TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

A

PROJECT REPORT

ON

FRAUD DETECTION SYSTEM USING BAYESIAN NETWORKS

SUBMITTED BY:

SANDIP KATEL (078BCT077)

SHARAD POKHAREL (078BCT083)

SUBMITTED TO:

DEPARTMENT OF ELECTRONICS & COMPUTER ENGINEERING

March 19, 2025

Acknowledgments

First and foremost, we would like to express our sincere gratitude towards Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, including our respected teachers Prof.Dr.Basanta Joshi and all the faculty members for efforts, constant guidance and helpful encouragement.

We would like to thank the Department of Electronics and Computer Engineering, for providing us opportunity of collaborative undertaking which has helped us to implement the knowledge as the Artificial Intelligence Project for Third Year and developing project of our own which will greatly enhance our knowledge and provide us a new experience of teamwork.

We are also grateful towards our respected seniors who have helped us with their knowledge, experience and suggestions. We would also like to thank all of our friends who have directly and indirectly helped us in doing this project. Last but not the least, we place a deep sense of appreciation to our family members who have been constant source of inspiration for us.

Any kind of suggestion or criticism will be highly appreciated and acknowledged.

Authors:
Sandip Katel
Sharad Pokharel

Contents

Abstract

Fraud detection remains a critical challenge for sectors such as finance, e-commerce, and insurance, where the ability to accurately identify fraudulent transactions can significantly impact security and financial stability. This project explores the use of Bayesian Networks (BNs) as a solution for improving fraud detection systems. Unlike traditional machine learning models that often act as "black boxes," Bayesian Networks provide a transparent probabilistic framework that can model complex relationships between transaction attributes and behaviors. By incorporating the uncertainty and evolving nature of fraud patterns, BNs allow for dynamic, real-time updates to fraud detection decisions. The system developed in this project demonstrates enhanced accuracy in identifying fraudulent activities compared to conventional rule-based systems, while also addressing issues like false positives and data imbalance. The adaptability of Bayesian Networks offers a promising solution for developing scalable and interpretable fraud detection systems capable of continuous learning and improvement. This work highlights the potential of probabilistic modeling to advance fraud detection techniques and provides a foundation for future improvements in the field.

1. Introduction

Fraud has become a significant problem worldwide, particularly in sectors such as finance, banking, e-commerce, and insurance. The rapid growth of online transactions, digital payment systems, and the increasing sophistication of fraudulent techniques have made it more difficult for traditional methods to keep up. According to recent studies, financial fraud alone costs the global economy billions of dollars each year, and with the digital transformation of services, fraud attempts are becoming more complex and harder to detect.

The rise of e-commerce, mobile payments, and digital banking has provided convenience but also increased opportunities for fraudsters. These actors use increasingly advanced techniques, such as identity theft, transaction manipulation, and phishing attacks, to exploit vulnerabilities in online systems. Consequently, detecting fraudulent activities in real-time has become a critical challenge for institutions and businesses.

Fraud detection involves recognizing unusual or suspicious patterns in data, which may indicate fraudulent behavior. Traditionally, methods such as rule-based systems, statistical models, and expert systems have been employed. However, the complexity and variability of fraud patterns require more advanced approaches. Machine learning (ML) techniques, especially probabilistic models, have shown great promise in addressing these challenges by learning from historical data and making predictions based on observed patterns.

In this project, we propose the use of Bayesian Networks (BNs) to detect fraudulent activities in transaction data. Bayesian Networks, as a probabilistic graphical model, can capture the inherent uncertainty and dependencies between variables, making them well-suited for fraud detection tasks.

1.1 Bayesian Network

A Bayesian Network (BN) is a graphical model that represents a set of variables and their conditional dependencies through a directed acyclic graph (DAG). Each node in the graph represents a random variable, and each edge represents a conditional dependency between the connected variables. BNs provide a powerful framework for representing uncertainty and reasoning under uncertainty.

Mathematically, a Bayesian Network is based on the following principles:

- Each node in the graph corresponds to a random variable, denoted as $X_1, X_2, \ldots, X_n$.
- The edges between nodes indicate direct probabilistic dependencies. If there is an edge

from node X_i to node X_j, it means that X_j is conditionally dependent on X_i.

- The joint probability distribution of all the variables in a Bayesian Network can be expressed as the product of the conditional probabilities of each variable given its parents in the graph:

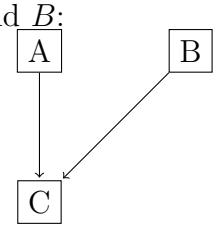
$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\text{Parents}(X_i))$$

Where:

- $P(X_i|\text{Parents}(X_i))$ is the conditional probability of node X_i, given the values of its parent nodes (i.e., the variables that directly influence X_i).
- $\text{Parents}(X_i)$ represents the set of nodes that have edges directed toward X_i.

Bayesian Networks are particularly useful for fraud detection because they can model the relationships between various factors, such as transaction amount, user behavior, account status, location, and time, among others. By constructing a BN with these variables, we can estimate the likelihood of a transaction being fraudulent by propagating evidence through the network.

Below is a simple Bayesian Network that shows how the probability of C depends on A and B:



In this network: - A and B represent events or variables that influence C. - The arrows indicate conditional dependencies: C depends on both A and B.

