On the Development of Smart-Watch Detection Using $C_{5.0}$ and $S_{5.24.1}$ Algorithms

Philip Seths Onyekachi., Amanze Bethran Chibuike., Igbe Chukwudi M
Department of Computer Science, Faculty of Physical Sciences,
Imo state University, Owerri, Nigeria
amanzebethran@yahoo.com

DOI: 10.56201/ijcsmt.v10.no1.2024.pg25.36

Abstract

Smart-Watch as an emerging IoT solution in order to recognize human activities with respect to payment transactions. To implement this, a web application was developed specifically for payment transactions. The Smart-Watch is embedded within the transaction platform to monitor transactions on the IoT platform, and detect when a transaction is suspected to be a fraudulent. For the purpose of classifying transactions, we considered the C5.0 algorithm and $S_{5.24.1}$ to present reliability of the Smart-Watch. The result from the software developed was tested using Confusion Matrix with a dataset of 100 transactions. The result obtained shows almost 93% of accuracy in detecting fraud using Smart-Watch. The Smart-Watch system has a high accuracy, and it will be a reliable system and also a profitable investment if put into action in the business world.

Keywords: *Smart watch, $C_{5.0}$, $S_{5.24.1}$ and IOT*

INTRODUCTION

Fraudsters can gather your personal information, and then contact your cell phone provider (example: MTN) to impersonate you and have your phone number ported to another device. Once they can access your phone number, they can intercept confirmation codes for two-factor authentication sent via SMS. This is the reason for Smart-Watch Detection System using *C5.0 algorithm* and *S5.24.1 algorithm*. SIM swapping happens when scammers contact your mobile phone's carrier and trick them into activating a SIM card that the fraudsters have. Once this occurs, the scammers have control over your phone number. Anyone calling or texting this number will contact the scammers' device, not your smartphone. This is known as SIM swap fraud, and it means scammers could potentially enter your username and password when logging onto your bank's website. The bank will then send a code by text (two-factor authentication) to your smartphone number, a code that you'll then have to enter to access your online account. After a SIM swap, that number now goes to the smartphone or other device possessed by scammers. They can then use that code to enter your bank account. The first sign that you could be a victim of SIM swapping is when your phone calls and text messages aren't going through. This likely means

fraudsters have deactivated your SIM card and are using your phone number. A SIM swap scam happens when criminals take over control of your phone by tricking your carrier to connect your phone number to a SIM card in their possession. These scammers basically take over control of your mobile phone's number. To steal your number, scammers start by gathering as much personal information on you as they can find and then engaging in social engineering. First, the scammers call your mobile carrier, impersonating you and claiming to have lost or damaged their so-called SIM card. They then ask the customer service representative to activate a new SIM card in the fraudster's possession. This ports your telephone number to the criminal's device, which contains the scammer's own SIM card. Once your carrier completes this request, all phone calls and texts that are supposed to go to you will instead go to the scammer's device. That is where the data scammers have collected from you through phishing emails, malware, or social engineering becomes useful. Scammers might send you an email claiming to be from your smartphone provider. This email might say that you need to click on a link to keep your account active. When you do, you're taken to a new page that asks you to provide personal information, including your name, birthdate, and passwords. Once you fill this out and click "Send," you've given the scammers access to the information they need to trick your mobile phone carrier into a SIM swap scam. Other scammers trick you into clicking on email links that fill your computer with malware that records your keystrokes, including any passwords or security question answers you type; this is called **keylogger**. Again, this provides the fraudsters with the information they need to pull off a successful SIM swap. Fraudsters might also buy your personal and financial information on the Dark web. This, too, would arm these con artists with the information they need to successfully work their scam. Once scammers provide your smartphone provider with the information, they've gotten from you or the Dark Web, they use it to convince your provider to switch your number to a new SIM card. These criminals then gain access to, and control over your cellphone number, something that fraudsters can use to access your phone communications with banks, in particular, your text messages. They can then receive any codes or password resets sent to that phone via call or text for any of your accounts. Skimming is a method used by identity thieves to capture payment and personal information from a card holder. Skimming can occur anytime a bank customer uses an electronic payment card at a brick-and-mortar location, like in a fuel station or POS stand. Fraudsters can obtain information in various ways, and the technology that they use is becoming more sophisticated and challenging to detect. Skimming allows identity thieves to capture information from a cardholder that can be used to make fraudulent transactions. Some fraudsters may simply photocopy or take digital photos of information that can be used fraudulently. Other more advanced technologies also exist, such as *skimming devices* designed for use in many different situations. At brick-and-mortar locations, a fraudster can use a small skimming device that allows them to swipe a card and obtain information from its magnetic strip. Some skimmers may also include a touchpad that allows the thief to enter a security code. Thieves can also build skimming devices that can be used at automated teller machines (ATMs) and other POS locations such as fuel stations. Skimming devices can be installed on an ATM with cameras and overlay touchpads can be added to capture individual personal identification numbers. Fuel stations and shopping malls like Shoprite are another target where skimming devices can be easily installed since ATMs are now installed within the environment. Skimming technology is becoming more

