IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

Credit Card Fraud Detection Using AI (Python)

Arshee Naz[1], Praveen Kumar[2], Dr. Ashad Ullah Qureshi[3]
Junior Research Fellow,National Institute of Technology, Kurukshetra, Haryana[1]
Website Administrator, Shri Vishwakarma Skill University, Dudhola[2]
Technical Officer, Indian Institute of Information Technology, Sonepat, Haryana[3]
arshee_jrf@nitkkr.ac.in[1], praveen.kumar@svsu.ac.in[2], aqureshi@iiitsonepat.ac.in[3]

Abstract: *Credit card fraud poses a formidable obstacle in the financial sector, resulting in considerable monetary damages for both individuals and institutions. The necessity for efficient fraud detection systems has become paramount due to the rising number of online transactions and the advancement of fraudulent methods. Machine learning methods have demonstrated potential in tackling this issue by utilizing past transaction data to detect fraudulent behavior. This study provides a thorough examination and evaluation of different machine learning techniques used in the detection of credit card fraud.*

The aim of this study is to assess and compare the efficacy of several machine learning algorithms in identifying fraudulent credit card transactions. The dataset utilized for experimentation is acquired from a prominent financial institution, including of both authentic and deceitful transactions. The dataset has been preprocessed to address missing values, outliers, and feature scaling.

A variety of machine learning algorithms, such as "logistic regression, decision trees, random forests, support vector machines (SVM), and artificial neural networks (ANN)," are utilized and trained on the preprocessed information. The evaluation of each method is conducted using criteria like as "accuracy, precision, recall, and F1-score. In addition, several evaluation methods, such as k-fold cross-validation", are used to assure the reliability of the findings.

The empirical findings suggest that machine learning algorithms has the capability to accurately identify fraudulent credit card transactions. The algorithms demonstrate varying performance across different parameters, with certain algorithms displaying higher accuracy but worse precision or recall. The "Support Vector Machine (SVM)" algorithm gets the maximum accuracy rate of 98%, while the "Artificial Neural Network (ANN)" model displays the optimal balance between precision and recall..

Keywords: Credit card fraud

I. INTRODUCTION

Fraud is generally viewed as a crime of gain. Fraud can encompass an array of acts. Simply, fraud is "any intentional or deliberate act to deprive another of property or money through deception or unfair means" (Cepeda, Gerardo, Perez, & Rivera, 2015, p. 21). There are many types of fraud, ranging from relatively minor acts to massive multi-national coordinated efforts. These acts of fraud are very costly for individuals, organizations, and governments alike (King & Doig, 2016).

Individuals and organizations become victimized by fraud. The term fraud can be used to refer to something as simple as the unauthorized use of someone's credit card to make an illegal purchase, or as complex as a massive conspiracy enacted by a large number of actors to alter financial statements to bilk investors out of billions of dollars (Hamid, 2015). While the "Enron's" and "Bernie Madoff's" of the world make headlines and capture the world's attention, it should be noted that the smaller everyday fraudulent events add up as well, costing an estimated five percent of gross revenues globally (King & Doig, 2016).

II. LITERATURE SURVEY

Autonomous credit card fraud detection using machine learning approach J Femila Roseline, GBSR Naidu , Dr. V. Samuthira Pandi ,S Alamelu alias Rajasree , Dr.N. Mageswari ,2022

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

The incidence of credit card fraud has increased significantly in recent years due to the growing number of individuals utilizing credit cards for making purchases. This is due to technological improvements and the expansion of online transactions, which have led to significant financial losses as a result of fraudulent activities. In order to mitigate such financial losses, it is imperative to develop and deploy a robust fraud detection system. Machine learning algorithms employed for credit card fraud detection operate autonomously and do not consider the intricacies of deceptive techniques or behavioral anomalies, potentially resulting in false warnings. The objective of this study is to ascertain methods for detecting instances of credit card fraud. A "*Long Short-Term Memory-Recurrent Neural Network (LSTM-RNN)"* is suggested as a means to identify instances of fraud. Furthermore, an attention technique has been incorporated to further enhance performance. Models with a complex connected structure have demonstrated high efficiency in cases such as fraud detection, when the information sequence consists of vectors with intricate features. "*LSTM-RNN*" is evaluated alongside other classifiers, including "*Naive Bayes, Support Vector Machine (SVM), and Artificial Neural Network (ANN)"*. The results of our studies demonstrate that our suggested model yields robust outcomes and exhibits a high degree of accuracy..

