



A Model for Predicting the Impact of Alcoholism and Drug Abuse on Students' Academic Performance using Machine Learning Techniques

Ogwo Eme
Department of Computer
Science,
Akanu Ibiam Federal
Polytechnic, Unwana,
Nigeria

Goodluck Ikwudito
Emereonye
Department of Computer
Science,
Akanu Ibiam Federal
Polytechnic, Unwana,
Nigeria

Malachy Amaechi
Eziechina
Department of Computer
Science,
Akanu Ibiam Federal
Polytechnic, Unwana,
Nigeria

Taiwo Adisa
Oyeniran
Department of Computer
Science,
Akanu Ibiam Federal
Polytechnic, Unwana,
Nigeria

ABSTRACT

Nowadays, the illicit consumption of drugs and alcohol by Nigerian youths has a substantial impact on both their academic performance and society at large. In order to help find solutions that can shield students from the disturbing problem of alcoholism and drug abuse, machine learning (ML) techniques that are capable of predicting the risks that Nigerian students might fall prey to drug and alcohol addictions, which could affect their academic performance was deployed. Data regarding the impact of alcoholism and drug abuse among students were gathered through a field survey from different tertiary institutions across Nigeria. Two renowned machine learning methods - Support Vector Machine (SVM) and Random Forest (RF) were applied to the preprocessed dataset collected for our survey. An optimizer was employed to achieve the best optimization function for the deployed machine learning models. The performance and effectiveness of both ML classifiers were evaluated in order to determine which of them has the best prediction accuracy and error rate using several well-known ML evaluation metrics.

General Terms

Higher Institution, ML Model, Random Forest, Support Vector Machine

Keywords

Academic Performance, Alcoholism, Drug Abuse, Impact, Machine Learning, Prediction.

1. INTRODUCTION

Students' academic performance in school is a measure of how well the students have completed their educational objectives. The most popular ways to assess students' academic achievement are through exams or ongoing evaluations, but there is no consensus on the optimal methods or the most crucial elements. Parents are concerned about their children's academic achievement because they think that a strong academic performance would lead to lucrative employment with appropriate job security. Numerous criteria are used to measure academic performance in schools. Students may showcase their knowledge through, among other things, completing written and oral tests or exams as part of continuous assessments, giving presentations, submitting homework, and taking part in class exercises and conversations in groups. Countless young adults in school age are abusing drugs,

dropping out of school, wallowing in hospital wards, and roaming the streets of Nigerian schools with mental health issues as a result of alcoholism and drug abuse. Alcohol addiction and drug abuse have a significant negative impact on student's academic performance in school. The effect of these vices has led to increase in crimes including rape, murder, theft, armed robbery, and thuggery.

A drug refers to a substance that could bring about a change in the biological function through its chemical actions. It is also considered as a substance that modifies perceptions, cognition, mood, behaviour and general body functions [1]. This could thus be considered as chemical modifiers of the living tissues that could bring about psychological and behavioural changes [2]. Drug abuse is a major public health problem all over the world [3]. The use and abuse of drugs by young people have become one of the most disturbing health related phenomena in Nigeria and other parts of the world. The alarming evidence in the prevalence of drug abuse, the effects and consequences of substance abuse among students has called for concern among parents, government and to the general public to mount strategies of equipping youths with skills of living devoid of substance abuse.

Drug abuse is a worldwide issue and it is raising serious concern both to governments and individuals. The problem is more prevalent among young people who in most cases are not aware of the implications of what they do. Many of such youths abuse drugs out of frustration or due to poverty, which results from unemployment. Others abuse drugs out of ignorance of its dangers, while some do so out of deviant behaviour. Hence, they just want to be disobedient to authorities and to their parents [4]. Most intuitively, alcohol and drug consumption may have some detrimental effects on student's cognitive abilities, for instance, by decreasing their ability to concentrate. Concerning the indirect channels, drug and alcohol consumption may for instance be responsible for shifting individuals' resources away from schooling. Additionally, it may undermine students' progress by making them less likely to attend classes or keep up with their studies. Experts in the field of Psychology argue that alcoholism may lower individuals' expectations about their academic performance. This effect could be driven by a shift in students' peers when they engage in abusive alcohol consumption [5].



2. REVIEW OF RELATED LITERATURE

A review of recent studies on the impact of alcoholism and drug abuse on students' academic performance was done in the cause of conducting this research. Understanding the techniques and approaches used by these authors whose works were consulted provided an avenue to find research gap and fill the research gap discovered.