sophisticated each year, and it is difficult for authorities to stay one step ahead. Some skimmers are as thin as a credit card, and can be inserted into ATM machines or POS devices like Moni point machines. Session hijacking, also known as TCP session hijacking, is a method of taking over a web user session by secretly obtaining the session ID and masquerading as the authorized user. Once the user's session ID has been accessed, the attacker can masquerade as that user and do anything the user is authorized to do on the network. A byproduct of this type of attack is the ability to gain access to a server without having to authenticate to it. Once the attacker hijacks a session, they no longer have to worry about authenticating to the server as long as the communication session remains active. The attacker enjoys the same server access as the compromised user because the user has already authenticated to the server prior to the attack. With Wireshark, or OWASP Zed, you can capture network traffic containing session IDs. Applications use sessions to store parameters that are relevant to the user. The session is kept "alive" on the server as long as the user is logged on to the system. The session is destroyed when the user logs-out from the system or after a predefined period of inactivity. When the session is destroyed, the user's data should also be deleted from the allocated memory space. A session ID is an identification string (usually a long, random, alpha-numeric string) that is transmitted between the client and the server. When cybercriminals have hijacked a session, they can do virtually anything that the legitimate user was authorized to do during the active session. The most severe examples include transferring money from the user's bank account, buying merchandise from web stores, accessing personally information for identity theft. Banks relied on traditional rule-based systems to detect fraudulent activities in the past. These systems were designed to detect fraud based on predefined rules, such as specific transaction thresholds or patterns. While these systems were effective in detecting known fraud patterns, they were limited in their ability to detect new and evolving fraud patterns. With the advent of ML, banks now have a more effective tool for detecting unauthorized activity. ML algorithms (like the *C5.0* algorithm) can analyze large amounts of data and detect patterns that may not be apparent using traditional rule-based systems. By analyzing transactional data and customer behavior, the analytics provided by these algorithms can identify potentially fraudulent activities in real-time. One of the key advantages of ML over traditional rule-based systems is its ability to adapt to new and evolving fraud patterns. Machine learning algorithms can learn from past fraud cases, estimate probability, and adapt to new patterns as they emerge, making them more effective in detecting and preventing scammers. Another advantage of ML is its ability to reduce false positives. Traditional rule-based systems often generate false positives, flagging legitimate transactions as potentially fraudulent activities. This can be a time-consuming process for banks and frustrating for customers. Machine learning algorithms can reduce false positives by identifying patterns that are indicative of fraud while taking into account the individual behavior of each customer. Machine learning algorithms can analyze vast amounts of data, detect new and evolving fraud patterns, and reduce false positives, making them a critical component of any bank's fraud detection strategy. Finally, this thesis aims at developing a Smart-Watch Detection System using the *C5.0 algorithm* for detecting and preventing fraudulent transactions. Financial fraud in IoT environment is one of the fast-growing problems since the mobile network can aid practically any type of payments. Due to the fast increase in mobile commerce and the growth of the IoT environment, financial fraud in mobile payment platforms

has increased, and it's more common. The problems currently experienced in IoT environment as related to financial transactions are described below:

1. Successful authorization of fraudulent transactions by banks using passwords, pins, and/or biometrics as means of final authentication.
2. A lot of measures have been put in place by financial institutions to detect fraudulent transactions, but these measures only detect fraudulent transactions after it has taken place. With respect to this, there is need for an enhanced model to support the existing models in detecting and preventing fraudulent transactions from taking place.
3. Huge financial losses are experienced whenever fraudulent transactions occur using the existing system.

