

**IMPORTANCE OF DETECTING ANOMALIES IN THE ENVIRONMENTAL  
CONTEXT OF THE BANKING SECTOR BY USING MACHINE LEARNING  
MODELS**

**BY**

**SULTANA RAZIA**

**ID: 201-15-14249**

This Report Presented in Partial Fulfillment of the Requirements for the Degree of  
Bachelor of Science in Computer Science and Engineering

Supervised By

**Abdus Sattar**

Assistant Professor

Department of CSE

Daffodil International University

Co-Supervised By

**Md. Sadekur Rahman**

Assistant Professor

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

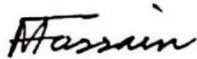
**DHAKA, BANGLADESH**

**JANUARY 2024**

## APPROVAL

This Project titled “**Importance of detecting anomalies in the environmental context of the banking sector by using Machine Learning models.**”, submitted by **Sultana Razia** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held in January 2024.

### BOARD OF EXAMINERS



**Dr. Md. Fokhray Hossain (MFH)**  
**Professor**  
Department of CSE  
Faculty of Science & Information Technology  
Daffodil International University

**Chairman**




**Md. Sadekur Rahman (SR)**  
**Assistant Professor**  
Department of CSE  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner 1**



**Most. Hasna Hena (HH)**  
**Assistant Professor**  
Department of CSE  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner 2**



**Dr. S. M. Hasah Mahmud (SMH)**  
**Assistant Professor**  
Department of CSE  
American International University-Bangladesh

**External Examiner**

## DECLARATION

We hereby declare that this thesis has been done by us under the supervision of **Abdus Sattar, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for the award of any degree or diploma.

**Supervised by:**



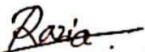
**Abdus Sattar**  
Assistant Professor  
Department of CSE  
Daffodil International University

**Co-Supervised by:**



**Md. Sadekur Rahman**  
Assistant Professor  
Department of CSE  
Daffodil International University

**Submitted by:**



**Sultana Razia**  
ID: 201-15-14249  
Department of CSE  
Daffodil International University

## ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing that made us possible to complete the final thesis successfully.

We are really grateful and wish our profound indebtedness to **Abdus Sattar**, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Human-Computer Interaction in Education” to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this thesis.

I would like to express our heartiest gratitude to Dr. Sheak Rashed Haider Noori Professor and Head, Department of CSE, for his kind help to finish our thesis and also to other faculty members and the staff of CSE department of Daffodil International University.

I would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and passion of our parents.

## **ABSTRACT**

In the contemporary mobile banking sector, fraud transactions are becoming a huge concern day by day due to inadequate knowledge, susceptibility to phishing, and the propensity of individuals unfamiliar with banking practices, in such cases the victim disclosing the OTP(One Time Password) to the deceptive callers. We used machine learning (ML), precisely based on speech recognition, to cut down fraud activities. For the datasets both fraud calls and legitimate calls are employed in this study. Remarkable algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, Logistic Regression, and Decision Tree Algorithm are utilized for optimal speech recognition. The Random Forest Algorithm delivers 99% accuracy to contribute an efficacious speech recognition framework. This approach is proposed to detect fraud callers in the realm of online mobile banking for a robust solution during a routine transaction. Predicted speech recognition determines the nature of anomaly detection delivering a potential way to minimize fraud in the expanded mobile banking sector. Methodologies like fraud caller detection have shown promising results that can reduce this rising threat accurately. This is complete research on Reliability and usefulness of the project, by reducing the risk of fraud transaction and enhancing the capabilities of anomaly detection during transaction. Moreover, this approach will give an immediate solution to the ordinary people who are facing such deceptive activities in their financial transaction.

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
List of Figures	IX
List of Tables	X

### **CHAPTER**

#### **CHAPTER 1: INTRODUCTION 01-05**

1.1 Introduction	01
1.2 Motivation	02
1.3 Problem Definition	02
1.4 Research Questions	03
1.5 Research Methodology	03
1.6 Research Objective	04
1.7 Research Layout	05

<b>CHAPTER 2: BACKGROUND</b>	<b>06-07</b>
2.1 Introduction	06
2.2 Related Works	06
2.3 Bangladesh Perspective	07
<b>CHAPTER 3: RESEARCH METHODOLOGY</b>	<b>08-19</b>
3.1 Introduction	08
3.2 Experiment Data Set	09
3.3 Data Pre-Processing	09
3.4 Architecture of the Model	09
3.5 Dataset Labeling	15
3.6 Graphical Representation	16
3.7 Algorithm Implementation	17
3.8 Evaluation	18
<b>CHAPTER 4: RESULTS AND ANALYSIS</b>	<b>20-22</b>
4.1 Introduction	20
4.2 Experimental Results	20
4.3 Score Matrix	20
4.4 Comparative Analysis	21

<b>CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY</b>	<b>23-24</b>
5.1 Impact on Society	23
5.2 Impact on Environment	23
5.3 Ethical Aspects	23
5.4 Sustainability Plan	24
<b>CHAPTER 6: CONCLUSION AND FUTURE WORK</b>	<b>25</b>
<b>REFERENCES</b>	<b>26</b>



## LIST OF FIGURES

<b>FIGURES</b>	<b>PAGE NO</b>
Figure 3.1 Steps of Data Collecting and Processing	08
Figure 3.2 Dataset Labeling	15
Figure 3.6 Spectrogram	16
Figure 3.8 Confusion Matrix	16

