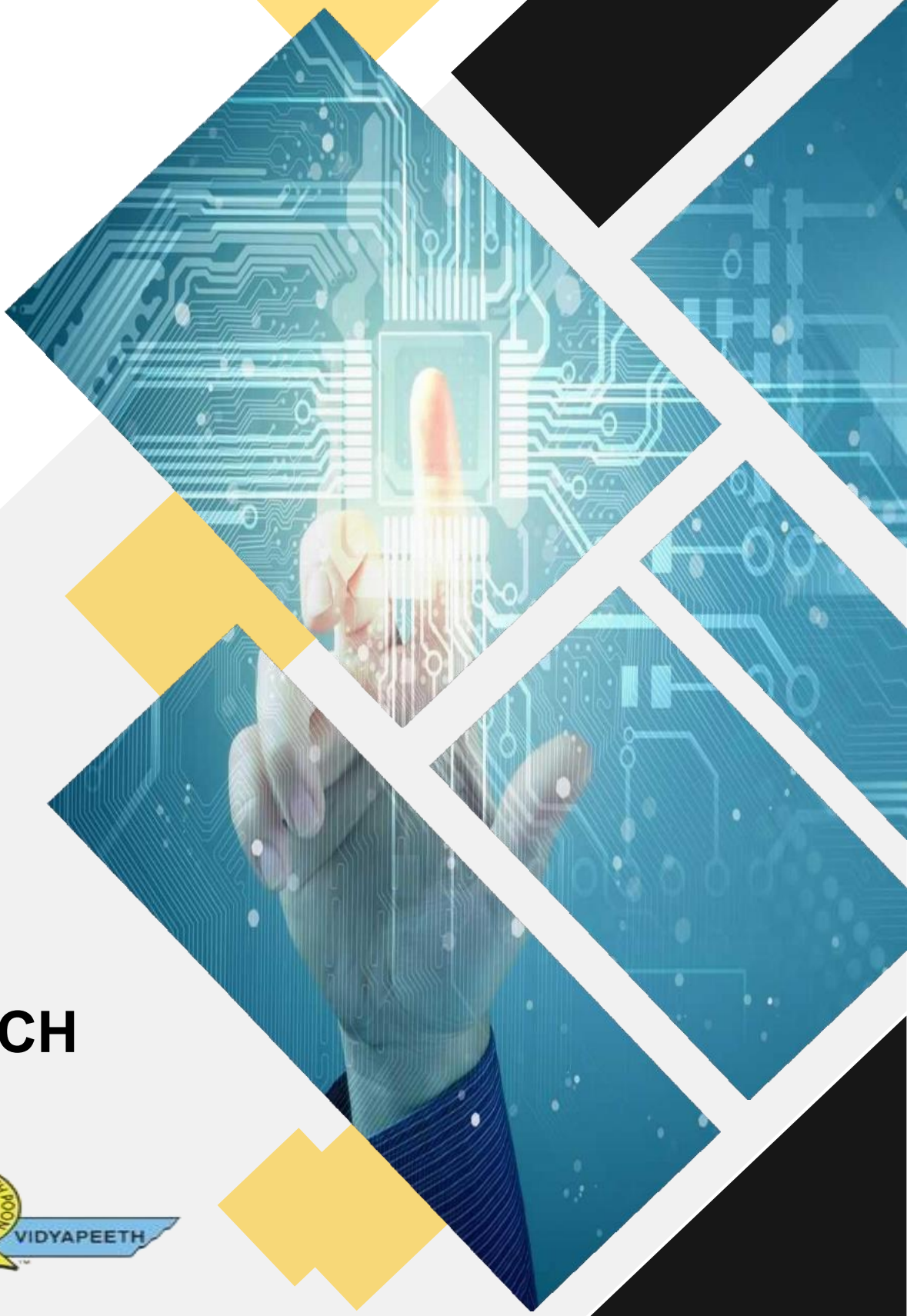
Annual Magazine
2023-24

Department of Information Technology



INFO TECH

BHARATI VIDYAPEETH
PUNE
BHARATI VIDYAPEETH POONA

VISION MISSION

Vision

Aiming to empower the students to demonstrate technical and operational excellence through a commitment to professionalism and bridge the gap between academy and industry to contribute to the betterment of the society and the nation.

Mission

M1: To create a culture that fosters excellence and combines rigorous academic study with the support of a diverse campus community.

M2: To enable students to recognize the importance of development by pursuing higher education for challenging and rewarding careers in Computer and Information Sciences and business in the evolving global community.

M3: To create competent and trained professionals and entrepreneurs in Information Technology who shall contribute towards the advancement of engineering, science and technology & development of the nation.

Programme Educational Objectives (PEOs)

PEO 1. Provide socially responsible, environment friendly solutions to Information technology related broad-based problems adapting professional ethics.

PEO 2. Adapt state-of-the-art Information Technology broad-based techniques to work in multidisciplinary work environments.

PEO 3. Solve broad-based problems individually and as a team member communicating effectively in the world of work.

Program Outcomes (POs):

PO 1. Basic and Discipline specific knowledge: Apply knowledge of basic mathematics, sciences and engineering fundamentals and engineering specialization to solve the engineering problems.

PO 2. Problem analysis: Identify and analyse well-defined engineering problems using codified standard methods.

PO 3. Design/ development of solutions: Design solutions for well-defined technical problems and assist with the design of system components or processes to meet specified needs.

PO 4. Engineering tools, Experimentation and Testing: Apply modern engineering tools and appropriate technique to conduct standard tests and measurements.

PO 5. Engineering practices for society, sustainability and environment: Apply appropriate technology in context of society, sustainability, environment and ethical practices.

PO 6. Project Management: Use engineering management principles individually, as a team member or a leader to manage projects and effectively communicate about well-defined engineering activities.

PO 7. Life-long learning: Ability to analyse individual needs and engage in updating in the context of technological changes.

Programme Specific Objectives (PSOs)

PSO 1. Modern Information Technology: Use latest technologies for operation and application of information.

PSO 2. Information Technology Process: Maintain the information processes using modern information and commu-

P . N. TANDON

PRINCIPAL

Information Technology Engineering is becoming a demanding field in the industries, more and more students are getting attracted to Information Technology Engineering., but many of them doesn't know what Information Technology Engineering actually is? This year's magazine committee has taken a step to highlight the insights of Information Technology Engineering for which I congratulate them. Globalization demands a new dimension to our approach to education. We thus have to recognize the need for a globally relevant education. In Bharati Vidyapeeth, we open the gates for a student to expand his limits beyond the curriculum. The new technological advancements in the fields of engineering are made known to them so they can walk with the technology. Department of Information Technology Engineering has come up with an excellent activity which wakes student's curiosity up. Allowing students to search for various inventions, reaches, requirements, problems and solutions. The purpose of the magazine is to make student aware of the responsibilities they withhold as a Information Technology Engineer I hereby congratulate the students and teaching and non- teaching faculty of Information Technology Engineering department for their efforts in their work and bringing this wonderful magazine every year.

RANJIT PAWAR

HEAD OF DEPARTMENT

Welcome to the department of Information Technology, BVIT, Navi Mumbai. The department has a team of highly experienced and motivated faculty members who are in process of tuning the young minds to make them globally competitive. During this course, students are able to mold their career and inculcate team spirit with good oral & written communication through innovative teaching learning process a teamwork approach & leadership building experience our student gain vital communication & critical-thinking skills The magazine plays an instrumental role in providing exposure to the students to develop written communication skills and command over the language. It is a step towards building professional and ethical attitude in them. The entire journey of creating InfoTech is an outcome of rigorous effort made by students. Students not only gain the knowledge. about the latest technological developments and advancements through reading and writing articles but they also develop verbal and written communication skills.

PRATIBHA TAMBEWAGH

MAGAZINE CO-ORDINATOR

This is the annual magazine of Information Technology department magazine. As the magazine coordinator of the Information Technology department magazine InfoTech, this issue is particularly special to me as it was a challenge to not only live up to the standards set by the previous issue but also set new ones. InfoTech is all about the emerging trends in computer that inspires students to do something, that leaves an everlasting mark on the world of technology. Thus it was our job to ensure inspiring technological developments are being brought to the students of Bharati Vidyapeeth Institute of Technology, Kharghar. It gives me an immense pleasure to write and interact with you through this channel. We are always eager to assist you by keeping you abreast of the latest developments, to keep track of important achievements and to exhibit the talent of our stars. We, at BVIT always strive hard to offer our students salubrious learning ambience so that they can be given plentiful opportunities to groom their overall personality. We create the favorable environment to learn not just academics but lessons of life, character, ethics and values, so that they emerge as responsible individuals. We are confident enough to get you the targeted goal. Best wishes for all your future endeavors.

PRERANA KALE

TECHNICAL EDITOR

On behalf of the entire Information Technology Department and all the reader, we extend our whole heart gratitude to Hon. H.O.D Prof. Ranjeet Pawar for their guidance and Inspiration towards our department. The faculty Of our department boost our confidence for publishing InfoTech Magazine As the name suggest InfoTech means "Information Technology". We are glad to get the Golden Opportunity for publishing InfoTech magazine on behalf of the entire Information Technology Department. This magazine includes articles on emerging new Technologies in Computer Science. Students have properly researched and made the articles on exciting topics. Our endeavor with each edition is to update you on the latest trends of technologies coming up and flashing some light on the innovative minds of the youth today. Lastly, quoting my special thanks to Pratibha Tambewagh madam for her support and guidance all along, the Departmental faculty members and also to all my team members without whom this issue wouldn't have been possible. We hope all the readers will enjoy this issue as much as we enjoyed creating it.