The aim is to develop a Smart-Watch Detection System using IoT and Machine Learning Technique

1. Avoiding fraudulent transactions based on passwords, pins, and/or biometric authentications using *C5.0 Decision tree algorithm* to learn customers' transaction pattern.
2. Developing a Smart-watch model codenamed *S5.24.1 algorithm* that will be able to generate an encoded string pattern to support the *C5.0 algorithm* in detecting fraudulent transactions before they happen.
3. To compare the Smart-Watch Detection system to existing fraud detection systems.

Litereature Review

An intelligent system is a system that learns during its existence i.e. it senses its environment and learns from each situation and considers that action that is relevant to the existing condition. Thus we can say that artificial intelligent system is that system that functions in a same way a biological brain works. Artificial intelligence has a wide range of applications in today's world including medicine, stock trading, science, robotics, discovery, telecommunication, banking and many more. Artificial intelligence can be defined as the intelligence artificially created and not existing naturally. We can also define it as a field that seeks to explain and emulate human behavior i.e. intelligence in terms of computational processes and this is based on results from philosophy, psychology and brain sciences (Bhalinder and Shivinder, 2020). It is composed of two things- first to study the thinking ability of humans and secondly to represent it by the use of machines. Intelligence can thus be recognized as interlinking and processing of multiple processes and these process models can be further be simulated on machines. Thus intelligent systems are a development of AI that provide certain features of human cognition i.e. knowledge gained by thought and perception and reasoning which are the abilities found in humans (Bhalinder and Shivinder, 2020). AI can be used for information processing either by exact formulizations by or by exemplary realisation via implementations. Thus the basic idea of AI is to make machines smarter and useful. Basically intelligence refers to the learning due to past experience, to create

understanding of ambiguous message, adaptive to new conditions, recognizing different elements in a situation and developing solutions for problems. If we compare AI and Natural intelligence we say AI is more superior due to its lesser cost, ease of duplication and documentation, permanence and consistency. But on the other hand natural intelligence is creative and we can use its direct benefit and also through the sensory experiences, it enables us to sense the relationship between things (Bhalinder and Shivinder, 2020). Machine intelligence can be divided in two parts Hard Computing based on Artificial intelligence and soft computing techniques based on computational Intelligence (Bhalinder and Shivinder, 2020). Hard computing is related with design of physical processes and systems and is based on binary logic, crisp systems, numerical analysis, differential equations; mathematical programming etc. soft computing is based on fuzzy logic, artificial neural networks, genetic algorithms and parts of machine learning etc. Imprecision and uncertainty is undesirable in hard computing whereas tolerance for imprecision and uncertainty is exploited to achieve approximate solutions. Combination of soft computing techniques is quite effective in many applications; like the use of neuro-fuzzy control is very popular in chemical process control of consumer goods. Fuzzy logic is mainly concerned with imprecision, neural network with learning and evolutionary algorithms with optimization (Bhalinder and Shivinder, 2020).

Machine Learning Techniques

Machine learning is a data analytics technique that teaches computers to do what comes naturally to humans and animals: learn from experience. Machine learning algorithms use computational methods to directly "learn" from data without relying on a predetermined equation as a model (Bhalinder and Shivinder, 2020). As the number of samples available for learning increases, the algorithm adapts to improve performance. Deep learning is a special form of machine learning. Machine learning uses two techniques: supervised learning, which trains a model on known input and output data to predict future outputs, and unsupervised learning, which uses hidden patterns or internal structures in the input data.