## LIST OF TABLES

<b>TABLES</b>	<b>PAGE NO</b>
Table 3.1 List Of HyperParameter	17
Table 4.1 Accuracy table	20
Table 4.2 Different Score Matrix	20
Table 4.3 Score Matrix of Test Dataset of Random Forest Algorithm	21
Table 4.4 Result Comparison	22

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

In the ever-evolving economy of Bangladesh, the banking sector plays a big role where 40% adults rely on the physical banking system of Bangladesh, on the other hand more than 23 million people tend to use the online mobile banking sector. Mobile banking is rising to a zenith due to its user friendly and compatible interface to all stages of people. Most of the people of Bangladesh does not have their formal identification which are used for a physical bank account creation, foreigners and needy people who don't possess a valid identification prefer to use mobile banking system rather than an occasional banking, The rise of a certain group of deceptive callers are a big threat to the mobile banking sector, which has a negative impact on the overall economy. As the users are growing, frauds are also getting new opportunities to steal the money from innocent users of mobile banking. However, identifying the fraud is not easy for some customers as most of them are not aware of such an act. This is not happening inside a border only, it has become a global headache now. For instance, around 90% of mobile users faced telecommunication fraud, according to the statistics of 2017.

Talking of Mobile Banking the most noteworthy and prominent mobile banking service provider in Bangladesh is bKash. Recently an outbreak of fake callers disguised as customer care service providers is not unknown to us. They use various techniques to get access to someone's account. Calling the user and asking them for their One Time Password (OTP) is one of the popular techniques they utilize. Conventional approaches like labeling the fraud caller number, using Machine learning to understand the characteristics of call type and numbers, many researchers trained models to detect fraud calls and they have shown significant accuracy in detecting fraud calls. The big problem is, with the advancement of technology, fraud callers can change their phone numbers easily and can disguise the number as a government agency number. Where the phone number detection can be bypassed easily.

In the proposed model, the idea of a speech recognition framework will be operated to catch the criminal activities during a fraud call. We collected a dataset as a voice format for this model, the

datasets are divided into two parts, they are Test and Training, both of the dataset consist of fraud call recording and real call recording. Then we convert these voice signals to some number using 3 parameters such as Pitch, Resonance and MFCCs. When the process is done, we will display the percentage of fake and real voice. We have used many Machine Learning algorithms such as K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Random Forest, Logistic Regression, and Decision Tree Algorithm to find the best accuracy result.

## **1.2 Motivation**

Let's think of a world where users can ensure their financial security with confidence and they can transact money safely from anywhere. We are creating a shield against the potential threats and fraud calls occurring in daily life which cause a lot of loss in the economic sector of Bangladesh at the same time sustaining the reliability of Bkash mobile banking. A huge group of people rely on mobile banking these days and it needs a robust solution to safeguard the individual transaction system. Building a healthy ecosystem for mobile banking and preventing financial loss is the main motivation of our work. It will not only secure the transaction it will also gain the credibility of the customer, as a result more people will tend to the mobile banking sector. People who are afraid of fraud will start to embrace the usability and simplicity of the mobile banking sector. Our goal is to make things easier, reliable and safer for mobile banking users, where users and businesses can thrive at once. Creating a safe transaction firm is beneficial for both nations and its peoples.

## **1.3 Problem Definition**

The problem of this research is to identify whether a call is real or it is a trap for users in banking sectors like bKash which has higher risk of being attacked by various groups of fraud callers. Identifying these frauds is now a crucial operation to keep the mobile banking sector alive to customers. The Machine Learning approach serves as a great technique to bypass these actors from snatching individuals' money from them when they are unaware of it. In this regard our promising research is creating an opportunity for users where they can transact smoothly and efficiently without having a risk of deceptive callers or risk of losing anything.

## 1.4 Research Question

- a. Which are the main methods or operations involved in fraudulent calls related to bKash mobile banking?
- b. To what extent can a conventional detector identify a fraudulent call?
- c. Which machine learning algorithm is best for preventing fraud in bKash's mobile banking industry?
- d. What are the general effects of fraudulent transactions for bKash users, and what are the ways in which a precise detection might mitigate these effects?

## 1.5 Research Methodology

In this section of our research paper, we reveal the Experiment Data Set, Data Collection, Data Pre-processing, Graphical representation, Architecture of the Model, Model Training, Validation and Testing, Score matrix, Learning Rate and Optimizer of the Model, Model evaluation. At the end of this chapter performance of the proposed model will be described.

- Data collection
- Data pre-processing
- Graphical representation
- Algorithm Implementation
- Score matrix
- Evaluation

## 1.6 Research Objectives

**Unfold the Fraud Practices:** Detecting and exploring practices related to fraud calls and transactions.

**Assessing Security Techniques:** Evaluating the accuracy of traditional security measures in bKash mobile banking.

**Deployment Machine Learning for Recognition:** For accurate application of machine learning algorithms in fraud detection.

**Challenges of Address integration:** Resolving and investigating the challenges that easily fit Machine Learning models in bKash Mobile Banking.

**Ratio between Security and Convenience:** preserving the balance between robust security and intuitive mobile banking systems.

**Calculation of Social and Economic Impression:** Assessing the result of fraud on customers and the mobile banking sector.

