**Informatics, Electrical and Electronics Engineering (Infotron)**
**Volume 5, Number 2, (2025) pp. 96-106**
ISSN 2798-0197
Sinta 4 Accreditation – No SK. 177/ E/ KPT/ 2024
http:// riset.unisma.ac.id/index.php/INFOTRON/article/view/24546/version/22105
DOI: 10.33474/infotron.v5i2

# Hidden Markov Model with Baum–Welch Algorithm for Fraud Detection in Academic Business Processes

*Siti Nunung Purwasih[1], Abd. Charis Fauzan[2]*
*Department of Computer Science,Universitas Nahdlatul Ulama Blitar, Jawa Timur, Indonesia*
*[1]sitinunungpurwasih@gmail.com; [2]abdcharis@unublitar.ac.id**

** corresponding author*

**Article History:**

Received 19 October 2025
Reviewed 14 November 2025
Revised 22 November 2025
**Accepted 30 November 2025**



**A B S T R A CT**

This study investigates fraud detection in academic business processes, where fraudulent behavior is often latent and embedded in sequential activities recorded in event logs. Fraud is defined as deviations from established academic procedures, including course registration (KRS) without payment of the Operational Education Fee (DOP), role violations, and irregular activity sequences. To address this problem, this study employs a Hidden Markov Model (HMM) to model the sequential and hidden nature of academic process behavior. The dataset consists of academic event logs from the odd semester of the 2024/2025 academic year, which were processed through data preprocessing, event-log transformation, feature construction, and rule-based labeling. Violation-based features, including wdecision, skippedactivity, wresource, wduty, wthroughput, and wpattern, were used as observation symbols in the HMM. Model parameters were estimated using the standard Baum–Welch algorithm, while the Viterbi algorithm was applied to infer the most probable hidden state sequence for each case. Model performance was evaluated using a confusion matrix and standard classification metrics. Experimental results show that the proposed approach achieves an accuracy of 96%, a precision of 87%, a recall of 100%, and an F1-score of 93%. These results demonstrate that HMM is effective for detecting fraudulent patterns in academic business processes by capturing sequential deviations that are difficult to identify using non-sequential models.

**Keywords**: Baum–Welch; Event Logs; Fraud Detection; Hidden Markov Model; Viterbi

## 1. Introduction

In the digital era, academic business processes play a crucial role in managing data, operational activities, and administrative functions within higher education institutions. These processes include various activities such as student registration, course enrollment (KRS), grading, and financial management. Although digitalization has significantly improved efficiency, it has also increased the risk of fraud, defined as deliberate actions by individuals or organizations to gain personal benefits through data manipulation or unauthorized access [1]. In the academic context, fraud may take the form of deviations in business process flows, such as data manipulation, skipping mandatory activities, validation delays, incorrect process sequences, or improper decision-making [2]. Detecting fraud in academic processes is particularly challenging because fraudulent behavior cannot be directly observed and must be inferred from event logs. However, not all academic activities generate event logs, making it essential to focus on processes that do, in order to identify deviations from established academic guidelines and calendars.

Academic guidelines and academic calendars function as formal standards for process execution; however, deviations such as delayed grade submission, late KRS validation, or incomplete

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

administrative procedures frequently occur in practice. These deviations often resemble legitimate academic behavior, making it difficult to distinguish fraud from non-fraud activities. The complexity of academic processes, the involvement of multiple actors, and variations in user behavior further complicate fraud detection. Consequently, a systematic analytical approach is required to identify anomalous patterns embedded within sequential academic activities. To address this challenge, this study employs the Hidden Markov Model (HMM) to analyze sequential academic process behavior using event log data. HMM is a probabilistic model designed to represent systems with hidden states, where the actual condition cannot be directly observed but is inferred through observable sequences and state transition probabilities [3]. Its capability to model time-dependent observations makes HMM particularly suitable for capturing dynamic behavioral patterns in academic processes [4].