Analysis of the Proposed System

This work focused on developing a Smart-Watch Detection System for detecting fraudulent financial transactions. To do this, a dataset of banking transaction records will be used. Here the pattern of current fraudulent usage of the bank application is analyzed with the previous transactions using the *C5.0* Decision tree algorithm. The new system will allow account holders to run their financial transactions as usual, but with the introduction of the *C5.0* Decision tree algorithm and the *S5.24.1* algorithm, acting as the Intelligent Agent that will intelligently filter out fraudulent transactions and abort the operation. The user will enter the account he/she wants to transfer money to, the amount, and the PIN for authentication. Once the PIN is verified, the *C5.0* Decision tree algorithm and the *S5.24.1* algorithm will monitor, and classify the transaction as

being valid or fraudulent. The main goal of the proposed system is to apply a set of Classification algorithm (Decision Tree) to obtain a classification model to be used as a scanner for transaction authentication. The implementation involves tasks such as data preprocessing, feature extraction, training models, etc. See Figure 3.5 below for a detailed pictorial illustration. In this research, the proposed system uses Machine Learning classifier to classify the transactions. The Decision Tree algorithm will be used for data classifications. Decision trees are a type of Supervised Machine Learning that uses a tree-like model to represent decisions and their possible consequences. In fraud detection, Decision Trees can be used to identify the most important factors that contribute to fraudulent activities, such as transaction amount, frequency of usage, and location of transaction – automatically.

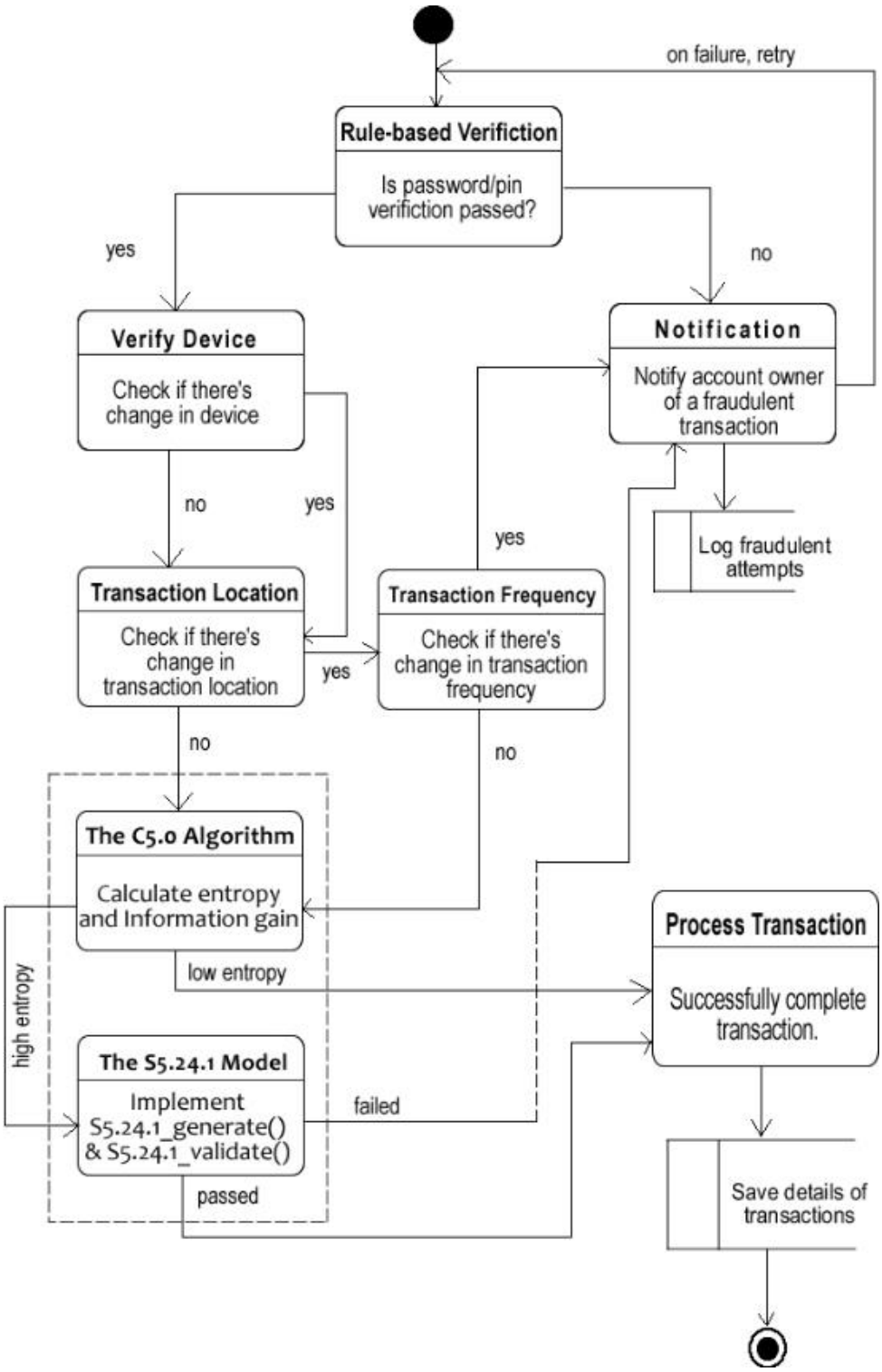
on failure, retry
Rule-based Verification
Is password/pin verifiction passed?
yes
no
Verify Device
Check if there's change in device
Notification
Notify account owner of a fraudulent transaction
no
yes
yes
Log fraudulent attempts
Transaction Location
Check if there's change in transaction location
Transaction Frequency
Check if there's change in transaction frequency
yes
no
no
The C5.0 Algorithm
Calculate entropy and Information gain
high entropy
low entropy
Process Transaction
Successfully complete transaction.
The S5.24.1 Model
Implement S5.24.1_generate() & S5.24.1_validate()
failed
passed
Save details of transactions

