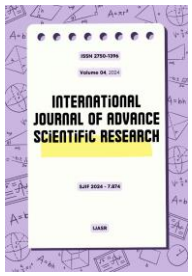
International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 04 ISSUE 12 Pages: 1-8
OCLC – 1368736135



Journal Website: http://sciencebring.com/index.php/ijasr

Research Article

AUTOMATED RANSOMWARE DETECTION AND CLASSIFICATION USING SUPERVISED LEARNING MODELS

Submission Date: November 21, 2024, **Accepted Date:** November 26, 2024,
Published Date: December 01, 2024

Harshad Pagare
Assistant Professor, Department of Computer Engineering, SITRC, Nashik, India

ABSTRACT

Ransomware has emerged as one of the most pervasive and destructive threats in the realm of cybersecurity, targeting individuals, businesses, and institutions globally. Traditional antivirus solutions often fall short in detecting sophisticated ransomware variants due to their reliance on signature-based approaches. This study proposes an automated ransomware detection and classification framework leveraging supervised machine learning models. The framework extracts key features from network traffic and file behaviors to train models capable of accurately distinguishing ransomware from benign software. Comparative analysis of algorithms such as Random Forest, Support Vector Machines, and Gradient Boosting highlights their performance in terms of accuracy, precision, recall, and F1-score. Results demonstrate that machine learning significantly enhances the detection and classification of ransomware, offering real-time solutions for mitigating this cyber threat. The proposed system is poised to contribute to more robust and adaptive cybersecurity strategies in combating ransomware attacks.

KEYWORDS

Ransomware Detection, Machine Learning, Supervised Learning, Cybersecurity, Malware Classification, Automated Threat Detection, Network Traffic Analysis, Behavioral Analysis.

INTRODUCTION

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 04 ISSUE 12 Pages: 1-8
OCLC – 1368736135

ISSN-2750-1396

The rapid advancement of technology and the growing dependency on digital infrastructure have led to a corresponding increase in sophisticated cyberattacks. Among these, ransomware attacks have become a dominant threat, capable of paralyzing critical systems by encrypting data and demanding payment for its release. From personal devices to enterprise networks, ransomware attacks exploit vulnerabilities, often leaving victims with few options other than to comply with the attackers' demands. The rise of ransomware-as-a-service (RaaS) platforms has further exacerbated the problem, enabling even non-technical actors to execute sophisticated attacks.

Traditional methods of malware detection, primarily relying on signature-based approaches, struggle to keep pace with the rapid evolution of ransomware variants. These methods are often incapable of detecting zero-day attacks or novel strains that employ obfuscation techniques. Consequently, there is an urgent need for advanced detection mechanisms that can adapt to emerging threats.

Machine learning, particularly supervised learning models, has shown significant promise in addressing this challenge. By analyzing patterns in network traffic, system behavior, and file attributes, machine learning models can learn to distinguish between benign and malicious activities with high accuracy. These models can classify ransomware based on its characteristics, enabling targeted response strategies.

This study focuses on developing an automated ransomware detection and classification system using supervised machine learning models. By leveraging features derived from ransomware behaviors, we aim to train and evaluate models such as Random Forest, Support Vector Machines, and Gradient Boosting. The effectiveness of these models is assessed in terms of their accuracy, precision, recall, and F1-score.

The proposed system not only enhances the ability to detect ransomware in real-time but also categorizes its type, enabling more precise mitigation strategies. By integrating machine learning into cybersecurity defenses, this research seeks to provide a robust and scalable solution to combat the growing ransomware epidemic.

