International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 11 | Nov 2024 www.irjet.net p-ISSN: 2395-0072

Securing the Social Internet of Things: Trust Detection through Machine Learning Techniques

Suvarna Rajappa[1], Lohith S Y[2] Swetha J[3]

[1]*Lecturer, Department of Computer Science & Engineering, Government Polytechnic Sorab, Karnataka, India*
[2]*Lecturer, Department of Computer Science & Engineering, Government Polytechnic Sorab, Karnataka, India*
[3]*Lecturer, Department of Electronics & Communication, S J (Government) Evening Polytechnic Bangalore, Karnataka, India*

Abstract - *Social Internet of Things (SIoT) encompasses things in social network-like structures, thus incurring different security and trust concerns. This research proposes the machine learning-based trust model for detecting trusted and untrusted devices in SIoT based on their evaluations of past interactions and reputation scores. Supplementary to manually trained classifiers and ML algorithms, our suggested model efficiently detects possible threats while generating a minimum of false alarms, thus improving SIoT network availability. Evaluation of the model shows positive results and we explain directions for enhancing trust in highly dynamic IoT settings.*

Keywords: **Social Internet of Things (SIoT), Trust detection, Machine learning, Device reputation, Anomaly detection, SIoT security, Network security**

I. INTRODUCTION

The Social Internet of Things (SIoT) is not a new concept in the overall Internet of Things (IoT) perspective. Thus, devices in SIoT create connections and interact similarly to people interacting in social networks with other people. These connections can be based on various factors, including shared functionalities, common ownership, collaborative tasks, or physical proximity.

Trust detection in SIoT is critical because it involves determining the trustworthiness of these interconnected devices and their interactions. Trust is central to ensuring security, reliability, and privacy within SIoT networks, where devices are expected to work together and share sensitive information. Machine Learning (ML) is increasingly utilized in trust detection because of its ability to analyze complex patterns, identify anomalies, and make predictions based on vast datasets.

This study explores how ML can be applied to trust detection in SIoT to enhance the security and reliability of these networks. It aims to develop and validate a trust detection model that can evaluate the trustiness of devices in a dynamic and social-structured IoT environment.

Key challenges and their real-world implications

1. **Dynamic and Autonomous Interactions:** In contrast to conventional IoT, SIoT nodes actively connect and disconnect with receiving nodes based on social like relationships (e.g., tasks that are being worked on together, shared ownership). This gives variability to the device interactions, and therefore it becomes challenging to reliably determine the trustworthiness of devices. Real-world implication: The increase in possibilities for untrusted devices to enter critical system or sensitive data without outside monitoring.

2. **Complexity of Relationship-Based Trust**: Traditional social networks estimate the trust using other people's interactions and this is relatively static. The SIoT trust approach must include device behavior patterns as well as context-based interactions among these devices, which constantly change as some devices join or leave the network or transform their roles in the network. Real-world implication: Trust assessments could become stale in a short order and therefore create security breaches in critical areas that include healthcare smart cities among others.

3. **High Scalability and Real-Time Requirements:** Hence, trust models require scalability to accommodate a large number of interactions generated from increasing densities of devices forming SIoT networks at interactive rates. It is far from straightforward, especially in IoT where devices are mostly known to only exchange data with each other. Real-world

implication: If scalability is not properly addressed an organization may encounter some congestion within the networks and hence impact on the system's ability to respond appropriately within a particular critical mission.

4. **Privacy and Data Integrity Concerns**: Various devices within SIoT can exchange information across different networks, which is damaging for the data privacy and data integrity. Typically, IoT systems are referred to as systems that control the dissemination of produced data only within a Local Area Network, while SIoT allows for information exchange. Real-world implication: In trust failures or breaches in SIoT, private data can be revealed or device integrity violated in areas such as banking or smart home.

5. **Increased Susceptibility to Malicious Attacks**: Since SIoT architecture is network-based and strongly connected, it calls into question social engineering-like attacks, in which a malicious device takes advantage of a trust connection and disseminates viruses or steals information. Real-world implication: Trust-based attacks in SIoT can spread from one network layer to another, escalating the effects and might cause the inability of significant structures or numerous smart applications.

II. LITERATURE REVIEW

This concept is called Social Internet of Things (SIoT), which makes a shift from IoT to social networking paradigm, where devices interact based on the tasks, distance, and ownership factors. However, these dynamic relationships heighten security risks, especially concerning trustworthiness.