A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detectio Zhenchuan Li, Mian Huang , Guanjun Liu, Changjun Jiang,2021

Detecting electronic fraud transactions is significantly challenging due to the simultaneous occurrence of class imbalance and overlap. Scammers have utilized elaborate tactics to imitate fraudulent transactions in order to closely resemble genuine ones, with the aim of avoiding discovery. Therefore, a substantial quantity of deceptive transaction data aligns with authentic transactions, posing a difficulty in distinguishing between the two. However, the primary emphasis has been on rectifying the disparity in class representation rather than addressing the overlapping issues in machine learning techniques employed for detecting fraudulent transactions. This paper introduces a novel hybrid way to tackle the problem of class imbalance with overlap, employing a divide-and-conquer strategy. Firstly, a model for detecting anomalies is trained using the minority samples. This model is used to identify and exclude a small number of outliers from the minority class, as well as a significant number of majority samples from the original dataset. Therefore, the remaining samples form a subset that overlaps and has a low imbalance ratio and less interference in learning from both the minority and majority classes, as compared to the original dataset. Afterwards, the difficult subset that overlaps is tackled by employing a non-linear classifier to accurately distinguish them. In order to assess the quality of the overlapping subset, we propose the utilization of a novel evaluation criterion known as "Dynamic Weighted Entropy (DWE)". This criterion allows for the analysis of the subset's favorable properties. The trade-off is intentionally designed to achieve a balance between the exclusion of outliers in the minority class and the ratio of class imbalance in the overlapping group. The application of DWE substantially decreases the duration of searching for optimal hyper-parameters. The efficacy of our approach has been rigorously evaluated on both the Kaggle fraud detection dataset and a substantial real electronic transaction dataset. The experimental findings demonstrate that our approach outperforms the present state-of-the-art approaches significantly.

[3] Comparison of Poisson process and machine learning algorithms approach for credit card fraud detection Anastasiia Izotovaa , Adel Valiullinb,2021

This article discusses the identification of financial fraud in datasets that include an unequal distribution of fraudulent and non-fraudulent transactions. We evaluate different methodologies for the credit card fraud detection challenge. One approach involves utilizing both homogeneous and heterogeneous Poisson processes to calculate the likelihood of detecting fraud using different intensity parametric functions. Alternatively, we address classification problems by employing machine learning algorithms and various types of ensemble methods such as boostings. A comparison is made between the findings of both procedures. The essay also addresses the issue of "false positives".

[4] Fraud Detection in Supply Chain with Machine Learning, Mahdi Seify, Mehran Sepehri , Chain hOSSIEN-FAR, Aryana Darvish,2022

Fraudulent activities within Supply Chains can be identified through the detection of counterfeit physical components or the manipulation of digital data. Our approach involves employing supervised machine learning techniques to

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

identify and uncover instances of fraud and disinformation inside supply networks. The study focuses on a car manufacturer's efforts to address a growing issue of fraud, encompassing many forms such as fake invoicing and inflated costs. Pattern recognition is facilitated by the provision of big data. The code provided is a high-level implementation with specific Python algorithms. The research is ongoing, while the current work is being presented with encouraging outcomes.