INDEX

TECHNICAL ARTICLES

"The Impact of Artificial Intelligence on Healthcare"

In today's rapidly evolving technological landscape, Artificial Intelligence (AI) stands out as one of the most transformative innovations of our time. From powering virtual assistants to driving autonomous vehicles, AI has permeated nearly every aspect of our lives, revolutionizing industries, enhancing efficiency, and reshaping human interaction with technology. AI refers to the simulation of human intelligence processes by machines, enabling them to perform tasks that typically require human cognition, such as learning, problem-solving, and decision-making.

AI systems can analyze vast amounts of data, recognize patterns, and adapt their behavior autonomously, mimicking human-like intelligence. Artificial Intelligence (AI) has emerged as a game-changer in the field of healthcare, promising to revolutionize patient care, diagnosis, treatment, and operational efficiency. With its ability to analyze vast amounts of data, recognize patterns, and make predictions, AI is reshaping the healthcare landscape in unprecedented ways. AI-powered diagnostic tools are transforming the way diseases are detected and treated. Machine learning algorithms can analyze medical images, such as X-rays, MRIs, and CT scans, with incredible accuracy, aiding healthcare professionals in early detection and precise diagnosis of conditions ranging from cancer to neurological disorders. AI technologies are improving patient care and monitoring by enabling real-time analysis of patient data and vital signs. Additionally, AI-driven treatment planning algorithms can personalize treatment regimens based on individual patient data, optimizing outcomes and minimizing side effects. Furthermore, virtual health assistants powered by AI can provide patients with personalized care recommendations, medication reminders, and lifestyle management tips, enhancing patient engagement and adherence to treatment plans. AI is optimizing healthcare operations by streamlining administrative tasks, enhancing workflow efficiency

By Sakshi Mohite (TYIF)

“Navigating the Impact of Social Media on Student Mental Health"

Social media is a big and integral part of our lives and especially for college students. While it can be fun and useful, it’s important to know how it affects our mental health.it’s offering students avenues for connection, self-expression, and information sharing. However, it's essential to recognize the potential impact it can have on mental well-being.

Social media provides platforms for students to stay connected with friends and family, discover new interests, and engage in meaningful discussions. It can enhance social support networks and foster a sense of belonging within the college community. While social media offers benefits, it also presents challenges. Constant exposure to curated images and lifestyles can lead to feelings of inadequacy and self-comparison. Cyberbullying and negative interactions online can exacerbate stress and anxiety among students. Encouraging students to establish healthy boundaries with social media can be empowering. This includes limiting screen time, taking breaks from social platforms, and unfollowing accounts that trigger negative emotions. It’s important to Educate students about the impact of social media on mental health is crucial. Teaching digital literacy skills can help them critically evaluate online content. Navigating the impact of social media on student mental health requires awareness, boundaries, and balance. By fostering digital mindfulness and seeking support when needed, students can navigate the online world with resilience and prioritize their mental well-being.

Encouraging students to find a balance between online and offline activities is key. Suggesting alternative ways to spend time, such as pursuing hobbies, exercising, or engaging in face-to-face interactions, can help maintain overall well-being. Social media can be a great way to connect and share, but it’s important to take care of our mental health too. By being mindful of how it affects us and taking breaks when needed, we can enjoy social media while staying happy and healthy.

By Sakshi Mohite (TYIF)

Securing Our Digital Frontier: Navigating the Complex Landscape of Cybersecurity

In an age where digital technology permeates every aspect of our lives, the importance of cybersecurity cannot be overstated. From personal information to critical infrastructure, our digital assets are constantly under threat from cyberattacks. As our reliance on interconnected systems grows, so too does the need for robust cybersecurity measures to protect against malicious actors seeking to exploit vulnerabilities for their own gain. In this article, we delve into the multifaceted world of cybersecurity, exploring its challenges, advancements, and the imperative for collective action in safeguarding our digital frontier.

The Evolving Threat Landscape

Cyber threats come in various forms, ranging from simple phishing emails to sophisticated nation-state attacks. With the proliferation of internet-connected devices through the Internet of Things (IoT) and the increasing sophistication of cybercriminal tactics, the threat landscape is constantly evolving. Threat actors employ a myriad of techniques, including malware, ransomware, social engineering, and zero-day exploits, to infiltrate networks and compromise sensitive data.

Moreover, the rise of interconnected systems poses new challenges, as vulnerabilities in one area can have cascading effects across entire networks. For instance, a breach in a smart home device could provide a gateway for hackers to infiltrate corporate networks, potentially leading to devastating consequences. As such, cybersecurity must extend beyond traditional boundaries to encompass a holistic approach that addresses vulnerabilities at every level of the digital ecosystem.

The Role of Technology in Cyber Defense

Advancements in technology have both fueled the proliferation of cyber threats and provided innovative solutions for defense. Artificial intelligence (AI) and machine learning (ML) algorithms are being leveraged to detect anomalies and identify suspicious behavior in real-time, enabling proactive threat mitigation. Similarly, block chain technology is being explored for its potential to enhance data integrity and secure transactions, particularly in industries such as finance and healthcare.

Additionally, the emergence of quantum computing presents both opportunities and challenges for cybersecurity. While quantum computing holds the promise of revolutionizing encryption techniques, it also has the potential to render existing cryptographic protocols obsolete, necessitating the development of quantum-resistant algorithms. As such, staying ahead of the curve in cybersecurity requires continual innovation and adaptation to emerging technologies.

The Human Element: Education and Awareness

Despite the advancements in technology, humans remain the weakest link in cybersecurity. Social engineering tactics, such as phishing and pretexting, rely on exploiting human psychology to gain unauthorized access to systems or sensitive information. Therefore, education and awareness are paramount in building a cyber-resilient society.

Organizations must invest in comprehensive cybersecurity training programs to empower employees with the knowledge and skills needed to recognize and respond to potential threats effectively. Furthermore, fostering a culture of cybersecurity awareness encourages individuals to take proactive measures to protect their digital assets, whether at home or in the workplace.

The Imperative for Collaboration

Cybersecurity is not a problem that any single entity can solve in isolation. It requires collaboration and cooperation among governments, businesses, academia, and civil society to effectively combat cyber threats on a global scale. Information sharing, best practice dissemination, and joint research efforts are essential components of a collaborative cybersecurity ecosystem.

Furthermore, international cooperation is crucial in addressing the transnational nature of cybercrime and establishing norms of responsible behavior in cyberspace. Initiatives such as the Budapest Convention on Cybercrime and the Paris Call for Trust and Security in Cyberspace exemplify international efforts to promote cybersecurity cooperation and build mutual trust among nations.

Conclusion

As our society becomes increasingly reliant on digital technology, the importance of cybersecurity cannot be overstated. The evolving threat landscape, coupled with rapid technological advancements, underscores the need for continual vigilance and adaptation in defending against cyber threats. By embracing a holistic approach that combines technological innovation, education, awareness, and collaboration, we can navigate the complex landscape of cybersecurity and secure our digital frontier for generations to come.

-By Manasi Rugale TYIF

Unlocking the Potential of Mobile 5G: Revolutionizing Connectivity in the Digital Age

In today's fast-paced digital world, connectivity is paramount. From streaming high-definition videos to powering the Internet of Things (IoT), the demand for faster, more reliable mobile networks continues to escalate. Enter 5G – the fifth generation of cellular technology poised to revolutionize the way we connect, communicate, and consume content on our mobile devices. In this article, we explore the transformative potential of mobile 5G, its impact on various industries, and the opportunities it presents for innovation and economic growth.

A Quantum Leap in Connectivity

Mobile 5G represents a quantum leap in wireless connectivity, offering significantly higher speeds, lower latency, and increased network capacity compared to its predecessors. With theoretical download speeds of up to 20 gigabits per second (Gbps) and latency as low as one millisecond, 5G promises to deliver seamless connectivity and near-instantaneous responsiveness, enabling a host of innovative applications and services.

One of the defining features of 5G is its ability to support a massive number of connected devices simultaneously. This is particularly crucial in the era of the IoT, where billions of interconnected devices – from smart appliances to autonomous vehicles – rely on seamless connectivity to function efficiently. 5G's enhanced network capacity and scalability make it ideally suited to meet the demands of the IoT, laying the foundation for a truly interconnected and intelligent world.

Transforming Industries

The impact of 5G extends far beyond faster download speeds and smoother streaming. Across various industries, from healthcare to manufacturing, 5G is poised to revolutionize operations, drive innovation, and unlock new business opportunities.

In healthcare, for example, 5G-enabled remote monitoring and telemedicine solutions have the potential to transform patient care delivery, particularly in underserved areas where access to healthcare services is limited. Real-time monitoring of vital signs, remote consultations with healthcare providers, and even surgical procedures conducted via augmented reality (AR) are just a few examples of how 5G can enhance healthcare delivery and improve patient outcomes.

