RESEARCH METHODOLOGY

INTRODUCTION

Anushka Kavli - 86092300003

Anil Vhatkar - 86092300004

Abdul Awwal Choudhari - 86092300026

Charul Sankhe - 86092300011

Om Muddebihal - 86092300033

Yash Dudeja - 86092300013

GROUP - 7

TOPIC:

FRAUD DETECTION

Why ?

Choosing fraud detection as a topic for research project can be relevant and compelling. As we know fraud is a significant problem that affects not only individuals but companies, government also and as technology is advancing fraudsters have become more advance in making a fraud and giving a challenge to detect the fraud.

Research Types

Analytical Research

The research is to know how frauds are been done.

Applied Research

The research is to know more about fraud detection and minimize fraud risks .

Quantative Research

The research takes a measurement of data .

Research Types *(contd)*

Empirical Research

The research relies on some previous data .

One-time research

Research is confined to a single time-period,

Literature Citation

1) Anushka Kavli - 86092300003

Cite : Fanai, H., & Abbasimehr, H. (2023). A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. *Expert Systems with Applications*, *217*, 119562.

The above-mentioned research paper demonstrates *Automatic Fraud Detection Systems* which financial institutions must make use of, for reducing the risk of fraudulent transactions.

Such automated systems are very useful for institutions issuing credit cards or managing online transactions.

A Binary Classification Model is built to know whether the transactions are normal or fraudulent.

The model uses Deep Autoencoder with supervised deep learning methods combined with unsupervised neural network thus making the model accurate to detect fraud.

After testing and comparing the model with trained data set’s model, the model created by Deep Autoencoder showed better results.

2) Anil Vhatkar - 86092300004

Credit Card Fraud Detection

Fraud detection is the process of using tools and procedures to prevent the theft of money, information, and assets. It is a security barrier that protects against various forms of fraud, including minor infractions and felony crimes

Cite : Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 international conference on computing networking and informatics (ICCNI)* (pp. 1-9). IEEE.

The above research paper investigated the performance of 3 machine learning algorithms for credit card fraud detection. They found that k-nearest neighbor performed the best, with an accuracy of 97.69%, followed by naïve bayes with 97.92% and logistic regression with 54.86%. The authors concluded that k-nearest neighbor is a good choice for credit card fraud detection, especially when the data is highly skewed.

3) Charul Sankhe - 86092300011

Cite : Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 international conference on computing networking and informatics (ICCNI)* (pp. 1-9). IEEE

Conclusion:

Research paper is all about Credit card fraud detection using machine learning techniques: A comparative analysis

Data mining had played an imperative role in the detection of credit card fraud in online transaction. Credit card fraud detection which is the data mining problem becomes challenge due to two major reasons first the profiles of normal and fraudulent behaviour change constantly and secondly credit card fraud data sets are highly skewed. The work is implemented in python. The performance of the techniques is evaluated based on accuracy, sensitivity, specificity, precision, Matthews Correlation Coefficient and balanced classification rate. The comparative result show that k-nearest neighbour performs better than naive bayes and logistic regression techniques. Using AVS and CVV checks can verify the authenticity of card not present transaction and minimize the risk of fraud. Make sure you verify that a website is authentic before making a purchase.

4) Yash Dudeja - 86092300013

A survey on statistical methods for health care fraud detection

Cite: Li, J., Huang, K. Y., Jin, J., & Shi, J. (2008).
A survey on statistical methods for health care fraud detection. Health care management science, 11, 275-287.

Summary: So this paper tells us about the fraud happening in the health care sector by saying how US lost at least 3%, or more than $60 billion of their annual expenditure due to fraud.
Making a statistical method so that can be applied for the fraud detection, using the data characteristics identifying the frauds.
Using Machine learning, Neural Networks , Fuzzy Logic comparing all the methods so that the fraud can be prevented
Based on the survey, the thing which are being missed or has been lacking in the existing research has been pinpointed in the paper

Steps Involved in identifying the Detection:

1) **Classification of the fraudulent:** So they have stated that there are three major aprties involving in the fraud which are

a)Service Providers : i)Billing services that are not actually performed

ii)Falsifying patients' diagnosis and/or treatment histories to justify tests, surgeries, or other procedures that are not medically necessary

b)Insurance Subscribers : Filing claims for medical services which are not actually received;

c)Insurance Carriers: Falsifying reimbursements & Falsifying service statements

2) **Data Preprocessing:** Where the data is been taken from the health care system and is been processed cleaned so that a statistical method can be formed from the raw data and carried out and another is to set a goal where the step is that the objective is to identify types of fraud on which detection should be focused

3) **Statistical Modelling for fraud detection:** Using supervised and unsupervised methods, neural networks and decision trees

The overall objective of the research in this area is to develop fraud detection methods and algorithms that are **Scalable, Accurate, and Fast :**

i) Scalable refers to the capability of handling the immense volume of health care data.

ii) Accurate refers to low errors

iii) Fast refers to the capability of catching frauds so there is no severe loss and damage.

Yash Dudeja -

5) Abdul Awwal Choudhari - 86092300026

Online transaction fraud

With the popularization of online shopping, transaction fraud is growing seriously. Therefore, the study on fraud detection is interesting and significant. An important way of detecting fraud is to extract the behavior profiles (BPs) of users based on their historical transaction records, and then to verify if an incoming transaction is a fraud or not in view of their BPs.

Zheng, L., Liu, G., Yan, C., & Jiang, C. (2018). Transaction fraud detection based on total order relation and behavior diversity. *IEEE Transactions on Computational Social Systems*, *5*(3), 796-806.

The above research paper states that the LGBP is a new model for fraud detection that is more flexible and accurate than Markov chain models.

6) Om Muddebihal - 86092300033

A Model Based on Convolutional Neural Network for Online Transaction Fraud Detection

Zhang, Z., Zhou, X., Zhang, X., Wang, L., & Wang, P. (2018). A model based on convolutional neural network for online transaction fraud detection. *Security and Communication Networks*, *2018*.

Conclusion:

This research paper presents a fraud detection model for online transactions. The model utilizes an innovative convolutional neural network approach. By rearranging the transaction features and e-liminating the need for complex feature engineering, the model achieves excellent performance. It exhibits notable improvements in precision and recall compared to existing methods, showing its potential in combatting online transaction fraud. Its design for structured data highlights its relevance for financial institutions.

Hypothesis

Hypothesis 1:

In the context of fraud detection, the null and alternative hypotheses are framed to test the effectiveness of a fraud detection system or method.

Here's how you can frame the null and alternative hypotheses:

Null Hypothesis:

The null hypothesis represents the default or no-effect scenario:

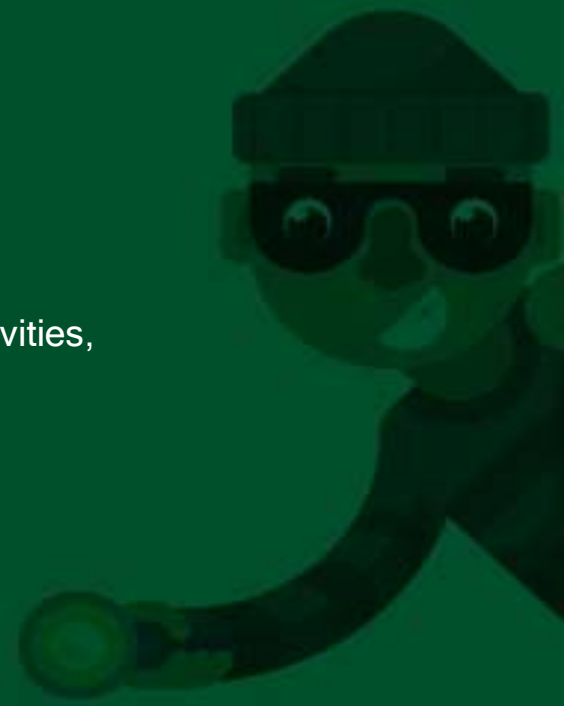
"There is no significant difference between the observed fraudulent activities and the expected level of fraudulent activities based on random chance or the current fraud detection system."

Alternative Hypothesis:

The alternative hypothesis represents what you're trying to show or demonstrate:

"There is a significant difference between the observed fraudulent activities and the expected level of fraudulent activities, indicating that the fraud detection system/method is effective in identifying and detecting fraudulent transactions."

In this you would typically be looking to reject the null hypothesis in favor of the alternative hypothesis, showing that the fraud detection system/method is indeed effective at identifying instances of fraud.