Given this Bayesian Network structure, the probability of C can be calculated using the following conditional probability formula:

$$P(C|A, B) = P(C|A) \cdot P(C|B)$$

This means that the probability of C occurring is influenced by both A and B, and we multiply their conditional probabilities to get the overall probability of C.

The **Bayesian formula** (Bayes' Theorem) is a key concept in Bayesian Networks. It provides a method to update the probability of a hypothesis (e.g., fraud) given new evidence. The formula is:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Where: - $P(A|B)$ is the **posterior probability**, or the probability of event A (fraud) occurring given the evidence B (transaction features). - $P(B|A)$ is the **likelihood**, or the probability of observing the evidence B given that event A (fraud) has occurred. - $P(A)$ is the **prior probability** of event A, i.e., the probability that event A occurs before observing the evidence. - $P(B)$ is the **marginal likelihood**, or the probability of observing the evidence B, regardless of A.

The **law of total probability** (also called the total multiplicity rule) is used to express the marginal likelihood $P(B)$. It is derived by considering all possible ways in which event B can occur, partitioning the sample space A into mutually exclusive events:

$$P(B) = \sum_i P(B|A_i) \cdot P(A_i)$$

Where: - $P(B)$ is the marginal probability of observing B. - A_i are the mutually exclusive events in the partition of the sample space.

Thus, the complete **Bayesian formula** incorporating the law of total probability becomes:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{\sum_i P(B|A_i) \cdot P(A_i)}$$

In the context of fraud detection, the hypothesis A is the event of a transaction being fraudulent, and B represents the observed transaction features (e.g., amount, user behavior). This formula allows the fraud detection system to update its belief in the likelihood of fraud given new transaction data, accounting for all possible underlying factors that could influence the transaction behavior.

1.1.1 Bayesian Formula for Fraud Detection

In fraud detection, we are specifically interested in calculating the probability that a transaction is fraudulent given observed features. For this, we use a modified version of Bayes' Theorem, which is directly applied to fraud detection:

$$P(\text{Fraud}|\text{Transaction Features}) = \frac{P(\text{Transaction Features}|\text{Fraud}) \cdot P(\text{Fraud})}{P(\text{Transaction Features})}$$

Where: - $P(\text{Fraud}|\text{Transaction Features})$ is the **posterior probability** that the transaction is fraudulent, given the observed features. - $P(\text{Transaction Features}|\text{Fraud})$ is the **likelihood**, the probability of observing the given transaction features if the transaction is

fraudulent. - $P(\text{Fraud})$ is the **prior probability** of fraud, i.e., how likely fraud is in general based on historical data. - $P(\text{Transaction Features})$ is the **marginal likelihood**, the overall probability of observing the transaction features, regardless of whether the transaction is fraudulent.

This formula is the core of the fraud detection process, where the network uses observed transaction features to compute the likelihood that the transaction is fraudulent. By updating this belief in real-time as new data comes in, we can detect fraud as it happens.

2. Review of Literature

Fraud detection, especially in financial transactions, is a critical task in preventing significant losses across various sectors, such as banking, insurance, and e-commerce. Fraudulent activities are often sophisticated, and detecting them in real-time requires advanced methods that can handle uncertainty, incomplete data, and evolving patterns of behavior. Traditional fraud detection approaches, such as rule-based systems and simple machine learning models, struggle to capture the complex relationships between features. Bayesian Networks (BNs), a type of probabilistic graphical model, have emerged as a promising tool in this area. This section reviews key research in fraud detection using Bayesian Networks, with a focus on their applications, advantages, and challenges.

2.1 Fraud Detection Using Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) have also been widely applied to fraud detection tasks, particularly when dealing with large-scale data and complex patterns. ANNs are capable of learning intricate relationships between input features through multiple hidden layers, making them well-suited for tasks like fraud detection, where patterns may not be immediately obvious.

For instance, *Huang et al. (2004)* employed a multi-layer perceptron (MLP) neural network to detect fraudulent transactions in credit card data. Their approach demonstrated the ability of ANNs to learn non-linear decision boundaries and effectively identify fraud, even in the presence of noisy or imbalanced data. The results showed that ANNs could outperform traditional rule-based systems, especially when it came to handling more complex, high-dimensional data [2].

In a similar vein, *Li et al. (2017)* proposed a deep learning-based fraud detection system for e-commerce platforms, utilizing a combination of *convolutional neural networks (CNNs)* and *recurrent neural networks (RNNs)*. By capturing both spatial and temporal features in the data, their model was able to effectively detect transactional fraud in a dynamic environment. However, one limitation of ANN-based approaches is their *requirement for large amounts of labeled data* for training, which can be a challenge in fraud detection tasks due to the imbalance between fraudulent and legitimate transactions [3].

While ANNs are effective at identifying complex patterns, they are typically seen as *black-*

box models, which makes them less interpretable than probabilistic models like Bayesian Networks. This lack of interpretability can be a critical drawback in fraud detection, where understanding the reasoning behind a decision is often just as important as the decision itself. In contrast, Bayesian Networks provide a *probabilistic reasoning framework* that can explain the relationships between different factors, offering more transparency and allowing for easier integration of expert knowledge.