Previous studies have demonstrated the effectiveness of HMM in fraud detection across various domains. For example, HMM has been successfully applied to identify suspicious user behavior in online transactions [5] and to detect credit card fraud with higher accuracy compared to traditional classifiers such as Naïve Bayes [6]. In addition, HMM has shown robustness in handling incomplete and noisy sequential data [7] and scalability in large transaction datasets [8], [9]. However, most existing studies focus on financial or commercial transactions, while the application of HMM in academic business processes remains limited. Academic process data are characterized by rule-driven activities, sequential dependencies, and institutional constraints that differ from financial transaction patterns. Therefore, applying HMM to the academic domain offers a distinct analytical perspective for detecting fraud based on process deviations rather than isolated transactional anomalies. Although prior studies have analyzed academic processes using process mining and business process perspectives [10], [11], [12] the integration of sequential probabilistic models such as HMM for fraud detection in academic business processes remains limited.

Other machine learning approaches, such as Random Forest (RF), Support Vector Machine (SVM), and Naïve Bayes, have been widely applied for fraud detection in various domains [13]. These approaches generally rely on feature-based classification and assume independence between instances, which may limit their ability to represent temporal dependencies inherent in sequential process data [14]. In academic business processes, activities are executed in a structured sequence and governed by procedural rules, making temporal relationships between events an important analytical aspect. In this context, Hidden Markov Models (HMM) provide a suitable framework by explicitly modeling sequential behavior and hidden states derived from event logs. Therefore, this study focuses on developing an HMM-based fraud detection approach to analyze academic business process activities, aiming to identify deviations from normal behavior and support transparent academic governance.

## 2. Method

In this study, the fraud detection process follows a systematic sequence of stages, as illustrated in Fig. 1. The research framework is designed to ensure that each phase is conducted in a structured and sequential manner. The process begins with problem identification within academic business processes, followed by data collection, which includes activities such as course registration KRS and academic advising. The next phase involves data preprocessing and transformation to minimize errors and ensure data quality. Subsequently, the Hidden Markov Model (HMM) is implemented by defining the initial probabilities, transition matrix, and emission matrix. The model is then trained using the Baum–Welch algorithm, and fraud pattern detection is performed using the Viterbi algorithm to determine the most probable sequence of hidden states. Finally, model evaluation is carried out using a confusion matrix to measure the accuracy of the model in identifying activities that deviate from academic guidelines and calendars. The results of this analysis are used to identify fraud patterns and assess the effectiveness of the applied algorithm.

DOI: 10.33474/infotron.v5i2.24546

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

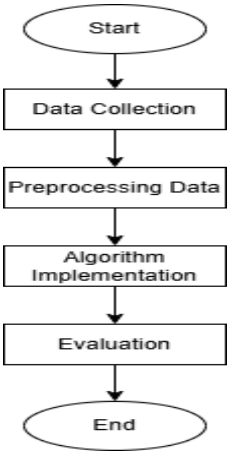

Fig. 1. Research Flow

### 2.1 Data Collection

The data in this study were obtained through a structured data collection process from the Academic Information System of XYZ University. The data include both academic and administrative information of students, which play a crucial role in analyzing academic business processes. The primary data sources consist of three main tables, the KRS–KHS table (Study Plan Card and Study Result Card), the Re-registration and DOP (Operational Education Fund) table, and the Academic Advising table. These tables were selected because each represents an essential process in the student academic cycle, including course registration, academic validation by advisors, and payment or financial administration processes. The data collection process began with the identification of problems to determine the focus of fraud detection in academic processes, followed by interviews with the IT department of XYZ University to obtain relevant and accurate data. All data retrieved from the Academic Information System were stored in CSV format, containing student activity logs recorded during the odd semester of the 2024/2025 academic year. The dataset includes various academic activities such as KRS submission and validation, tuition payment, and KHS processing. The detailed number of data entries for each activity is presented in Table 1.