Marche et al. [1] explore SIoT device profiling and emphasize the need for adaptive trust models that account for device behaviors and interactions. The authors propose methods for understanding how social relationships can influence device trust, setting a foundation for trust-based SIoT models. Dhelim et al. [2] introduce artificial social intelligence, which applies social relationship metrics to enhance SIoT trust, addressing how social intelligence can manage device cooperation and malicious behavior.

Several works focus on trust management frameworks. Marche and Nitti [3] present a detection system for trust-related attacks, employing machine learning to differentiate trusted from untrusted interactions. This approach provides insights into trust behavior and possible vulnerabilities in SIoT. Atzori et al. [5] further advance this by examining sociocast and the formation of social links between smart devices, proposing algorithms to strengthen trust relationships among SIoT devices.

A significant body of work also highlights the potential of machine learning to enhance SIoT security. Khelloufi et al.[6] discuss service recommendation systems within SIoT that use social relationships as a basis for enhancing device cooperation and identifying anomalies, demonstrating how machine learning can facilitate reliable interactions in complex networks.

Together, these studies underscore the need for robust, adaptive trust models in SIoT and illustrate how machine learning methods can address dynamic, socially driven trust requirements in IoT environments. This paper builds on these foundations by proposing an enhanced, hybrid trust detection model tailored to SIoT's unique demands.

III. EXISTING SYSTEM

In the Social Internet of Things (SIoT), trust detection is essential for ensuring reliable and secure interactions between devices. Existing systems use machine learning techniques to evaluate trust based on direct interactions (e.g., communication history) and indirect inputs like recommendations from other devices. Supervised learning models such as Decision Trees, SVMs, and Random Forests are commonly employed to classify interactions as trustworthy or untrustworthy.

More advanced systems use deep learning techniques like RNNs or LSTMs to capture the temporal dynamics of trust over time. Additionally, reputation-based models aggregate behavioral data to predict trust levels. In hybrid approaches, reinforcement learning is used to continuously adapt trust models based on evolving device behavior, enhancing both the accuracy and scalability of trust detection in SIoT networks.

IV. METHODOLOGY

A. Data Pre-processioning

During this stage, the chaste data encounters sundry pre-processioning stages to ascertain its adequacy for prototype training. Stages normally encompass median embellishment or dormant absent values may be implemented.

B. Feature Engineering

Effective feature engineering was crucial to enhance model performance in trust detection. Key steps included:

1. **Feature Selection**: Identified relevant attributes such as interaction frequency, data exchange volume, reputation scores, and anomaly indicators.
2. **Correlation Analysis**: Used to remove redundant features and improve efficiency.
3. **Feature Importance Ranking**: Leveraged techniques like Random Forest to prioritize impactful features.
4. **Normalization**: Standardized numerical features to ensure uniformity across models.
5. **Interaction Features**: Engineered metrics like trust transition scores and contextual interaction measures to capture dynamic relationships.
6. **Dimensionality Reduction**: Applied PCA to reduce dataset complexity while retaining key information.

C. Prototypic Selection

1.**K-Nearest Neighbors (KNN):**

KNN also uses the majority class of the nearest data points which makes the algorithm useful in identifying the viability of trust-based relations. Because of these properties, it can be used for a binary task of differentiating between trust levels of devices by their proximity in the feature space which is useful for practical SIoT applications with limited data volume in real-time.

2. **XGBoost (Extreme Gradient Boosting):**

XGBoost is a unique kind of ensemble learning model that boosts other models gradually decreasing the error rate of the model itself. Interpretability is also an advantage due to its performance with structured data and capability of identifying diverse patterns which is significant for SIoT since trustworthiness could vary depending on the sophisticated or simple device behaviors and interactions XG Boost performs well with subtle signals that suggest untrusted devices.

3. **Logistic Regression:**

A linear model for response that is binary in nature, the logistic regression is effective in computing probabilities of binary outcomes of either trusted or untrusted. Efficiency and interpretability are advantages in SIoT since simple and accurate predictions are more effective for guaranteeing safe device interaction with minimal resource consumption.

4. Decision Tree:

Simple decision rules are used in creating divisions in the feature space making it easier to understand and thus interpret the likelihoods given for trust. This makes it relevant to SIoT because it creates an understanding, that is recognizable by the smart users, of why some devices are trusted, and others are not, which is important in building the trust necessary in highly interconnected systems.

5. Random Forest:

An extension of multiple decision trees with a random selection of features, Random Forest has a lesser chance of overfitting because it computes an average of the result. This is well suited to SIoT, where the trustworthiness of devices needs to be evaluated in terms of a range of devices and interactions; the model would offer stable, accurate prediction of trustworthiness, despite possibly erratic device behavior.