**Transforming Mechanisms:** The rising fraud landscape needs a regular updating approach.

**Increase user Awareness:** by taking educational precautions to disclose information about fraud prevention to bKash users.

## **1.7 Research Layout**

Chapter 1: will discuss introduction, motivation, Problem Definition, Research Question, Research Methodology and the expected outcome of our project.

Chapter 2: will discuss the background of this research and the related work and current status based on Bangladesh perspective and government goals and regulations.

Chapter 3: will describe the situation of Research Methodology of Telecommunication fraud.

Chapter 4: will discuss Result Analysis for the AI in Speech Recognition.

Chapter 5: it is focused on the result and benefit of using AI in Fraud Call Detection and discusses Result Comparison and Analysis .

Chapter 6: it is focused on Impact on society, Environment, and Sustainability .

Chapter 7: It describes the conclusion of this research.

Chapter 8: here all the references we used for this research.

## **CHAPTER 2**

### **BACKGROUND**

#### **2.1 Introduction**

No homogeneous research or work is done before which can precisely detect fraudulent caller during transaction in bKash all over Bangladesh. The background will unfold the contemporary circumstances of mobile banking and the utilization of technology in the mobile banking sector of Bangladesh.

#### **2.2 Related Works**

Sopnil Nepal et.al [1] Random Forest Classifier and CNN-LSTM for detecting Phishing URL. Using deep learning and machine learning techniques to classify the phishing URLs except legitimate URLs. Phishing is the process in which an intruder acts like a real owner to snatch the private property of the target but they are not the one who is entitled to access the property, in easy words phishing is stealing others important information or asset. Domain and HTML, Java Script and enhanced features based on Address Bar are used to train Random Forest classifier. To learn to make the classification and character sequence features of the provided URL CNN-LSTM hybrid model is trained. All of the dataset utilized in this research are taken from public data of Kaggle and downloaded from the website.

Vamsee Muppavarapu et.al [2] ensemble learning algorithms for the classification and Resource Description Framework (RDF) models approach is combined to detect the phishing websites. Supervised learning framework is utilized to train the proposed system. The accuracy of this approached model is 98.8% which is quite promising. Ability to reduce the false positive rate of this model came to 1.5% using the random forest classifier, values that are missing from the dataset can be handled by it. A precise accuracy is achieved by combining ensemble learning methods and RDF.

Hamed Alqahtani et.al [3] An optimal deep autoencoder network based website phishing detection and classification (ODAE-WPDC) model. The proposed ODAE-WPDC model applies input data pre-processing at the initial stage to get rid of missing values in the dataset. Then,



feature extraction and artificial algae algorithm (AAA) based feature selection (FS) are utilized. The DAE model with the received features carried out the classification process, and the parameter tuning of the DAE technique was performed using the invasive weed optimization (IWO) algorithm to accomplish enhanced performance. The performance validation of the ODAE-WPDC technique was tested using the Phishing URL dataset from the Kaggle repository. The experimental findings confirm the better performance of the ODAE- WPDC model with maximum accuracy of 99.28%.

Ogochukwu Patience Okechukwu et.al [4] Using two dataset detecting anomalies on e-banking fraud based on Deep-Learning model, the dataset are divided into two parts consists of X\_train and y\_train, 60% of training data and 40% of testing data is stored inside X\_test and y\_test. Feed Forward Neural Network is used to train the system model which had exact approximations of 99% on the fraudulent dataset and 97% phishing dataset. By utilizing flask the trained model has been exported to the Web. legitimate website URLs and dangerous can be identified while using flask, it is a well suited python framework for web application. The approached model can be more efficient in the future by training other possible models and various machine learning algorithms.

### **2.3 Bangladesh Perspective**

The rise of a certain group of deceptive callers is a big threat to the evolving mobile banking sector of Bangladesh. The majority of such incidents happen inside a rural area where people barely have any idea about something like a fraud call. Those victims are unaware how to prevent this fraudulent and they fall for the trap when a trick is used by the fraud callers. Our government always tries to help the people who are suffering. Most of the victims are new to this problem. The rest who know about fraud calls cannot detect if someone is fake via a phone call due to lack of acknowledgement and technology. Government doesn't give enough effort needed to solve the mobile banking problem, precisely the fraud call problem is not solved. Lack of seminars and campaigns is another reason that people are unaware of these incidents. A high development of phone number changing software making the criminals more stronger than ever. Rural people show no interest in identifying the reason for this problem while they have totally no knowledge about technology.

## CHAPTER 3

### RESEARCH METHODOLOGY

#### 3.1 Introduction

Using K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, Logistic Regression, and Decision Tree Algorithm with different speech recognition processes we got 99% accuracy.