Similarly, in manufacturing, 5G-powered technologies such as augmented reality (AR), virtual reality (VR), and robotics are driving the emergence of smart factories. With ultra-low latency and high reliability, 5G enables real-time monitoring and control of manufacturing processes, predictive maintenance of equipment, and the seamless integration of automation and robotics into production workflows. This not

only improves operational efficiency and productivity but also enables manufacturers to respond quickly to changing market demands and customer preferences.

Empowering Innovation

The rollout of 5G is paving the way for a new era of innovation and entrepreneurship, catalyzing the development of groundbreaking technologies and applications that were previously unimaginable. From autonomous vehicles to immersive gaming experiences, the possibilities enabled by 5G are limited only by our imagination.

For startups and innovators, 5G represents a fertile ground for experimentation and disruption. With high-speed, low-latency connectivity at their disposal, entrepreneurs can develop and deploy cutting-edge solutions that leverage the full potential of 5G technology. Whether it's creating immersive AR experiences, delivering ultra-high-definition content on mobile devices, or building next-generation IoT devices, 5G empowers innovators to push the boundaries of what's possible and drive meaningful change in society.

Challenges and Opportunities Ahead

Despite its immense potential, the widespread adoption of 5G presents several challenges, including infrastructure deployment, spectrum allocation, and cybersecurity concerns. Building out the necessary infrastructure, including small cell networks and fiber-optic backhaul, requires substantial investment and coordination among stakeholders. Moreover, ensuring equitable access to 5G technology and bridging the digital divide will be essential in realizing its full societal benefits.

However, with the right investments, policies, and collaborations, these challenges can be overcome, paving the way for a future where mobile 5G powers innovation, economic growth, and social progress. By harnessing the transformative potential of 5G, we can create a more connected, intelligent, and prosperous world for generations to come.

-By Vaibhavi T. TYIF

Cyber Law and Today's Generation

In today's digital age, where virtually every aspect of our lives is intertwined with technology, the importance of cyber law cannot be overstated. From safeguarding personal privacy to combating cybercrime, cyber law plays a crucial role in shaping the digital landscape and protecting the rights and interests of individuals and organizations alike. In this article, we explore the evolving field of cyber law and its significance in the context of today's generation, where technology is ubiquitous and its impact profound.

Defining Cyber Law

Cyber law, also known as internet law or information technology law, encompasses a broad range of legal issues pertaining to the use of technology and the internet. It encompasses laws and regulations governing online behavior, digital transactions, intellectual property rights, cybersecurity, and data protection, among other areas. As technology continues to evolve at a rapid pace, cyber law must adapt accordingly to address emerging challenges and regulate novel forms of online activity.

The Digital Generation: Navigating Cyberspace

For today's generation, technology is an integral part of daily life. From social media and e-commerce to online education and remote work, the internet permeates virtually every aspect of modern existence. This widespread reliance on digital technology presents both opportunities and challenges, particularly concerning legal rights and responsibilities in cyberspace.

On one hand, the internet has democratized access to information, facilitated global communication, and spurred innovation and entrepreneurship. However, it has also given rise to new forms of cybercrime, privacy violations, and ethical dilemmas that require robust legal frameworks to address effectively. As digital natives who have grown up in the era of smartphones, social media, and cloud computing, today's generation must navigate the complexities of cyberspace while understanding the legal implications of their online activities.

Key Legal Issues in Cyberspace

Several key legal issues are of particular relevance to today's generation in cyberspace:

1. Privacy and Data Protection: With the proliferation of online services and digital platforms, protecting personal privacy and data has become paramount. Laws such as the General Data Protection Regulation (GDPR) in the European Union and the California Consumer Privacy Act (CCPA) in the United States aim to safeguard individuals' rights to privacy and control over their personal information.
2. Cybersecurity: As cyber threats continue to evolve in sophistication and scale, ensuring the security of digital infrastructure and networks is essential. Cybersecurity laws and regulations impose obligations on organizations to implement measures to protect against cyber attacks and data breaches, as well as to disclose incidents promptly and mitigate harm to affected individuals.

3. Intellectual Property Rights: The internet has transformed the way intellectual property is created, distributed, and consumed. Copyright, trademark, and patent laws govern the protection of creative works, brands, and inventions in the digital realm, while also addressing issues such as online piracy, infringement, and fair use.
4. Cybercrime: The rise of cybercrime poses significant challenges for law enforcement and regulatory authorities worldwide. Cybercriminal activities such as hacking, phishing, identity theft, and ransomware attacks target individuals, businesses, and governments, underscoring the need for robust legal frameworks to deter, detect, and prosecute cyber offenders.

The Role of Education and Awareness

In navigating the complexities of cyberspace, education and awareness are key. Today's generation must be equipped with the knowledge and skills to understand their rights and responsibilities in the digital world, as well as the potential legal consequences of their online actions. Educational initiatives, cybersecurity awareness campaigns, and digital literacy programs can empower individuals to make informed decisions and protect themselves against online threats and vulnerabilities.

Looking Ahead: Shaping the Future of Cyber Law

As technology continues to evolve and shape our world, the field of cyber law will likewise evolve to address emerging challenges and opportunities. Collaboration among governments, businesses, academia, and civil society is essential in developing comprehensive legal frameworks that promote innovation, protect fundamental rights, and ensure a safe and secure digital environment for all.

By staying informed, advocating for responsible digital citizenship, and actively engaging in the legal and regulatory discourse surrounding cyberspace, today's generation can help shape the future of cyber law and contribute to a more equitable, inclusive, and sustainable digital society.

-By Ayush Gole SYIF

The Impact of Artificial Intelligence on Student Life: Navigating the Future of Education

In recent years, artificial intelligence (AI) has emerged as a transformative force in various aspects of our lives, including education. From personalized learning experiences to intelligent tutoring systems, AI has the potential to revolutionize how students learn, teachers teach, and educational institutions operate. In this article, we explore the profound impact of AI on student life and the opportunities and challenges it presents for the future of education.

Personalized Learning Experiences

One of the most significant benefits of AI in education is its ability to provide personalized learning experiences tailored to individual students' needs, preferences, and learning styles. Through the analysis of vast amounts of data, AI-powered systems can identify each student's strengths and weaknesses, adapt learning materials accordingly, and provide targeted support and feedback in real-time. This personalized approach not only enhances student engagement and motivation but also improves learning outcomes by addressing each student's unique learning needs.

Intelligent Tutoring Systems

AI-powered intelligent tutoring systems offer students personalized support and guidance outside the traditional classroom setting. These systems use algorithms to assess students' knowledge levels, track their progress, and deliver adaptive learning activities and exercises. By providing immediate feedback and scaffolding, intelligent tutoring systems help students master difficult concepts, overcome learning challenges, and achieve their academic goals more effectively.

Enhanced Accessibility and Inclusivity

AI has the potential to enhance accessibility and inclusivity in education by removing barriers to learning for students with diverse learning needs and abilities. For example, AI-driven speech recognition and natural language processing technologies can facilitate communication and comprehension for students with speech or language impairments. Similarly, AI-powered translation tools can help students overcome language barriers and access educational resources in their native languages, fostering greater inclusivity and diversity in the learning environment.

Streamlined Administrative Processes

In addition to transforming teaching and learning, AI can streamline administrative processes within educational institutions, freeing up valuable time and resources for educators and administrators. AI-powered systems can automate routine administrative tasks such as grading assignments, scheduling classes, and managing student records, allowing educators to focus more on teaching and providing personalized support to students. Moreover, predictive analytics algorithms can help identify at-risk students early on and intervene proactively to prevent academic difficulties and dropout.

Ethical and Societal Implications

While the potential benefits of AI in education are vast, it is essential to consider the ethical and societal implications of its widespread adoption. Concerns about data privacy, algorithmic bias, and the depersonalization of education must be addressed to ensure that AI technologies are used responsibly and ethically. Moreover, it is crucial to promote digital literacy and critical thinking skills among students to empower them to navigate the complex ethical and societal issues surrounding AI effectively.

Looking Ahead: Embracing the Future of Education

As AI continues to evolve and reshape the educational landscape, it is essential for educators, policymakers, and stakeholders to embrace innovation while also safeguarding the principles of equity, inclusivity, and ethical responsibility. By harnessing the power of AI to personalize learning experiences, enhance accessibility, and streamline administrative processes, we can create a more responsive, equitable, and student-centered education system that prepares students for success in the digital age. Ultimately, the successful integration of AI into education will depend on our ability to leverage its potential while also addressing its ethical, social, and cultural implications responsibly.

-By Jay Patil SYIF

STUDENT ACHIEVEMENT

BHARATI VIDYAPEETH
INSTITUTE OF TECHNOLOGY

Sector-7, CBD Belpada, Opp. Kharghar Rly. Station, Navi Mumbai-400 614.