[5] An Innovative Sensing Machine Learning Technique to Detect Credit Card Frauds in Wireless Communications, G. Sasikala, M. Laavanya ,B. Sathyasri,1 C. Supraja,1 V. Mahalakshmi,S. S. Sreeja Mole, Jaison Mulerikkal,S. Chidambaranathan,C. Arvind, K. Srihari ,and Minilu Dejene, 2022

The prevalence of credit card fraud has risen in tandem with the growing use of e-commerce. The detection of bank fraud is crucial due to the vital role financial transactions play in our economy. There is a need for experiments to be conducted on the implementation of automated and real-time fraud detection in this context. Several machine learning approaches can be used to detect credit card fraud, with the most common ones being support vector machine (SVM), logistic regression, and random forest. The accuracy of these detection procedures becomes critical when models uniformly penalize all errors during training. This paper employs a novel sensing method to evaluate the classification algorithm. It takes into account the misclassification cost and utilizes SVM hyperparameter optimization through grid search cross-validation. Additionally, it applies the theory of reproducing kernels, such as linear, Gaussian, and polynomial, to separate the hyperplane. The paper ensures the robustness of the approach. As a result, credit card fraud has been found to be considerably more effective than before.

[6] Detection and Analysis of Credit Card Application Fraud Using Machine Learning Algorithms Yaodong Han, Shun Yao, Tie Wen, Zhenyu Tian, Changyu Wang,Zheyuan Gu, 2020

Fraud is a pervasive issue in the financial sector that has severe consequences. Efficiently preventing and minimizing fraud is crucial. Conventional methods, like expert systems, are limited in their ability to handle intricate problems and large volumes of data. However, the recent advancements in different machine learning techniques offer new options. While numerous research studies have concentrated on addressing credit card transaction or insurance fraud, only a small number have addressed the issue of identity fraud in credit card applications. This article introduces several machine learning algorithms designed to identify instances of fraudulent activity. Initially, we analyze and sanitize the data. We established 331 expert variables with the assistance of professional consultation and then narrowed down the selection to 30 in order to reduce the dimensionality of our data. Various models, including logistic regression and decision trees, are constructed and trained using the training set. Ultimately, we determined that the random forest model exhibits the highest level of effectiveness in detecting fraud, achieving a 54% detection rate in the out-of-time test. The acquired model can be utilized in anti-fraud surveillance systems, or a comparable model development procedure can be executed in associated business domains to identify fraud and mitigate the frequency of such actions.

III. RESEARCH METHODOLOGY

Detecting Fraud

A long-standing approach for detecting fraud has relied heavily on the act of manual auditing. In manual auditing, a fraud investigator reviews a transaction or event to determine whether fraud has occurred. These events are referred to the investigator via a heuristic model, or, in some cases, through a random sampling of events (Seeja & Zareapoor, 2014). While rule-based models have been used for decades and are well accepted in the field of fraud detection, the lack of efficiency in this method has driven the desire for an enhanced detection system (Akila & Bhuvaneswari, 2018). Data analytics and predictive modeling, with the ability to review millions of records in near real-time, are solutions to credit card fraud.

Fraud Investigation

Fraud investigation and examination describes the holistic approach to addressing fraud from start to finish. It covers all aspects of fraud including detection, investigation, research, and disposition (Wells, 2014). Fraud investigations a conducted by different types of professionals based on the type and level of fraud. While Certified Public Accountants

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

(CPA), who are trained and Certified in Financial Forensics (CFF), may be employed to investigate suspected large-scale fraud, however, less experienced professionals such as Certified Fraud Examiners (CFE), search for day to day fraud that is commonly found with credit card transactions (Clements & Knudstrup, 2016).

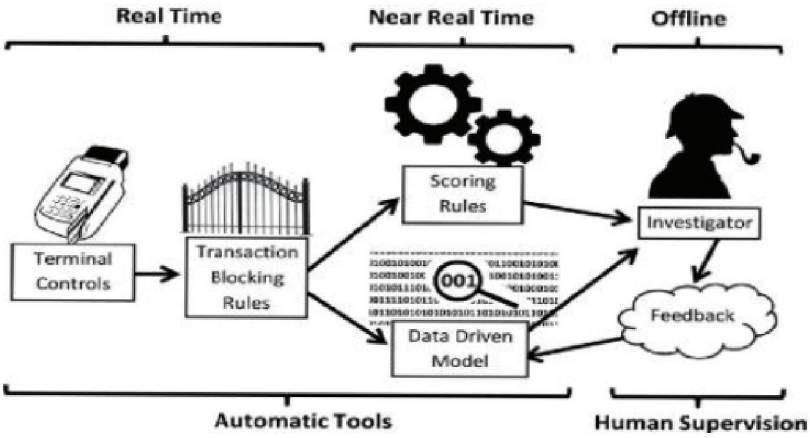

Figure : A graphic description of a proposed automated real-time fraud detection system.