Table 1. Data collection

| Data Sources | Data Volume |
|---|---|
| KRS-KHS | 19.399 |
| Validation of KRS | 3.800 |
| DOP | 18.963 |

### 2.2 Preprocessing Data

The pre-processing stage is an initial step aimed at preparing raw data to be optimally utilized in analysis and modeling. This stage is essential to ensure the quality, consistency, and relevance of the data with respect to the requirements of the applied algorithm. According to [15], pre-processing plays a crucial role in improving analytical accuracy by correcting data errors and inconsistencies. The pre-processing process in this study includes several steps: data cleaning, data reduction, and data transformation. In the data cleaning stage, missing values, duplicate entries, and inaccurate records were removed to ensure that the data used is valid and representative. This process ensures that the analyzed data is consistent and suitable for further processing [16]. In addition, only records of students with active status were retained to ensure that the analysis accurately reflects real academic activities. Next, data reduction was performed to simplify the data structure without eliminating essential information. This stage aims to maintain dataset efficiency and quality in the model-building proces [16]. Reduction was applied to several primary data sources, namely the KRS–KHS tables, SPP tables,

DOI: 10.33474/infotron.v5i2.24546

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

and Academic Advising tables. From each table, only attributes directly related to academic and administrative activities were selected. The data transformation stage involves converting raw data from its initial CSV format into an event log format suitable for sequential activity analysis. This transformation was conducted to refine the data structure and ensure that each activity is recorded chronologically and accurately represents the academic business process being analyzed. According to Pingkan et al.[17], data transformation is necessary to align the structure and distribution of data for optimal use in subsequent analytical stages. The transformation procedure in this study follows the approach described by [16], where data is converted into an event log format consisting of four main attributes: case_id, activity, resource, and timestamp. The case_id column identifies individual process flows (in this case, student ID or NIM), the activity column represents the academic stages performed, the resource column indicates the user or party executing the activity, and the timestamp column records the execution time. The results of this transformation serve as the foundation for the creation of violation features and data labeling in subsequent stages. The final dataset structure after completing all pre-processing stages is presented in Table 2, which also illustrates the data division of 80% for training and 20% for testing in the Hidden Markov Model process.

Table 2. Dataset Distribution for Training and Testing

| Class | Data *Training* | Test Data |
|---|---|---|
| Fraud | 412 Case | 103 Case |
| Non-Fraud | 1.453 Case | 364 Case |
| Total | 1865 Case | 467 Case |

### 2.3 Algorithm Implementation

The next stage after pre-processing is the implementation of the Hidden Markov Model (HMM) algorithm to detect potential fraud. This model consists of two main components, hidden states (fraud and non-fraud) and observations, which are derived from violation attributes such as wdecision, skipseq, skipdec, wresource, wduty, tmax, tmin, and wpattern. Each sequence of student activities (case) is considered as one observation sequence to estimate the most probable hidden states. Parameter estimation is carried out using the Baum-Welch algorithm, which adjusts the values of initial probabilities, state transition probabilities, and emission probabilities based on the training data. The relationship structure between states is visualized through a trellis diagram, illustrating the possible transitions from the initial state to the subsequent states throughout the sequence. Through this approach, suspicious activity patterns can be systematically mapped and classified.

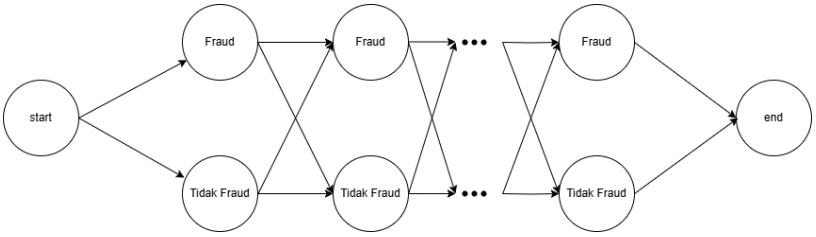


Fig. 2. HMM Trellis Diagram for State Transitions

The trellis diagram in Fig. 3 shows the final results of the probability computation between states. Each node represents a hidden state of the violation attributes within a case, while the arrows indicate the direction and likelihood of transitions. This process aims to determine the most probable sequence of states based on the actual observations.