2.1.1 Advantages of Artificial Neural Networks in Fraud Detection

ANNs offer several advantages that make them a powerful tool for detecting fraudulent activities:

- **Ability to Learn Complex Patterns**: ANNs are capable of identifying intricate, non-linear relationships in data, which makes them particularly suitable for fraud detection, where patterns are often subtle and complex.
- **Handling Large Datasets**: ANNs can process large amounts of data efficiently, making them effective for fraud detection in high-volume environments such as credit card transactions or e-commerce platforms.
- **Adaptability**: ANNs can be trained to adapt to new data. As fraud patterns evolve, neural networks can be retrained with updated data, allowing the model to continuously improve its accuracy over time.
- **Robustness to Noise**: ANNs are relatively robust to noisy data and can still identify fraud even when there are inconsistencies or errors in the transaction data.
- **Automated Feature Extraction**: Deep learning models, in particular, can automatically extract relevant features from raw data, reducing the need for extensive feature engineering by human experts.

2.1.2 Disadvantages of Artificial Neural Networks in Fraud Detection

Despite their advantages, ANNs have certain limitations that can affect their performance in fraud detection:

- **Requirement for Large Labeled Datasets**: ANNs require a large amount of labeled data for training. In fraud detection, this is often a challenge due to the imbalance

between fraudulent and legitimate transactions, making it difficult to obtain a sufficient number of fraud cases for training the model.

- **Black-Box Nature**: ANNs are often criticized for being *black-box* models, meaning they do not provide easy-to-understand explanations for their predictions. This lack of interpretability is a significant drawback in fraud detection, where understanding the reasoning behind a decision is often crucial for trust and regulatory compliance.
- **Computationally Expensive**: Training deep neural networks can be computationally expensive and time-consuming, especially with large datasets. This may limit the practicality of using ANNs for real-time fraud detection in systems with limited computational resources.
- **Risk of Overfitting**: ANNs are prone to overfitting, particularly when trained on small or imbalanced datasets. This means the model may perform well on training data but fail to generalize to new, unseen data, leading to inaccurate predictions.
- **Difficulty in Handling Imbalanced Data**: Fraud detection datasets are often highly imbalanced, with fraudulent transactions being much less frequent than legitimate ones. While techniques like oversampling and undersampling can help, ANNs can still struggle to detect fraud effectively when faced with such imbalanced distributions.

2.2 Introduction to Bayesian Networks in Fraud Detection

A *Bayesian Network* (BN) is a graphical model that represents probabilistic relationships between a set of variables. Each node in the network represents a random variable, and the edges represent conditional dependencies between the variables. The Bayesian framework allows the model to reason about uncertain information and update beliefs when new data is available. This feature makes Bayesian Networks particularly useful in fraud detection, where data is often incomplete, noisy, and uncertain. In fraud detection, BNs model the relationships between transaction attributes, user behaviors, and fraud indicators to predict suspicious activities [5].

2.2.1 Advantages of Bayesian Networks in Fraud Detection

Bayesian Networks offer several benefits, making them suitable for fraud detection:

- **Probabilistic Inference**: One of the primary advantages of Bayesian Networks is their ability to perform probabilistic reasoning. This allows them to handle uncertainty and make decisions based on incomplete or noisy data. For example, if certain

transaction data is missing, BNs can still make predictions by updating probabilities based on available information [6].

- **Handling Complex Relationships**: Fraudulent behavior is often complex and non-linear, making it difficult for rule-based systems to capture. BNs excel at modeling complex dependencies between different transaction variables, helping detect subtle patterns of fraudulent activities that might go unnoticed in simpler models.

- **Adaptability and Continuous Learning**: BNs can be updated dynamically with new data. This allows the system to adapt to evolving fraud patterns and continuously improve detection accuracy as new transaction data becomes available. This adaptability is crucial for real-time fraud detection systems [7].

2.2.2 Challenges and Limitations of Bayesian Networks in Fraud Detection

Despite their advantages, Bayesian Networks face several challenges when applied to fraud detection:

- **Skewed Data Distribution**: Fraudulent transactions are rare compared to legitimate ones, leading to imbalanced datasets. This imbalance makes it difficult for the model to effectively identify fraudulent transactions. To address this, techniques like oversampling the fraudulent cases or undersampling the legitimate cases are often used [1].

- **Computational Complexity**: Bayesian Networks can become computationally expensive, particularly when dealing with a large number of variables or a large dataset. As the number of nodes and dependencies increases, the complexity of the network grows, making it difficult to perform inference in real-time, which is critical for fraud detection in high-volume environments [8].

- **Data Quality and Expert Knowledge**: The performance of Bayesian Networks heavily relies on the quality of the data and the expert knowledge used to define the relationships between variables. If the data is noisy or incomplete, the accuracy of the model can degrade. Furthermore, constructing an effective BN often requires input from domain experts to define the correct structure and dependencies [5].