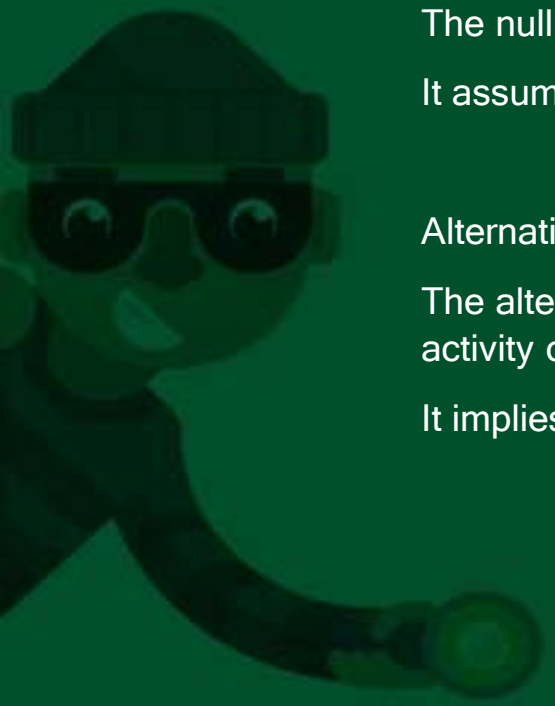


Hypothesis 2:

In the context of fraud detection, the null and alternative hypotheses are used to formulate a hypothesis test that helps determine

whether there's enough evidence to support a particular claim about fraudulent activities.

Here's how you can frame the null and alternative hypotheses:



Null Hypothesis:

The null hypothesis represents the default assumption that there is no fraudulent activity or no significant difference.

It assumes that any observed variations are because of random chance or noise.

Alternative Hypothesis:

The alternative hypothesis represents the claim you are testing, which suggests that there is indeed some form of fraudulent activity occurring.

It implies that the observed variations are not because of chance and that there's a real effect or difference.

Hypothesis 2 (contd):

Here are a couple of examples of how you could structure null and alternative hypotheses for fraud detection scenarios:

Example 1: Credit Card Fraud Detection

Null Hypothesis:

There is no significant difference between the spending patterns of legitimate transactions and potentially fraudulent transactions.

Alternative Hypothesis:

There is a significant difference between the spending patterns of legitimate transactions and potentially fraudulent transactions.

These hypotheses will form the basis for conducting statistical analyses and tests to determine whether the evidence supports rejecting the null

hypothesis in favor of the alternative hypothesis

Hypothesis 3:

In the context of fraud detection, the null and alternative hypotheses are framed if the transaction is fraudulent or not.

Here's how you can frame the null and alternative hypotheses:

Null Hypothesis:

The transaction is not fraudulent

Alternative Hypothesis:

The transaction is fraudulent

Here are some examples of hypotheses that can be used for fraud detection:

* The mean transaction amount for fraudulent transactions is higher than the mean transaction amount for non-fraudulent transactions.

* The distribution of transaction amounts for fraudulent transactions is more skewed than the distribution of transaction amounts for non- fraudulent transactions.

* Fraudulent transactions are more likely to occur during certain times of the day or week.

* Fraudulent transactions are more likely to involve certain types of products or services.

* Fraudulent transactions are more likely to be initiated by certain types of customers.

Hypothesis 4:

In the context of fraud detection, the null and alternative hypotheses are framed if the k-nearest neighbor algorithm demonstrate superior performance compared to Naive Bayes and Logistic Regression technique in the domain of fraud detection.

Here's how you can frame the null and alternative hypotheses:

Null hypothesis:

There is jo significant difference in the performance of the k-nearest neighbor, Naive Bayes and Logistic Regression:

technique in the context of fraud detection

Alternate Hypothesis:

The k- nearest neighbor algorithm exhibit significantly better performance than the Naive Bayes and Logistic Regression techniques for