The findings of a study [6] revealed that students under the influence of drugs can: fight a teacher, sneak out of school and break school regulations and damage school properties on purpose. Their study also showed that students who abuse drugs often perform below average in their respective classes. The authors in [7] stated that the foremost reasons why students abused substances were mainly to supposedly 'enhance' their academic performances while some other students combined prescription stimulants with alcohol for recreational purposes but at the end it was discovered that this drug abuse impaired their educational pursuits and affected their academic performance. In a study conducted by [8], the authors discovered that 62.5% students who abused drugs and depended on alcohol failed in their first year of study in higher institution in comparison with 50 % among the students who did not report these drinking problems. The findings of a related study conducted by [9], showed that increased odds of good academic performance were observed among students reported to be non-smokers. This is an indication that academic performance of those who abuse drugs were affected [10]. Their study showed that there is significant difference between the academic performance of students who abuse drugs and those who do not and there is significant gender difference among adolescent substance abusers on the basis of academic performance. Reduction or discontinuation of smoking is of high importance for good academic achievement among these target groups.

The authors in [11] introduced Linear Regression framework for prediction in the form of classification that derives data features based on the classification of two significant attributes. A generic disease prediction system that utilizes machine learning techniques was put forth by [12]. In [13], an in-depth analysis of various machine learning methods employed for coconut sugar quality prediction and analysis. In a study [14], a DT machine-learning algorithm was employed to predict the daily smoking habit among students. The highest level of accuracy was 84.11% with a maximum depth of five using the extreme gradient boosting (XGBoost) decision tree technique. A research [15] employed smoking-associated deoxyribonucleic acid (DNA) and machine learning classifiers for predicting the prognosis and mortality of human immunodeficiency viruses (HIV). An approach for predicting obstructive pulmonary disease exacerbations using ML features was proposed by [16]. In a research by [17], the authors used statistical methods and machine learning to predict the prevalence of smoking. With 83.44% accuracy, 83% precision, 83.4% recall, and 83.2% F-measure in their work, logistic regression performed the best. An approach that predicts alcohol use disorder by using a machine learning classifier to verify the treatment-seeking state was developed by [18]. Cognitive, mood, impulsivity, personality, aggression, early-life stress, and childhood trauma were the data collection domains. In a paper presented by [19], the authors predicted how students' performance was affected by alcohol and drug addiction using data mining. Two machine learning algorithms namely, Decision tree (DT) and Naïve Bayes were deployed by

the authors for predicting alcohol consumptions among students. The result of their research showed that the DT algorithm produced highest value for accuracy as compared to other classification algorithms. The authors concluded that there was a strong relation that showed that students' academic performance was affected by addiction to alcohol and drug abuse.

In another research [20], a proposed model that predicts students' alcohol consumption using Data Mining was developed. The authors used three machine learning algorithms namely, K-Nearest Neighbor, J48, and Random Forest for predicting alcohol consumptions among students and what the impact it has on their academic performance. The result of their research showed that the KNN algorithm performed better than the Random Forest and J48 for this classification problem as it produced the highest value for accuracy as compared to the other classification algorithms used for the research. The authors concluded that there was a strong relation that showed that students' academic performance is affected by addiction to alcohol.

3. METHODOLOGY

In this work, Machine Learning techniques were chosen for the design of the model that predicts the impact of drug abuse and alcoholism on students' academic performance. Two different machine learning algorithms, which include support vector machine (SVM) and random forest (RF) classifiers were combined for the purpose of this study. A set of metrics for evaluating the numerous facets of the model including F1 score, Recall and Precision was adopted to evaluate the performance of the ML model. The system implementation was broken down into various steps:

3.1 Data collection

The dataset used for this research comprise of 1,470 records from a well-structured self-study questionnaire. The dataset was obtained from an online survey collected from alcohol/drug addicted and non-addicted students across different colleges, polytechnics and universities in Nigeria which was collected an online survey using Google online form.

3.2 Pre-processing

Preprocessing is a set of procedures used to change the data before feeding it to the algorithm. It changes data into a format that can be used for data mining, machine learning, and other data science operations more quickly and efficiently. This technique was adopted to remove incomplete and redundant and irrelevant data for the study. Data integration, data reduction, data cleaning, and data transformation were data pre-processing procedures used for this data preparation step.

3.3 Model building

Support Vector Machine (SVM) and Random Forest (RF) machine learning algorithms that can pick up delicate data from the training set, including the drug and alcohol consumption of different students were used for building of the model. In order to encode the correlations in the data, machine learning algorithms work by evaluating a large amount of data and adjusting their parameters.

3.4 Prediction model

As part of the decision-making process, students' preferences and restrictions were taken into account, and the model created in the preceding stage was used to produce predictions of the



impact of alcoholism and drug abuse on the students’ academic performance.