DOI: 10.33474/infotron.v5i2.24546

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

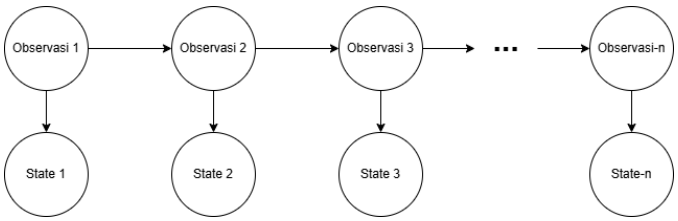


Fig. 3. HMM Trellis Diagram for Viterbi Probability Calculation

a. Defining HMM Parameters

The Hidden Markov Model (HMM) is defined by three main parameters: the initial probability ($\pi$), the transition matrix (A), and the emission matrix (B). The initial probability ($\pi$) indicates the likelihood of the system starting in a specific hidden state at time t = 1, as expressed in Equation (1). The transition matrix (A) describes the probability of moving from one hidden state to another ($S_1 \rightarrow S_2$), as shown in Equation (2). Meanwhile, the emission matrix (B) represents the probability that an observation ($O_k$) is generated from a given hidden state ($S_i$), as presented in Equation (3).

$$\pi_i = \frac{Number\ of\ cases\ from\ state\ S_i}{Total\ number\ of\ case} \quad (1)$$

$$\pi_i = \frac{Number\ of\ transitions\ from\ state\ (S_i \rightarrow S_j)}{Total\ transition\ from\ S_i} \quad (2)$$

$$\pi_i = \frac{Number\ of\ observation\ O_k of\ S_i}{Total\ observation\ of\ state\ S_i} \quad (3)$$

b. Calculating Total Probability per Case

After all parameters of the Hidden Markov Model (HMM) are established, the next step is to calculate the probability for each case. This process is based on the computation of inter-observation probabilities to determine the most likely state for each observation. The initial probability value is calculated using Equation 4:

$$a_1(t) = \pi_i * emit_{prob_{a(s)}} \quad (4)$$

Here, $\pi_i$ represents the initial probability matrix, which is applied only to the initial state, while $emit\_prob_{a(s)}$ denotes the emission probability matrix (B) that reflects the relationship between observation attribute and hidden state s. To compute the probability for the subsequent variable (t+1), the initial probability vector is updated using the result from the previous step, adjusted according to the sequence of states. At this stage, the computation also incorporates transition probabilities from matrix $A$. The probability for time (t+1) is determined using Equation 5:

$$a_1(t) = a_1(t-1) * trans_{prob_{ij}} * emit_{prob_{a(s)}} \quad (5)$$

In this equation, $a_1(t-1)$ represents the probability value obtained from the previous observation, $trans_{prob_{ij}}$ refers to the transition probability matrix (A) indicating the likelihood of transitioning from state $i$ to state $j$, denotes the $emit\_prob_{a(s)}$ emission probability matrix (B), which expresses the probability of observation $a$ occurring given state. Where $\pi_i$ denotes the initial probability of state i, $a_{ij}$ represents the transition probability from state i to state j, $b_j(o_t)$ denotes the emission probability of observation $o_t$ given state j, and $\alpha_t(i)$ represents the forward probability of being in state i at time t.

DOI: 10.33474/infotron.v5i2.24546

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

### 2.4 Evaluation Model

The performance evaluation of the HMM-based fraud detection model in this study was conducted using a Confusion Matrix. The Confusion Matrix is a table used to evaluate the performance of a classification model by comparing the predicted results against the actual data [18]. It has been proven effective and is widely used for assessing classification models. In a study by [19], the Confusion Matrix was praised for its ability to evaluate the efficiency of intrusion detection systems by providing a comprehensive analysis of the false alarm and successful detection ratios. A similar approach was applied in this research, where the True Positive Rate (TPR) and True Negative Rate (TNR) served as key indicators to assess the HMM model's ability to consistently distinguish between fraud and non-fraud activities. The Confusion Matrix consists of four main elements, as shown in Table 3: True Positive (TP), representing the number of fraud cases correctly identified; True Negative (TN), representing non-fraud activities accurately recognized; False Positive (FP), referring to non-fraud activities incorrectly classified as fraud; and False Negative (FN), representing fraud activities that the model failed to detect [18].

