OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

Profile Identification in Social Network using ML and NLP

Ms. K. Hareesh[1], Ms. M. Anitha[2], Mr. D. Anil[3]

[1] Asst. Professor, Dept. of MCA, SRK Institute of Technology, Vijayawada-521108, Andhra Pradesh, India.

[2] Asst. Professor and Head, Dept. of MCA, SRK Institute of Technology, Vijayawada-521108, Andhra Pradesh, India.

[3] M. Tech Student, Dept. of MCA, SRK Institute of Technology, Vijayawada, Andhra Pradesh-521108, India.

harish@gmail.com [1], anitha3harshak@gmail.com [2], iamanil1252@gmail.com [3]

Abstract: ***In the contemporary landscape of pervasive social media usage, the identification of fake profiles stands as a crucial element in preserving online security. This research delves into the realm of machine learning algorithms, focusing on the comparative analysis of LightGBM and SVM for the task of detecting fake accounts on social media platforms.***

Our study harnesses the power of these two algorithms, both known for their robust capabilities, and evaluates their performance against each other. The research utilizes a diverse set of features derived from user profiles, posting behavior, and linguistic patterns, with Natural Language Processing (NLP) techniques applied to extract nuanced insights from textual content.

The outcomes of our investigation reveal that LightGBM and SVM exhibit superior accuracy in the identification of fake profiles. Further exploration involves fine-tuning hyperparameters, optimizing feature engineering, and experimenting with ensemble methods to enhance the overall efficacy of the models.

As the social media landscape continually evolves, our findings contribute valuable insights into the strengths of Light GBM and SVM in addressing the dynamic challenges of fake profile detection. This research not only provides a foundation for practitioners to make informed decisions in implementing detection systems but also offers a pathway for future advancements in refining machine learning models to combat emerging threats in the realm of online social interactions.

I. INTRODUCTION

It is now simple to use the Internet to obtain information from anywhere in the world. Growing need for social networks lets people get a lot of user data. These sites' large volumes of data draw false users.Twitter is fast turning into a source of real-time user data. Twitter allows users to express news, opinions, and emotions. Politics, news, and significant events all provide room for debate. Tweets are transmitted quickly to a user's followers, hence enabling them to disseminate information more broadly. The growth of OSNs has made it more important to track and study social media users' actions. Anyone unacquainted with OSNs might be easily duped by fraudsters. It is also required to control OSN spammers and advertising.

Recent social media spam detection piqued scholars' interest. Social network security finds spam detection to be difficult. To safeguard consumers from damaging attacks and safeguard their privacy, OSN site spam must be recognized Spammers' dangerous strategies undermine real-world communities. Twitter spammers push spontaneous comments, gossip, and false information. Spammers maintain mailing lists and send out spam messages haphazardly to further their goals via ads and other techniques. These actions disturb non-spammers. It also damages the standing of OSN systems. Developing a spam detection system is absolutely essential to stop their harmful actions

Twitter spam detection has been the subject of several research. Many surveys have been held to gauge the present state of false user identification on Twitter. Tingmin et al. examine novel Twitter spam detection mechanisms. The given study contrasts current techniques. By comparison, looked into Twitter spammers' conduct. The literature review of the study recognises Twitter spammers. Though many studies have been done, the literature still falls short. We investigate sophisticated Twitter spammer detection and false user identification to close the gap. This study also offers a taxonomy of Twitter spam detection techniques and information on present developments.

This study classifies Twitter spam detection techniques and offers a taxonomy by means of identification. We discovered four spamming reporting techniques that might help identify false user IDs for classification. False information, URL-based spam detection, trending topic identification, and phoney user identification help to identify spammers. By contrasting their goals and outcomes, Table 1 helps consumers grasp the relevance and

effectiveness of current techniques. Table 2 contrasts characteristics of Twitter spam detection. Readers should be able to get different spammer detection data all in one location thanks to this poll.