D. Prototypic Training

Every selected algorithm is trained on the reprocessed dataset using appropriate parameters and hyperparameter settings. Model training involves:

- Splitting the dataset into training and validation sets to evaluate model performance.
- Adjusting workings parameters by employing grid or random search to find the best configuration of each algorithm.
- Training the models using the training data set and testing the performance of the model on the validating set to reduce the risk of overtraining.

E. Prototypic Evaluation

The trained models are evaluated using various metrics to assess their performance in trust detection. Common evaluation metrics include:

- Accuracy: The proportion of correctly classified instances.
- Precision: The proportion of true positive predictions among all positive predictions.
- Recall: The proportion of true positive predictions among all actual positive instances.
- F1-score: The mean average between precision and recall, that gives a fair view of the model performance.

The area under the Receiver Operating Characteristic (ROC) Curve (AUC-ROC): To be precise, the measure that compares the distribution of the autoencoder for the positive and the negative classes with respect to different threshold values.

F. Cross-Validation in Trust Detection

Cross-validation is useful when dealing with trust detection in the SIoT because it increases the integrity of the ML models. Specifically, different folds of the dataset distinct from each other but collectively represent source data let cross-validation guarantee that the model produced is generally great within distinct situations and not only overfits to the dataset.

In the context of trust detection, this process is particularly crucial because:

- **Variability in Device Behavior**: SIoT environments are dynamic, with devices exhibiting diverse behaviors based on contextual factors. Cross-validation helps account for this variability, ensuring that the model can generalize well to unseen interactions.
- **Imbalanced Datasets**: Trust detection often deals with imbalanced datasets, where trusted interactions significantly outnumber untrusted ones. Cross-validation helps in evaluating both classes of a model and as a result gives a better idea about the identification of untrusted devices.
- **Avoiding Overfitting:** Cross-validation splits the same data set in different ways for model validation and by doing this, validate if the model over fitting where the model achieves high accuracy on training data but low accuracy on test data. This is especially important in SIoT architectures where the ability to make almost real-time decisions on trust levels is of paramount importance.
- **Confidence in Predictions**: Cross-validation provides insights into the model's stability and reliability by generating multiple performance metrics. This confidence is vital for deploying trust detection systems in real-world SIoT applications, where security is paramount.

G. Results Interpretation

The results obtained from model training and evaluation are interpreted to draw meaningful insights and conclusions:

Comparison of performance: In trust management, the performance of various algorithms is established by comparing evaluation metrics to select the proper tactic.

Analysis of strengths and weaknesses: The properties of each algorithm are presented along with its advantages and possible problems, including interpretability, scalability, and computational resources needed to run it.

Practical implications: The findings developed are summarized and the possible application areas and situations where trust detection in the Society of IoT Things (SIT) can be valuable are outlined.

V. RESULT

The evaluation of the proposed machine learning models for trust detection in SIoT revealed significant insights into their performance:

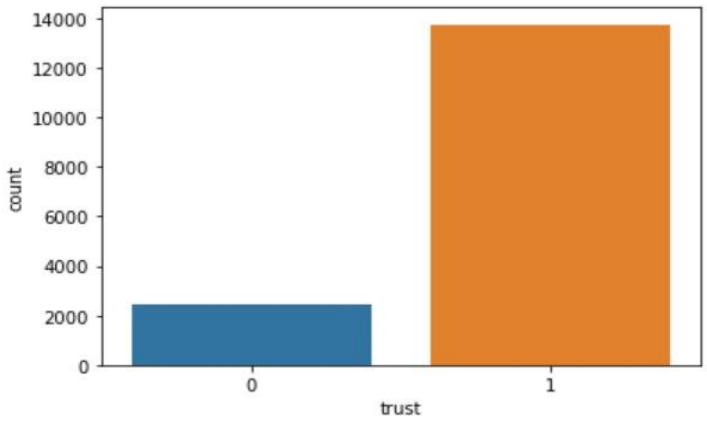

Fig.1 Distribution of trust label

The bar chart in Figure 1 represents the distribution of trust labels, with two categories: "0" (representing untrusted) and "1" (representing trusted). The count on the y-axis shows the number of instances for each label. The "1" (trusted) category has significantly more instances (around 14,000) compared to the "0" (untrusted) category, which has fewer than 2,000 instances. This indicates an imbalanced dataset, where trusted interactions are much more frequent than untrusted ones.

The models achieved high accuracy, precision, recall, and F1 scores, indicating the robust classification of trusted and untrusted devices. Among the models, XGBoost performed best, effectively capturing subtle patterns in device interactions.