Figure 3: Proposed System Architecture

An overview of the Proposed System Architecture with Illustration

"*Consider a situation where a customer makes regular transactions of amount less than or equal to N100,000 from within a given locality, say Owerri. We can build a Decision Tree to predict the probability of fraud based on transactions made.****"*** As soon as a transaction is initiated, in the Decision Tree using the *C5.0* algorithm, we can check if the transaction amount is greater than *N100,000*. If the result is '*yes'* then we could check the transaction *location* where the transaction is being initiated. Any of these parameters can be checked/evaluated first. Now *frequency* of the transactions will be considered as a parameter for classification. This simply means checking the number of times a customer does carry out transactions in a week or month, or even day. After that, as per the probabilities calculated for these conditions above, the transaction will be predicted as '*fraud'* or '*non-fraud'*. Now, if the *amount is greater than N100,000* and location is equal to the IP address of the customer, then there is only a *25 percent chance of 'fraud'*, and a *75 percent chance of 'non-fraud'*. Similarly, if the *amount is greater than N100,000* and the *location is not equal* to the customer's location, then there is a *75 percent chance of 'fraud'* and a *25 percent chance of 'non-fraud'.* But if the location (IP address) of the customer is not the same as the "*learned*" location, then the new system can go further to launch the additional security layer referred to as the ***S5.24.1*** algorithm which is difficult to bypass. At this stage, it's assumed that a customer moved out from the "*learned*" location (IP address). If the *S5.24.1* model is successfully passed, then the system will trigger the final stage of credential authentication by sending *S5.24.1 string pattern* to the customer's mobile number in the form of OTP. The *S5.24.1 string* is an encrypted intelligent code which can only be valid for authentication once, when properly decrypted by the user/customer. Before this stage, the customer must have learnt how to decrypt the *S5.24.1 string* pattern for his/her account. This string pattern is not the same for every customer, and it varies during transaction.