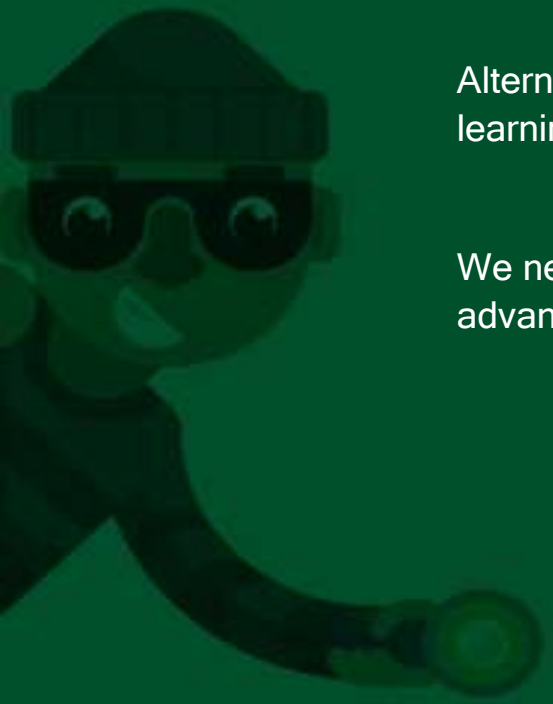
detecting instances of fraud

Hypothesis 5:

In the context of fraud detection, the null and alternative hypotheses are compared to traditional methods the involvement of advanced machine learning algorithms into fraud detection significantly enhances the accuracy in identifying and preventing fraudulent transactions.

Here's how you can frame the null and alternative hypotheses:

Null Hypothesis: There is no significant association between the traditional methods and advanced machine learning methods.



Alternative Hypothesis: There is significant association between the traditional methods and advanced machine learning methods

We need to collect the empirical data for analyzing and comparing the performance of fraud detection systems using this advanced machine learning algorithms against those using rule-based methods.

Hypothesis 6:

In the context of fraud detection, the null and alternative hypotheses are frames as the number of **credit card users** have increased and resulted in the **growth of online businesses**. So As credit card users increase.

Prose - Growth in online businesses Increase.

Cons - Frauds increases also become vital issue.

Here's how you can frame the null and alternative hypotheses:

Null hypothesis: There is no significant relationship between the usage of credit cards and online business growth.

Alternate hypothesis: There is a significant relationship between the usage of credit cards and online business growth.

Research Design

Sampling Design

Sampling Design is divided onto parts which are as follows:

-Probability :- In probability sampling, every person in the target population (either random or representative) has an equal chance of being selected for the sample.

- Non Probability :-In non-probability sampling, some individuals in the group will be more likely to be selected than others.

The method of selecting items to be observed for a given study

Observational Design

Simply observe behaviors or phenomena and record them rather than conducting an experiment.

Observing through Interviews where the conversation between researcher and respondent is able to flow more naturally

The conditions under which observations are to be made

Statistical Design

The statistical design is a method for planning and conducting experiments

How many items are to be observed and how the information and data gathered are to be analyzed

Operational Design

The technique by which procedures specified in the sampling, statistical, observational designs can be carried out

Questionnaire :-

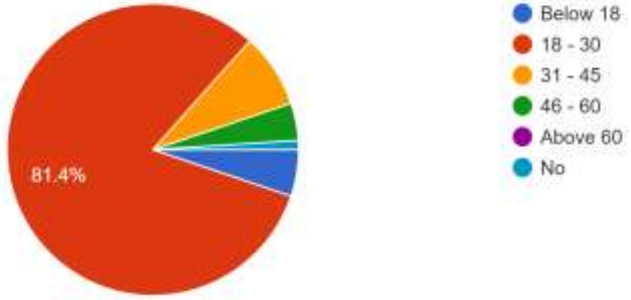
1. Your Age
   - Below 18
   - 18-30
   - 31-45
   - 46-60
   - Above 60
2. Gender
   - Male
   - Female
   - Other
3. According to you what is fraud?
   - Tricking people for money
   - Accidental mistakes in money matters
   - Honest business practices
   - Online
4. Which of the following do you think is the most common type of fraud?
   - Identity theft
   - Phishing emails/messages
   - Credit card fraud
   - Investment scams
   - Lottery scams
   - Other:

5. Do you know how to protect yourself from fraud?
    - Yes
    - No
6. Do you regularly review your bank and credit card statements for suspicious transactions?
    - Yes
    - No
7. What do you think can help prevent fraud?
    - Learning about online safety
    - Being careful with sharing personal information
    - Asking for help from adults if something seems suspicious
    - Ignoring all online messages
    - Other
8. If you saw someone doing something dishonest, do you know how to tell someone about it?
    - Yes
    - No
9. How do you think people can be encouraged to report fraud?
    - Making sure they won't get in trouble for reporting
    - Giving them a reward for reporting
    - Having a special phone number or website to report fraud
    - Talking about it in school or on TV
    - Other
10. Any other comments or suggestions about fraud prevention?