II.LITERATURESURVEY

The most important aspect of software development is literature study. The next step is to select a language and operating system for tool development after deciding the time element, economy, and corporate traffic redundancy removal. Programmers need much of outside assistance once they begin building the tool. Senior programmers, books, and websites provide this help. We must understand the following ideas before creating the suggested system

Today, Twitter spam is a hot problem, and C. Chen et al. presented statistical models to identify dispersed spam. Late efforts have used tweet metrics to relate AI algorithms for Twitter spam placement. Tweets operate as a data index,however spam tweets' factual properties fluctuate with period, reducing the presentation of AI-built classifiers. "Twitter Spam Drift" describes this issue. We initially analyse over one million spam and non-spam tweets' quantifiable properties to resolve this issue. A new Lfun conspiracy is suggested. Unlabelled spam tweets will be consolidated into classifier preparation. Numerous tests evaluate the proposed plan. The findings demonstrate the Lfun strategy can increase spam detection.

C. Buntain, J. Golbeck suggest Automatically spotting false news in well-liked Twitter threads Web-scale data makes it more difficult for experts to assess and handle a major portion of the false material, or "phoney news", By learning how to forecast accuracy assessments in two validity-cantered Twitter datasets—CREDBANK improves accuracy, for example—this article creates a method for computerizing counterfeit news location on Twitter.

Both non-rumors and rumours This lets us Twitter place BuzzFeed's false news dataset and models against publicly reinforced experts beat models based on journalists' assessment and models on a pooled dataset of openly supported workers and writers. All three released balanced datasets. A component investigation then locates expected variables for journalistic accuracy assessments and publicly supported ones, hence enabling comparisons to prior results. Chen et al.'s study on machine learning-based streaming spam tweet identification and Twitter virality.

Twitter attracts more spammers. Spammers promote harmful sites or services on Twitter by sending negative tweets. Scientists propose many components to fight spammers. AI approaches for Twitter spam location are the focus of recent development. Regardless, tweets are retrieved beautifully, and Twitter offers designers and analysts the Issuing API to access tweets continually.

A comprehensive evaluation of AI-generated gushing spam identification algorithms is lacking. Presenting valuations on three separate data, feature, and ideal shares let us overcome any hurdle. To find spam in tweets, here are 12 lightweight features. Spam placement became a component space duplicate layout issue that can be explained by routine AI computations. We evaluated components' effects to Spam recognition employs non-spam to spam %, discretisation preparing data size, time-related data, data testing, and artificial intelligence computations. Results indicate that a decent location system should examine, include, and model since leaking spam tweet detection remains a significant problem. M. Bouguessa, F. Fathaliani propose Model-based spam detection in interpersonal organisations We address the issue of detecting spammers in informal groups from a mix displaying perspective in this research and provide a principled unassisted technique to identify spammers.

Every informal community client is first interviewed with an element vector that reflects its behaviour and relationships with other members. Next, using the analysed clients Highlight vectors, we present a quantifiable Dirichlet circulation method to differentiate spammers. The suggested system may naturally identify spammers from legitimate customers, while solo alternatives require human intervention to define casual edge settings. Additionally, our strategy is adaptable to other online social media platforms. We examined Instagram and Twitter data to prove the method's efficacy.

C. Meda et al. suggest Twitter traffic spam detection: Uneven backwoods and component inspection system Law enforcement has to investigate publicly available data and use effective methods to handle problematic data. In real life, law enforcement agencies monitor Twitter and profile accounts. Unfortunately, some online users use micro blogs to harass others or propagate malware.

Characterising clients and spammers helps reduce Twitter traffic by useless content. Popular Twitter client datasets are analysed. The Twitter dataset has 54 characteristics identifying clients as spammers or real customers. Exploratory results show better highlight testing is viable.

III.EXISTING SYSTEM:

Twitter is fast turning into a source of real-time user data. Tweets reach a user's followers instantly, hence enabling them to disseminate information more broadly. Studying and monitoring social media users' activity has grown increasingly crucial with OSNs' growth. Those unacquainted with OSNs might be easily duped by fraudsters. It is also required to control OSN spammers and advertising.

Twitter has rapidly become an online source for acquiring real-time information about users. When a user tweets something, it is instantly conveyed to his/her followers, allowing them to outspread the received information at a much broader level. With the evolution of OSNs, the need to study and analyze users' behaviors in online social platforms has intensied. Many people who do not have much information regarding the OSNs can easily be tricked by the fraudsters. There is also a demand to combat and place a control on the people who use OSNs only for advertisements and thus spam other people's accounts.