V. MACHINE LEARNING

A branch of artificial intelligence, machine learning, is a system of statistical models and computer algorithms designed to learn from data with minimal human intervention. Developed to find patterns and make predictions from data, machine learning is the driving force behind predictive and prescriptive analytics (Lee, Wu, & Kim, 2015). In the arena of fraud detection, machine learning is employed to help predict which transactions could potentially be fraudulent (Perols et al., 2017).

The data used by machine learning algorithms to develop models is training data. In all machine learning examples discussed in this paper, the data is limited to data that consists of rows and columns as would be found in a spreadsheet or relational database. In machine learning terminology, the rows are referred to as either records or observations, and the columns are features. In fraud detection, training data often consists of transaction records (Randhawa et al.,2018). The data are the source of learning for the algorithm. The algorithm uses the training data to create a model. The model can then be given new transactions and produce a prediction (Kleinet al., 2016).

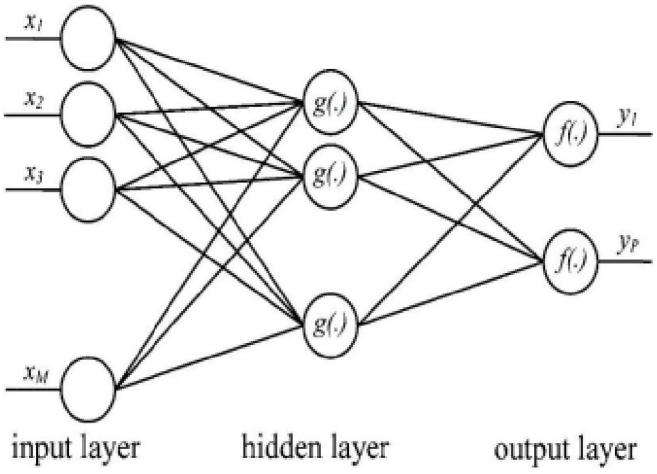

Figure : Diagram Showing Layers of a Neural Network Reprinted from "Support Vector Machines and Feed-Forward Neural Networks

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

VI. ENSEMBLE METHODS

Boosting

AdaBoost. AdaBoost creates a sequence of weak learning models to create a stronger learner. With each iteration, AdaBoost evaluates the performance of the model built. If the model incorrectly predicts an observation, that observation is given a higher weight for the next pass (Wang & Pineau, 2016).

Bootstrapping

Bagging. Bagging is the most common name given to a process known as bootstrap aggregating. It is a machine learning ensemble algorithm designed to improve accuracy and reduce variability in machine learning classifiers and regression models. Bagging is commonly used to reduce overfitting in models (Wang et al., 2016).

VII. RESULTS

What effect will the reduction of false positives have on need for manual research performed by fraud auditors? Manual investigation of potential fraud events is a manual process every fraud department must endure, no matter how automated or advanced an organization's data system. Methods have been devised to improve the ability of machine learning classifier algorithms to detect the rare fraudulent events (true positives) found in the data set because of the imbalanced nature of data sets such as the credit card fraud data set used in this study. As seen in figures below, as the true positive detection improve, so do the false positives (see Figures 25 & 26). Therefore, the choice is to increase the number of false positives that fraud investigators must sift through, or risk capturing less than half the cases of actual fraud.