Certificate of Appreciation

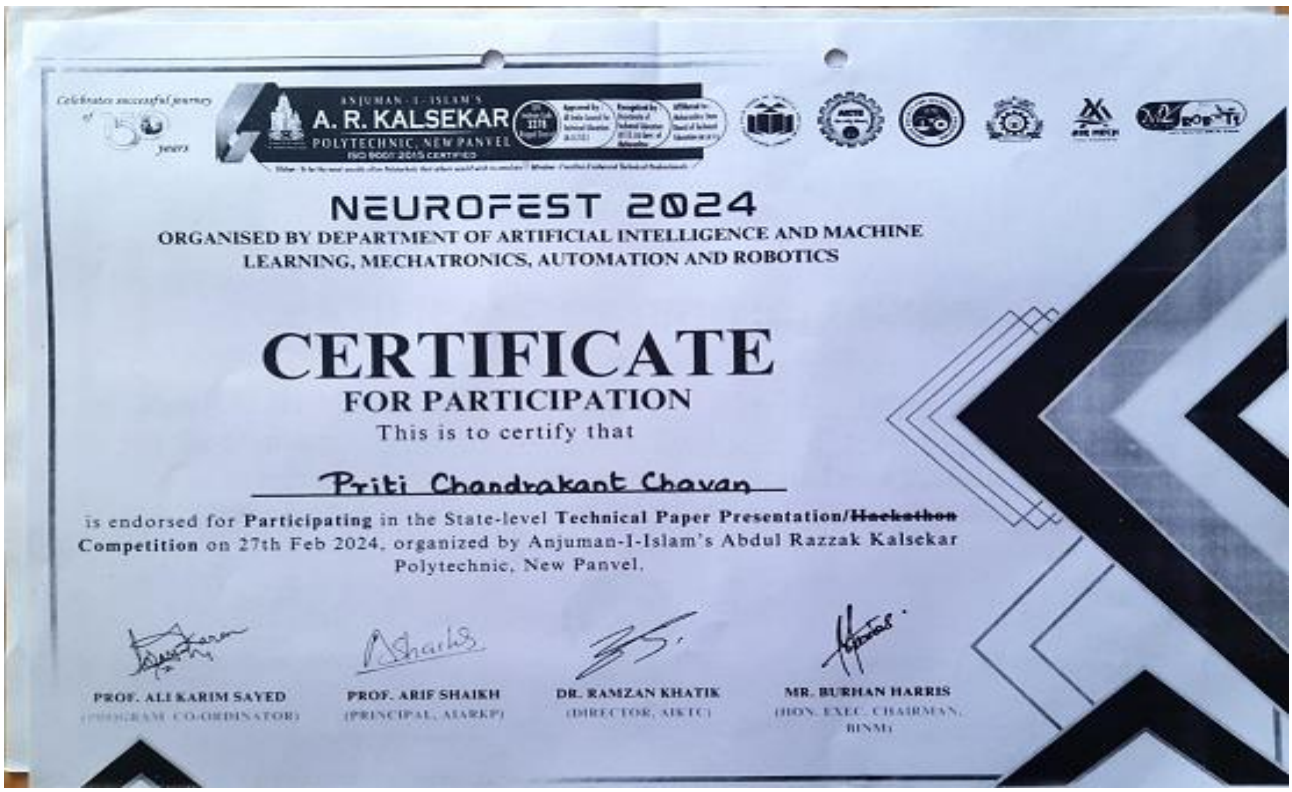
This is to certify that ~~Mr./~~Ms. Divya Dashrath Patil
of Institute Bharati Vidyapeeth Institute of Technology *has participated in* ***State Level Technical Paper Presentation Competition*** *for students of computer engineering group.*

Organized by Department of
Computer Technology and Information Technology
In association with
ISTE Student Chapter [MH-149]
on Wednesday, 13th October 2023.

Convener
ISTE Student Chapter [MH-149]

Program
Co-ordinator

Principal
BVIT, Navi Mumbai



ANJUMAN-I-ISLAM'S
A. R. KALSEKAR
POLYTECHNIC, NEW PANVEL

NEUROFEST 2024

ORGANISED BY DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING, MECHATRONICS, AUTOMATION AND ROBOTICS

CERTIFICATE

FOR PARTICIPATION

This is to certify that

Priti Chandrakant Chavan

is endorsed for **Participating** in the State-level **Technical Paper Presentation/~~Hackathon~~ Competition** on 27th Feb 2024, organized by Anjuman-I-Islam's Abdul Razzak Kalsekar Polytechnic, New Panvel.

PROF. ALI KARIM SAYED
(PROGRAM CO-ORDINATOR)

PROF. ARIF SHAIKH
(PRINCIPAL, AIARKP)

DR. RAMZAN KHATIK
(DIRECTOR, AIKTC)

MR. BURHAN HARRIS
(HON. EXEC. CHAIRMAN, BINM)

Maharashtra State Board of Technical Education
State Level Technical Quiz Competition - Computer Engineering Group 2023-24
Organized by
PILLAI HOC COLLEGE OF ENGINEERING & TECHNOLOGY
(Diploma Section) , RASAYANI

CERTIFICATE OF PARTICIPATION

This is to certify that Mr./Ms. Vaibhavi Anil Talawnekai
of the institute Bharati Vidyapeeth Institute of Technology, Kharghar
has participated in the MSBTE sponsored '**State Level Technical Quiz Competition 2024**' held at PHCET(Diploma Section), Rasayani on 20 February 2024.

Principal , PHCET

Dy.Secretary, RBTE,MUMBAI

ANJUMAN-I-ISLAM'S
A. R. KALSEKAR POLYTECHNIC, NEW PANVEL
ISO 9001:2015 certified

NEUROFEST 2024

ORGANISED BY DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING, MECHATRONICS, AUTOMATION AND ROBOTICS

CERTIFICATE
FOR PARTICIPATION

This is to certify that

Swati Dattatray Mane

is endorsed for **Participating** in the State-level **Technical Paper Presentation/~~Hackathon~~ Competition** on 27th Feb 2024, organized by Anjuman-I-Islam's Abdul Razzak Kalsekar Polytechnic, New Panvel.

PROF. ALI KARIM SAYED (PROGRAM CO-ORDINATOR)
PROF. ARIF SHAIKH (PRINCIPAL, AIARKP)
DR. RAMZAN KHATIK (DIRECTOR, AIKTC)
MR. BURHAN HARRIS (HON. EXEC. CHAIRMAN, BENM)

ANJUMAN-I-ISLAM'S
A. R. KALSEKAR POLYTECHNIC, NEW PANVEL
ISO 9001:2015 CERTIFIED

NEUROFEST 2024

ORGANISED BY DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING, MECHATRONICS, AUTOMATION AND ROBOTICS

CERTIFICATE

FOR PARTICIPATION

This is to certify that

Swati Dattatray Mane

is endorsed for **Participating** in the State-level **Technical Paper Presentation/~~Hackathon~~ Competition** on 27th Feb 2024, organized by Anjuman-I-Islam's Abdul Razzak Kalsekar Polytechnic, New Panvel.

PROF. ALI KARIM SAYED	PROF. ARIF SHAIKH	DR. RAMZAN KHATIK	MR. BURHAN HARRIS
(PROGRAM CO-ORDINATOR)	(PRINCIPAL, AIARKP)	(DIRECTOR, AIKTC)	(HON. EXEC. CHAIRMAN, BINM)

AGNEL POLYTECHNIC, VASHI
(APPROVED BY AICTE & AFFILIATED TO MSBTE)

TECHNOCRATZ 2023 -2024

STATE-LEVEL TECHNICAL PAPER PRESENTATION/QUIZ COMPETITION

1st & 2nd March 2024

CERTIFICATE OF APPRECIATION

~~Mr.~~/Ms. Priti Chandankant Chavan of Institute Bharati Vidyapeeth has participated in State Level Technical ~~Paper Presentation~~ / Quiz competition of Computer group organized in the academic year 2023-24

Mrs. SALY ANTONY
PRINCIPAL

AGNEL POLYTECHNIC, VASHI

(APPROVED BY AICTE & AFFILIATED TO MSBTE)

TECHNOCRATZ 2023 -2024

STATE-LEVEL TECHNICAL PAPER PRESENTATION/QUIZ COMPETITION

1st & 2nd March 2024

CERTIFICATE OF APPRECIATION

Mr./Ms. Kshitija Sadip Khilari of Institute BVIT, Navi Mumbai has participated in State Level Technical Paper Presentation / ~~Quiz competition~~ of Computer group organized in the academic year 2023-24

AGNEL POLYTECHNIC VASHI

Mrs. SALY ANTONY
PRINCIPAL

SARASWATI Education Society's

SARASWATI Institute of Technology

Learn Live Achieve and Contribute

Kharghar, Navi Mumbai - 410 210

Plot No.46,Near MSEB Sub Station , Sector 5, Kharghar , Navi Mumbai, Maharashtra

INSTITUTION'S INNOVATION COUNCIL

CERTIFICATE OF PARTICIPATION

THIS IS TO CERTIFY THAT

Mr./~~Mrs.~~ Aayush Arun Gole

has actively participated in Technical Paper Presentation Competition based on " Innovative Horizon of Current Era" organized by Saraswati Institute of Technology,Kharghar

Under ISTE Student Chapter MH-168

22nd MARCH 2024
Date

Dr.D.R.Suroshe
(Principal)

SARASWATI Education Society's

SARASWATI Institute of Technology

Learn Live Achieve and Contribute

Kharghar, Navi Mumbai - 410 210.