Identifying fraudulent profiles on social media platforms, particularly Twitter, has emerged as an essential endeavor in the context of misinformation and digital manipulation. By leveraging

a blend of Machine Learning (ML) and Natural Language Processing (NLP), researchers and developers have created systems capable of effectively detecting dubious or counterfeit Twitter accounts.

A prevalent method involves gathering data through the Twitter API or utilizing publicly accessible datasets, such as the 'Fake Profiles Dataset' available on Kaggle, which encompasses user details including tweet content, follower count, account age, and engagement statistics.

The initial phase involves data preprocessing, where NLP techniques are employed to refine tweets by eliminating extraneous elements like URLs, hashtags, and emojis, followed by processes such as tokenization and lemmatization.

IV.PROPOSED SYSTEM

This paper assesses techniques for spotting Twitter spam. Techniques for detecting spam on Twitter are categorised in a taxonomy, classifying them by their ability to detect fake content, URL spam, trending topic spam, and phoney users. The presented methodologies are compared using user, content, graph, structural, and temporal factors.

In the existing system no accurate spam detection system that why lot of spam account could not be identified in this way lots of carpeted data was coming in to the social network.

Although systems for detecting fake profiles utilizing Machine Learning (ML) and Natural Language Processing (NLP) have demonstrated considerable potential, they are accompanied by various drawbacks and constraints. A significant issue is the scarcity of high-quality labeled datasets, which complicates the effective training of models, particularly when addressing sophisticated fake profiles that closely resemble genuine users.

Furthermore, models frequently experience bias and overfitting, especially if the training data lacks diversity or fails to represent real-world behaviors across different regions, languages, or user demographics.

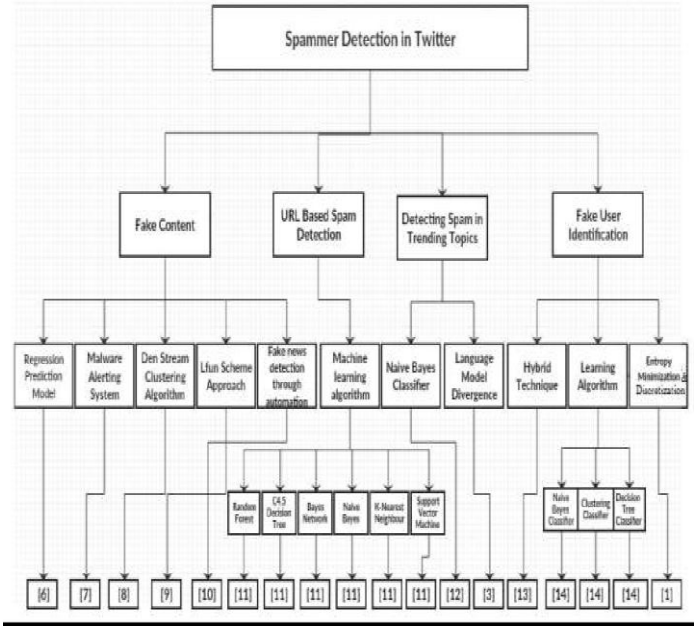
NLP-based systems also encounter challenges in processing multilingual content, sarcasm, coded language, or evolving slang that fake profiles may employ to evade detection.

Another critical issue is the evasion strategies utilized by bots and malicious users many fake accounts are crafted to seem authentic by mimicking real user behaviors or employing AI to produce realistic text, which can deceive even advanced models.

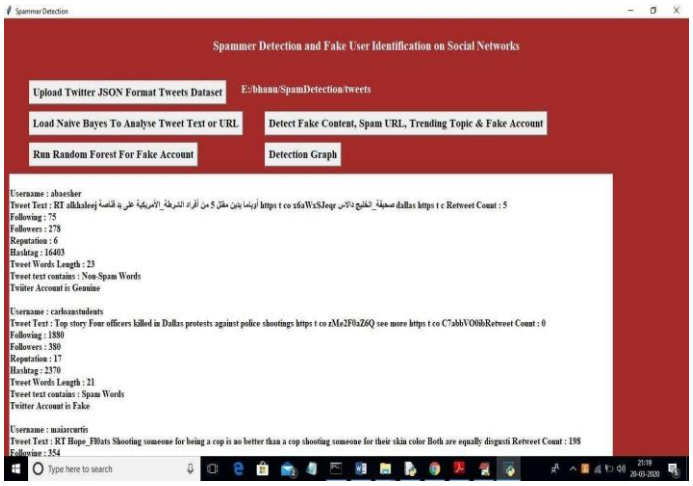
Additionally, these detection systems often lack transparency, meaning they may identify a profile as fake without offering clear justification, complicating the rationale behind decisions in sensitive situations such as moderation or user bans.