3.5 Model evaluation

The prediction accuracy, standard deviation, mean score, confusion matrix, classification report and AUC, ROC curve are employed to evaluate the performance of SVM and RF classifiers. The following evaluation metrics were used in the model evaluation:

3.5.1 Classification accuracy

The Classification accuracy is the ratio of correctly classified data points to the total number of points in the dataset which ranges from 0-100%.

$$\text{Accuracy} = \frac{\text{Number of correct classifications}}{\text{Total number of classifications}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad 1$$

3.5.2 Precision

Precision is a metrics used to measure the positive classifications represented as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad 2$$

3.5.3 Recall

Recall is a metric used to measure the false negative classifications represented as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad 3$$

3.5.4 F1-score

F1-score takes into consideration the true positive and false positive regardless of false negative and false positive classifications. The F1-score is sensitive to which class is positive and negative as given below in equation 4.

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \quad 4$$

4. RESULTS AND DISCUSSION

4.1 Feature correlation map

Figure 1 displays the feature correlation matrix of our proposed system’s dataset, highlighting the groupings of highly correlated features that make significant contributions to model predictions. It displays a graphical representation of the linked features. Correlation matrix is used to measure the relationship between database attributes (variables). The matrix depicts a linear correlation between all possible pairs of dropping out, alcohol intake rate, age, performance, smoking in the future, smoking cigarette and drug alcohol addictions values. There is a positive and negative relationship between attributes in the database as shown in the main diagonal and other pairs. The positive relation reveals indicates that the independent and dependent variables moves in opposite direction while negative correlation shows that, both variables are moving in same direction.

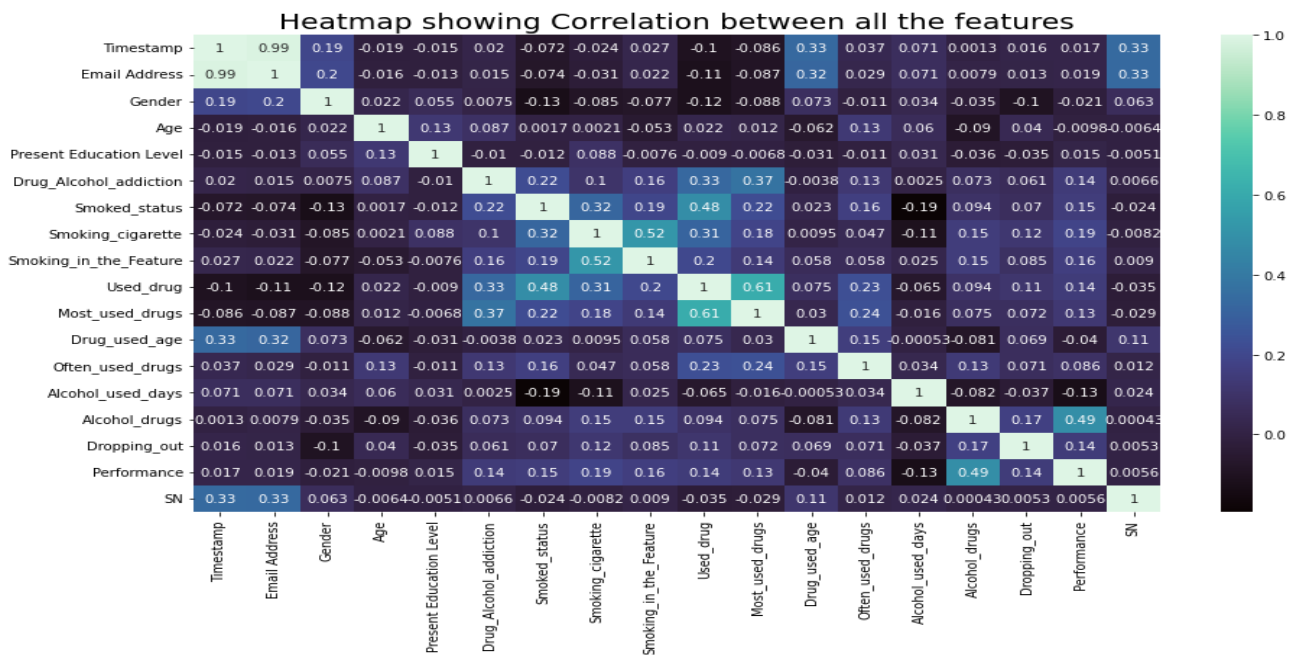


Fig 1: Feature correlation map

4.2 Addiction to drugs

Figure 2 displays the overall number of students that are addicted to using drugs on a daily basis. The visualization

illustrates that 69 students were addicted to drugs, while 1401 are unfamiliar with heavy substances. The aggregate of both target classes yielded 1470 items for training and testing purposes.