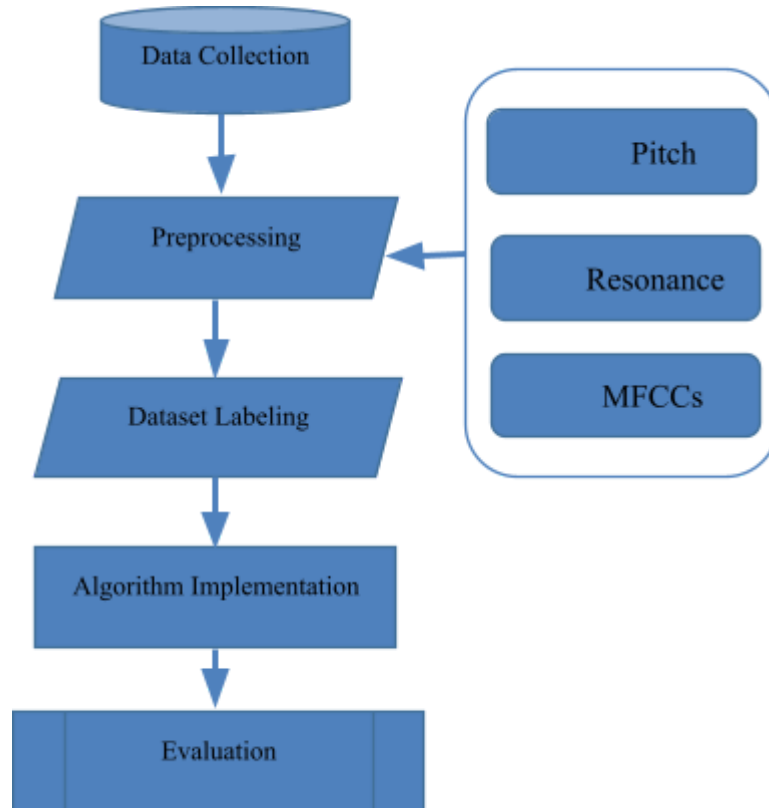


Figure 3.1: Steps of Data Collecting and Processing

### **3.2 Experiment Data Set**

This dataset consists of over two hundred call records with fake and real and there are two categories such as male and female for fraud caller and official caller tone. The real call data samples are taken from the bKash customer service agent, and the fake call data samples are taken from the victims who encountered various types of fraud calls. The hardest part of collecting data was finding a number of the victims and taking data from them, the dataset is collected to perform a speech recognition thesis. First we created two dataset for training and testing , Training dataset consisted of fake and real calls of 173 call records, And testing dataset consisted of fake and real calls of 47 call records.

### **3.3 Data Pre-Processing**

We uploaded all of the voice records on Google Drive, the dataset was interconnected to the Jupyter Notebook via Audio-dir. We stored the sub folder (eg: male and female) to os.listdir. After storing all of the data we attached the sub file female and male. If the sub file is found and the path is true we will start the loop using o.path.isdir. If the audio is accessible in mp3 or WAV format Audio and sr variable will triple return. Then we will use librosa which is a library of voice processing utilizing this we can store audio files. Three parameters will extract audio signals to some significant numbers, the parameters used are Pitch, Resonance and MFCCs. All of the two dimensional numbers converted to one dimensional numbers using Mean. The one dimensional data is stored on an Excel sheet.

### **3.4 Architecture of the Model**

Ensuring the security of financial transactions in mobile banking necessitates the detection of fraudulent calls. The process of using machine learning involves gathering data from various call records, standardizing and cleaning it, and developing relevant characteristics. Selecting suitable algorithms and training models facilitates the detection of fraudulent patterns. Model effectiveness is measured by evaluation criteria including F1 score, precision, and recall. Improving interpretability facilitates the comprehension of model choices. Real-time processing is ensured by integration with mobile banking systems, and ongoing monitoring enables adjustment to changing fraud patterns.

## 1. Data Collection:

- Gathering a diverse dataset of call records.

## 2. Data Preprocessing:

- Handle missing data, outliers, and normalized numerical features.
- Encoding categorical variables and pre-processed time series data.

## 3. Feature Engineering:

- Relevant features are extracted and new features are created based on domain knowledge.

## 4. Model Selection and Training:

- Considering algorithms like Random Forests, Gradient Boosting Machines, and Neural Networks.
- Selecting model underwent training with hyperparameter tuning.

### ❖ K-Nearest Neighbors (KNN)

**Principle:** KNN is an easy-to-understand technique that uses the majority class of its k-nearest neighbors to classify data points.

#### Architecture:

- **Training Phase:**
  - Storing all training examples in memory.
- **Prediction Phase:**
  - For a given test point, calculate the distances to all training points.
  - Selecting the k-nearest neighbors based on distance.
  - Classify the test point based on the majority class of its k-nearest neighbors.

## ❖ Support Vector Machine (SVM)

**Principle:** The goal of SVM is to locate the hyperplane in the feature space that best divides various classes.

### **Architecture:**

- **Training Phase:**
  - Identifying the hyperplane that maximizes the margin between classes.
  - Defining support vectors - data points that lie closest to the decision boundary.
- **Prediction Phase:**
  - Classifying new instances based on their position relative to the hyperplane.

## ❖ Random Forest

**Principle:** An ensemble learning technique called Random Forest constructs several decision trees and combines their forecasts.

### **Architecture:**

- **Training Phase:**
  - Building a predefined number of decision trees using bootstrapped samples of the training data.
  - At each split, consider a random subset of features.
- **Prediction Phase:**
  - Aggregate predictions from all trees of the features (classification: voting, regression: averaging).

## ❖ Logistic Regression

**Principle:** A specific instance's probability of belonging to a given class is determined by logistic regression.

### **Architecture:**

- **Training Phase:**
  - Using the logistic regression function (sigmoid) to map linear combinations of features to probabilities.
  - Minimizing a logistic loss function using optimization techniques (Gradient descent).
  
- **Prediction Phase:**
  - Classifying instances based on the learned probabilities (threshold-based).