Table 3. Confusion Matrix

| Label | Positif | Negative |
|---|---|---|
| Positif | TP | FP |
| Negative | FN | TN |

In addition to the Confusion Matrix, several evaluation metrics such as accuracy, precision, recall, and F1-score were also applied. Accuracy measures how closely the system's predictions match manual classifications [20] and is calculated using Equation 6:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} * 100\% \quad (6)$$

The accuracy value is derived from True Positive (TP), which represents fraud cases correctly detected as fraud by the HMM model; True Negative (TN), which represents non-fraud cases correctly identified as non-fraud; False Positive (FP), which represents non-fraud cases incorrectly classified as fraud; and False Negative (FN), which represents fraud cases the model failed to detect, classifying them as non-fraud. Other evaluation metrics include precision, recall, and F1-score, calculated using the following formulas:

$$\text{Presition} = \frac{TP}{TP+FP} * 100\% \quad (7)$$

$$Recall = \frac{TP}{TP+FN} * 100\% \quad (8)$$

$$F1 = \frac{2*Presisi*recall}{Presisi+Recall} \quad (9)$$

## 3. Results and Discussion

### 3.1 *Probability ($\pi$)*

Table 4 presents the results of the initial probability matrix. From the calculations, it can be observed that the frequency of occurrence for the Non-Fraud state is 77.91%, while the probability value for the Fraud state is 22.09%. These results indicate that most of the data in the initial observation phase are in the Non-Fraud condition. This occurs because the probability values are derived from the overall frequency distribution of the states before being influenced by transitions from previous states.

DOI: 10.33474/infotron.v5i2.24546

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

Table 4. Initial State Probability Matrix (π)

| Matriks Probability | |
|---|---|
| **Label** | **Frequency** |
| Non-fraud | 0.7791 |
| Fraud | 0.2209 |

### 3.2 *Transition Matrix (A)*

The transition matrix describes the probability of the system moving between the Fraud and Non-Fraud states across consecutive process attributes. In general, the results indicate that the system tends to remain stable within the same state, with relatively low transition probabilities between states. For instance, in the Tmax → Tmin transition (Table 5), the Non-Fraud state has a 66.67% probability of remaining unchanged and a 33.33% probability of shifting to Fraud. Meanwhile, the Fraud state shows a 53.68% probability of persisting and a 46.32% chance of moving to Non-Fraud.

Table 5. State Transition Matrix (Tmax → Tmin)

| State | Non-Fraud | Fraud |
|---|---|---|
| Non-fraud | 0.6667 | 0.3333 |
| Fraud | 0.4632 | 0.5368 |

Similarly, in the wresource → wduty transition (Table 6), the Non-Fraud state exhibits a strong stability with a 99.70% probability of remaining unchanged, while the Fraud state has a 73.96% chance of transitioning to Non-Fraud and only a 26.04% probability of persisting as Fraud. Overall, the transition results across all attributes demonstrate that the Non-Fraud state generally dominates the system's behavior, whereas transitions to the Fraud state occur with relatively lower probabilities. This indicates a strong stability pattern in business process activities modeled using HMM.

Table 6 . State Transition Matrix (wresource → wduty)

| State | Non-Fraud | Fraud |
|---|---|---|
| Non-fraud | 0.9970 | 0.0030 |
| Fraud | 0.7396 | 0.2604 |

### 3.3 *Emission Matrix (B)*

The emission probability matrix illustrates how frequently each observed attribute takes the value 1 or 0 within the Fraud and Non-Fraud states. As shown in Table 7, several attributes exhibit a high probability of being 1 in the Fraud state. For instance, the attribute Tmax has a probability of 99.03%, indicating that it almost always occurs when the system is in the Fraud state. Similarly, attributes such as skipdec, wresource, and wpattern also demonstrate very high probabilities of being 1—98.31%, 97.58%, and 98.31%, respectively. Conversely, attributes like wduty and wdecision show very low probabilities of being 1—only 0.48% and 1.69%, respectively—indicating that they predominantly take the value 0 in the Fraud state. In contrast, within the Non-Fraud state, the distribution changes slightly. Although Tmax still shows a high probability of being 1 (98.83%), other attributes such as skipseq, skipdec, and wdecision have very low probabilities (around 0.07%). These results suggest that most attributes tend to take the value 0 when the system is in the Non-Fraud state. However, certain attributes such as wduty (27.42%) and wresource (77.80%) still exhibit a relatively high probability of being 1, even in Non-Fraud conditions. This indicates that the presence of these