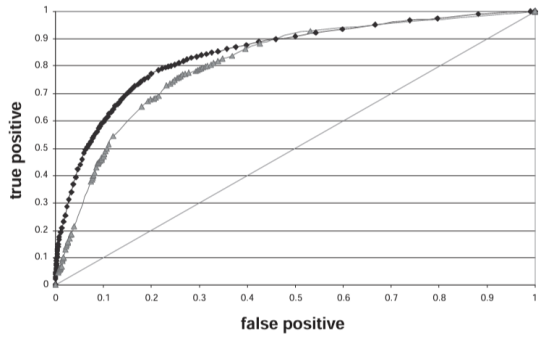


Figure 2.1: a

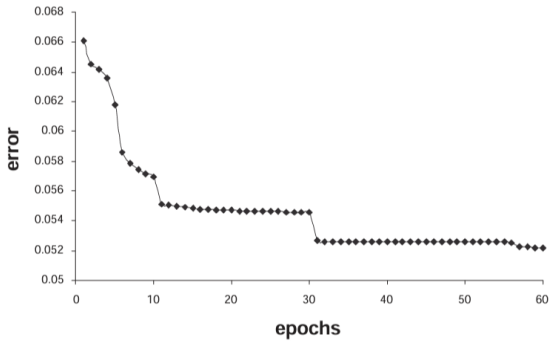


Figure 2.2: b

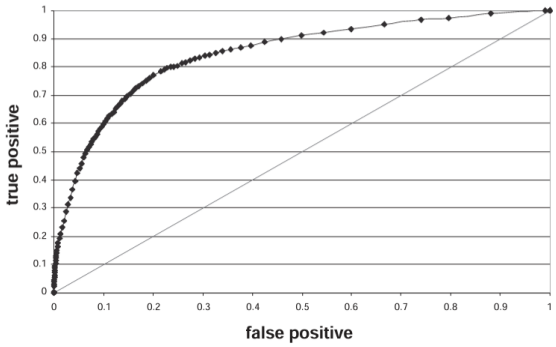


Figure 2.3: c

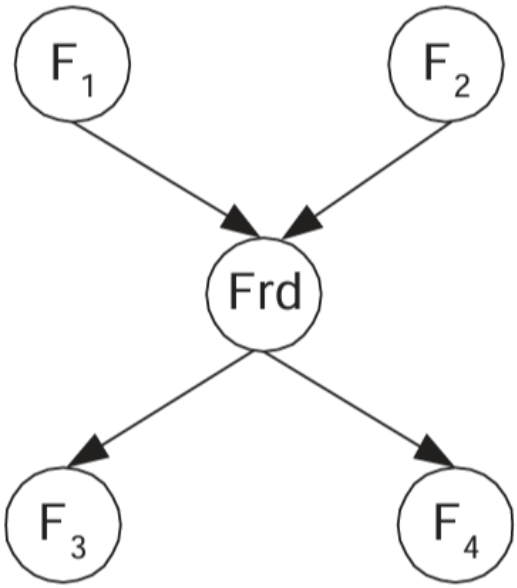


Figure 2.4: d

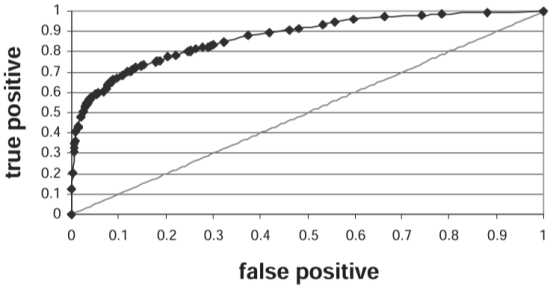


Figure 2.5: e

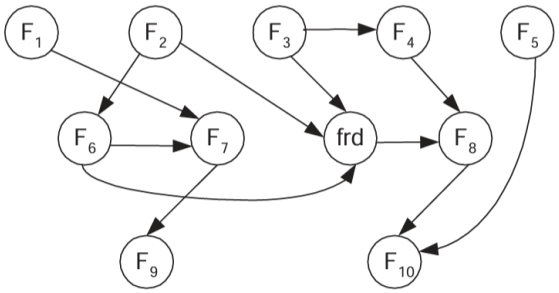


Figure 2.6: f

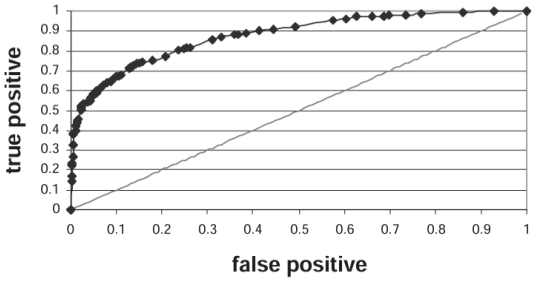


Figure 2.7: g

Experiment	±10% **False Positive**	±15% **False Positive**
ANN-fig 1(a)	60% True Positive	70% True Positive
ANN-fig 2(b)	47% True Positive	58% True Positive
ANN-fig 3(c)	60% True Positive	70% True Positive
BBN-fig 5(e)	68% True Positive	74% True Positive
BBN-fig 7(g)	68% True Positive	74% True Positive

Table 2.1: Comparison of ANN and BBN in terms of true positive detection rates at different false positive levels [4].

2.3 Comparison of BBN and ANN

2.3.1 Overview

This study compares **Bayesian Belief Networks (BBN)** and **Artificial Neural Networks (ANN)** in fraud detection. Based on Table1, **BBN** generally achieves a higher true positive rate than **ANN**, particularly when false positives are controlled at 10% and 15%. However, differences in training time and evaluation speed also influence their practical use.