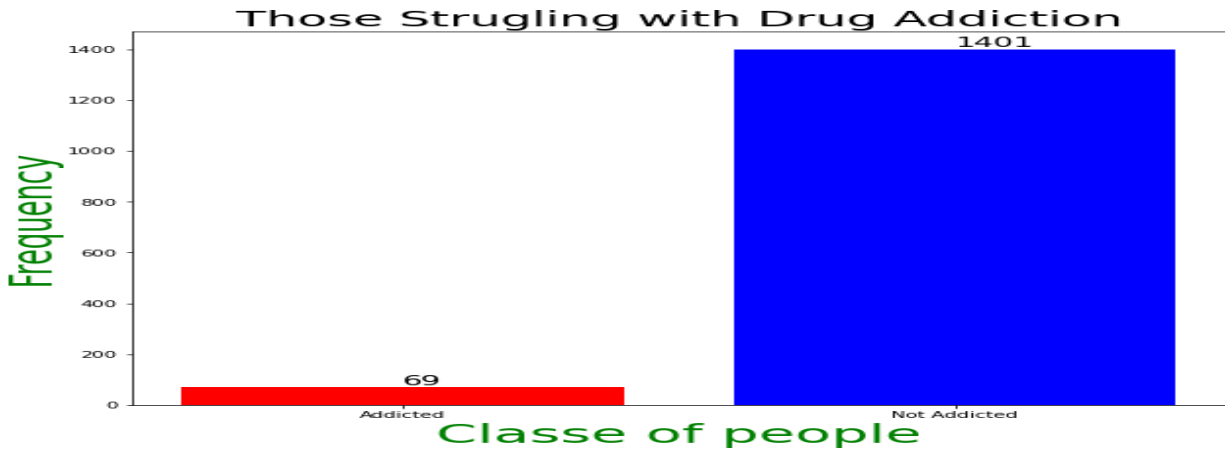


Fig 2: Those addicted to drugs

4.3 Those Anticipating to Smoke Cigarette

Figure 3 displays individuals who intend to smoke cigarettes if offered by friends in the coming years. Those who never intended to use a cigarette yielded 92.9%, while those who responded 'yes' contributed 6.7% are committed to smoke a

cigarette in the coming years and 0.4% was not too sure but may still smoke the cigarettes. The result shows that a percentage number of students are prone to smoking cigarettes and perhaps abuse any other drugs if offered by their friends. This is an indication that students' addiction to alcoholism and drug abuse can be influenced by peer pressure.

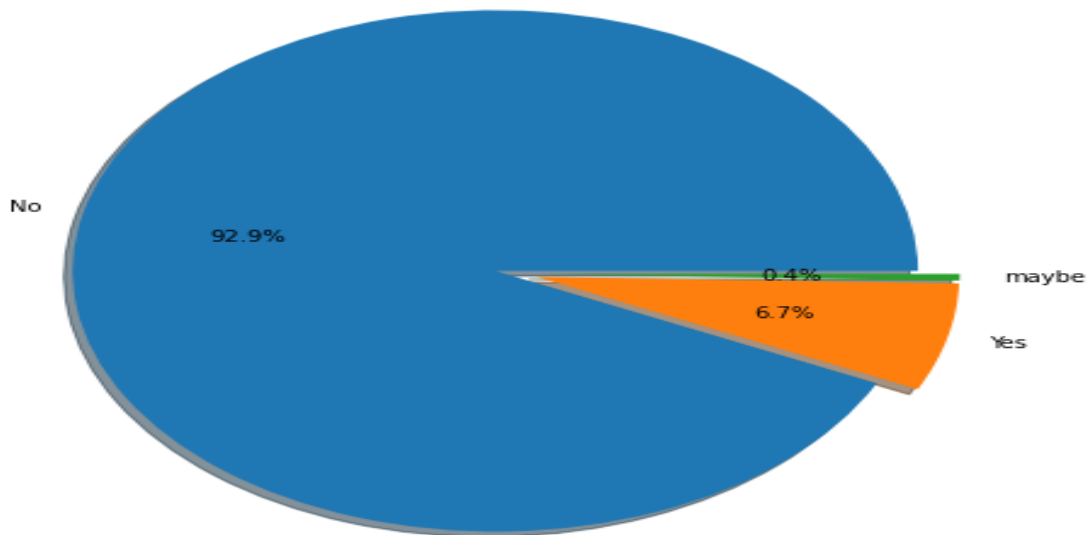


Fig 3: Those anticipating smoking cigarettes

4.4 The Most Used Drugs by Students

Figure 4 shows the most frequently used hard drugs by students in the study area. Marijuana intake recorded over 35, morphine,

tobacco, and crack were all used at exactly the same frequency in the horizontal axis by students in the research area. This suggests that marijuana is the most commonly used drugs by students in higher institutions in Nigeria.