DOI: 10.33474/infotron.v5i2.24546

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

attributes does not necessarily imply fraudulent behavior, as some may also occur naturally within normal process activities.

Table 7. Emission Probability Matrix (B)

| Fraud | Probability | Non-fraud | Probability |
|---|---|---|---|
| tmax_1 | 0.9903 | tmax_1 | 0.9883 |
| tmax_0 | 0.0097 | tmax_0 | 0.0117 |
| skipdec_1 | 0.9831 | skipdec_1 | 0.0007 |
| … | … | … | … |
| wdecision_0 | 0.9831 | wdecision_0 | 0.9993 |
| wpattern_1 | 0.9831 | wpattern_1 | 0.0447 |
| wpattern_0 | 0.0169 | wpattern_0 | 0.9553 |

### 3.4 *Viterbi Decoding Results*

The results of the maximum path search for each case are presented in Table 8. To obtain the value of the first attribute, the computation involves multiplying it by the specific parameter results from the previously initialized initial probability matrix. For the subsequent attributes, the highest probability value from the state of the previous attribute is used. The resulting trellis diagram selects the highest probability as the optimal path, which also represents the hidden state of the PBF attribute. The final results of each attribute then determine the final state of every case processed through the iterative computation of the Hidden Markov Model.

Table 8. Viterbi Decoding Results per Case

| case | step | prob_fraud | prob_non_fraud | state_max |
|---|---|---|---|---|
| 1 | 1 | 0.0021344109 | 0.009102751905 | Non-Fraud |
| 1 | 2 | 0.003166174576 | 0.002727697478 | Fraud |
| 1 | 3 | 0.00315489632 | 0.003160361731 | Non-Fraud |
| 1 | 4 | 0.00004178717683 | 0.002469714464 | Non-Fraud |
| 1 | 5 | 0.001872291281 | 0.001492711313 | Fraud |
| 1 | 6 | 0.00000235543891 | 0.0001337063683 | Non-Fraud |
| 1 | 7 | 0.000131117838 | 0.0001332812707 | Non-Fraud |
| 1 | 8 | 0.00003316177183 | 0.000001506933273 | Fraud |
| … | … | … | … | … |
| 467 | 1 | 0.2187771173 | 0.7699857199 | Non-Fraud |
| 467 | 2 | 0.1860835054 | 0.1603131781 | Fraud |
| 467 | 3 | 0.1854206559 | 0.1857418709 | Non-Fraud |
| 467 | 4 | 0.03980649829 | 0.00002782895597 | Fraud |
| 467 | 5 | 0.03817853053 | 0.03043838586 | Fraud |
| 467 | 6 | 0.02809979808 | 0.02049308065 | Fraud |
| 467 | 7 | 0.02749295541 | 0.02794658674 | Non-Fraud |
| 467 | 8 | 0.006953402588 | 0.0003159756894 | Fraud |

### 3.5 *Algorithm Performance Evaluation*

Furthermore, to evaluate the performance of the HMM model in detecting business process activities, an analysis of the resulting confusion matrix was conducted. This evaluation aims to measure the extent to which the model can accurately distinguish between non-fraudulent and fraudulent

DOI: 10.33474/infotron.v5i2.24546

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

activities. Fig. 4 presents the confusion matrix of the HMM model's classification results in detecting non-fraudulent and fraudulent business process activities. Based on these results, the model successfully classified 103 fraud cases correctly (True Positive) and 349 non-fraud cases correctly (True Negative). However, there were 15 non-fraud cases that were incorrectly classified as fraud (False Positive), while there were no fraud cases incorrectly classified as non-fraud (False Negative). With these results, the model demonstrates a fairly high level of accuracy in distinguishing between fraudulent and non-fraudulent activities.