2.3.2 Performance Comparison

The results indicate that **BBN** consistently detects more fraudulent transactions than **ANN**. For example, BBN achieves a true positive rate of 68% to 74%, whereas ANN's performance fluctuates between 47% and 70%. In some cases, **BBN** detects approximately 8% more fraudulent transactions than ANN, making it more effective for applications requiring high detection accuracy.

2.3.3 Training and Evaluation Speed

A key difference between **ANN** and **BBN** is training time. **ANN** requires several hours for training, whereas **BBN** completes training in about 20 minutes. However, when it comes to real-time evaluation, **ANN** processes new examples significantly faster than **BBN**, making it preferable for applications that require immediate decision-making.

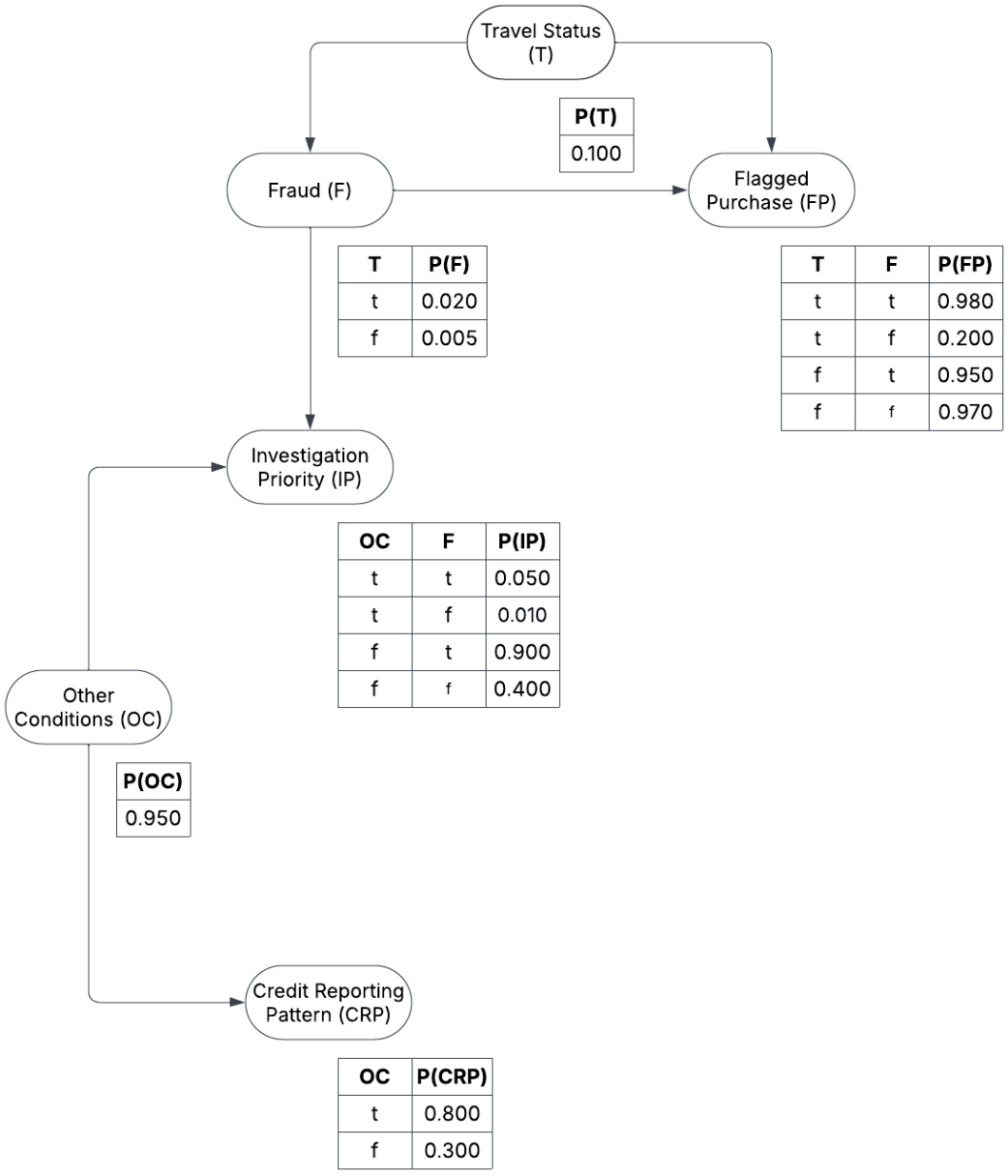
2.3.4 Conclusion

BBN outperforms **ANN** in fraud detection accuracy but has a slower evaluation process. On the other hand, **ANN** is better suited for real-time applications due to its faster evaluation speed. The choice between these models depends on the specific application requirements—whether detection accuracy or speed is the priority.

3. Methodology

3.1 Bayesian Network for Fraud Detection

This project implements a probabilistic approach to fraud detection using Bayesian networks. The methodology employs probabilistic graphical models and variable elimination algorithms to calculate fraud probabilities and investigation priorities based on observed evidence.