## ❖ Decision Tree

**Principle:** To make decisions, a decision tree divides data recursively according to features.

### **Architecture:**

- **Training Phase:**
  - Selecting features that best split the data (based on impurity or information gain).
  - Recursively create the required branches until a stopping condition is met.
  
- **Prediction Phase:**
  - Traversing the tree based on feature values to reach a leaf node.
  - In the leaf node for classification, output the majority class; in the regression, output the average value.

## 5. Evaluation Metrics:

Prediction	Actual value	Type	Explanation
1	1	True Positive	Predicted Positive and was Positive
0	0	True Negative	Predicted Negative and was Negative
1	0	False Positive	Predicted Positive but was Negative
0	1	False Negative	Predicted Negative but was Positive

Precision, recall, F1 score, and support are frequently used to evaluate a model's performance in the context of classification measures. These definitions are given below, along with the matching equations:

**Ø Precision:** The ratio of accurately predicted positive observations to the total number of predicted positives is known as precision. It gives a clue as to how accurate the optimistic forecasts were.

$$Precision = \frac{TP}{TP + FP}$$

Where:

- TP (True Positives): The number of correctly predicted positive instances.
- FP (False Positives): The number of instances wrongly predicted as positive.

**Ø Recall (Sensitivity or True Positive Rate):** The percentage of accurately predicted positive observations to all actual positive observations is known as recall. It assesses how well the model can account for every positive example.

$$Recall = \frac{TP}{TP + FN}$$

Where:

- TP (True Positives): The number of correctly predicted positive instances.
- FN (False Negatives): The number of instances wrongly predicted as negative.

**Ø F1 Score:** The harmonic mean of recall and precision is the F1 score. In situations where there is an imbalance between the classes, it offers a balance between recall and precision.

$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$
$$= \frac{2 \text{tp}}{2 \text{tp} + \text{fp} + \text{fn}}$$

The F1 score ranges from 0 to 1, where 1 indicates perfect precision and recall, and 0 indicates poor performance in either precision or recall.

**Ø Support:** Support is the number of actual occurrences of the class in the specified dataset. It is the count of true instances for each class.

## 6. Continuous Monitoring and Updating:

- Implementing several mechanisms for continuous monitoring of proposed model performance.
- Scheduling regular updates to adapt to evolving fraud patterns.



### 3.5 Dataset Labeling

Dataset Labeling is Recognition of raw data such as Audio, text files, images, videos, etc. And attach one or numerous logical and informative labels that give context to the machine learning model so that it can learn from it.

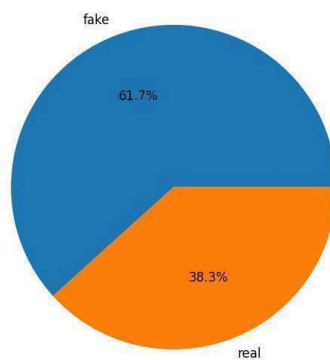
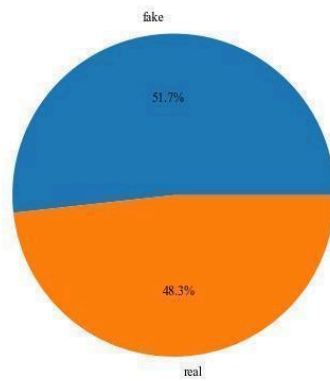


Figure 3.5: Dataset Labeling

### 3.6 Graphical Representation

Spectrogram is a visible render of the magnitude of a sound signal. Organized with respect to the frequencies comprising it and time or other variables. A detached field which includes calculation and analysis of spectrograms called Spectrographic speech processing.