Privacy and ethical dilemmas also emerge when surveilling users' activities and personal information, particularly if data is gathered without consent.

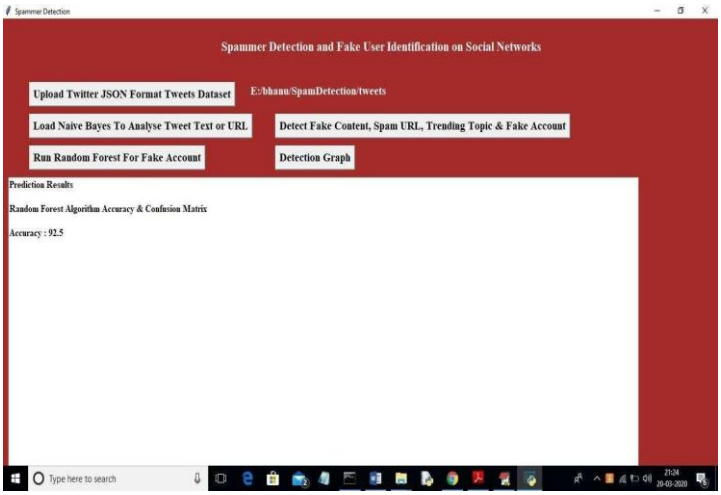
V.SYSTEM ARCHITECTURE



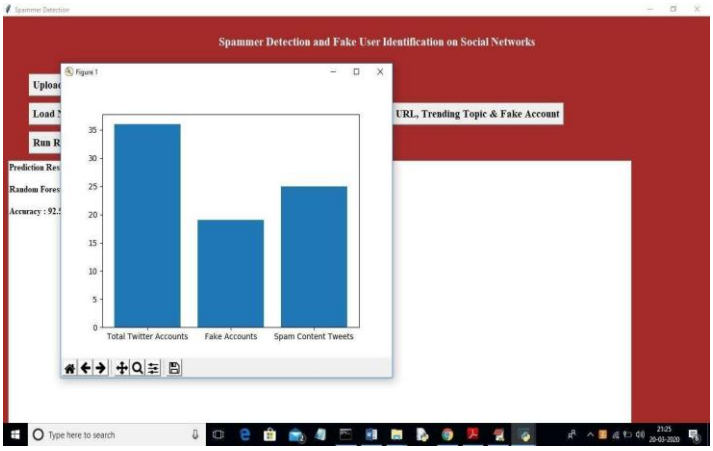
VI.RESULTS:



Above screen shows all twitter dataset characteristics and their analysis to find spam tweets. The upper text box shows values like TWEET TEXT, FOLLOWERS, etc. with phoney or genuine accounts and spam or non-spam words separating tweet records with empty lines. Train a random forest classifier using extracted tweet characteristics by clicking 'Run Random Forest Prediction' to forecast/detect false or spam accounts for future tweets. Above, scroll down to get tweet details.



The next panel displays 92% random forest prediction accuracy. Total tweets, spam, and fake account graphs may be seen by clicking 'Detection Graph'.



The x-axis shows total tweets, false account, and spam words content, while the y-axis counts them.

VI.CONCLUSION

Apart from NLP, this suggestion applied machine learning techniques. These techniques let us easily locate fake social media accounts. Our study found fake accounts using Instagram data. We used NLP pre-processing and ML techniques including Random Forest and Gradient Boost to structure the profiles and examine the dataset. In this investigation, these learning strategies improved detection accuracy.

In summary, utilizing Machine Learning and Natural Language Processing to identify fraudulent profiles on Twitter presents a robust and scalable method to address the increasing risks posed by bots, spammers, and malicious entities on social media platforms. By analyzing user metadata alongside tweet content, it becomes feasible to uncover significant patterns that differentiate authentic users from automated or misleading accounts.