Plot No.46,Near MSEB Sub Station , Sector 5, Kharghar , Navi Mumbai, Maharashtra

INSTITUTION'S INNOVATION COUNCIL (Ministry of HRD Initiative)

CERTIFICATE OF PARTICIPATION

THIS IS TO CERTIFY THAT

Mr./~~Mrs.~~ Jay Janardhan Patil

has actively participated in Technical Paper Presentation Competition based on " **Innovative Horizon of Current Era**" organized by Saraswati Institute of Technology,Kharghar

Under ISTE Student Chapter MH-168

22nd MARCH 2024
Date

Dr.D.R.Suroshe
(Principal)

AGNEL POLYTECHNIC, VASHI

(APPROVED BY AICTE & AFFILIATED TO MSBTE)

TECHNOCRATZ 2023 -2024

STATE-LEVEL TECHNICAL PAPER PRESENTATION/QUIZ COMPETITION

1st & 2nd March 2024

CERTIFICATE OF APPRECIATION

~~Mr.~~/Ms. Swati Dattatray Mane of Institute Bharati Vidyapeeth has participated in State Level Technical ~~Paper Presentation~~ / Quiz competition of Computer group organized in the academic year 2023-24

Mrs. SALY ANTONY
PRINCIPAL

AGNEL POLYTECHNIC, VASHI

(APPROVED BY AICTE & AFFILIATED TO MSBTE)

TECHNOCRATZ 2023 -2024

STATE-LEVEL TECHNICAL PAPER PRESENTATION/QUIZ COMPETITION

1st & 2nd March 2024

CERTIFICATE OF APPRECIATION

Mr./Ms. Jay Janardan Patil of Institute BVIT, Navi Mumbai has participated in State Level Technical Paper Presentation / ~~Quiz competition~~ of Computer group organized in the academic year 2023-24

AGNEL POLYTECHNIC VASHI SECTOR-9A, VASHI, NAVI MUMBAI DIST THANE

Mrs. SALY ANTONY
PRINCIPAL

SARASWATI Education Society's

SARASWATI Institute of Technology

Learn Live Achieve and Contribute

Kharghar, Navi Mumbai - 410 210.

Plot No.46,Near MSEB Sub Station , Sector 5, Kharghar , Navi Mumbai, Maharashtra

INSTITUTION'S INNOVATION COUNCIL (Ministry of HRD Initiative)

CERTIFICATE OF ACHIEVEMENT

THIS IS TO CERTIFY THAT

Mr.Ayush Arun Gole

has actively participated in Technical Paper Presentation Competition based on " **Innovative Horizon of Current Era**" organized by Saraswati Institute of Technology,Kharghar & Secured **1st Rank** Under ISTE Student Chapter MH-168.

22nd MARCH 2024
Date

Dr.D.R.Suroshe
(Principal)

SARASWATI Education Society's

SARASWATI Institute of Technology

Learn Live Achieve and Contribute

Kharghar, Navi Mumbai - 410 210.

Plot No.46,Near MSEB Sub Station , Sector 5, Kharghar , Navi Mumbai, Maharashtra

INSTITUTION'S INNOVATION COUNCIL (Ministry of HRD Initiative)

CERTIFICATE OF PARTICIPATION

THIS IS TO CERTIFY THAT

Mr./~~Mrs.~~ Jay Janardhan Patil

has actively participated in Technical Paper Presentation Competition based on " **Innovative Horizon of Current Era**" organized by Saraswati Institute of Technology,Kharghar

Under ISTE Student Chapter MH-168

22nd MARCH 2024
Date

Dr.D.R.Suroshe
(Principal)

SARASWATI Education Society's

SARASWATI Institute of Technology

Learn Live Achieve and Contribute

Kharghar, Navi Mumbai - 410 210.

Plot No.46,Near MSEB Sub Station , Sector 5, Kharghar , Navi Mumbai, Maharashtra

INSTITUTION'S INNOVATION COUNCIL (Ministry of HRD Initiative)

CERTIFICATE OF ACHIEVEMENT

THIS IS TO CERTIFY THAT

Mr.Ayush Arun Gole

has actively participated in Technical Paper Presentation Competition based on " **Innovative Horizon of Current Era**" organized by Saraswati Institute of Technology,Kharghar & Secured **1st Rank**

Under ISTE Student Chapter MH-168.

22nd MARCH 2024
Date

Dr.D.R.Suroshe
(Principal)

STUDENT AND FACULTY PAPERS

Deceptive Opinion Spam Detection Using Bidirectional Long Short-Term Memory with Capsule Neural Network

Sandeep A. Shinde[1*], Ranjeet R. Pawar[2], Asmita A. Jagtap[3], Pratibha A. Tambewagh[4], Punam U. Rajput[5], Mohan K. Mali[6], Satish D. Kale[7], Sameer V. Mulik[8]

[1*]Lecturer, Bharati Vidyapeeth Institute of Technology, Navi Mumbai, India,
Email:Sandeep.Shinde@bharatividyapeeth.edu

[2]HOD, Bharati Vidyapeeth Institute of Technology, Navi Mumbai, India,
Email:Ranjeet.Pawar@bharatividyapeeth.edu

[3]Lecturer, Bharati Vidyapeeth Institute of Technology, Navi Mumbai, India,
Email:Asmita.Jagtap@bharatividyapeeth.edu

[4]Lecturer, Bharati Vidyapeeth Institute of Technology, Navi Mumbai, India,
Email: Pratibha.Tambewagh @bharatividyapeeth.edu

[5]Lecturer, Bharati Vidyapeeth Institute of Technology, Navi Mumbai, India,
Email:Poonam.Rajput@bharatividyapeeth.edu

[6]Lecturer, Bharati Vidyapeeth Institute of Technology, Navi Mumbai, India,
Email:Mohan.Mali@bharatividyapeeth.edu

[7]Lecturer, Bharati Vidyapeeth Institute of Technology, Navi Mumbai, India,
Email:Satish.Kale@bharatividyapeeth.edu

[8]Lecturer, Bharati Vidyapeeth Institute of Technology, Navi Mumbai, India,
Email:Sameer.Mulik@bharatividyapeeth.edu

Abstract: Product reviews are becoming a more popular tool for businesses and individuals when making judgements about purchases. Spammers create synthesized reviews to either promote certain items or denigrate those of rivals to make money. As a result, in recent years, both the business and research sectors have paid close attention to the detection of false opinion spam. Customers' decision-making is severely harmed by false opinion spam in service or product evaluations. It's becoming difficult to identify false opinion spam. Accordingly, the article proposed to detect deceptive opinion spam based on a hybrid deep learning technique. Initially, the model was tested using deceptive reviews gathered from several online forums. To identify deceptive reviews, many researchers at the moment create models based on a single text attribute. On the contrary, deceptive reviewers will decisively copy the wording style of legitimate evaluations while submitting reviews. These text-feature-based techniques may or may not be successful. As a result, the research suggested an ensemble multiple-feature selection technique of the Extra tree classifier to extract information based on a variety of features, including text, behaviour, and deceptive scoring features. In addition, a data resampling approach is used that integrates the Borderline-SMOTE algorithm to reduce the effects of the high dimensional imbalanced class category distribution. For detecting deceptive reviews, the article developed a hybrid technique of Bidirectional Long Short-Term Memory (Bi-LSTM) with a Capsule Neural Network to detect the positive and negative false opinions spam. The model optimizes the dynamic routing algorithm and changes the structure of the conventional capsule network without sacrificing classification performance, leading to high model accuracy. The model performance is evaluated using Python software. The study assesses the suggested model using data from two distinct domains (hotel and restaurant) as a standard benchmark. The experimental results demonstrate the advantage of neural models with higher accuracy of 99% respectively, showing that the suggested neural model greatly outperforms the state-of-the-art techniques.

Keywords: Deceptive Opinion, Spam Detection, Social Media, Bidirectional Long Short-Term Memory, Capsule Neural Network, Borderline-SMOTE, Extra Tree Classifier.

1. INTRODUCTION

Numerous online reviews detailing consumers' perceptions of numerous items and services have been published as a result of e-growing commerce prominence and the quick growth of internet shopping [1]. Because of this, a growing number of buyers use online product evaluations to assess the worth of a good or service, which has an impact on their purchasing choices. Higher percentages of favorable ratings encourage consumers to buy particular items, strengthening financial profits for manufacturers, while negative reviews encourage consumers

to explore alternatives and risk financial consequences for manufacturers [2] [3]. For financial advantage, individuals or businesses may make false statements that mislead consumers and undermine societal welfare [4]. In the meanwhile, buyers and producers research online reviews before making a choice. Therefore, by creating fake reviews, this feature was abused. In this contemporary day, weeding out incorrect facts becomes crucial [5]. For a range of deception detection tests (such as recognizing written or spoken lies on a person's status, biographical information, or any non-personal events), research suggests that humans frequently perform at the level of chance. Furthermore, when considering online evaluations, the sheer volume of them makes it unlikely for all but the most diligent customers to find dishonesty. In light of this, the use and verification of computer algorithms have become the focus of research [6]. If the product receives many positive reviews, sales would be high and vice versa. The suppliers' revenues will be significantly impacted by bogus reviews on platforms, and buyers will be misled into believing misleading information. Since both users and network operators are concerned about the detection of spam reviews, this problem must be addressed. Finding a spam review can be challenging as the false reviewer often follows the same trends as a legitimate user [7].

