Team Details	
Team Name:	Future Crusaders
Team Leader Name:	Ansh Gupta
Problem Statement:	The insurance industry faces significant challenges in detecting and preventing various types of fraud, including identity theft, premium fraud, and claim inflation. With fraudsters employing increasingly sophisticated techniques and the vast amount of data generated by insurers, it has become difficult to identify and mitigate fraudulent activities effectively.

BRIEF ABOUT THE IDEA :

The proposed solution is an AI/ML-driven fraud detection system for the insurance industry. It leverages machine learning models (including LSTM, Random Forest, and CNN) to identify complex fraud patterns, detect anomalies, and authenticate documents in real time. The system is designed to automatically detect emerging and unknown fraud patterns by analyzing large and diverse datasets, both structured (e.g., customer data, claims, policy info) and unstructured (e.g., scanned documents, signatures). The machine learning model will learn from historical data to predict trends, spot forgery in documents, and enable proactive fraud prevention.

OPPORTUNITIES:

A. How different is it from any of the other existing ideas?

- **Advanced multi-modal fraud detection:** Integrates time-series analysis, anomaly detection, and document authentication for a comprehensive approach.
- **Emerging fraud pattern detection:** Adapts dynamically using LSTM, Random Forest, and CNN models to identify new fraud patterns.
- **Real-time fraud prevention:** Operates in real-time to flag suspicious claims proactively.

B. How will it be able to solve the problem?

- Analyzes historical and real-time data to detect evolving fraud patterns.
- LSTM models capture time-based trends (e.g., spikes during holidays).
- Random Forest detects anomalies in high-value claims and activities.
- Authenticates documents using AI to prevent forgery.
- Deep learning models adjust to new data, ensuring adaptability.