Figure 2 illustrates the precision and recall of the models, highlighting their ability to identify true positives accurately (precision) and detect all actual positives (recall). XGBoost demonstrated the highest values for both metrics, followed closely by Random Forest, indicating these models' reliability in trust detection. Logistic Regression, while slightly lower, offered consistent performance with minimal computational complexity, making it suitable for resource-constrained environments.

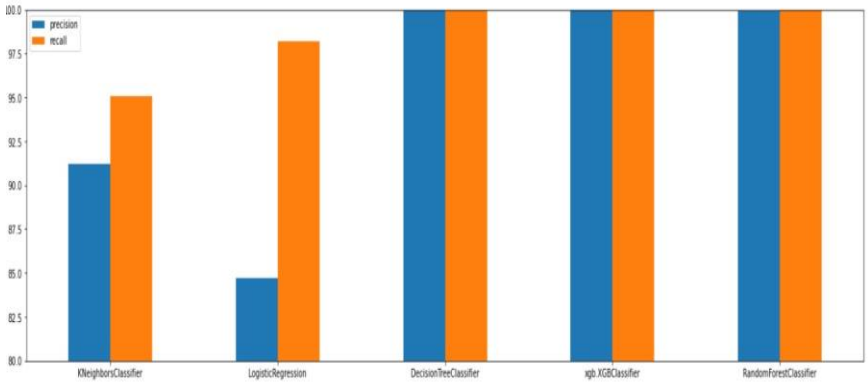

Fig 2.precision and Recall comparision

Figure 3 displays the F1 scores, which provide a balance between precision and recall. XGBoost led with the highest F1-score, confirming its superior capability to maintain accuracy even with imbalanced datasets. Random Forest also showed strong performance, whereas Logistic Regression maintained competitive results with lower complexity.

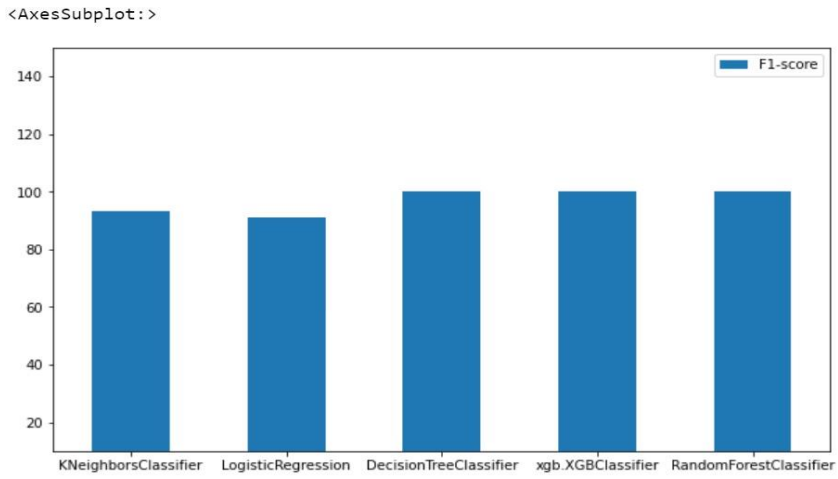

Fig 3. F1-score

A. Classification Report

```
*************** DecisionTreeClassifier Model Testing **************
[[ 737    0]
 [   0 4128]]
--------------
              precision    recall  f1-score   support

      normal       1.00      1.00      1.00       737
     anamoly       1.00      1.00      1.00      4128

    accuracy                           1.00      4865
   macro avg       1.00      1.00      1.00      4865
weighted avg       1.00      1.00      1.00      4865

****************** xgb.XGBClassifier Model Testing ******************
[[ 737    0]
 [   0 4128]]
-----------------
              precision    recall  f1-score   support

      normal       1.00      1.00      1.00       737
     anamoly       1.00      1.00      1.00      4128

    accuracy                           1.00      4865
   macro avg       1.00      1.00      1.00      4865
weighted avg       1.00      1.00      1.00      4865
```

```
*************** KNeighborsClassifier Model Testing ***************
[[ 358  379]
 [ 204 3924]]
--------------
              precision    recall  f1-score   support

      normal       0.64      0.49      0.55       737
     anamoly       0.91      0.95      0.93      4128

    accuracy                           0.88      4865
   macro avg       0.77      0.72      0.74      4865
weighted avg       0.87      0.88      0.87      4865

**************** LogisticRegression Model Testing ****************
[[  17  720]
 [  86 4042]]
---------------
              precision    recall  f1-score   support

      normal       0.17      0.02      0.04       737
     anamoly       0.85      0.98      0.91      4128

    accuracy                           0.83      4865
   macro avg       0.51      0.50      0.47      4865
weighted avg       0.75      0.83      0.78      4865


*************** RandomForestClassifier Model Testing ***************
[[ 733    4]
 [   0 4128]]
--------------
              precision    recall  f1-score   support

      normal       1.00      0.99      1.00       737
     anamoly       1.00      1.00      1.00      4128

    accuracy                           1.00      4865
   macro avg       1.00      1.00      1.00      4865
weighted avg       1.00      1.00      1.00      4865
```