Customers make buying selections based only on reviews because there is no other factor involved. Customer reviews, therefore, offer decision-makers useful information about the products in addition to helping them make purchasing decisions. Dishonest reviews can be identified using reviewer behavioral features, content-related features, or review features. However, not all of the retrieved traits may be necessary for recognizing misleading. A hybrid filter-wrapper strategy is introduced in [8] to choose the best attributes for detecting false online customer reviews. Deceptive communication, especially scamming and fake consumer reviews, can affect news sources. According to a recent study, during the 2016 presidential election, there were roughly 312% more fake news items than usual [9].

Zhong *et al* [10] investigated the elements that influence customer purchasing decisions in online review platforms that are inundated with false reviews, according to the findings of false reviews. As a result, a misleading review influence model that is based on three crucial components of the online review system has been developed. This includes sentimental aspects, review volume, and characteristics of internet sellers. Text mining is employed to measure the signals of the 3 important elements based on them. The deceptive review detection framework (DRDF) [11] combines the chunk-level, sentence-level, word-level, and topic sentiment frameworks to extract the unique elements of the reviews. To detect fake reviews, these are processed as the multiple-classifier vector input. Increasingly advanced machine learning techniques and vision were used to simulate non-verbal deceitful behavior and create a real-time deception detection method. As the foundational models for fraud detection, random forests, artificial neural networks, TextCNN, long short-term memory (LSTM), Bi-directional LSTM (BiLSTM), and support vector machines were chosen [12] [13].

False and misleading evaluations have been able to dodge the frequently non-existent means of authentication of internet platforms, resulting in the proliferation of new businesses. To deal with this issue, the most recent classification methods, built on deep and machine learning techniques, aim to teach automated systems how to distinguish between true and erroneous judgements [14]. However, many avaricious businesses use false reviews to mislead consumers, enhance their perception of the brand, or disparage a rival. Different models are put out to identify false opinion reviews. The majority of these models use conventional techniques that centre on feature extraction and conventional classifiers. Nevertheless, those approaches ignore the context of the view while failing to capture the semantic aspect [15]. It is impractical to manually identify deceit in every text message given the number of text messages produced by news organizations, social media platforms, reviewers, businesses, and other groups. As a result, to solve this issue, automated techniques and solutions must be developed and implemented. Section 2 presents relevant work in the following sections. Details of the suggested neural model are provided in section 3. Experimental findings are presented in section 4. This work is concluded in section 5.

1.1 Problem statement

The problem statement of Deceptive Opinion Spam Detection is to develop a model or algorithm that can automatically identify and differentiate between genuine and deceptive opinions or review content posted on online platforms, such as product reviews, social media posts, or online forums. Deceptive opinion spam refers to fake, misleading, or manipulated reviews and opinions that are designed to mislead readers or consumers. This can include fraudulent positive reviews to artificially boost the reputation of a product or service, as well as negative reviews meant to harm the reputation of a competitor. The goal of the problem is to create a reliable system that can accurately detect and filter out deceptive opinion spam, helping users make informed decisions based on trustworthy and authentic reviews. The detection of deceptive opinions is crucial for maintaining the credibility and trustworthiness of online platforms and protecting consumers from making decisions based on misleading information.

1.2 Objective

The objective of this research is to address the growing problem of deceptive opinion spam in product reviews, which can mislead customers and affect their purchasing decisions. The researchers aim to develop a reliable and

effective model for automatically detecting deceptive reviews from genuine ones, utilizing a hybrid deep learning technique. By leveraging multiple features, including text, behavior, and deceptive scoring, the goal is to achieve high accuracy in identifying deceptive opinions and improve the overall trustworthiness of online reviews. The novelty of the article lies in the comprehensive and innovative approach to tackling the deceptive opinion spam detection problem. The integration of multiple features, including text, behavior, and deceptive scoring), data resampling, and a hybrid deep learning technique contributes to achieving higher accuracy in identifying false opinion spam and outperforming existing state-of-the-art methods. The technique described, which combines Bi-LSTM with a Capsule Neural Network, has been used separately in various applications, but its specific combination for deceptive opinion spam detection may be a novel application.

1.3 Contribution

- The research proposes a novel approach by combining Bidirectional Long Short-Term Memory (Bi-LSTM) with a Capsule Neural Network. This hybrid deep learning model enables the simultaneous detection of positive and negative deceptive opinion spam, providing a comprehensive solution to the problem.
- Instead of relying solely on text-based attributes, the study introduces an ensemble technique using the Extra tree classifier to extract information from various features, including text, behaviour, and deceptive scoring. This integration allows the model to capture diverse patterns and characteristics associated with deceptive reviews, enhancing detection performance.
- To mitigate the challenge of imbalanced class distribution, the research employs a data resampling approach that combines the Borderline-SMOTE algorithm. This technique ensures a more balanced representation of deceptive and genuine reviews during training, improving the model's ability to handle imbalanced data effectively.
- The researchers optimize the dynamic routing algorithm and modify the conventional capsule network's structure without compromising classification performance. This refinement leads to a high-performing neural model with improved accuracy in detecting deceptive opinion spam.
- The suggested model is thoroughly evaluated using data from two distinct domains, hotels, and restaurants, which are considered standard benchmarks. The experimental results demonstrate the model's superiority over state-of-the-art techniques, achieving a remarkable accuracy of 99% in both domains.

The research contributes a powerful solution to the critical problem of deceptive opinion spam detection in product reviews. By leveraging a hybrid deep learning technique with multiple-feature selection and an optimized dynamic routing algorithm, the model achieves exceptional accuracy, outperforming existing methods. The proposed approach provides businesses and consumers with a valuable tool to distinguish genuine reviews from deceptive ones, thereby enhancing customer decision-making and fostering trust in online product evaluations.

2. LITERATURE SURVEY

To mislead or deceive the customer's purchases, deceptive opinion spam is purposefully written to appear credible and authentic. Because of their nature, it is difficult for both people and algorithms to discern these opinions. The majority of studies rely on creating sparse features and conventional machine learning. However, these models do not account for the reviews' semantic content. When compared to the vast majority of conventional machine learning classifiers, only a small number of studies take into account incorporating contextual information with neural networks. Fahfouh *et al* [16] suggested a capsule neural network, bidirectional long short-term memory, attention mechanism, and paragraph vector distributed bag of words detect deceptive opinion spam. Our model provides a powerful representation of the opinions since it centers on the preservation of their contexts and the relationships between them. The results show that our model significantly outperforms the existing state-of-the-art models. Existing work mainly uses traditional discrete models with rich features from the viewpoint of linguistics and psycholinguistics. The drawback is that these models fail to capture the global semantic information of a sentence or discourse. Accordingly, Ren *et al* [17] explore a hierarchical neural network model with an attention mechanism, which can learn a global review representation from the viewpoint of the user and product, to identify deceptive reviews. Experimental results show that the proposed neural model achieves 91.7% accuracy on the Yelp datasets, outperforming traditional discrete models and neural baseline systems by a large margin. Cao *et al* [18] introduced a new feature fusion technique to fully utilize various aspects for better feature representation of identifying false reviews, and they verified its effectiveness by contrasting it with previous feature fusion strategies. TextCNN, Self-Attention, and Bidirectional Gated Recurrent Unit (GRU) were used to find the local semantic features, temporal semantic features, and weighted semantic features of reviews. The outcomes of several datasets from the part of speech perspective were used to enhance the model's interpretability. Sharmila et al [19] TFIDF victimizer is used to integrate and convert the preprocessed data, sentiment scores, and membership values. To forecast the reviews' sentiment, a pre-trained machine learning model loaded from a pickle file is employed. By combining web scraping, natural language processing, and machine learning, the program successfully demonstrates how to gain insightful information about the tone of product reviews. With a strong tool to fight fake reviews and help consumers make smart choices, our project tries to create a fair and

reliable purchasing environment.

Sparsh Kotriwal et al [20] studied the manual labelling of a small dataset and then used it as a base to check the authenticity of other reviews, but this is not considered a viable option considering the amount of manual work required. This categorizes customer reviews into fake and not fake based on review-centric features so that a reviewer may know whether a review they are reading is potential spam. Various review-centric features include the rating of the review, the product to which the review is related and the authenticity of the review. Ramadhani Ally Duma et al [21] proposed Deep Hybrid Model for fake review detection, which jointly learns from latent text feature vectors, aspect ratings, and overall ratings. Initially, it computes contextualized review text vectors, extracts aspects, and calculates respective rating values. Then, contextualized word vectors, overall ratings, and aspect ratings are concatenated. Finally, the model learns to classify reviews from such unified multi-dimensional feature representation. Extensive experiments on a publicly available dataset demonstrate that the proposed approach significantly outperforms state-of-the-art baseline approaches.

Opinion spamming harms organizations' finances and reputations in addition to both and is also very important in politics. It is increasingly harder to detect such spamming actions as technology becomes more sophisticated. A technique to recognize false product reviews was put forth by Jacob *et al* [22]. The model uses the SVM and Naive Bayes classifier to distinguish between fake and legitimate reviews. This extracts the characteristics for categorization using several criteria, including the review length, review nature, use of personal pronouns, confirmed purchase status, review rating, and product type. The experimental findings demonstrate that the model was operating effectively with a high rate of classification accuracy.