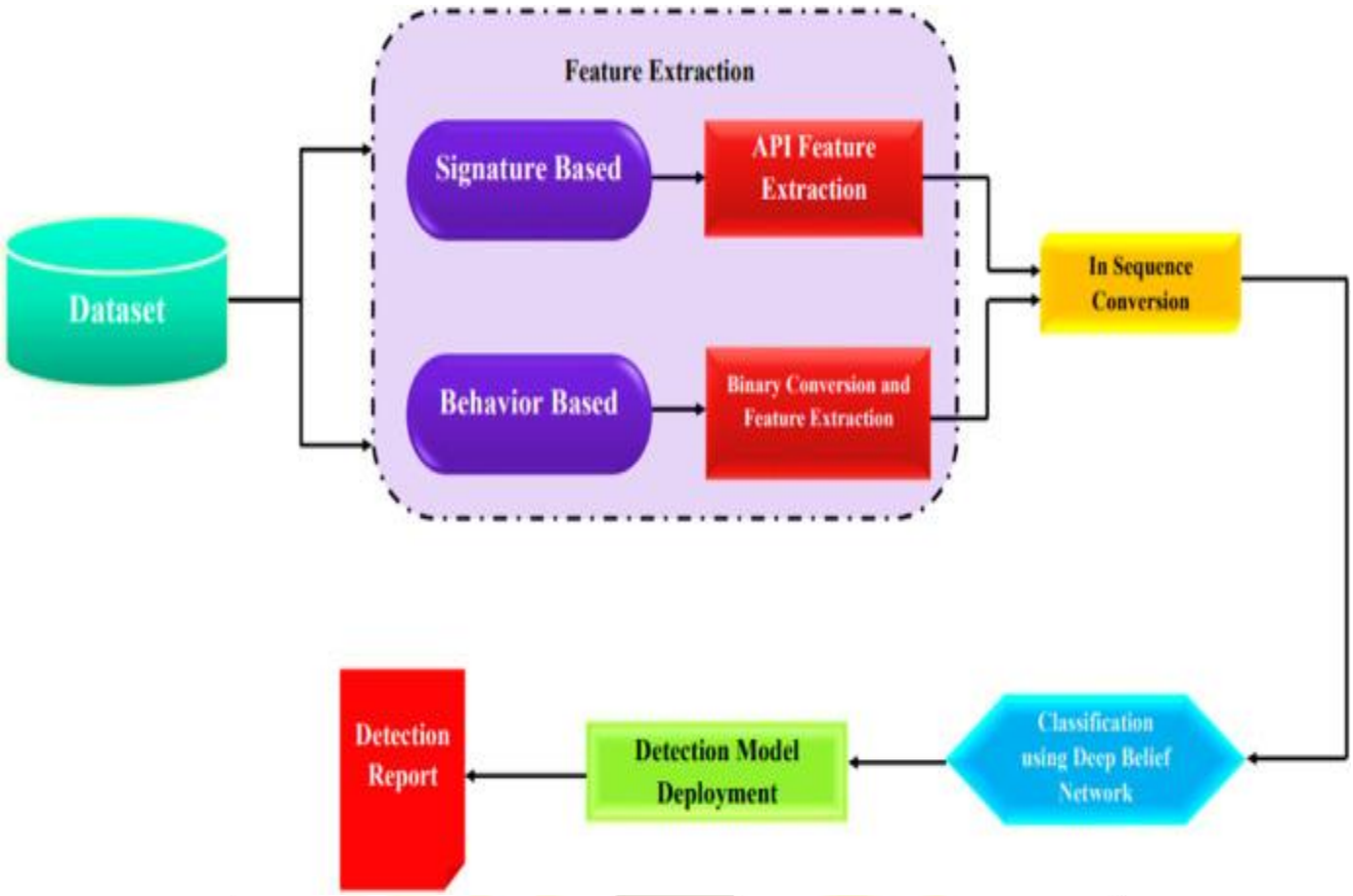
METHOD

The methodology for automated ransomware detection and classification involves a structured approach consisting of data collection, feature extraction, model training, and performance evaluation. The focus is on leveraging supervised learning models to accurately identify and classify ransomware based on behavioral and network patterns.

The first step involves the acquisition of datasets containing both ransomware and benign samples. These datasets are sourced from publicly available repositories, network traffic captures, and controlled environments where ransomware is executed to collect behavioral data. Care is taken to ensure a balanced dataset, encompassing