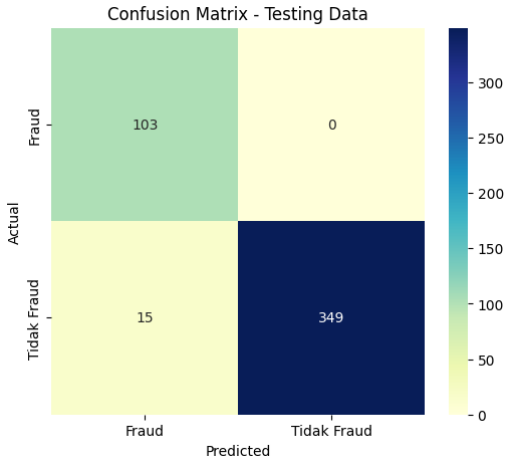


Fig. 4. Confusion Matrix

Further evaluation was carried out by calculating the model's performance metrics, namely Accuracy, Precision, Recall, and F1-Score. Based on the evaluation results, the model achieved an accuracy of 96%, indicating a high overall level of correctness in distinguishing between fraudulent and non-fraudulent activities. The precision value of 87% indicates that most of the fraud predictions made by the model were indeed fraud, showing a very low false positive rate. Meanwhile, the recall value reached 100%, meaning that all fraud cases were successfully identified without any being missed. The perfect recall may be influenced by clearly defined rule-based labeling and the relatively limited scope of the dataset, which reduces ambiguity between fraud and non-fraud cases. The combination of precision and recall resulted in an F1-score of 93%, demonstrating that the model has an excellent balance between sensitivity and accuracy in detecting fraudulent activities.

## 4. Conclusion

Based on the results of this study, it can be concluded that the Hidden Markov Model (HMM) is effective for detecting fraudulent activities in academic business processes characterized by sequential and latent behavior. By utilizing violation-based features derived from academic event logs as observation symbols, the model is able to infer hidden states representing fraud and non-fraud activities. Model parameters were estimated using the standard Baum–Welch algorithm, producing the initial state probabilities, state transition matrix, and emission matrix, while the Viterbi algorithm was applied to determine the most probable state sequence for each case. Experimental evaluation showed that the proposed approach achieved an accuracy of 96%, with high recall and acceptable precision, indicating its ability to identify fraudulent patterns embedded in academic process flows. Nevertheless, this study has several limitations, including the use of data from a single academic semester and reliance on rule-based labeling, which may affect generalizability. Future work should involve multi-semester datasets, validation of labeling rules by academic audit experts, and comparative evaluation with other machine learning models. From a practical perspective, the proposed HMM-based approach can be

DOI: 10.33474/infotron.v5i2.24546

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

integrated into academic information systems as an early warning mechanism to support academic governance, enhance data integrity, and assist institutions in monitoring procedural compliance in a transparent and systematic manner.

## *5.* References

[1] D. M. Elisabeth and W. A. Simanjuntak, "Analisis Review Pendeteksian Kecurangan (Fraud)," *METHOSIKA J. Akunt. dan Keuang. Methodist*, vol. 4, no. 1, pp. 9–18, 2020, doi: 10.46880/jsika.vol4no1.pp9-18.

[2] A. H. Koosasi, R. Sarno, and A. Munif, "Deteksi Fraud Menggunakan Metode Model Markov Tersembunyi Pada Proses Bisnis," *J. Tek. ITS*, vol. 6, no. 1, pp. 11–15, 2017, doi: 10.12962/j23373539.v6i1.22328.

[3] M. Mulyami and A. I. Sofiyat, "Hidden Markov Model Dan Aplikasinya," *Mat. Sains*, vol. 2, no. 1, pp. 46–52, 2024, doi: 10.34005/ms.v2i1.3803.