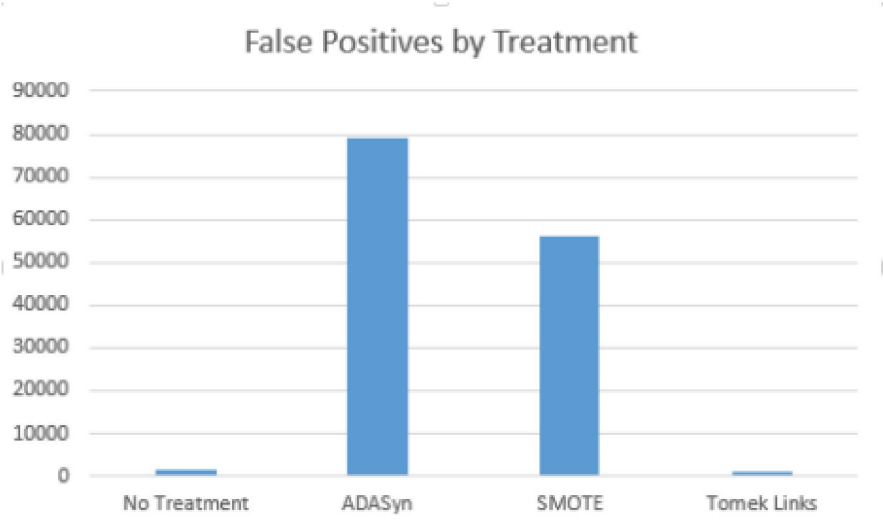

Figure . Average of false positives across all tested algorithms by data sampling treatment

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

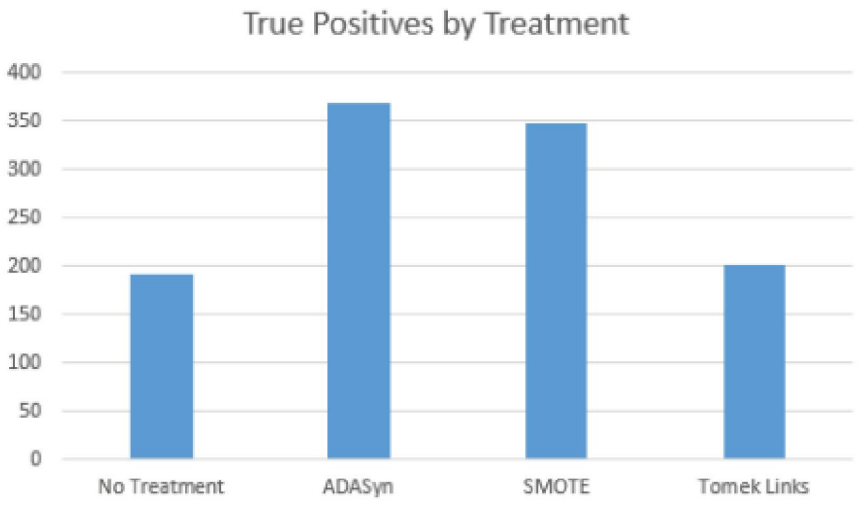

Figure . Average of true positives across all tested algorithms by data sampling treatment

ML Algorithm	TP	FP	FN	TN	DOR
Decision Tree	352	865	141	283,449	2,665.96
Naïve Bayes	378	21,240	113	263,074	209.55
Perceptron	169	40,611	322	243,704	20.47
SVM	201	38,761	290	245,553	23.75

Table: Confusion Matrix and DOR Averages Across All Algorithm, Data Sampling Treatment, and Ensemble Method

VIII. CONCLUSION

In conclusion, credit card fraud is a significant concern in today's digital age, with increasingly sophisticated fraudsters exploiting vulnerabilities in payment systems. Machine learning methods have emerged as powerful tools for detecting and preventing credit card fraud, enabling financial institutions and businesses to identify suspicious transactions in real-time and protect their customers from financial losses.

Through this study, we have explored various machine learning techniques employed in credit card fraud detection, including supervised learning algorithms like "*logistic regression, decision trees, random forests, support vector machines, and neural networks"*. Additionally, unsupervised learning techniques such as "*clustering and anomaly detection algorithms "* have also been investigated. These methods leverage the power of data analysis and pattern recognition to identify fraudulent patterns and behaviors.