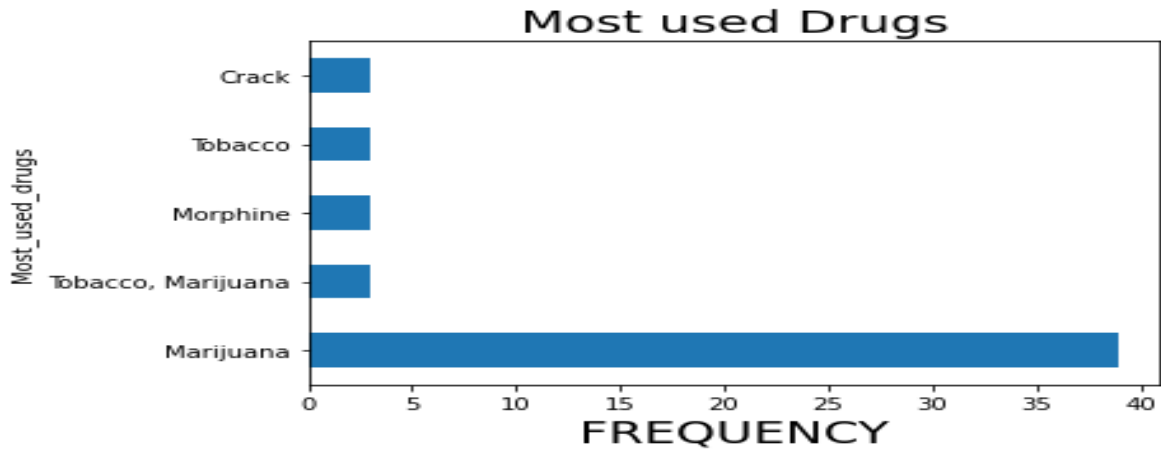


Fig 4: Most Used Drugs by Students

4.5 Impact of Alcoholism and Drug Abuse on Students’ Academic Performance

Figure 5 presents the respondents' reactions regarding the impact of alcoholism and drug abuse on students' academic performance. According to the above data, "agreed" had the

highest number of responses (recorded more than 700), followed by "Strongly Agreed" while few number of respondents "Disagreed," and others "Strongly Disagreed" on the impact of alcoholism and drug abuse on students’ academic performance. The results indicate that the school dropout rates in the research area are influenced by alcoholism and drug use.

Alcoholism and Drug abuse Affect academic Performance

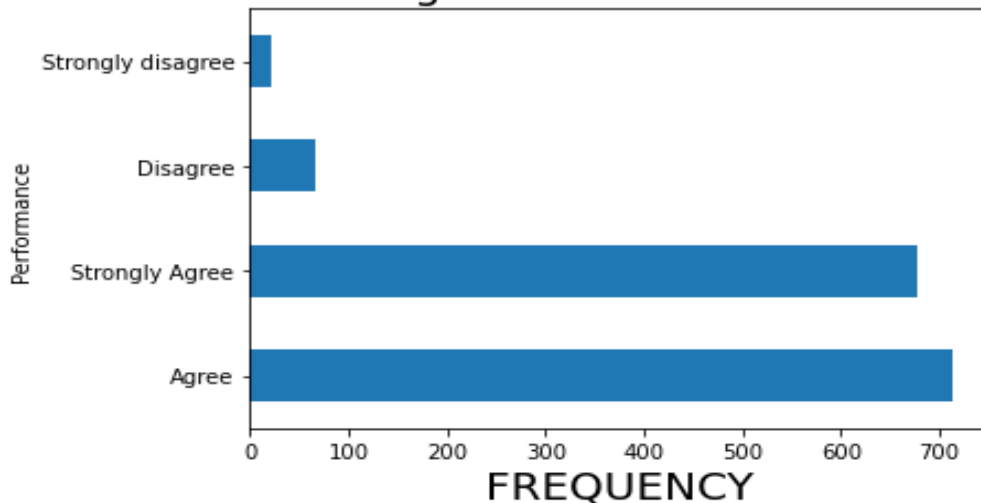


Fig 5: Impact of Alcoholism and Drug Abuse on Students’ Academic Performance

4.6 Drop-out Rate Caused by Alcoholism and Drug Abuse on Students’ Academic Performance

Figure 6 illustrates how drug addiction can have a detrimental effect on a student's academic performance, ultimately leading to high dropout rates. From the results above, “agreed” yielded

the highest response or frequency of drop-outs with over 700 in number. This was followed by those that “Strongly Greed”, then came respondents with “Disagree”. Those that strongly “Disagreed” come a distant last. The findings show that alcoholism and drug abuse is a contributing factor in school dropout rate among students in higher institutions in Nigeria.