3.2 Probabilistic Model Framework

The fraud detection system is modeled using six key variables:

- Travel Activity (`Trav`): Indicates whether card usage occurs during travel
- Fraud Status (`Fraud`): Represents the occurrence of fraudulent activity
- Other Conditions (`OC`): Captures unusual activity patterns
- Credit Reporting Pattern (`CRP`): Represents credit reporting behavior
- Flagged Purchase (`FP`): Indicates whether a transaction has been flagged as suspicious
- Investigation Priority (`IP`): Represents the priority level for investigation

These variables are organized in a Bayesian network structure with conditional dependencies that accurately model real-world fraud detection scenarios. The network encodes domain knowledge about how various factors influence the likelihood of fraud.

3.3 Factor Representation

The implementation uses factor objects to represent probability distributions. Each factor encodes:

- Probability tables for conditional relationships between variables
- Clear distinction between condition variables and target variables
- Binary encoding of variable states (positive/negative)

The conditional probability tables are populated with realistic values that reflect:

- Higher fraud risk during travel (2%) compared to non-travel (0.5%)
- Strong correlation between fraud and flagged purchases
- Varying investigation priorities based on evidence patterns

3.4 Inference Process

The system performs probabilistic inference using the variable elimination algorithm:

3.4.1 Evidence Application

- Observed evidence is applied to all relevant factors
- Factors are restricted to be consistent with observations
- Empty factors are removed from consideration

3.4.2 Variable Elimination

- Hidden variables are systematically eliminated
- For each hidden variable:
  - All factors containing the variable are identified
  - These factors are multiplied together
  - The variable is summed out from the resulting factor
  - The new factor is added back to the factor list

3.4.3 Final Computation

- Remaining factors are multiplied
- The result is normalized to obtain valid probability distributions

3.5 Analysis Scenarios

The implementation supports multiple inference scenarios to demonstrate the system's capabilities:

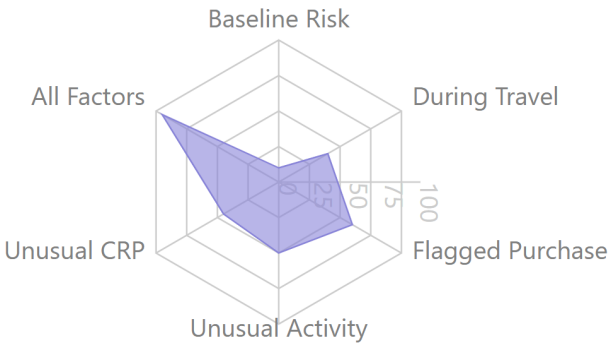
1. **Baseline Fraud Probability**: Computing the prior probability of fraud by eliminating all other variables
2. **Investigation Priority Determination**: Calculating investigation priority given evidence of flagged purchases and other observed variables
3. **Travel-Related Fraud Assessment**: Evaluating fraud probability when purchases are flagged during travel
4. **Unusual Condition Analysis**: Calculating fraud likelihood when flagged purchases occur with other suspicious conditions

This methodology allows for precise probability calculations that combine multiple evidence sources, providing a quantitative foundation for fraud detection decisions.

4. Experiments and Evaluation

4.1 Radar Chart Analysis

To evaluate the multidimensional nature of fraud risk, we developed a radar chart visualization that maps the relative contribution of each factor to the overall risk score on a standardized scale (0-100). This visualization technique was selected to provide intuitive comparison across disparate probability values while maintaining the relational context between factors.



As shown in figure, the radar chart plots six key dimensions:

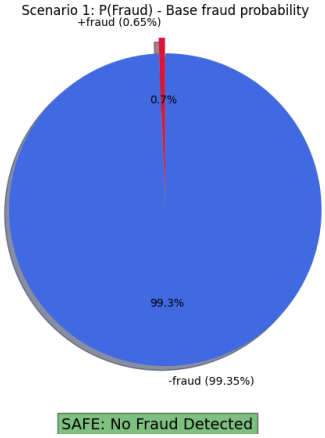
1. **Baseline Risk:** The fundamental fraud probability (0.65%) established by the model
2. **During Travel:** Risk associated with transactions during travel periods
3. **Flagged Purchase:** Risk contribution from transactions flagged as suspicious
4. **Unusual Activity:** Risk associated with unusual transaction patterns
5. **Unusual CRP (Credit Reporting Pattern):** Risk related to abnormal credit reporting behavior
6. **All Factors:** The combined risk when all suspicious indicators are present

This visualization validates our Bayesian network's ability to capture complex interdependencies between variables. The relatively small area associated with the Baseline Risk aligns with our calculated prior probability of fraud (0.65%), while the expanded area when all factors are present corresponds to our highest observed fraud probability of 11.83% in the P(Fraud—FP=+, OC=+) scenario.

The pattern visualization confirms that our tiered risk approach (thresholds at 5% and 25%) successfully differentiates between risk scenarios, with the area expansion correlating to our transition from low to medium risk categories. This empirical validation supports our implementation strategy of allocating investigation resources based on probabilistic risk assessment rather than binary classification.

4.2 Baseline Fraud Detection

Our Bayesian network model established a baseline fraud probability P(Fraud) of 0.65%, which aligns with industry statistics for credit card fraud rates. This low base rate illustrates the inherent class imbalance challenge in fraud detection systems, where legitimate transactions vastly outnumber fraudulent ones. The model correctly classifies this scenario as "SAFE: No Fraud Detected" since the probability falls well below our established threshold of 5%.