The Proposed System Architecture of *C5.0* and *S5.24.1* Algorithms

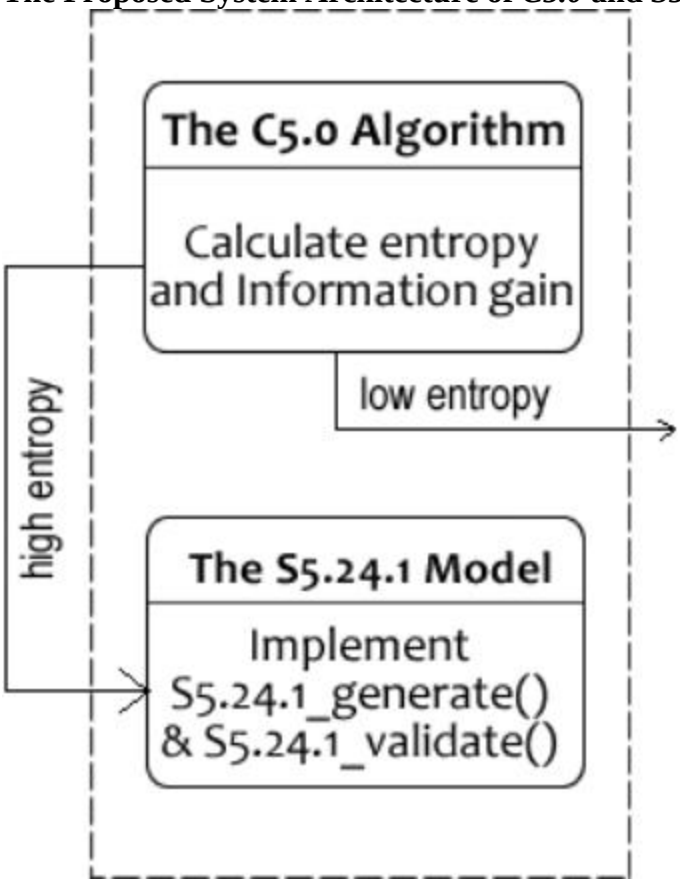

Figure 4: The *C5.0* and *S5.24.1* Algorithms

The Proposed System Architecture of *C5.0* Algorithm

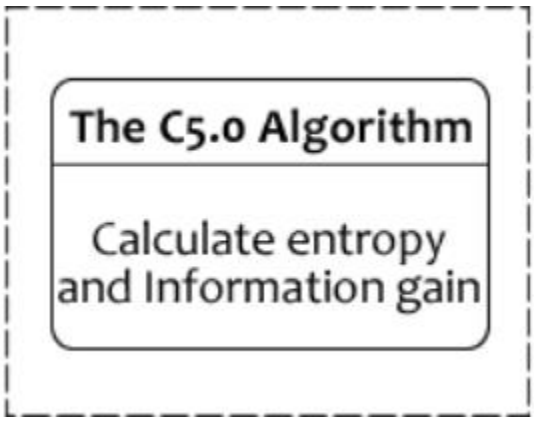

Figure 5: The *C5.0* Decision Tree Algorithm

C5.0 Decision Tree Algorithm

The *C5.0 algorithm* is a Decision tree algorithm used to measure the disorder in the collection of attributes, and effectiveness of an attribute using Entropy and Information gain, respectively. The operation of *C5.0* on the dataset can be categorized into two equations:

Calculating the entropy value of the data
Calculating the Information gain for an attribute

C5.0 algorithm uses the concept of **entropy** for measuring purity. The entropy of a sample of data indicates how mixed the class values are; the minimum value of 0, it indicates that the sample is completely homogenous/pure, while 1 indicates the maximum amount of disorder/impure. The definition of entropy can be specified as:

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Here **Pi** is the probability of an element/class **i** in our data. For simplicity let us say we only have two classes, a positive class and a negative class. Therefore, **i** here could be either **+** or **-**. So, if we had a total of **100 data points** in a dataset with *30 belonging to the **positive*** class and *70 belonging to the **negative*** class, then **P+** would be 3/10 and **P-** would be 7/10.