Dropping out of school casued by Drug Abuse

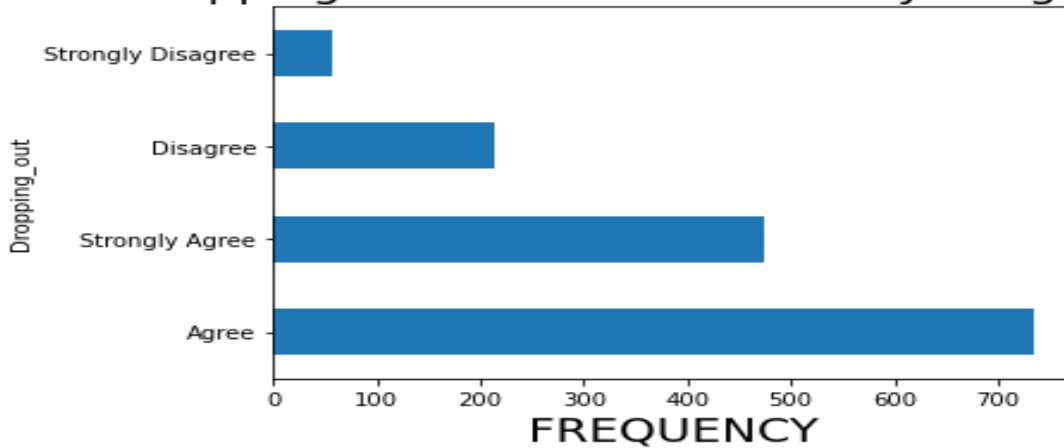


Fig 6: Drop-out Rate among Students

4.7 Predicting Student Academic Performance

Figure 7 shows how the model predicts the impact by alcoholism and drug abuse decision trees on students’ academic performance using the “If...Then...Else” conditional statement of the Random Forest algorithm. The Random Forest structure combines several voting Decision Trees to build a random

forest space.

This was accomplished by selecting samples at random from the original set with replacement in the growing trees. If a student’s academic performance as a result of alcoholism and drug abuse falls below a certain threshold, the branching nodes are moved to the left, and otherwise to the right node with its squared error term.

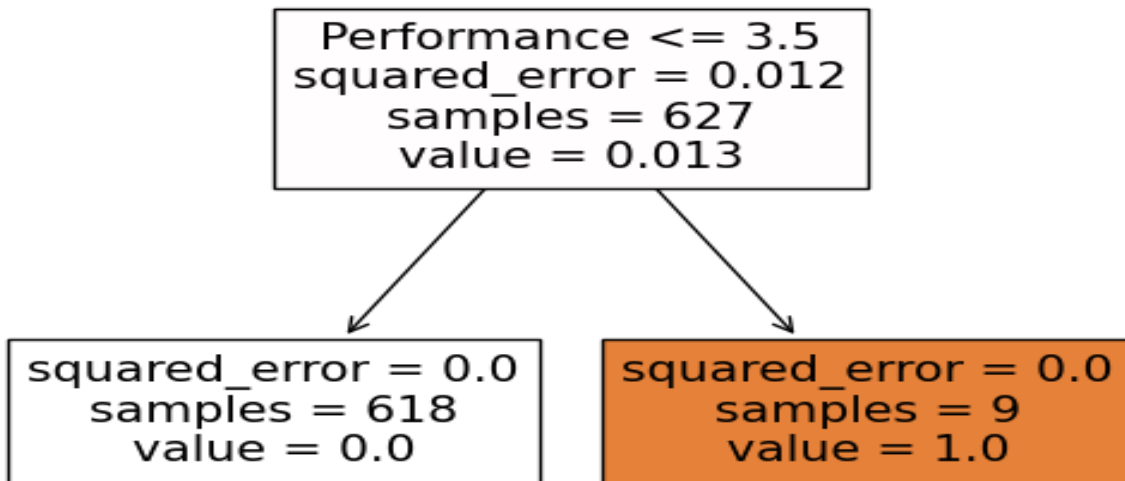


Fig 7: Predicting Student Academic Performance

4.7 Support Vector Machine optimization plot

Figure 8 is the Support Vector Machine optimization plot

demonstrating performance at various iterations or epochs. Based on the given data, the optimizer concluded that the best performing SVC model had hyper parameter values of C = 5.583 and degree = 3.