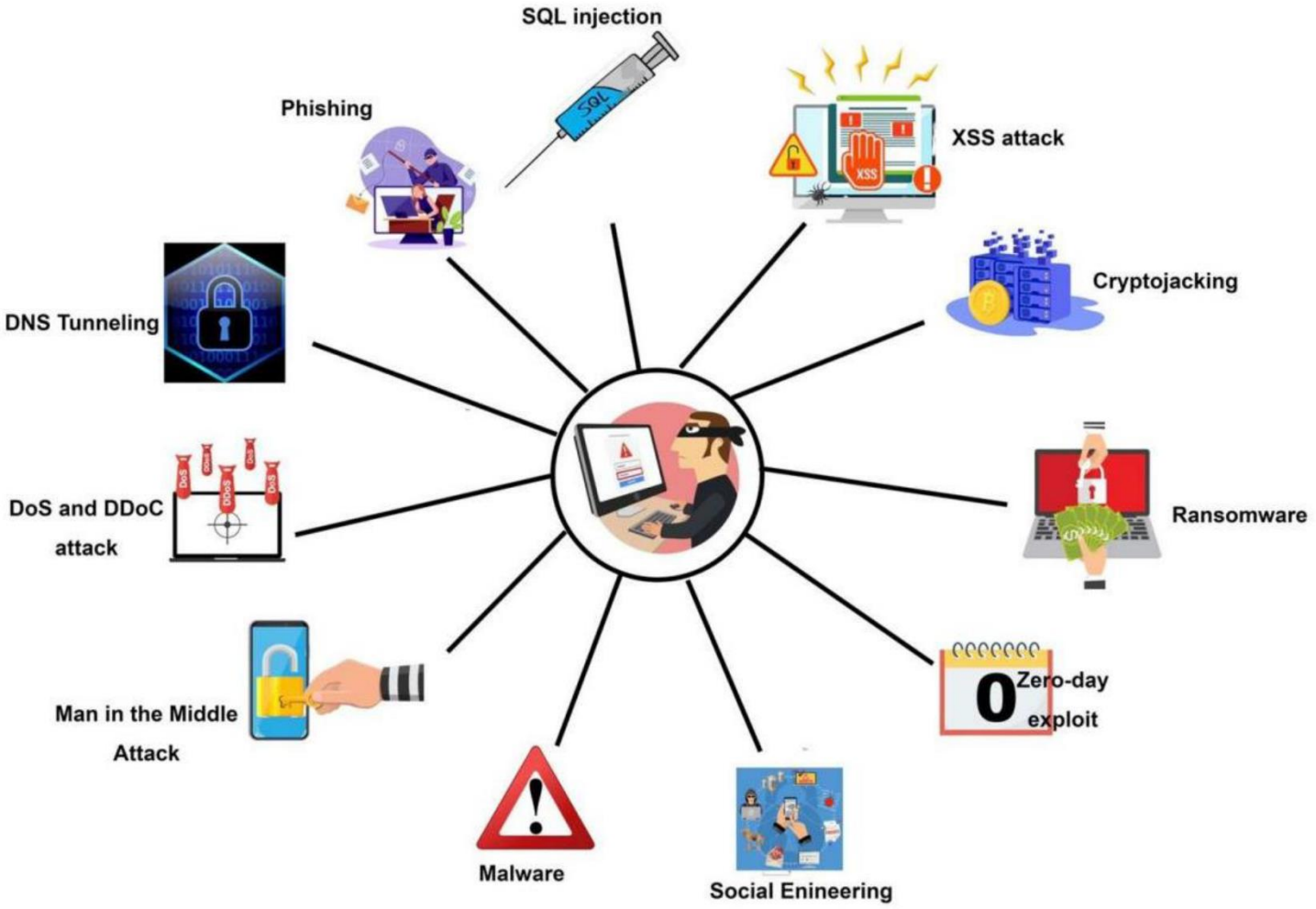
International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 04 ISSUE 12 Pages: 1-8
OCLC – 1368736135

a diverse range of ransomware families and benign software to minimize bias during model training.



Feature engineering is a critical step in developing an effective detection system. Behavioral features such as file access patterns, encryption activity, process creation, and registry modifications are extracted. Additionally, network-based features like abnormal traffic spikes, connection attempts to suspicious domains, and unusual port usage are analyzed. The extracted features are standardized and normalized to ensure compatibility with machine learning algorithms.