Employing supervised learning algorithms, NLP methodologies, and feature engineering can lead to high accuracy and dependability in classification tasks. This strategy not only bolsters platform security and user trust but also contributes to wider initiatives aimed at combating misinformation, safeguarding digital integrity, and fostering constructive online dialogue. Nevertheless, the ever-evolving and adaptive characteristics of fake accounts present persistent challenges, necessitating ongoing model enhancements, comprehensive datasets, and ethical supervision. As social media remains pivotal in global communication, such systems will be crucial in promoting safer and more genuine digital environment.

The identification of fraudulent profiles on Twitter through the utilization of Machine Learning (ML) and Natural Language Processing (NLP) signifies a significant and essential progression in the persistent struggle against digital manipulation and online deceit.

By methodically examining both user metadata such as account activity, creation date, and social connections and the linguistic characteristics of tweets, it becomes feasible to develop intelligent systems that can accurately differentiate between genuine and fraudulent users. The outcomes of such systems are not only advantageous for detecting bots and spammers but are also vital for safeguarding the wider information ecosystem.

In a time when social media shapes public perception, influences electoral results, and even affects global security, the identification and mitigation of inauthentic behavior is of utmost importance. The implementation of ML and NLP in this field enables large-scale, real-time detection, allowing platforms like Twitter to respond swiftly in curtailing the reach and impact of malicious entities

However, despite the encouraging results from these models, numerous challenges persist. Fake profiles and bots are becoming increasingly advanced, frequently imitating authentic user behavior to evade detection. This changing threat landscape necessitates that detection systems remain adaptable, continuously trained on current datasets, and robust against adversarial strategies. Furthermore, the ethical implications of such systems must be carefully considered. Incorrectly classifying a legitimate user as a fraudulent account can result in reputational damage, account suspensions, or the suppression of authentic voices. Consequently, transparency, fairness, and accountability must be integral to the creation and implementation of these models.

VII. REFERENCES:

[1] Praveen, S. P., Nakka, R., Chokka, A., Thatha, V. N., Vellela, S. S., & Sirisha, U. (2023). A novel classification approach for grape leaf disease detection based on different attention deep learning techniques. International Journal of Advanced Computer Science and Applications (IJACSA), 14(6), 2023.

[2] S. Gharge, and M. Chavan, "An integrated approach for malicious tweets detection using NLP," in Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT), Mar. 2017,pp. 435_438.

[3] Vellela, S. S., & Balamanigandan, R. (2024). Optimized clustering routing framework to maintain the optimal energy status in the wsn mobile cloud environment. Multimedia Tools and Applications, 83(3), 7919-7938.

[4] Elovici, Yuval, F. I. R. E. Michael, and Gilad Katz. "Method for detecting spammers and fake profiles in social networks." U.S. Patent 9,659,185, issued May 23, 2019

[5] Vellela, S. S., & Balamanigandan, R. (2023). An intelligent sleep-awake energy management system for wireless sensor network. Peer-to-Peer Networking and Applications, 16(6), 2714-2731.

[6] Vellela, S. S., & Krishna, A. M. (2020). On Board Artificial Intelligence With Service Aggregation for Edge Computing in Industrial Applications. Journal of Critical Reviews, 7(07).

[7] T. Wu, S. Wen, Y. Xiang, and W. Zhou, "Twitter spam detection: Survey of new approaches and comparative study," Comput. Secur., vol. 76, pp. 265_284, Jul. 2018.

[8] Vellela, S. S., & Balamanigandan, R. (2024). An

efficient attack detection and prevention approach for secure WSN mobile cloud environment. Soft Computing, 28(19), 11279-11293.

[9] Polasi, P. K., Vellela, S. S., Narayana, J. L., Simon, J., Kapileswar, N., Prabu, R. T., & Rashed, A. N. Z. (2024). Data rates transmission, operation performance speed and figure of merit signature for various quadurature light sources under spectral and thermal effects. Journal of Optics, 1-11.

[10] Vellela, S. S., Rao, M. V., Mantena, S. V., Reddy, M. J., Vatambeti, R., & Rahman, S. Z. (2024). Evaluation of Tennis Teaching Effect Using Optimized DL Model with Cloud Computing System. International Journal of Modern Education and Computer Science (IJMECS), 16(2), 16-28.