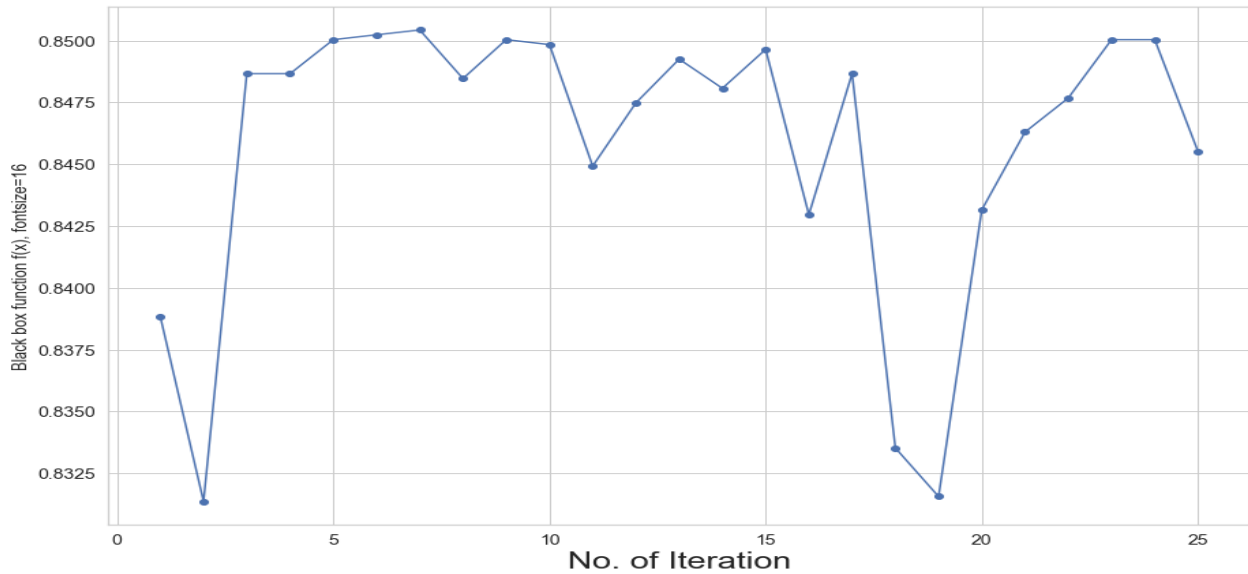


Fig 8: SVM Optimization Plot

4.9 Comparison of SVM performance and RF Models Using Different Evaluation Metric

Table 2 displays the results of SVM and RF across various diagnostic tools, comprising precision, recall, and f1-score terms. SVM recall metrics beat RF for both students with poor academic performance due to drug use and those who were

unaffected. In terms of performance, the SVM scored higher, with a precision score of 0.97, recall score of 1.00 and F1 score of 0.99 which higher than the RF algorithm. The results show that the SMV machine learning technique performs better in predicting the impact of alcoholism and drug abuse in students

Table 1: Evaluation Comparison of SVM and RF Models

| MODELS | Performance | Precision | Recall | F1-score |
|--------|-------------|-----------|--------|----------|
| SVM | High | 0.97 | 1.00 | 0.99 |
| SVM | Low | 0.00 | 0.00 | 0.00 |
| RF | High | 0.99 | 0.99 | 0.99 |
| RF | Low | 0.99 | 0.99 | 0.99 |

5. CONCLUSION

This paper focused works on the deployment of Machine Learning techniques to predict the impact of alcoholism and drug abuse on the students’ academic performance. The dataset used for the research was collected from students across several higher institutions in Nigeria using an online survey. The results of the research show that alcoholism and drug impact significantly on students’ academic performance. The result also indicates that alcoholism and drug abuse among students lead to high dropout rate among students. The result of the evaluation of the performance and effectiveness of both ML classifiers of the Machine Learning Model revealed that the SVM algorithm performs better than RF algorithm in predicting the impact of alcoholism and drug abuse on students’ academic performance. It is recommended that all stakeholders in the educational sector including the government in all levels, teachers, parents and the generally public should put hands on deck to nibble this ugly trend in the bud.

Future research direction will focus on predicting the impact of stakeholders including the government, educators and parents in mitigating the negative impact of alcoholism and drug abuse on students’ academic performance. This will go a long way to significantly reduce this menace among students in higher institutions across Nigeria.