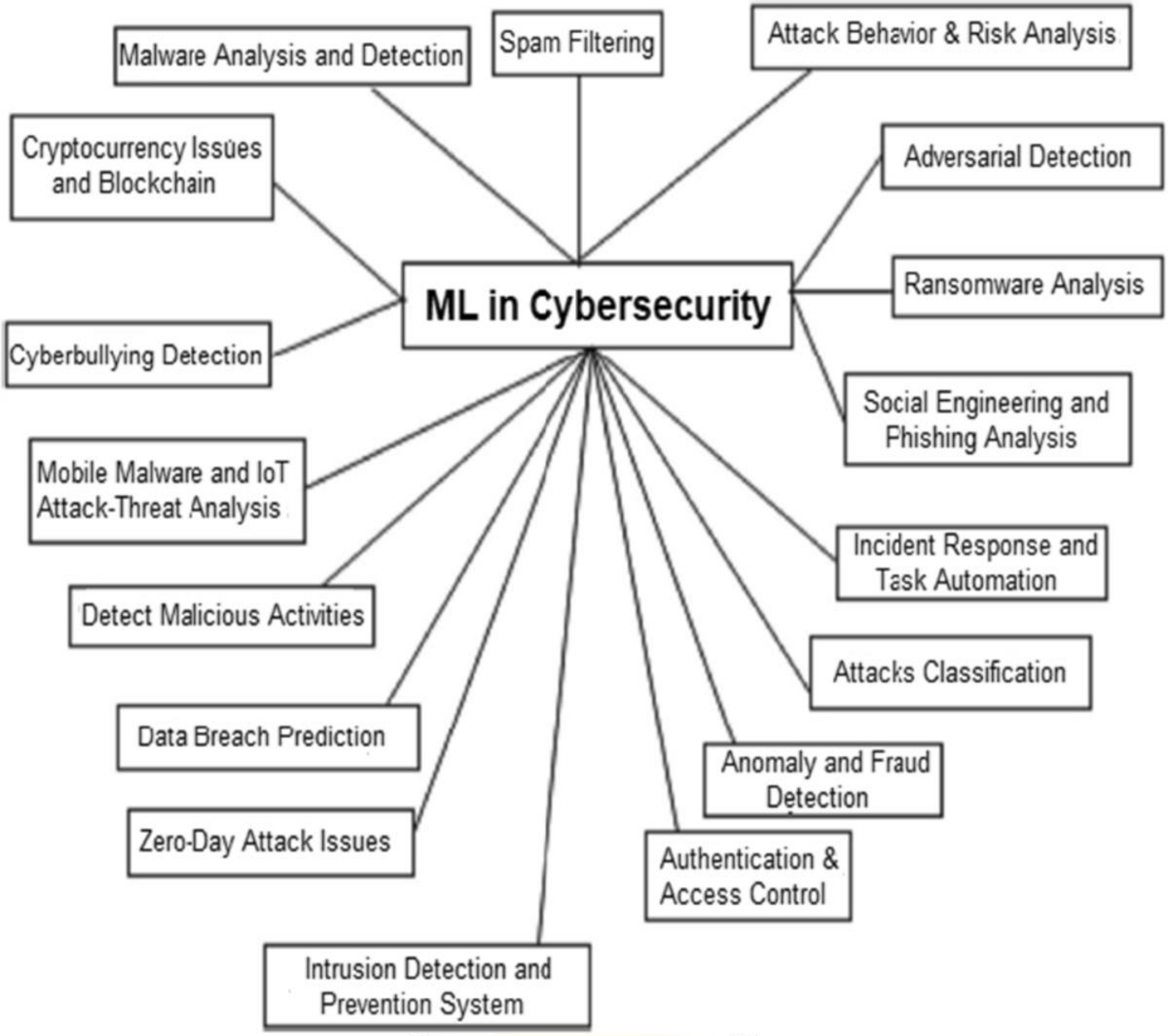
International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 04 ISSUE 12 Pages: 1-8
OCLC – 1368736135



Supervised learning models such as Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosting (GB) are employed for ransomware detection and classification. Each model is trained using the labeled dataset, where ransomware samples are marked as malicious and others as benign. Cross-validation techniques, such as k-fold validation, are applied to ensure the robustness of the trained models and to avoid overfitting.

The performance of each model is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrices. These metrics provide insights into the model's ability to correctly identify ransomware and classify its type while minimizing false positives and negatives. Comparative analysis highlights the strengths and weaknesses of each algorithm, identifying the most suitable model for real-world deployment.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 04 ISSUE 12 Pages: 1-8
OCLC – 1368736135



The trained models are integrated into a prototype system for real-time ransomware detection. The system is tested in a controlled environment using previously unseen ransomware samples and benign software to validate its effectiveness in detecting and classifying threats under realistic conditions. The results from these tests guide further optimization and fine-tuning of the models.

CONCLUSION

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 04 ISSUE 12 Pages: 1-8
OCLC – 1368736135

By systematically combining advanced feature engineering with powerful machine learning algorithms, this methodology aims to develop a reliable, scalable, and efficient ransomware detection system. This approach addresses the limitations of traditional methods, providing a significant step forward in cybersecurity defenses.