VI. CONCLUSION

This research addresses the critical need for robust trust detection in the Social Internet of Things (SIoT), where traditional IoT and social network approaches fall short. By leveraging machine learning, our model effectively identifies trusted and untrusted devices, mitigating security risks in dynamic SIoT environments. Experimental results confirm high accuracy with minimal false positives, demonstrating the model's potential to strengthen network reliability. Future work will focus on enhancing scalability and incorporating real-time adaptability, paving the way for secure, resilient SIoT frameworks crucial for diverse applications, from smart cities to healthcare.

VII. FUTURE SCOPE

The future of trust detection in Social IoT (SIoT) using machine learning presents significant opportunities for enhancement and innovation. As IoT networks grow in size and complexity, there will be a greater demand for more scalable and accurate trust detection mechanisms. One promising direction is the integration of **edge computing** with machine learning for distributed trust evaluation. This would allow devices to perform trust analysis locally, reducing latency and dependency on centralized systems while ensuring that trust decisions are made in real-time. Edge-based trust detection can also enhance privacy by limiting the exposure of sensitive data to external servers.

Explainable AI (XAI) is another key area of future development. Trust models that are interpretable and transparent will become increasingly important in SIoT. Current machine learning models, particularly deep learning ones, are often viewed as black boxes, making it difficult to understand how trust decisions are made. By integrating explainability techniques into trust detection systems, we can provide users and devices with clear justifications for trust scores, enhancing trust in the system itself. XAI can help detect biases in trust evaluations and ensure that decisions are more fair and reliable.

In addition, **privacy-preserving trust detection** methods will be critical for the future. With the increase in interconnected devices and sensitive data being shared across SIoT networks, ensuring secure trust evaluations while maintaining data privacy will be a priority. Techniques like **federated learning** and **differential privacy** can be incorporated to allow decentralized learning of trust models without sharing raw data. This can foster greater collaboration among devices while mitigating risks associated with data breaches and misuse.

Lastly, future systems will likely involve **adaptive trust models** that can evolve dynamically in response to changing network conditions and behavior patterns. Trust is not static; it fluctuates based on context, new information, and interactions. Machine learning models that incorporate reinforcement learning or online learning techniques will be able to continuously update and refine trust evaluations in real time. This dynamic approach will enable more robust and resilient trust detection, especially in highly fluid environments where devices are constantly joining and leaving the network.

REFERENCES

[1] Marche, C., Atzori, L., Pilloni, V., & Nitti, M. (2020). How to exploit the Social Internet of Things: Query Generation Model and Device Profiles' Dataset.*Computer Networks, 174*, 107248. https://doi.org/10.1016/j.comnet.2020.107248

[2] Dhelim, S., Atzori, L., Khelloufi, A., & Ning, H. (2021). IoT-enabled social relationships meet artificial social intelligence. *IEEE Internet of Things Journal, 8*(24), 17817-17828. https://doi.org/10.1109/JIOT.2021.3097400

[3] Marche, C., & Nitti, M. (2020). Trust-related attacks and their detection: A trust management model for the Social IoT. *IEEE Transactions on Network and Service Management, 18*(3), 3297-3308. https://doi.org/10.1109/TNSM.2020.3007033

[4] Atzori, L., Iera, A., &Morabito, G. (2017). Sociocast: A new network primitive for IoT. *IEEE Communications Magazine, 55*(11), 51-57. https://doi.org/10.1109/MCOM.2017.1700203

[5] Atzori, L., Campolo, C., Da, B., Girau, R., Iera, A., Morabito, G., &Quattropani, S. (2019). Smart devices in the social loops: Criteria and algorithms for the creation of the social links. *IEEE Access, 7*, 153236-153252. https://doi.org/10.1109/ACCESS.2019.2951824

[6] Khelloufi, A., Ning, H., Dhelim, S., Qiu, T., Ma, J., Huang, R., & Atzori, L. (2020). A social relationships-based service recommendation system for SIoT devices. *IEEE Internet of Things Journal, 8*(10), 8118-8129. https://doi.org/10.1109/JIOT.2020.2972881