6. REFERENCES

- [1] Balogun, S. K. (2006) .Chronic intake of separate and combined alcohol and nicotine on body Maintenance among albino rats. *Journal of Human Ecology*, 19, (1), 21 – 24
- [2] Nnachi, R. O., (2007). *Advanced psychology of learning and scientific enquiries*, Enugu: J. J. Classic Publisher Ltd, 1-12.
- [3] UNODC, (2005). *World Health Organization Expert Committee on Dependence Producing Drugs Fourteenth Report Urban Adolescent*. *Child Development*, 61, 2032-2046
- [4] Ya’u, H. U. (2015). *Combating Drug Abuse through Adult Education in Nigeria*, *International Journal of Humanities and Social Science Invention*, 4(3).36-40.
- [5] Akanbi, M. I., Godwin, A., Theophilus, A.B., Muritala, M. & Ajiboye, S. A. (2015). *Impact of Substance Abuse on Academic Performance among Adolescent Students of Colleges of Education in Kwara State, Nigeria*. *Journal of Education and Practice*.6 (28), 108-112
- [6] Atayi, S.G., Ethel, G.O., Ismaila, S., Olokoba, O., & Ogbonna, N.C. (2021). *Drugs Abuse, Civic Behaviour and Students' Academic Performance: Perspectives of Senior Secondary Schools In Zaria Education Zone, Kaduna*



- State, Nigeria. *Global Journal of Pure and Applied Sciences*. 27, 321-326
- [7] Manish,G., Deepali, T. & Neetu , S. (2020). Effect of the Drug Abuse on the Academic Performance of the Students/Adolescent *Biomedical Journal of Scientific & Technical Research, BJSTR*. MS.ID, 28(3), 1-10.
- [8] Aertgeerts, B. & Buntinx, F. (2002).The relation between alcohol abuse or dependence and academic performance in first-year college students. *Journal of Adolescent Health*. Available at <https://pubmed.ncbi.nlm.nih.gov/12225733/>
- [9] Muritala,I.A., Godwin, A., Anyio.,B Theophilus., M., Muritala, A. & Stephen, A. (2015). Impact of Substance Abuse on Academic Performance among Adolescent Students of Colleges of Education in Kwara State. *Journal of Education and Practice*, 6(28), 1-18.
- [10] Ethan, S., Acion, L. & Arndt. S.(2015), National analysis of differences among substance abuse treatment outcomes: College student and nonstudent emerging adults, *Journal of American College Health*, 63(2), 118-124.
- [11] Dahiwade, D., Patle, G. & Meshram,E. (2019) “Designing disease prediction model using machine learning approach,”*Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC*, 1211–1215.
- [12] Alonzo, L. M., Chioson, F. B., Co, H. & Bugtai, N. T.(2018), A Machine Learning Approach for Coconut Sugar Quality Assessment and Prediction, Conference: 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management, 1-14.
- [13] Zhand, Y., Liu, J. & Zhang, Z.(2019), Prediction of daily smoking behavior based on decision tree machine learning algorithm, Conference: 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), 1-15.
- [14] Zhang, X., Aouizerat, B. E., Pend, G. & Hu, Y.(2018), Mchine learning selected smoking-associated DNA methylation signatures that predict HIV prognosis and mortality, *Clinical Epigenetics* springer, 10(1), 1-12
- [15] Granero, M. A. F., Morillo, D. S., Lopez, M. A., Leon-Jimenez, A.(2015), A machine learning approach to prediction of exacerbations of chronic obstructive pulmonary disease, Conference: International Work-Conference on the Interplay Between Natural and Artificial Computation, 3-6.
- [16] Frank, C., Seetan, R., Wahbeh, A., & Habach, A.(2018), Predicting smoking status using machine learning algorithms and statistical analysis, *Advances in Science Technology and Engineering Systems Journal*. 3(2), 184-189.
- [17] Lee, M. R., Sankar, V., Hammer, A., Kennedy, W. G., Barb, J. J. McQueen, P. G. & Leggio,L.(2019) Using machine learning to classify individuals with alcohol use disorder based on treatment seeking status, *Lancet Eclinical Medicine*, 12, 70-78.
- [18] Trivedi, T. & Kotak, D. (2019).Exploring Prediction Modeling of Students Alcohol and Drug Addiction Affecting Performance using Data Mining Approach.*International Journal of Engineering Research & Technology (IJERT)*. Vol. 8 Issue 12, 169-173
- [19] Tincymo, M. T. & Joseph, G. (2021). Predicting Alcohol Consumption in Students Using Data Mining Tool.*Proceedings of the National Conference on Emerging Computer Applications (NCECA)*.Vol.3, Issue.1. DOI: 10.5281/zenodo.5093705. 259-263