Different attentions were purposefully used at the two layers of Liu *et al* [23] proposed hierarchical attention network capture significant, thorough, and multi-granularity semantic information. Numerous trials using freely accessible datasets demonstrate that the model surpasses the baselines in terms of detection performance, raising the mixed domain score to 89.3%. Because of the extra time, social media use increased, but so did the number of cyberattacks. Detection is a very important stage in stopping these activities. To review the current state of the art in the detection of malicious content, Sanjeev Rao et al [24] proposed a framework, the datasets are balanced using Near Miss and SmoteTomek techniques to feed several machine-learning models. Later, the baseline ML models and proposed voting-based ensemble models are evaluated on imbalanced and balanced datasets. For the proposed deep learning-based hybrid approaches, embedding are generated using GloVe and FastText word embedding on the balanced combined dataset and passed into the deep neural network comprised of Conv1D and Bi-directional recurrent neural network layers with the self-attention mechanism for improved context understanding and effective results. Jitendra Kumar Rout et al [25] study the effects of network scale on network-based review spammer detection models, specifically on the trust model and the SpammerRank model. We then evaluate both network models using two large publicly available review datasets, namely: the Amazon dataset (containing 6 million reviews by more than 2 million reviewers) and the UCSD dataset (containing over 82 million reviews by 21 million reviewers). It has been observed thatSpammerRank model provides a better scaling time for applications requiring reviewer indicators and in the case of the trust model distributions are flattening out indicating variance of reviews concerning spamming.

Aswathy Velutharambath et al [26] surveyed available English deception datasets which include domains like social media reviews, court testimonials, opinion statements on specific topics, and deceptive dialogues from online strategy games. We consolidate these datasets into a single unified corpus. For instance, training on OPSPAM and testing on BLTC achieves an F1 score of 0.76 on the deception label. Training on BLTC and testing on OPSPAM is however not as good (0.66). Similar observations can be made for DEREV2014 and DEREV2018, and CROSSCULTDE and DECOP. Some of the evaluation results are encouraging, but particularly between dissimilar domains, the generalization is limited and requires future research. In Maryam Tamimi et al [27] Generative Adversarial Networks (GAN) have been utilized to generate some data with a distribution close to the original ones. In this regard and along with the successful results of Generative Pre-trained Transformers (GPT) in textual tasks, it has also been used besides the GAN framework to detect deceptive reviews. Evaluation results in comparison between different methods have shown an increase in the accuracy of 1.4% on the TripAdvisor dataset and 3.8% on the YelpZip dataset by the proposed method.

Dheeraj Kumar Dixit et al [28] proposed a hybrid convolutional neural network-based Levy flight-based honey badger algorithm that detects fake news. The proposed model offers a precision, recall, and accuracy value of 95%, 97%, and 98% when evaluated with the ISOT dataset. When compared to the existing state-of-art methods, the proposed method yielded superior detection results and higher accuracy rates. In another study, Dheeraj et al [29] recognize the detection of fake news using the LSTM-LF algorithm to classify the news as fake or real optimally. Furthermore, this paper utilizes four different datasets namely the Buzzfeed dataset, GossipCop dataset, ISOT dataset as well as Politifact dataset for evaluation.

Dharmendra Dangi et al [30] presented an Artificial Rabbits optimized Robust Random Vector Functional Link Network (RRVFLN) for improving sentiment analysis accuracy. The simulation results obtained by comparing the proposed model with the various existing techniques display the effectiveness in terms of F1 score, precision, classification accuracy, error rate prediction, and kappa statistics. In another study, Dharmendra Dangi [31] proposed a method based on five different machine learning models such as Logistic Regression, Random

Forest Classifier, Multinomial NB Classifier, Support Vector Machine, and Decision Tree Classifier. Experimental analyses are made and these classifier models are used to calculate different values such as precision, recall, f1-score, and support.

From the literature review provided, the research gaps are not explicitly stated. However, some potential research gaps can be inferred from the content. Many existing studies focus on specific text-based attributes or linguistic features for deceptive opinion spam detection. There is a need for more comprehensive models that can effectively capture semantic content, contextual information, and behavioral patterns to improve detection accuracy. Some studies focus on detecting deceptive reviews in a specific domain, such as hotels or restaurants. There is a gap in research that explores the generalization of detection models across diverse domains, as different domains may exhibit unique characteristics of deceptive spam. Imbalanced class distributions are common in deceptive opinion spam detection datasets, with the majority of reviews being genuine. More research is needed to explore effective data resampling techniques and ensemble methods that address the imbalanced class problem and improve the model's performance on minority classes. While deep learning models have shown promise in detecting deceptive reviews, there is a potential gap in exploring hybrid approaches that combine deep learning techniques with traditional machine learning classifiers or linguistic features to harness the strengths of both methods. To effectively address these research gaps, the research study could focus on developing robust models that can generalize across domains, handle imbalanced data, and explore hybrid approaches. Additionally, the evaluation of real-world datasets and their interpretability can contribute to the practical applicability and reliability of deceptive opinion spam detection systems.

3. RESEARCH PROBLEM DEFINITION AND MOTIVATION

Humans now share their opinions and experiences through online reviews. They have a say in online marketing and learn the truth about goods and services. Online reviews, however, can occasionally be fake. They could have been created to discredit a brand's reputation or to advertise low-quality goods and services to deceive potential buyers. Such deceptive evaluations are referred to as spam and need urgent action. Unfortunately, a growing number of false opinions are being expressed as a result of the business that supports it. The truth is fake opinions that have been purposefully made to look legitimate to mislead the public by endorsing subpar goods (positive deceptive opinions) or criticizing goods that could be of high quality (negative misleading opinions). The removal of misleading comments from the opinions gathered is therefore crucial.

Business and scientific groups have given deceptive review detection a lot of attention recently. The majority of the currently available research makes use of rich, classic discrete models from linguistic and psycholinguistic perspectives. The flaw in these models is that they are unable to adequately represent the overall semantic content of a sentence or conversation. Unsupervised learning techniques are still in use, but they are ineffective because they rely on the features of reviewers to identify each fake review. Recently, methods for detecting false opinions have been proposed that require a large number of examples of false and true opinions that have been labelled by domain experts. The language and psychological characteristics of fraudulent reviews, as well as other constructed elements, are the foundation of traditional false review detection techniques. However, the suggested machine learning techniques surpass all current approaches and offer higher self-adaptability to extract the desired features implicitly. In light of this, the research suggested a mixed machine-learning approach for detecting false opinion spam.

4. PROPOSED RESEARCH METHODOLOGY

Deceptive reviews are made-up testimonials that are purposefully generated to sound real to mislead readers. Because more people are using them to research products and make business decisions, online opinions are significant to both consumers and businesses. Deceptive reviews steer consumers towards buying products that are at odds with internet facts, impeding fair competition between businesses. An efficient mechanism for automating the detection of false reviews is needed to safeguard the rights of buyers and sellers. The research then concentrates on the identification of both positive and negative false opinions. Figure 1 depicts the process structure diagram for the proposed task.

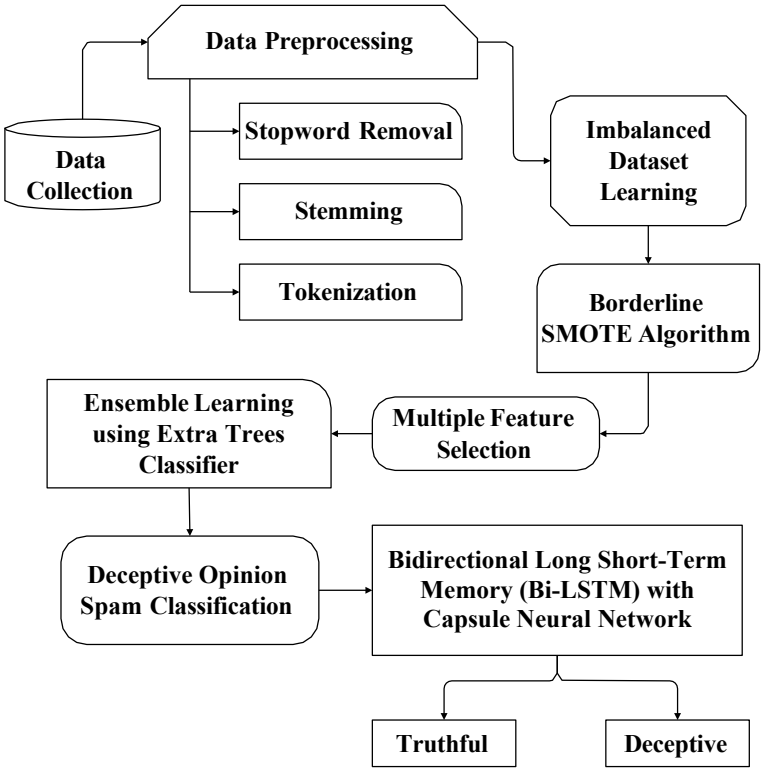


Figure 1: Block Diagram of the Proposed Work

The model was tested using the misleading reviews gathered from several online communities. To identify dishonest reviews, numerous researchers currently develop models based on single-text attributes. On the contrary, dishonest reviewers will purposefully copy the wording style of legitimate evaluations while submitting reviews. These text-feature-based techniques may or may not be successful. Consequently, the research proposed an ensemble feature selection technique which is extracted based on multiple features like text feature, behaviour feature, and deceptive score feature. In addition, a data resampling approach is exploited that integrates the Borderline-SMOTE algorithm to lessen the effects of the high dimensional imbalanced class category distribution. To detect the negative and positive deceptive opinions spam, this research study proposed a hybrid technique of Bi-LSTM with a Capsule Neural Network. The model optimizes the dynamic routing algorithm and changes the structure of the conventional capsule network without sacrificing classification performance, leading to high model accuracy.

