Machine Learning 1996 24:2*, vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1007/BF00058655.

[14] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.

[15] X. Wu and S. Meng, "E-commerce customer churn prediction based on improved SMOTE and AdaBoost," Aug. 2016. doi: 10.1109/ICSSSM.2016.7538581.

[16] "(PDF) A Proposed Churn Prediction Model | Mona Nasr - Academia.edu." https://www.academia.edu/26824075/A_Proposed_Churn_Prediction_Model (accessed Nov. 03, 2021).

[17] X. Hu, Y. Yang, L. Chen, and S. Zhu, "Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network," *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2020*, pp. 129–132, Apr. 2020, doi: 10.1109/ICCCBDA49378.2020.9095611.

[18] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/ACCESS.2019.2914999.

[19] "Ecommerce Customer Churn Analysis and Prediction | Kaggle." https://www.kaggle.com/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction (accessed Oct. 13, 2021).

[20] T. G. Dietterich, "Machine Learning Research: Four Current Directions".

[21] "The architecture of the stacking ensemble learning. In the... | Download Scientific Diagram." https://www.researchgate.net/figure/The-architecture-of-the-stacking-ensemble-learning-In-the-base-classifiers-the-training_fig1_335156833 (accessed Nov. 04, 2021).

[22] F. Guo and H. L. Qin, "The analysis of customer churns in e-commerce based on decision tree," *2015 International Conference on Computer Science and Applications, CSA 2015*, pp. 199–203, Jan. 2017, doi: 10.1109/CSA.2015.74.

[23] V. Naumovich. Vapnik, "Statistical learning theory," p. 736, 1998.

[24] M. A. Hassonah, A. Rodan, A. K. Al-Tamimi, and J. Alsakran, "Churn Prediction: A Comparative Study Using KNN and Decision Trees," *ITT 2019 - Information Technology Trends: Emerging Technologies Blockchain and IoT*, pp. 182–186, Nov. 2019, doi: 10.1109/ITT48889.2019.9075077.

[25] "E-commerceCustomerChurn | Kaggle." https://www.kaggle.com/ankitverma2010/e-commercecustomerchurn (accessed Nov. 16, 2021).

[26] F. Provost, F. Provost, and T. Fawcett, "Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions," *IN PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, pp. 43--48, 1997, Accessed: Oct. 13, 2021. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.9927

[27] X. Fan and K. Tang, "Enhanced maximum AUC linear classifier," *Proceedings - 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2010*, vol. 4, pp. 1540–1544, 2010, doi: 10.1109/FSKD.2010.5569339.