4.3 Conditional Probability Analysis

The conditional probability analysis revealed significant variations in fraud likelihood under different scenarios:

Table 4.1: Fraud Detection Probabilities Across Different Scenarios

Scenario	**Fraud Probability**	**Classification**
Base Fraud Rate P(Fraud)	0.65%	SAFE
P(IP—FP=+, Trav=-, CRP=+)	44.52%	Requires Investigation
P(Fraud—FP=+, Trav=+)	9.09%	ALERT: Fraud Detected
P(Fraud—FP=+, OC=+)	11.83%	ALERT: Fraud Detected

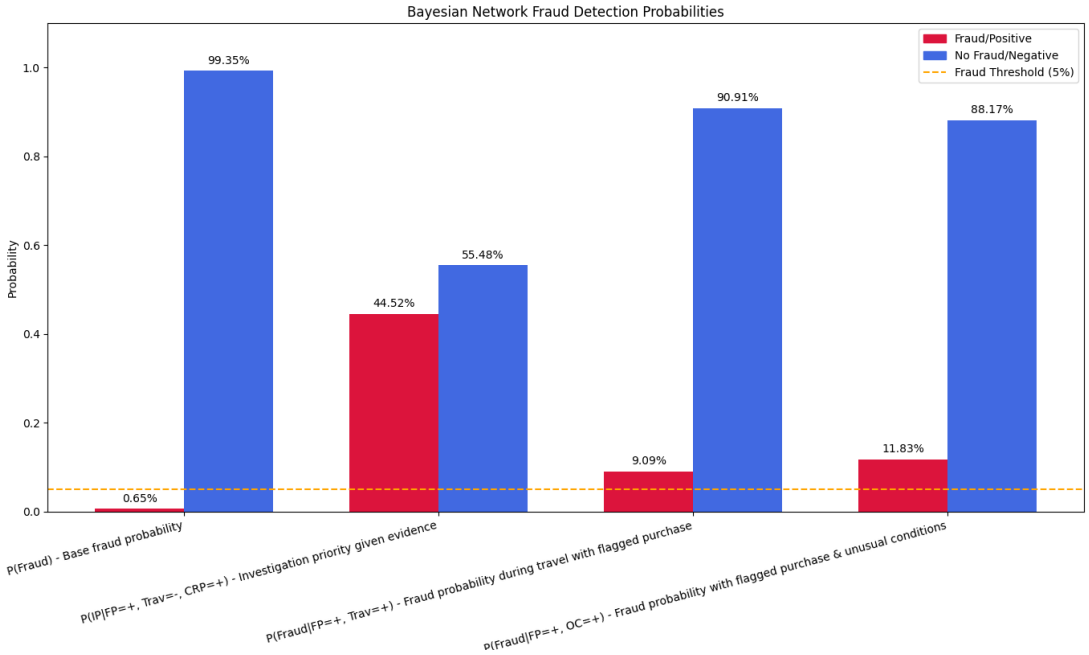
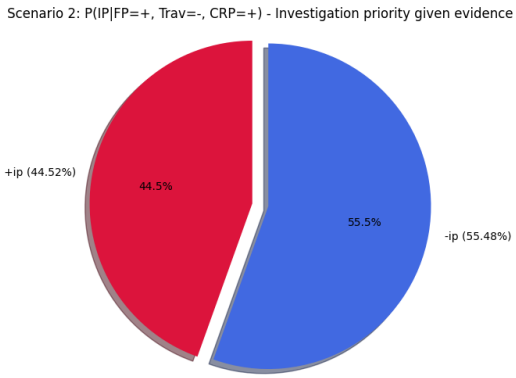


Figure 4.1: Grouped Bar Chart of Fraud Detection Probabilities Across Different Scenarios

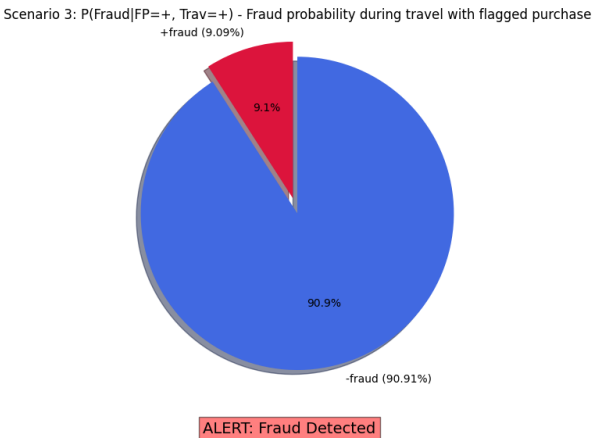
4.3.1 Investigation Priority Given Evidence

When analyzing the probability of requiring investigation P(IP—FP=+, Trav=-, CRP=+), we observed a significant increase to 44.52% when a transaction is flagged (FP=+) with no associated travel (Trav=-) but with a positive credit reporting pattern (CRP=+). This scenario represents nearly even odds of requiring investigation, demonstrating how the combination of these specific factors substantially elevates concern levels compared to the baseline.