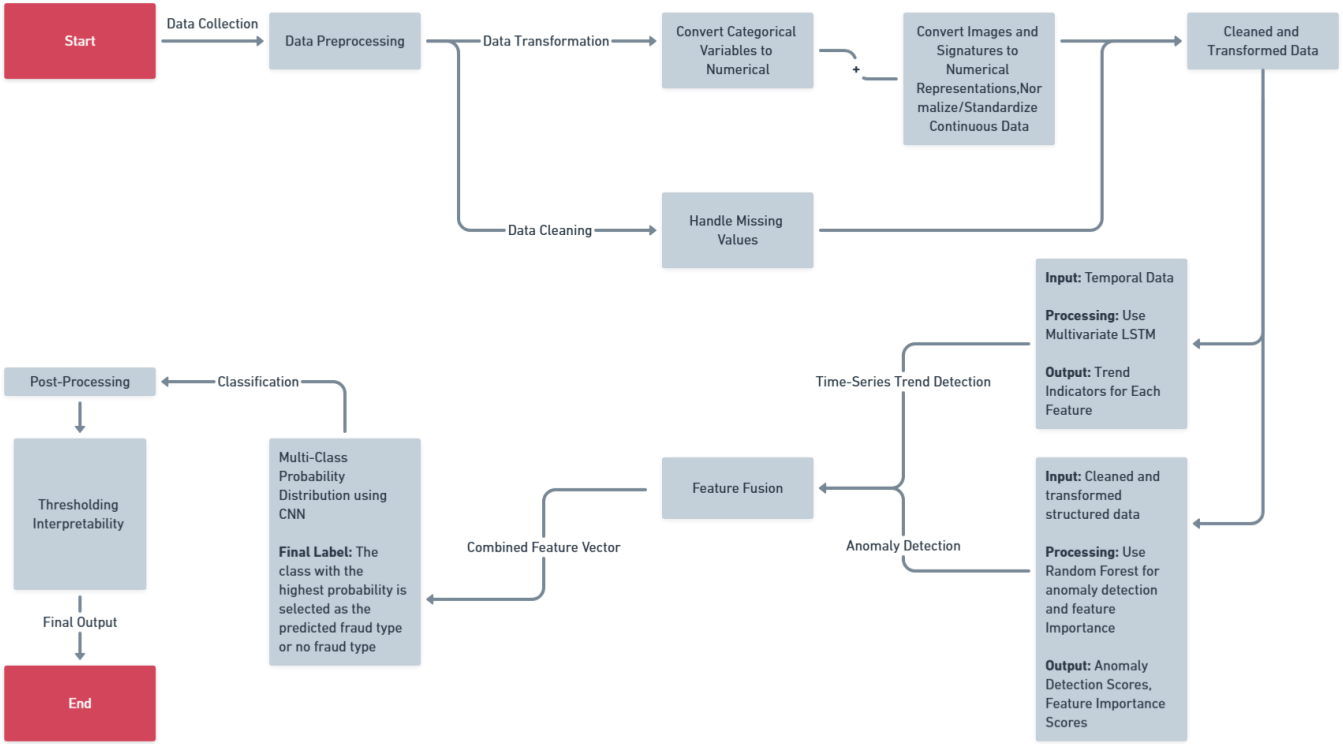
C. USP of the proposed solution

- **Emerging Fraud Detection -** Identifies hidden and unknown fraud patterns using dynamic learning from temporal trends and anomalies.
- **AI-Powered Document Authentication -** Prevents forgery through advanced verification methods like signature matching and image processing.
- **Proactive Fraud Prevention -** Detects and mitigates fraud early with real-time analysis and multi class classification.
- **Holistic Data Integration -** Handles diverse data types (numerical, categorical, and non-numerical) for comprehensive fraud detection.
- **Scalability and Transparency -** Processes large datasets efficiently and provides explainable outputs like feature importance and anomaly scores.

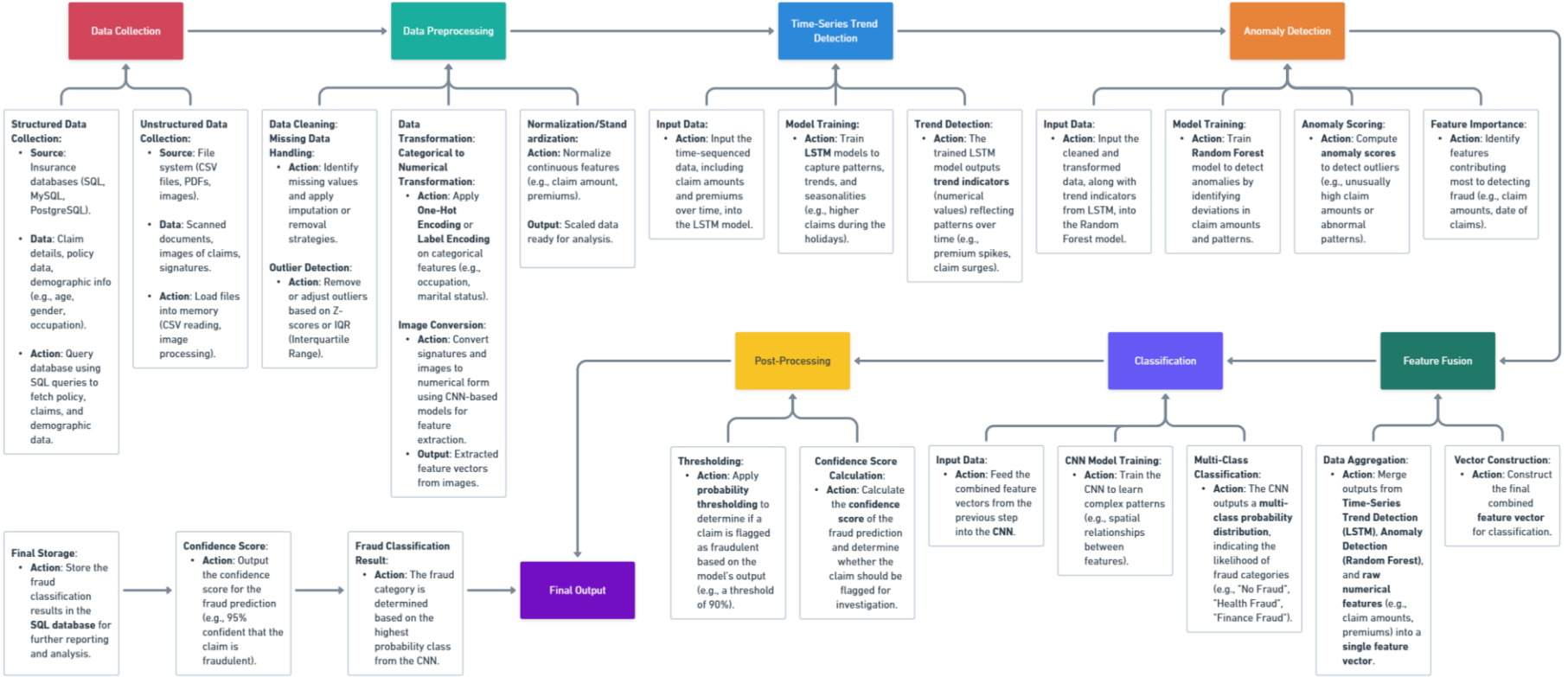
LIST OF FEATURES OFFERED BY THE SOLUTION:

- Emerging Fraud Pattern Detection using LSTM to detect previously unknown fraud patterns in customer behavior or claims.
- Time-Series Trend Detection to a and analyze seasonal or cyclical fraud trends (e.g., spikes during festivals or holidays).
- Anomaly Detection with Random Forest for detecting outliers in high-value claims and suspicious activities.
- Document Forgery Detection through image matching and signature verification (using AI techniques to convert documents to numerical data).
- Real-Time Fraud Prevention by flagging fraudulent claims during the claim submission process.
- Multimodal Data Integration to handle structured (numerical) and unstructured (images, documents) data.
- Multiclass Fraud Classification with CNN to classify fraud types (e.g., identity theft, premium fraud, claim inflation, no fraud).
- Probabilistic Scoring providing a confidence level for each prediction to aid decision-making.

PROCESS FLOW DIAGRAM



ARCHITECTURE DIAGRAM OF THE PROPOSED SOLUTION



TECHNOLOGIES TO BE USED IN THE SOLUTION:

- Data Processing & Preprocessing: Pandas, NumPy for data cleaning and transformation. OpenCV or TensorFlow for document and signature verification.

- Machine Learning Frameworks: LSTM (TensorFlow/Keras or PyTorch) for time-series analysis and fraud pattern detection. Random Forest (Scikit-learn) for anomaly detection in claims. CNN (TensorFlow/Keras or PyTorch) for multi-class classification and feature extraction.

- Data Storage: SQL Databases for structured data. NoSQL Databases (MongoDB) for storing unstructured data (e.g., scanned images, documents). Cloud storage solutions (AWS S3, Google Cloud Storage) for large datasets. (CSV for testing data)

- Cloud Platforms: AWS, Google Cloud, or Microsoft Azure for scalable deployment and storage.

ESTIMATED IMPLEMENTATION COST:

The model is being developed locally with pre-existing hardware and software tools, the costs will primarily involve **hardware, software tools, and time :**

1. Hardware:

- The model will be trained on local GPU's.
- If more computational power will be needed, our university's clustered computers can be used.

2. Software:

•Free/Open Source Tools:

- Python: Free
- Libraries: TensorFlow, PyTorch, Scikit-learn, Pandas, NumPy, Matplotlib, etc. (all free)

•Cloud Services: (if needed for large-scale training):

- Free tiers for Google Colab or AWS EC2 can be sufficient for small to medium datasets.
- Optional cost for extended use: **800rs-4000rs**.

3. Data Storage:

•Use local storage or free cloud storage options for datasets.

•If the data is huge, additional storage solutions may cost **5000rs-10000rs**.

ADVANTAGES:

1. **Hybrid Accuracy**: Combines LSTM, Random Forest, and CNN for superior fraud detection.
2. **Diverse Data Handling**: Processes numerical, non-numerical, and image data effectively.
3. **Explainability**: Provides interpretable outputs like feature importance and anomaly scores.
4. **Scalability**: Designed to handle large, multi-source datasets seamlessly.
5. **Trend Analysis**: Captures temporal patterns like seasonal fraud spikes.

FUTURE SCOPE:

1. **Real-Time Detection**: Extend for live fraud monitoring.
2. **Advanced Models**: Incorporate transformers for text and AutoML for feature optimization.
3. **Dynamic Trends**: Update fraud trends using online learning.
4. **Cross-Industry Use**: Adapt for banking, healthcare, and e-commerce fraud detection.

SBI Life
Apne liye. Apno ke liye.

PRESENTS

HACK-AI-THON 2024

POWERED BY H2S HACK2SKILL

Thank You!