**Fraud Detection in Blockchain Transactions Using Hybrid Models**

**Introduction**

* **Blockchain Technology**: The rise of blockchain has led to the development of secure and transparent transactions. However, fraud still remains a significant issue.
* **Objective**: To develop a model that effectively identifies fraudulent transactions using machine learning, combining techniques like LSTM, CNN, and LightGBM.

**Related Work**

* **Machine Learning in Fraud Detection**: Several studies have employed deep learning models (e.g., LSTM, CNN) and traditional models (e.g., LightGBM, XGBoost) for fraud detection in financial transactions.
* **Hybrid Approaches**: Recent research suggests combining models to enhance performance, addressing the challenges of imbalanced data and varying transaction patterns.

**Problem Formulation**

* **Data and Features**: A dataset with features representing blockchain transactions, with a target variable fraud indicating whether a transaction is fraudulent (1) or legitimate (0).
* **Challenges**:
  + **Imbalanced Classes**: Fraudulent transactions (1) are much less frequent than legitimate ones (0).
  + **Feature Engineering**: Extracting meaningful features from transaction data is crucial for model success.

**Contribution**

* **Model Development**: A hybrid model combining LSTM, CNN, and LightGBM is proposed for enhanced fraud detection.
  + **LSTM**: Captures sequential patterns in blockchain transaction data.
  + **CNN**: Detects local patterns and anomalies in the data.
  + **LightGBM**: A gradient-boosted tree model that handles categorical data and improves interpretability.

**Blockchain Transactions**

* **Data Preparation**:
  + **Dataset**: A cleaned dataset is used, with features representing transaction characteristics (e.g., amounts, time intervals).
  + **Scaling**: MinMaxScaler is used to normalize the features before feeding them into the models.
* **Model Inputs**:
  + For LSTM and CNN, reshaping the data for sequential and spatial pattern detection is necessary.

**Proposed Model**

* **LSTM Model**:
  + Two LSTM layers with 64 and 32 units, followed by dense layers for classification.
  + Trained on reshaped transaction data for capturing time-dependent patterns.
* **CNN Model**:
  + One convolutional layer with 64 filters, followed by a dense layer.
  + Trained to detect spatial patterns in transaction features.
* **LightGBM Model**:
  + A gradient-boosted tree model with 100 estimators and a learning rate of 0.05.
  + Handles categorical features and learns complex patterns for fraud detection.

**Results & Discussion**

* **Model Performance**:
  + **Accuracy**: 98.1% on the test set, indicating strong performance.
  + **Precision & Recall**:
    - **Class 0 (Legitimate Transactions)**: High precision (0.98) and recall (1.00), meaning very few legitimate transactions are misclassified.
    - **Class 1 (Fraudulent Transactions)**: High precision (0.99) but slightly lower recall (0.92), indicating some fraudulent transactions are missed, though the model is still strong in identifying most frauds.
  + **F1-Score**:
    - **Class 0**: 0.99, reflecting strong balance between precision and recall.
    - **Class 1**: 0.96, showing good performance in identifying fraud despite some false negatives

A graph of different colored squares

AI-generated content may be incorrect.

* **Conclusion**:
  + The hybrid model combining LSTM, CNN, and LightGBM effectively handles the fraud detection task.
  + While the model performs excellently on legitimate transactions, some fraud cases still go undetected, which is a common challenge in imbalanced classification problems.

A blue squares with white text

AI-generated content may be incorrect.

A graph with a line

AI-generated content may be incorrect.

**Future Work**

* **Improving Recall for Fraudulent Transactions**: Techniques like oversampling, undersampling, or using different loss functions could be explored to further reduce false negatives.
* **Real-Time Fraud Detection**: Investigating the model’s performance in a real-time blockchain transaction setting could be beneficial for deploying it in production systems.