Data Analysis

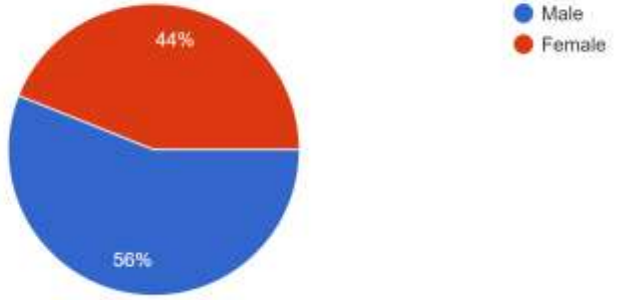
1. Your Age

97 responses



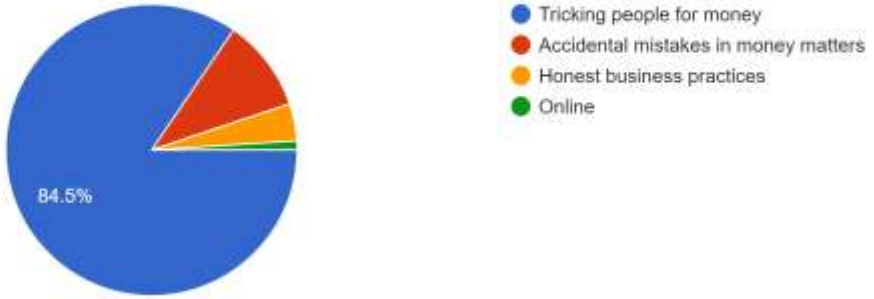
2. Gender

91 responses



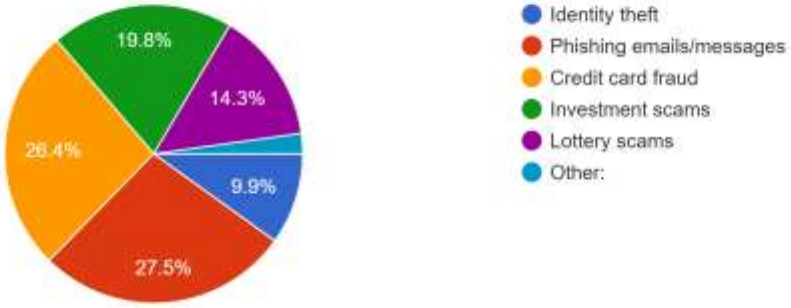
3. According to you what is fraud?

97 responses



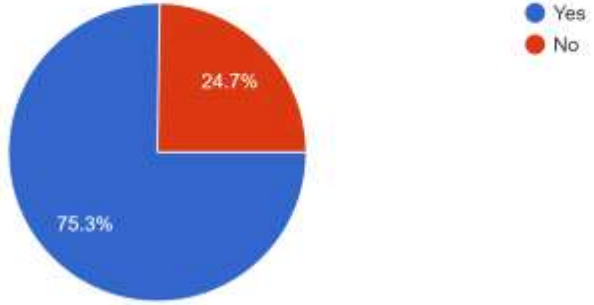
4. Which of the following do you think is the most common type of fraud?

91 responses



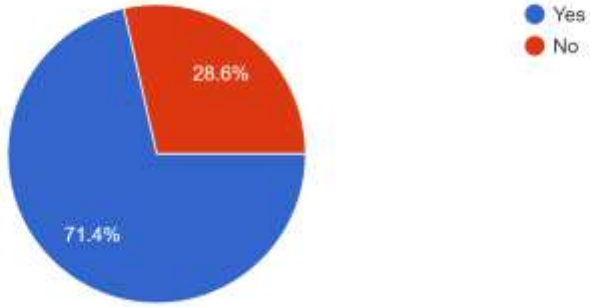
5. Do you know how to protect yourself from fraud?