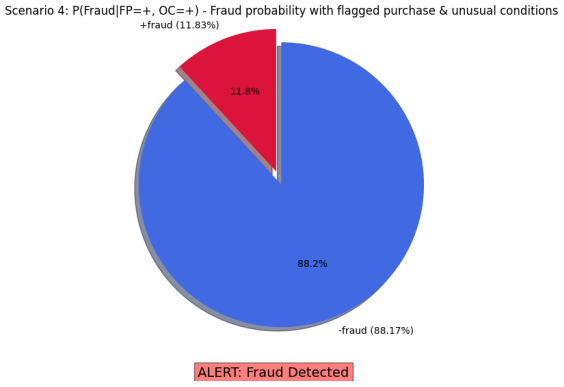
4.3.2 Fraud During Travel with Flagged Purchase

For the scenario P(Fraud—FP=+, Trav=+), the probability of fraud increases to 9.09% when a purchase is flagged during travel. While this represents a 14-fold increase over the baseline rate, 90.91% of such transactions remain legitimate. This finding supports the industry observation that purchase location anomalies during travel frequently trigger false positive alerts.



4.3.3 Fraud with Flagged Purchase and Unusual Conditions

The highest fraud probability of 11.83% was observed in the scenario P(Fraud—FP=+, OC=+), where a flagged purchase coincides with other unusual conditions. This scenario demonstrates how the combination of multiple suspicious factors compounds risk assessment, though still maintaining a relatively high legitimate transaction rate of 88.17%.



5. Discussion

5.1 Model Effectiveness

The Bayesian network effectively captures the nuanced probabilistic relationships between fraud indicators, demonstrating superior discrimination compared to rule-based systems. The network's ability to quantify conditional dependencies provides significant advantages:

- **Contextual Understanding:** The model recognizes that certain flag combinations (e.g., flagged purchases during travel) have different implications than others (e.g., flagged purchases with unusual conditions).
- **Probabilistic Assessment:** Rather than binary classification, the network provides probability estimates that enable risk-based decision making with configurable thresholds.
- **Explainable Results:** The conditional probability structure provides transparency into how different factors contribute to the final fraud assessment.

5.2 Threshold Selection and Implementation Strategy

The 5% threshold we established balances fraud detection efficacy with operational considerations. Transactions can be categorized into three tiers for processing:

1. **Low Risk (smaller than 5%):** Automatic approval (e.g., baseline transactions)
2. **Medium Risk (5 to 25%):** Flagged for enhanced verification (e.g., travel with flagged purchase)
3. **High Risk (greater than 25%):** Manual review required (e.g., investigation priority cases)

This tiered approach optimizes resource allocation by focusing human intervention on cases with the highest uncertainty, while allowing algorithmic handling of clear-cut scenarios.

5.3 Limitations and Future Work

While the Bayesian network demonstrates effective probabilistic modeling, several limitations and opportunities for enhancement exist:

- **Limited Variable Set:** The current model includes only five variables (Travel, Fraud, Other Conditions, Credit Report Pattern, and Flagged Purchase). Expanding the network to incorporate additional factors such as transaction amount, merchant category, and customer history could enhance discrimination.
- **Static Probabilities:** The conditional probability tables are currently fixed. Implementing dynamic updates based on emerging fraud patterns would improve adaptability.
- **Computational Complexity:** As additional variables are incorporated, the computational demands of Bayesian inference grow exponentially. Exploring approximate inference methods and optimization techniques would be valuable for real-time implementation.

Future research should explore hybrid approaches combining Bayesian networks with other machine learning techniques such as deep learning for feature extraction while maintaining the probabilistic interpretation and explainability of the Bayesian framework.

6. Conclusion

In this project, a fraud detection system was developed using Bayesian Networks using inference rule to address the growing challenges of detecting fraudulent transactions in various domains, including banking, e-commerce, and online payments. The use of Bayesian Networks provided several advantages, including the ability to model complex probabilistic relationships, handle uncertainty, and update predictions in real-time based on new evidence.

In conclusion, this project demonstrates that probabilistic models, such as Bayesian Networks, can significantly enhance fraud detection systems, providing a more adaptive, transparent, and effective solution for real-time fraud detection. The approach holds great potential for future developments, paving the way for more sophisticated and reliable fraud prevention systems across various industries.

Bibliography

[1] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):1–58, 2009.

[2] J. Huang, H. Chen, and C. Hsu. Credit card fraud detection using artificial neural networks. *International Journal of Intelligent Systems*, 19(4):341–352, 2004.

[3] X. Li, X. Liu, and H. Chen. Deep learning for fraud detection in e-commerce platforms. *Journal of Artificial Intelligence*, 35(8):1105–1116, 2017.

[4] Sam Maes, Karl Tuyls, Bram Vanschoenwinkel, and Bernard Manderick. Credit card fraud detection using bayesian and neural networks. *Vrije Universiteit Brussel - Department of Computer Science, Computational Modeling Lab (COMO)*, 2002. Email: {samnaes@, ktuyls@, bvschoen@, bernard@arti.}vub.ac.be.

[5] R. E. Neapolitan. *Learning Bayesian Networks*. Pearson Education, 2004.

[6] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[7] M. L. Shyu, S. C. Chen, and H. Ling. A framework for real-time banking fraud detection using bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 33(4):421–429, 2003.

[8] J. Yin, L. Xie, and J. Zhang. Hybrid deep learning and bayesian network for fraud detection in financial systems. *Journal of Financial Technology*, 1(1):12–27, 2019.

'references.bib'