[4] S. A. Kamal, "Implementasi Hidden Markov Model dengan Algoritma Forward-backward untuk Prediksi Buku Peminjamandi Perpustakaan (studi kasus: Perpustakaan SPS UIN Jakarta)," *Repository.uinjkt.ac.id*, 2023.

[5] Niki Patel, Yanyan Li, and Ahmad Hadaegh, "Online Transaction Fraud Detection using Hidden Markov Model & Behavior Analysis," *Int. J. Comput. Sci. Secur.*, vol. 15, no. 15, pp. 59–72, 2021.

[6] A. Srivastava, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection using Hidden Markov Model," *IEEE Trans. Dependable Secur. Comput.*, vol. 5, no. 1, pp. 37–48, 2019, doi: 10.1109/TDSC.2007.70228.

[7] L. Hoskovec *et al.*, "Infinite hidden Markov models for multiple multivariate time series with missing data," *Biometrics*, vol. 79, no. 3, pp. 2592–2604, 2023, doi: 10.1111/biom.13715.

[8] I. A. Amaefule and A. A. O, "Hidden Markov Model Application for Credit Card Fraud Detection Systems," *Int. J. Innov. Sci. Res. Technol.*, vol. 5, no. 1, 2020.

[9] A. Overcomer, I. Alex, O. Ogochukwu, E. O. Uzoamaka, and A. Angela, "The Application of Hidden Markov Model in Credit Card Fraud Detection System," vol. 10, no. 2, pp. 1–20, 2022.

[10] A. F. Q. Haq, K. S. Kinasih, D. A. Wiranti, M. A. Yaqin, and A. C. Fauzan, "Audit Pengelolaan Sekolah Menggunakan Proses Mining," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 2, no. 3, pp. 349–357, 2020, doi: 10.28926/ilkomnika.v2i3.148.

[11] F. U. Fatmawati, A. C. Fauzan, and V. A. Tricahyo, "Analisis Kompleksitas Proses Bisnis Penerimaan Mahasiswa Baru Universitas Nahdlatul Ulama Blitar menggunakan Control-Flow Complexity dengan Pemodelan Business Process Modelling Notation," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 2, pp. 1750–1760, 2024.

[12] A. C. Fauzan, R. Sarno, and N. F. Ariyani, "Structure-based ontology matching of business process model for fraud detection," *Proc. 11th Int. Conf. Inf. Commun. Technol. Syst. ICTS 2017*, vol. 2018-Janua, pp. 221–225, 2018, doi: 10.1109/ICTS.2017.8265674.

[13] I. Y. Hafez, A. Y. Hafez, A. Saleh, A. A. Abd El-Mageed, and A. A. Abohany, "A systematic review of AI-enhanced techniques in credit card fraud detection," *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-024-01048-8.

[14] Zaidan, "Deteksi Fraud pada Transaksi dalam Perbankan menggunakan Random Forest Program Studi Sarjana Informatika Fakultas Informatika Universitas Telkom Bandung," 2024.

[15] N. P. A. Widiari, I. M. A. D. Suarjaya, and D. P. Githa, "Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 8, no. 2, p. 137, 2020, doi: 10.24843/jim.2020.v08.i02.p07.

[16] M. JONATHAN, "Analisis Dan Penerapan Process Mining Untuk Menentukan Bottleneck Mata Kuliah Dan Proses Skripsi (Studi Kasus Program Studi …," 2023.

[17] W. Pingkan, J. Kaunang, N. Maringka, R. Banne, V. Dwi, and P. Humu, "KAJIAN MANAJEMEN DATA ( TRANSFORMASI DATA )," no. June, 2025.

DOI: 10.33474/infotron.v5i2.24546

Informatics, Electrical and Electronics Engineering (Infotron)
Volume 5, Number 2, (2025) pp. 96-106
ISSN 2798-0197

[18] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.

[19] M. K. Suryadewiansyah, T. Endra, and E. Tju, "Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert," *J. Nas. Teknol. dan Sist. Inf.*, vol. 8, no. 2, pp. 81–88, 2020.

[20] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.

DOI: 10.33474/infotron.v5i2.24546