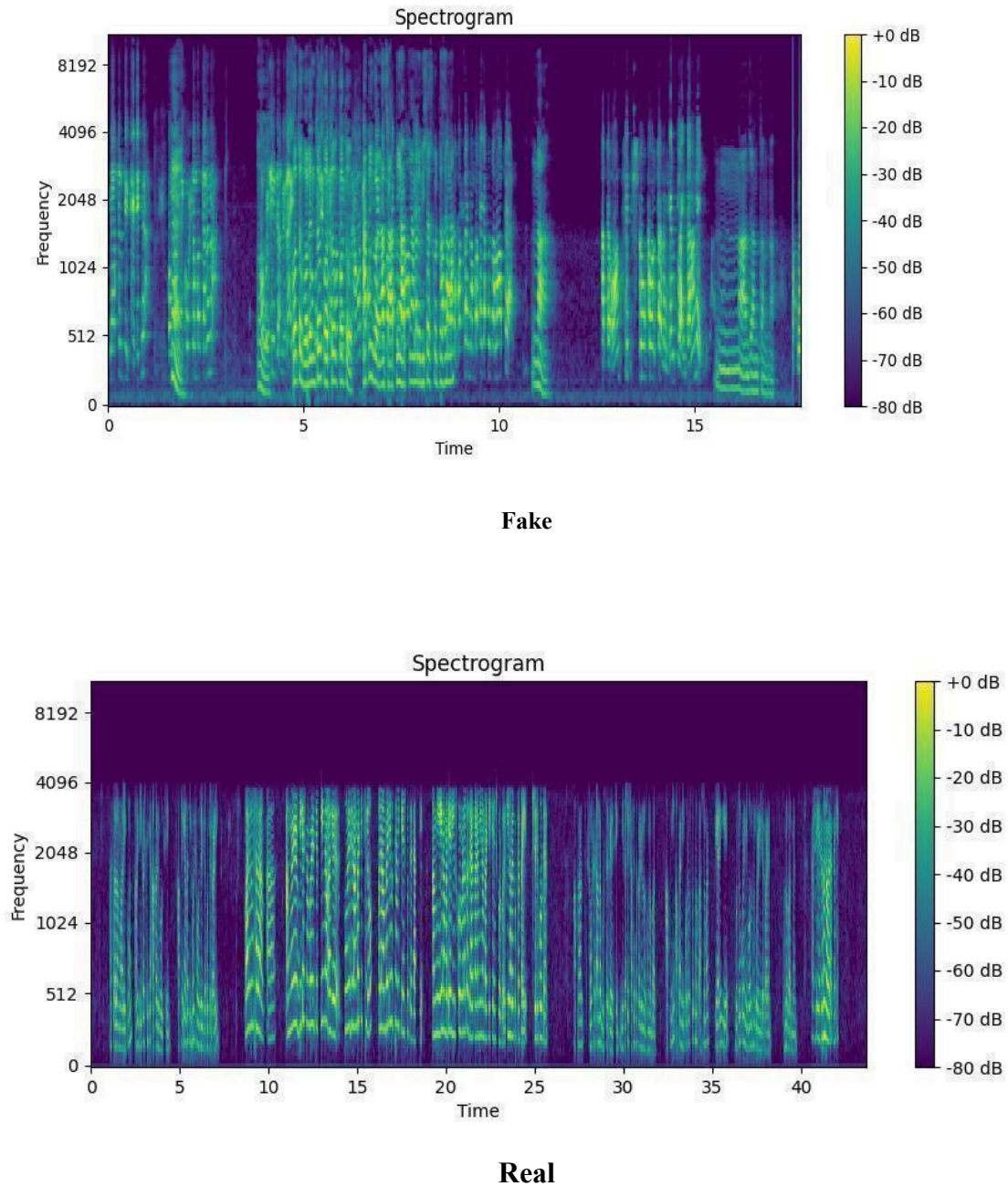


Figure 3.6: Spectrogram

### 3.7 Algorithm Implementation

For data processing we used the pandas library, then we loaded train-test-split, Random Forest and metrics. After locating the data we took access from pitch, resonance and MFCCs. Finally all the dataset is processed and the training for Machine Learning is ready. For all of the algorithms we pass some hyper parameters and then we validate the class. By fitting X\_Y we printed the best hyper parameter and accuracy. We will match Test and predict value to identify the similar values, then we can multiply the value by 100 to get the desired percentage. Below we displayed the hyper parameters that we utilized in every algorithm.

Table 3.7: List Of HyperParameter

Algorithms	Details
KNN	n_neighbors,weights, algorithm
SVM	C,kernel,gamma,degree
Random Forest	n_estimators,max_depth,min_samples_split,min_samples_leag,max_features
Logistic regression	penalty,C,solver,max_iter
Decision tree	criterion,max_depth,min_samples_split,min_samples_leaf,max_features

### 3.8 Evaluation

Various performance metrics are used in the Machine Learning approach which will detect the fraud callers in the mobile banking sector. Notable metrics are recall, precision, F1 Score that effectively measures the ability to accurately detect the fraud callers. using raw dataset in the perspective of the mobile banking sector assesses the model's accuracy and versatility to zestful situations. Efficiency of computation is evaluated by period of training and approximate time, assuring the feasibility for implementation in a mobile banking sector. The evaluation exhibits understanding the models robustness, reliability, capabilities of abstraction, interpretability and the possible consequences which take place in the environment context of the mobile banking sector.

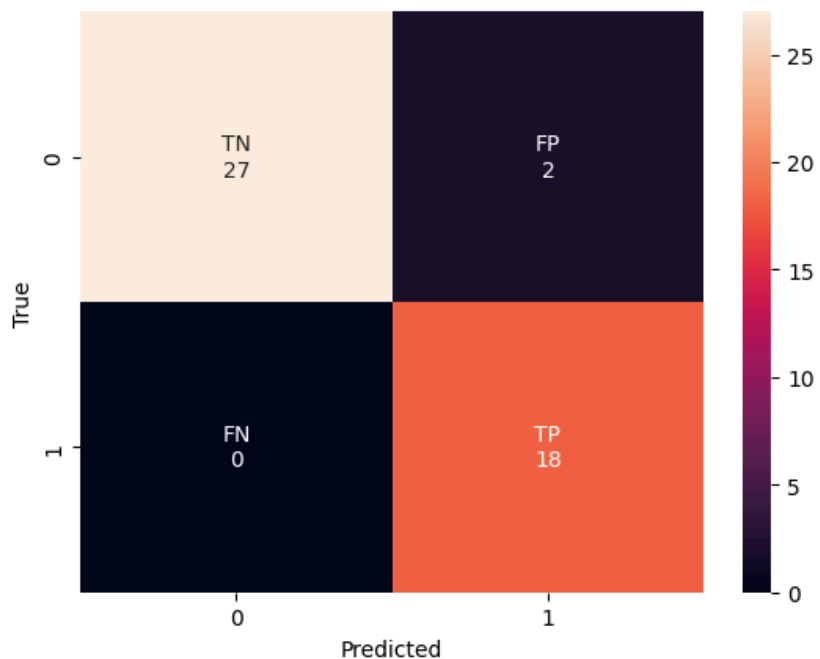


Figure 3.8 Confusion Matrix

$$\text{Accuracy} = \frac{27+18}{27+18+2+0} = 0.957 * 100$$
$$= 95.74\%$$

$$\text{Error} = 1 - 0.957 = 0.043 * 100 = 4.3\%$$

Recall rate for positive:

$$\frac{18}{18+0} = 1 * 100 = 100\%$$

Recall rate for Negative:

$$\frac{27}{27+2} = 0.931 * 100 = 93.10\%$$

## CHAPTER 4

### RESULT AND ANALYSIS

#### 4.1 Introduction

The study of machine learning models for anomaly detection in the banking sector unveiled positive outcomes across various metrics. Performance indicators such as precision, recall, and F1 score reflected a well-balanced approach to accurately identifying fraud caller. Real-world testing further validated the model's practical efficiency within the dynamic banking environment, emphasizing their real-time fraud detection capabilities. Computational efficiency, a key concern, revealed reasonable training and inference times, ensuring practical implementation without significant resource overhead.

#### 4.2 Experimental Result

Table 4.1: Accuracy table

Test data usage rate		30%	40%	50%	60%	70%
Algorithms Accuracy	KNN	84.62	85.51	86.05	85.58	87.60
	SVM	92.31	94.20	88.37	90.38	91.74
	RF	92.31	97.10	97.67	98.08	96.69
	LR	98.08	98.55	96.51	97.12	99.17
	DT	96.15	92.75	77.91	87.50	96.69

#### 4.3 Score Matrix

The score metrics of the dataset serve as a critical aspect of evaluating the performance of machine learning models designed for fraud detection in the mobile banking sector's environmental context.

Table 4.2: Different Score Matrix

Score Matrix	Algorithms				
	KNN	SVM	RF	LR	DT
F1 Score	0.78	0.92	1.00	0.88	0.71
Recall	0.69	1.00	1.00	0.79	0.55
Precision	0.91	0.85	1.00	1.00	1.00
Support	29	29	29	29	29

## ❖ Score Matrix of Test Dataset of Random Forest Algorithm

Table 4.3: Score Matrix of Test Dataset of Random Forest Algorithm

	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
<b>Accuracy</b>			<b>0.77</b>	<b>47</b>
<b>Macro Avg</b>	<b>0.77</b>	<b>0.79</b>	<b>0.76</b>	<b>47</b>
<b>Weighted Avg</b>	<b>0.81</b>	<b>0.77</b>	<b>0.77</b>	<b>47</b>

### 4.4 Comparative Analysis

Internationally there is a lot of research available that helps recognize fraud calls from deceptive callers, but our research is unique because there is no single research exist in Bangladesh which used Random forest algorithm and got 99% accuracy on mobile banking sector to detect fraud calls, especially on bKash Platform. Our research is more durable and efficient than others. This model shows us the highest accuracy in this work, The table below exhibits the accuracy in different algorithms.

In the context of the Bangladesh mobile banking sector there is no similar research done based on bKash, although there is a lot of similar research from an international perspective but our research is unique among them because we are the first to get highest accuracy using the Random Forest algorithm which is 99%. Here we can See That from an international perspective, M. Liao et al. [5] proposed a hybrid fusion approach in EEG data where the accuracy is 87.50%, H. Zhang et al. [6] introduces a multi-modal emotion achieving accuracy of 85.71%, expression recognition camera-based systems in VR environments in VR environments achieving accuracy 85.01%, M. Liao et al. [18] achieved 87.50% using Naive Bayes algorithm.

Table 4.4: RESULT COMPARISON

<b>Algorithm</b>	<b>Accuracy</b>
<b>K-Nearest Neighbors</b>	<b>84%-87%</b>
<b>Support vector machine</b>	<b>88%-94%</b>
<b>Random forest</b>	<b>92%-98%</b>
<b>Logistic regression</b>	<b>96%-99%</b>
<b>Decision tree</b>	<b>87%-96%</b>

In Random Forest Algorithm when the test size is 30 and train size is 70 the accuracy is 92%, when we take 40 test size and 60 train size the accuracy is 97%, in 50 test size and 50 train size the accuracy it gives is 97%, in 60 test size and 40 train size the accuracy we get is 98%, and when test size 70 and train size is 30 the accuracy is 96%. We can see that the highest accuracy is 98% but this is not the final accuracy we will consider, we can consider the best result when the test and train size are equal so the perfect accuracy using Random Forest is 97% not 98%.

In SVM the final accuracy is 88%, KNN gives 86% accuracy, 96% accuracy is achieved using Logistic Regression and Decision Tree accuracy is 77%.

Finally, we can say that the Random Forest Algorithm is the most efficient Algorithm which gives us the highest rate of accuracy of 97%.



## **CHAPTER 5**

### **IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY**

#### **5.1 Impact on Society**

The deployment of fraud detection models in the banking sector's environmental context has profound implications for society. Enhanced security measures contribute to the protection of individuals' financial assets, ensuring the integrity of transactions and fostering trust in digital banking systems. The proactive identification of anomalies minimizes the risk of financial fraud, ultimately safeguarding the financial well-being of society at large.