[11] Biyyapu, N., Veerapaneni, E. J., Surapaneni, P. P., Vellela, S. S., & Vatambeti, R. (2024). Designing a modified feature aggregation model with hybrid sampling techniques for network intrusion detection. Cluster Computing, 27(5), 5913-5931.

[12] Vellela, S. S., Malathi, N., Gorintla, S., Priya, K. K., Rao, T. S., Thommandru, R., & Rao, K. N. S. (2025, March). A Novel Secure and Scalable Framework for a Cloud-Based Electronic Health Record Management System. In 2025 3rd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) (pp. 131-135). IEEE.

[13] Vullam, N. R., Geetha, G., Rao, N., Vellela, S. S., Rao, T. S., Thommandru, R., & Rao, K. N. S. (2025, February). Optimized Multitask Scheduling in Cloud Computing Using Advanced Machine Learning Techniques. In 2025 International Conference on Intelligent Control, Computing and Communications (IC3) (pp. 410-415). IEEE.

[14] Vuyyuru, L. R., Purimetla, N. R., Reddy, K. Y., Vellela, S. S., Basha, S. K., & Vatambeti, R. (2025). Advancing automated street crime detection: a drone-based system integrating CNN models and enhanced feature selection techniques. International Journal of Machine Learning and Cybernetics, 16(2), 959-981.

[15] Vellela, S. S., Singu, K., Kakarla, L. S., Tadikonda, P., & Sattenapalli, S. N. R. (2025). NLP-Driven Summarization: Efficient Extraction of Key Information from Legal and Financial Documents. Available at SSRN 5250908.

[16] Vellela, S. S. (2024). A Comprehensive Review of AI Techniques in Serious Games: Decision Making and Machine Learning. A Comprehensive Review of AI Techniques in Serious Games: Decision Making and Machine Learning, International Journal for Modern Trends in Science and Technology, 10(02), 305-311.

[17] Mahmood S, Desmedt Y," Poster: preliminary analysis of google?'s privacy. In: Proceedings of the 18th ACM conference on computer and communications security", ACM 2011, pp.809– 812.

[18] Vellela, S. S., Manne, V. K., Trividha, G., Chaithanya, L., & Shaik, A. (2025). Intelligent Transportation Systems AI and IoT for Sustainable Urban Traffic Management. Available at SSRN 5250812.

[19] Vellela, S. S., Roja, D., Purimetla, N. R., Thalakola, S., Vuyyuru, L. R., & Vatambeti, R. (2025). Cyber threat detection in industry 4.0: Leveraging GloVe and self-attention mechanisms in BiLSTM for enhanced intrusion detection. Computers and Electrical Engineering, 124, 110368.

[20] Burra, R. S., APCV, G. R., & Vellela, S. S. (2024). Enhancing Ddos Detection Through Semi-Supervised Machine Learning: A Novel Approach for Improved Network Security. International Research Journal of Modernization in Engineering Technology and Science, 6.

[21] Ozbay, F.A. and Alatas, B., 2020. Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Statistical Mechanics and its Applications, 540, p.123174.

Author's Profile

Ms. M. Anitha is currently serving as an Assistant Professor and Head of the Department of Master of Computer Applications (MCA) at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District, Andhra Pradesh, India. She holds a Bachelor of Technology (B.Tech), a Master of Computer Applications (MCA), and a Master of Technology (M.Tech) in Computer Science and Engineering. With over 14 years of teaching experience at SRK Institute of Technology, she has been actively involved in both academic and administrative roles. Her primary areas of interest include Machine Learning using Python and Database Management Systems (DBMS). She is dedicated to fostering academic excellence and promoting practical learning in advanced computing technologies.

Mr. K. Hareesh Working as Assistant Professor, Dept. of MCA, in SRK Institute of technology in Vijayawada. He done with MCA, M. Tech. he has 4 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. His area of interest includes Machine Learning with Python, and DBMS.

Mr. D. Anil is an MCA Student in the Dept. of Computer Application at SRK Institute of Tech, Enikepadu, Vjd, NTR District. he has Completed Degree in B.Sc.(computers) from Nalanda degree College Vijayawada. His area of interest are DBMS and Machine Learning with Python