97 responses



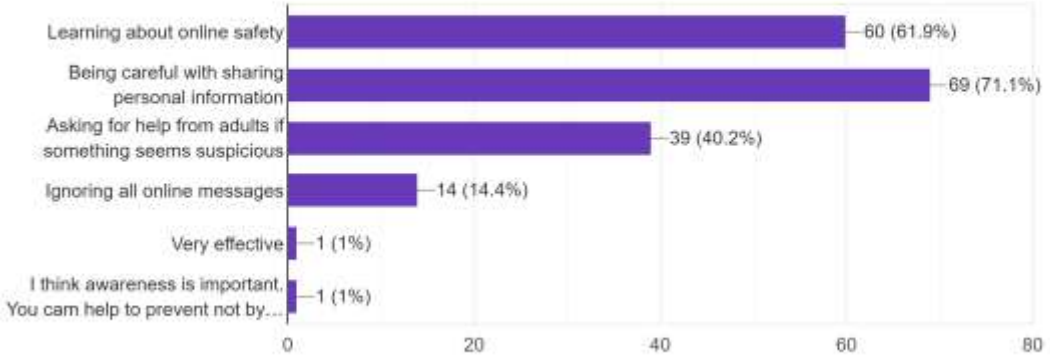
6. Do you regularly review your bank and credit card statements for suspicious transactions?

91 responses



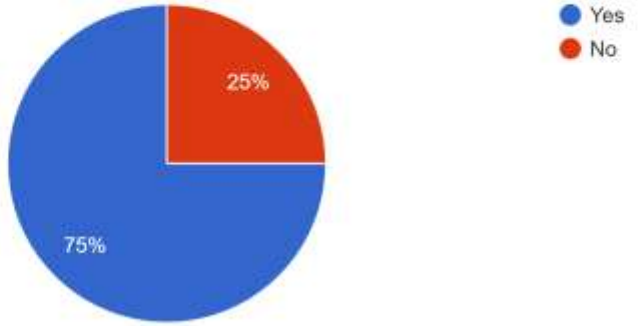
7. What do you think can help prevent fraud?

97 responses



8. If you saw someone doing something dishonest, do you know how to tell someone about it?

96 responses



9. How do you think people can be encouraged to report fraud?

96 responses

10. Any other comments or suggestions about fraud prevention?

there are generally multiple fraud cases but technical fraud it's has to very careful to we know about how to protect our data and some important things
when government also announce to all of the TV channel and social pages it's been to very helpful for everyone and all them which types of fraud there also teaching about them which kind of safety we can take it
properly
that is the my personal opinion about the fraud

Do not share anything without knowing about that thing related to financial things.

Never give out confidential information unless you know it's a trusted and verified source.

Just be cautious before putting your sensetive information

Awareness programs should be conducted for public and to educate them about fraud related matters, especially for village/old age peoples. It will also avoid them from falling in such kinds of traps.

Government should be create new landline for only fraud cases

Swdhan rahe satark rahe

Government should align with Truecaller block all the number permanently which causes Fraud

Be safe and alert

Security and privacy has to be confidential. Security should be max and priority should be given to the online bank transactions and tracking of identity thefts.

-

Stop fake accounts

Prepare short courses to avoid and how to overcome fraud

Sometimes Senior citizens get tricked very easily as they don't have much knowledge about this kind of frauds & this technology is new for them

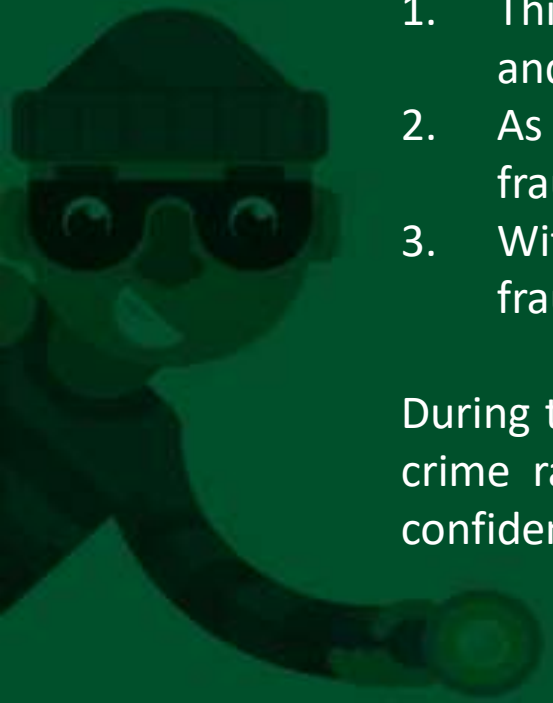
No suggestions

Conclusion :