The effectiveness of machine learning models in credit card fraud detection relies heavily on the quality and diversity of the data used for training. By utilizing large datasets that encompass both fraudulent and legitimate transactions, these models can learn to distinguish between normal and abnormal activities. Feature engineering techniques play a crucial role in capturing relevant information from the data, such as transaction amounts, location, time, and user behavior, which further enhance the accuracy of the models.

The evaluation of machine learning models for credit card fraud detection involves measures such as " *accuracy, precision, recall, and F1 score "*. It is essential to strike a balance between false positives and false negatives, as incorrectly flagging legitimate transactions as fraudulent can lead to customer dissatisfaction, while failing to detect actual fraud can result in financial losses. Various approaches, including "*cost-sensitive learning and ensemble methods*", have been explored to improve the performance of the models and optimize the trade-off between these metrics.

While machine learning methods have demonstrated considerable success in credit card fraud detection, it is worth noting that fraudsters continually evolve their tactics, adapting to new technologies and techniques. Consequently, the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

models must be regularly updated and refined to keep up with emerging fraud patterns. Continuous monitoring and feedback loops are vital to ensure the models remain effective in the face of evolving threats.

REFERENCES

[1]. Abdulhamid, O., Osho, O., & Shuaib, M. (2018). Evaluation of classification algorithms for phishing url detection. I-Manager's Journal on Computer Science, 6(3), 34 Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=shib&db=edb&AN= 135838191&site=eds-live

[2]. Akila, R., & Bhuvaneswari, M. (2018). Credit card fraud recognition using data mining techniques. International Journal of Advanced Research in Computer Science; Udaipur, 9(Special Issue 1), 86–87.

[3]. http://dx.doi.org.contentproxy.phoenix.edu/10.26483/ijarcs.v9i0.5618 Albrecht, W. S., Albrecht, C. O., Albrecht, C. C., & Zimbelman, M. F. (2016). Fraud Examination, 5e. Boston, MA: Cengage Learning.

[4]. Azim, M. A., & Bhuiyan, M. H. (2018). Text to emotion extraction using supervised machine learning techniques. Telkomnika, 16(3), 1394-1401. doi:http://franklin.captechu.edu:2123/10.12928/TELKOMNIKA.v16i3.8387

[5]. Badal-Valero, E., Alvarez-Jareño, J. A., & Pavía, J. M. (2018). Combining Benford's Law and machine learning to detect money laundering. An actual Spanish court case. Forensic Science International, 282, 24–34. https://doi.org/10.1016/j.forsciint.2017.11.008

[6]. Banerjee, A.V., Chassang, S., Snowberg, E. (2017). Decision theoretic approaches to experiment design and external validity. Handbook of Economic Field Experiments 1, p 141-174 https://doi.org/10.1016/bs.hefe.2016.08.005

[7]. Banerjee, P., Dehnbostel, F. O., & Preissner, R. (2018). Prediction is a balancing act: 109 importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical data sets. Frontiers in Chemistry, 6. Retrieved from https://doaj.org

[8]. Bang, S., & Kalavadekar, P. N. (2014). Determining K-most demanding products using data mining technique. International Journal of Computer Science and Network Security (IJCSNS), 14(6), 18. Retrieved from https://franklin.captechu.edu:2074/docview/1786151177?accountid=44888

[9]. Bangira, T., Alfieri, S. M., Menenti, M., & Adriaan, v. N. (2019). Comparing thresholding with machine learning classifiers for mapping complex water. Remote Sensing, 11(11) doi:http://franklin.captechu.edu:2123/10.3390/rs11111351

[10]. Boeren, E. (2018). The methodological underdog: A review of quantitative research in the key adult education journals. Adult Education Quarterly, 68(1), 63–79. https://doi.org/10.1177/0741713617739347

[11]. Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS ONE, 12(6), e0177678.

[12]. Canovas-Garcia, F., & Alonso-Sarria, F. (2015). Optimal combination of classification algorithm and feature ranking methods for object-based classification of submeter resolution Z/I-Imaging DMC imagery. Remote Sensing, (4), 4651. https://doi.org/10.3390/rs704404651