To calculate the entropy using the dataset as given in the illustration above, then it would be:

$$\left[-\frac{3}{10}\times \log_2\left(\frac{3}{10}\right)\right]+\left[-\frac{7}{10}\times \log_2\left(\frac{7}{10}\right)\right] = 0.88$$

The entropy here is approximately 0.88. This is considered **high entropy**; that's a high level of disorder (meaning low level of purity). Entropy is measured between 0 and 1. (Depending on the number of classes in your dataset, entropy can be greater than 1 but it means the same thing, a very high level of disorder.

Mathematical Workings of Entropy

To calculate $log_2(x)$ simply calculate log (x) / log (2).
$log_2(0.3)$ = log (0.3) / log (2) = -1.7368
$log_2(0.7)$ = log (0.7) / log (2) = -0.5145
*Therefore [-0.3 * -1.7368] + [-0.7 * -0.5145] à 0.52104 + 0.36015 = 0.88*

Pseudocode of the C5.0 Algorithm

```
function C5.0Algorithm(Data, Attributes)
    if all examples in Data belong to the same class (low entropy):
        return a leaf node with the class label
    else if Attributes is empty:
        return a leaf node with the majority class label in Data
    else (i.e high entropy):
        Select the best attribute, A, using Information Gain
        Create a decision node for A
            for each value v of A:
                Create a branch for v
                Recursively apply C5.0Algorithm function to the
                subset of Data where A = v
return the C5.0Decision tree
```

Conclusion

The innovations in technology have greatly influenced several improvements in commerce and our daily live activities. Looking at online transactions especially payment transactions, it creates avenue for frauds, and this research work focused on improving fraud detection by developing a Smart-Watch Detection System using IoT and Machine Learning technique. In this research work, we put forth fraud detection method based on Supervised Learning using Decision Tree algorithm. The application developed was able to use the algorithm to classify the present transactions based on the previous transaction history. It is recommended that all e-payment platforms integrate the Smart-Watch Detection System as developed in this research work so as to help detect payment frauds. More also bank users are advised to maintain secrecy about their account login credentials, and other sensitive details as well. This will help reduce if not eliminate completely fraudulent transactions.

References

Corp, k. (2016) "Mobile payments fraud survey report, Javelin strategy and research 2016

Panigrahi, S., Kundu, A., Sural, S. and Majumdar, A. K. (2019) "Credit card fraud detection: a fusion approach using dempstershafer theory and bayesian learning," Information Fusion, vol. 10, no. 4, pp. 354–363

Rani, S. L. (2018) Smart Banking Using IoT. Computer Science & Engineering Department Prof Ram Meghe College of Engineering and Management, Badnera Amravati, India

Bhalinder, K., Shivinder, K. (2020) Overview Of Intelligent Systems. International Journal of Computing & Business Research ISSN (Online): 2229-6166

Deutsche, B. (2020) The Use of Artificial Intelligence and Machine Learning in the Financial Sector. Directorate General Banking and Financial Supervision

Augustian, R. I., Prakhar, C., Pradip, G., Rohan, G. (2018) Secured E-Banking System using Artificial Intelligence. International Journal of Emerging Technologies in Engineering Research (IJETER) Volume 6, Issue 10, October (2018) www.ijeter.everscience.org

Alessandro, B. (2012): Fraud Detection in the banking Sector. *Amulti-agent Approach. International conference on mgt and service & science, 24-26 August, Wuhan, 1-4.*

Egu, J. (2010). The Role of Information and communication Technology (ICT) in Fraud Detection in Nigeria Banks.

Ojeigbede, F. (2000). Fraud in Banks.*A Paper Presented at the Effective Bank Institute Course Organized by FITC, Lagos.*

Jennings, N.B., Faratin, P., Normap, T.J., O'Brien, P., & Odgers, B. (2002). Autonomous agent's business process management.*International Journal of Applied Artificial Intelligence*, 14(2), 145-89.

Adekanye, F. (2008). Fraud in Banking Transactions. *The Nigerian Bankers, 6(1), 7-15.*

Akoroda, G.C.O. (2004). Frauds and Forgeries.*WAJFEM Regional course on Banking Supervision, Lagos.*