Fraud detection is essential in various fields, including finance, healthcare, and research, to prevent financial losses, ensure data security, and maintain the integrity of information. Advanced technologies like machine learning, data analytics, and artificial intelligence have significantly enhanced fraud detection capabilities. Continuous research and collaboration among experts are vital to staying ahead of evolving fraud tactics, ensuring a safer and more secure environment in different sectors.

Choosing Fraud detection as our topic has multiple reasons

1. This topic automatically combines with our studies and helps us to learn more about our subject and have a hands on learning
2. As in todays world everything is related to data and security is must with that the chances of fraud are increasing every organization need o protect their data or any loss
3. With that we get know about currently what's happening in the world, how different types of fraud are people doing



During the research we learned that Fraud detection is essential and various fields are using it as the crime rates are increasing in the cyber sector, as it prevents Financial/Private losses as well the confidential organization data. It helps

Research Limitations:

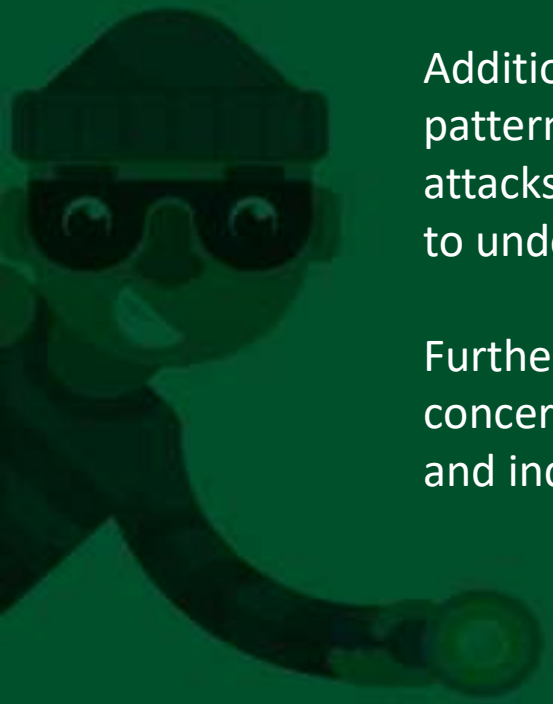
Research in fraud detection, like any other field, has its limitations. These limitations are important to acknowledge as they provide context for the findings and insights derived from such research. Some common limitations in fraud detection research include:

1. Cost and Resource Constraints: Developing and implementing fraud detection systems can be costly and resource-intensive, which can limit the scope and scale of research projects.
2. Evolving Fraud Techniques: Fraudsters are constantly developing new tactics, making it challenging for research to keep up with emerging threats. Research findings may quickly become outdated.
3. Interpretable Models: Some advanced machine learning models, like deep neural networks, can be difficult to interpret. This lack of transparency can be a limitation, especially in cases where regulatory compliance and model explain ability are critical.
4. Imbalanced Data: In fraud detection, the number of legitimate transactions typically far outweighs the number of fraudulent ones, leading to imbalanced datasets. Imbalanced data can lead to biased models that perform well on the majority class but poorly on the minority class

With this there can be a lot more limitations but here we presented this important ones

Future Research:

Future research on fraud detection is likely to focus on enhancing the accuracy and efficiency of detection methods. This may involve the development of more sophisticated machine learning algorithms, incorporating deep learning techniques, and leveraging big data analytics for real-time fraud detection.



Additionally, research efforts could concentrate on improving the detection of emerging fraud patterns, such as those related to cyber-physical systems, identity theft, and social engineering attacks. Collaboration between researchers, industry professionals, and policymakers will be crucial to understanding evolving fraud tactics and developing effective countermeasures.

Furthermore, exploring the ethical implications of fraud detection technologies, especially concerning user privacy and data protection, will be essential. Striking a balance between security and individual rights will likely be a focus of future research in the field of fraud detection.

Thank You