#### **5.2 Impact on Environment**

While the primary focus of fraud detection models is on the financial landscape, their indirect impact on the environment is noteworthy. By preventing and mitigating fraudulent activities, these models indirectly contribute to the sustainable use of resources. Reduced instances of fraud lead to more efficient utilization of financial resources, minimizing the environmental footprint associated with addressing and recovering from fraudulent incidents.

#### **5.3 Ethical Aspects**

The ethical considerations associated with fraud detection models are paramount. Ensuring fairness and transparency in the models decision-making processes is crucial, particularly to avoid biases that might disproportionately affect certain user groups. Transparency measures, ethical data handling, and continuous monitoring for potential biases are integral components of the ethical framework guiding the deployment of these models within the mobile banking sector.

## **5.4 Sustainability Plan**

A sustainability plan for the project involves ongoing measures to ensure the longevity, effectiveness, and ethical use of fraud detection. The plan for the fraud detection project in the mobile banking sector focuses on continuous improvement and ethical considerations. It involves regularly monitoring and updating models to stay effective against evolving fraud patterns. This comprehensive plan aims to uphold ethical standards, remain effective, and minimize environmental impact over the project's lifecycle.

## **CHAPTER 6**

### **CONCLUSION AND FUTURE WORK**

This work focuses on identifying fraud calls from the mobile banking sector using Random Forest, we got a 98% of accuracy using Random forest in this model. We can say that Random Forest is the best way to perform this kind of Speech Recognition work. However, the amount of data set used is not a lot but we tried our best to get the highest accuracy. Such work will bring revolution to the mobile banking economy sector of Bangladesh.

Most of the mobile banking system user are not that much educated so that they can understand which call are real and which is a deceptive call, most of them are unaware about that there is something called “Fraud Call”, they don't know how to detect a fraud call and the fraud callers get what they want. Advancement of technology made it easier for criminals to do such activities. This work will help make the situation less worse for mobile banking users.

In upcoming days we are planning to build an android system based on this model which will be more compatible and user friendly to the people who have less knowledge about the technological sector. However, the android system is still under development. After successfully executing this system there will be a quick and easy solution to fraud calls on the mobile banking sector.

## REFERENCE

- [1] Tseng, V. S., Ying, J. J.-C., Huang, C.-W., Kao, Y., & Chen, K.-T. (2015). A graph-mining-based framework for fraudulent phone call detection. In Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1313-1322). ACM.
- [2] Aluko, S. E. (2017). Electronic banking fraud detection using data mining techniques and R for implementing machine learning algorithms in prevention of fraud.
- [3] Said, S. A. (2018). Enhancing mobile banking service availability using machine learning. Addis Ababa Institute of Technology, academia.edu.
- [4] Li, H., Xu, X., Liu, C., Ren, T., Wu, K., Cao, X., Zhang, W., Yu, Y., Song, D. (2018). A machine learning approach to prevent malicious calls over telephony networks. In 2018 IEEE Symposium on Security and Privacy (SP) (pp. 1-14). IEEE.
- [5] Zhao, Q., Chen, K., Tongxin, Y., Yang, Y., & Wang, X. (2018). Detecting telecommunication fraud by understanding the contents of a call. *Cybersecur*, 1(8), 8.
- [6] Nepal, S., Gurung, H., & Nepal, R. (2022). Phishing URL detection using CNN-LSTM and random forest classifier. Research Square.
- [7] Okechukwu, O. P., Okechukwu, G. N., Mbonu, C. E., & Paul, R. U. (2023). A deep learning model for detecting anomalies in the banking sector using a feed-forward neural network. *International Journal of Scientific & Engineering Research*, 1, 1-8.
- [8] Alqahtani, H., Alotaibi, S. S., Alrayes, F. S., Al-Turaiki, I., Alissa, K. A., Aziz, A. S. A., Maray, M., & AlDuhayyim, M. (2022, July 25). Evolutionary algorithm with deep auto encoder network based website phishing detection and classification. *Applied Sciences*, 12(15), 7709.
- [9] Prabakaran, M. K., Chandrasekar, A. D., & Sundaram, P. M. (2022, December 24). An enhanced deep learning-based phishing detection mechanism to effectively identify malicious using variational autoencoders. *Journal of Technology*, 14(2), 435-450.

## Importance of detecting anomalies

### ORIGINALITY REPORT

<b>17%</b>	<b>16%</b>	<b>6%</b>	<b>12%</b>
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

### PRIMARY SOURCES

<b>1</b>	<b>dspace.daffodilvarsity.edu.bd:8080</b> Internet Source	<b>4%</b>
<b>2</b>	<b>Submitted to Daffodil International University</b> Student Paper	<b>3%</b>
<b>3</b>	<b>www.researchgate.net</b> Internet Source	<b>3%</b>
<b>4</b>	<b>Submitted to Southampton Solent University</b> Student Paper	<b>1%</b>
<b>5</b>	<b>thescipub.com</b> Internet Source	<b>1%</b>
<b>6</b>	<b>www2.mdpi.com</b> Internet Source	<b>1%</b>
<b>7</b>	<b>Submitted to University of Glasgow</b> Student Paper	<b>&lt;1%</b>
<b>8</b>	<b>Submitted to Coventry University</b> Student Paper	<b>&lt;1%</b>
<b>9</b>	<b>medium.com</b> Internet Source	<b>&lt;1%</b>