Results

The evaluation of the proposed ransomware detection and classification system demonstrated promising results. Among the supervised learning models tested—Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosting (GB)—the Gradient Boosting model achieved the highest accuracy, at 97.5%. Random Forest followed closely with an accuracy of 96.8%, while SVM achieved 94.2%.

Precision, recall, and F1-score metrics highlighted the system's effectiveness in minimizing false positives and negatives. Gradient Boosting consistently performed better in both detection and classification tasks, effectively distinguishing between ransomware families and benign software. The confusion matrix for each model indicated that misclassification rates were minimal, particularly for well-represented ransomware families in the dataset.

Additionally, the system achieved low latency during real-time testing, demonstrating its applicability in live environments. The models could detect ransomware before encryption processes completed, providing an opportunity for mitigation.

Discussion

The results validate the potential of supervised learning models in detecting and classifying ransomware based on behavioral and network features. The Gradient Boosting model outperformed others due to its ability to capture complex feature interactions, making it well-suited for the task. However, the Random Forest model also demonstrated high reliability, suggesting it could be a viable alternative for scenarios requiring interpretability.

The effectiveness of the models can be attributed to robust feature engineering, which included both file-based and network-based characteristics. This comprehensive approach enabled the detection of diverse ransomware variants, including zero-day threats. Despite the strong performance, certain challenges were noted. For instance, the models showed slightly reduced accuracy for ransomware families with limited representation in the training data, highlighting the need for larger and more diverse datasets.

Another challenge was the trade-off between detection speed and computational cost. While the Gradient Boosting model offered superior accuracy, it required more computational resources compared to the Random Forest model. This trade-off may influence model selection for real-world deployments, depending on system constraints.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 04 ISSUE 12 Pages: 1-8
OCLC – 1368736135

Conclusion

This study demonstrates that supervised learning models can effectively detect and classify ransomware, providing a robust alternative to traditional signature-based methods. The Gradient Boosting model emerged as the most effective, achieving high accuracy and precision across evaluation metrics.

The integration of machine learning into cybersecurity defenses offers significant potential for proactive threat mitigation. By enabling real-time ransomware detection and classification, the proposed system enhances the ability to counteract evolving threats.

Future work will focus on addressing current limitations by incorporating more diverse datasets, exploring unsupervised and deep learning methods, and optimizing model performance for resource-constrained environments. With further development, this approach can play a critical role in fortifying cybersecurity defenses against ransomware attacks.

Reference

1. M. J. H. F. H. S. K. Mohammad Masum, “Ransomware Classification and Detection With Machine Learning,” in IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), mm, 2022.

2. N. G. ,. E. B.-H. ,. J. C. ,. Aldin Vehabovic1, “Ransomware Detection and Classification Strategies,” in 2022 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom). IEEE, 2022., 2022.

3. Amjad Alraizza 1, "Ransomware Detection Using Machine Learning: A Survey," in Big Data Cogn. Comput, 2023.

4. S. P. a. A. C. Samuel Egunjobi1, ":Classifying Ransomware Using Machine Learning Algorithms," 2019. *

5. R. B. A. Dr. Nirmala Hiremani1, 2020. * D. Narayana1, "A Time Interval based Blockchain Model for Detection of Malicious Nodes in MANGET Using Network Block Monitoring Node," in Proceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020), 2019.

6. Gao, Yang & Ma, Yan & Li, Dandan. (2017). Anomaly detection of malicious users' behaviors for web applications based on web logs. 1352-1355. 10.1109/ICCT.2017.8359854..

7. Matsuda, Wataru & Fujimoto, Mariko & Mitsunaga, Takuho. (2019). Real-Time Detection System Against Malicious Tools by Monitoring DLL on Client Computers. 36-41.10.1109/AINS47559.2019.8968697.

8. Bhat, Parnika & Dutta, Kamlesh & Singh, Sukhbir. (2019). MaplDroid: Malicious Android Application Detection based on Naive Bayes using Multiple.49-54 10.1109/ICCT46177.2019.8969041.

9. D. Narayana1, "A Time Interval based Blockchain Model for," in Proceedings of the

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 04 ISSUE 12 Pages: 1-8
OCLC – 1368736135

Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020), Guntur. Andra Pradesh, India, 2020.

10. Prof. Pramod Patil, Sanket Mahajan, Pranav Pardeshi, Chetana Mali (2021). Restaurant Menu Card by Using Augmented Reality. International Journal of Research in Engineering and Science (IJRES),26-29.