4.1 Data Collection

This section, reports the efforts to gather standard opinion spam datasets. The datasets contain the following domains, namely hotels, and Restaurants. The article used the Yelp dataset, which includes favourable reviews for Chicago's top 20 hotels as seen on TripAdvisor (TA), for its project. A total of 800 reviews for 20 hotels were produced by 20 false reviews via AMT submissions and 20 honest reviews from 5-star reviews on TA. Deceptive reviews often include 115.75 words, while honest evaluations tend to have a similar average length. Figure 2 and 3 illustrates an example of the deceptive and positive opinion of the hotel's divine destination.

Figure 2: Deceptive Opinion of Hotel Divine Destination

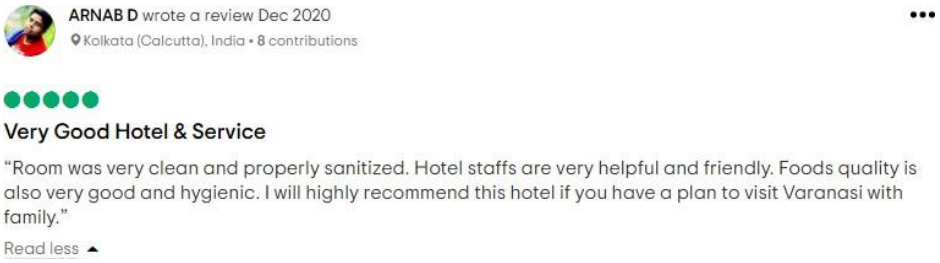


Figure 3: Positive Opinion of Hotel Divine Destination

Similarly, the article studies Yelp Restaurant's review data from the dataset. The information is arranged according to the date of the review and is linked to a Yelp product ID and user ID. Additionally, each review has a distinct rating number that ranges from 1 to 5. Moreover, labels created by Yelp's filtering algorithm are included in the dataset. Figures 4 and 5 illustrate the café raik restaurant's deceptive and positive review of visitors.

Figure 4: Deceptive Opinion of Cafe Raik

Figure 5: Positive Opinion of Cafe Raik

4.2 Data Preprocessing

Data preprocessing is a crucial step in preparing the raw online review data for machine learning algorithms. It involves transforming the textual information from the Deceptive Opinion Spam Corpus into a format that can be easily ingested by the classification algorithms. In this section, elaborate on the various preprocessing methods used in the study:

Coding the Classes: The first step in data preprocessing is to code the classes or labels of the reviews. In this case, the two classes are "Truthful" and "Deceptive." To facilitate the classification process, the classes are assigned numerical values, where "Truthful" is coded as 0, and "Deceptive" is coded as 1. This transformation allows the machine learning algorithms to process and understand the classes during training and evaluation.

Removing Numbers and Empty Text: To focus on the textual content of the reviews and avoid noise from numeric values or empty entries, all numbers and empty texts are removed from the data. This step ensures that the model is trained solely on textual information, which is the primary feature for deceptive opinion spam detection.

Lowercasing Words: To ensure uniformity and avoid duplication due to case sensitivity, all words in the reviews are converted to lowercase. For example, "This" and "this" are treated as the same word after lowercasing. This step helps in reducing the dimensionality of the data and enhances the efficiency of the model.

Stop Word Removal: Stop words are commonly used words in a language, such as articles, prepositions, and pronouns (e.g., "the," "is," "they," and "this"). These words do not carry significant meaning and may not

contribute much to the classification task. Therefore, they are removed from the text during preprocessing to focus on more informative words and improve model performance.

Stemming: Stemming is the process of reducing inflected or derived words to their common root form or base word. For example, words like "running," "run," and "ran" are stemmed from "run." This step helps to further reduce the dimensionality of the data and consolidate similar words into their base form, aiding in better feature representation and classification.

Tokenization: Tokenization is the process of breaking down the text into individual units, known as tokens. In the context of reviews, tokens can be individual words or phrases. By tokenizing the text, the model can work with discrete entities, making it easier to analyze and extract relevant features for classification.

Overall, data preprocessing is a crucial step in preparing the raw review data for further analysis. By applying techniques such as coding classes, removing numbers and empty text, lowercasing words, stop word removal, stemming, and tokenization, the textual information is transformed into feature vectors suitable for machine learning algorithms. This comprehensive preprocessing ensures that the model can effectively learn from the review data and make accurate predictions in detecting deceptive opinion spam.

4.3 Imbalanced Dataset Learning Using Borderline SMOTE Algorithm

To address the issue with unbalanced sample categories, the degree of unbalanced samples is primarily corrected at the data level by under or over-sampling samples. It is possible to erase some crucial information from the majority of samples without taking the distribution of the samples into account, but undersampling reduces sample imbalance by eliminating part of the data set's majority samples. By introducing minority class samples into the dataset, oversampling causes the dataset's sample category to approach equilibrium.

The feature vector space has a high degree of dimension for the textual features. As a result, the article has only selected terms with a frequency lower than 50. The performance of a classifier will be improved by utilising efficient feature selection approaches because some features may still be superfluous. Borderline SMOTE is an enhanced oversampling technique built on SMOTE that improves the sample category distribution by combining new samples with only a small number of class samples from the border. According to the classification of the data surrounding each minority sample, the b-nearest neighbour is determined. The Three kinds of data Danger sample, Safe sample, and Noise sample are used to categorize the original borderline SMOTE samples. The Danger samples were sparsely oversampled, before being summed upwards. The following are the algorithmic steps:

Step 1: Find the nearest m samples from the full dataset for each sample in a few classes x_i. The symbol m' stands for the number of additional categories in the most current samples m.

Step 2: Sort the samples by x_i.

The samples around x_i are all samples of different categories, and they are referred to as noise data if $m' = m$. It is recommended not to use these samples in the generation process since such data will have negative effects on the effect of generation. If $m/2 \leq m' < m$, then the majority of the m samples in the area around m are from distinct categories. Assign the boundary sample the label "Danger." If $0 \leq m' < m/2$, then more than half of the nearby m samples of x_i belong to the same categories and are therefore considered safe.

Step 3: This may perceive that $P \subseteq DANGER$ by examining the cases in DANGER, which represent borderline data from the minority class P. The model set

$$DANGER = \{x'_1, x'_2, \cdots, x'_{dnum}\}, \qquad 0 \leq dnum \leq pnum \tag{1}$$

The model determines its k nearest neighbours from P for each case in DANGER.

Step 4: Construct $s \times dnum$ synthetic positive examples in this stage using the information from DANGER. Where, s is an integer between 1 and k. For each x'_i, we randomly select s the nearest neighbours from its k nearest neighbours in P. Calculate the differences $di(j = 1,2, \cdots, s)$ between x'_i and its s nearest neighbours from P, then multiply $diff_j$ by a random number $r_j(j = 1,2, \cdots, s)$ between 0 and 1, finally, s new synthetic minority examples are generated between i'_p and its nearest neighbours:

$$synthetic_j = x'_i + r_j \times diff_j \quad j = 1,2, \cdots, s \tag{2}$$

For each x'_i in DANGER, repeat the technique above to get $s \times dnum$ synthetic examples.

Step 5: Subsequently marking, the Danger samples are expanded using the SMOTE algorithm. Determine the k-nearest neighbour samples of the same class x_{zi} for the selected x_i samples from the Danger dataset. New samples x_n are generated at random using the formula below.

$$x_n = x_i + (x_{zi} - x_i) \tag{3}$$

Where β is a random number between 0 and 1. A sample diagram of synthetic Borderline-SMOTE samples is shown in figure 6. The Noise, Safe and Danger samples are denoted as A, B, and C, respectively.

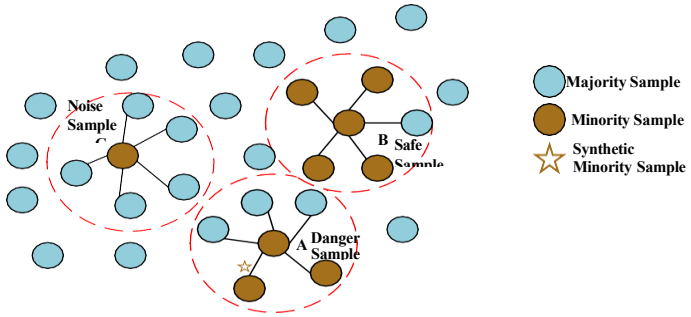


Figure 6: Synthetic Sample of Borderline SMOTE Algorithm

After locating all the Danger samples, Borderline-SMOTE determines the b-nearest neighbours of each boundary sample and synthesises additional samples similarly to SMOTE. Additionally, let the Borderline-SMOTE1 and Borderline-SMOTE2 randomly choose a few different kinds of samples from the K-nearest neighbour sample. Regardless of sample category, during Borderline-SMOTE2 and brand-new SMOTE for "Danger with any samples in the k-nearest neighbour", Borderline SMOTE, in contrast, increases the category distribution of samples by producing samples for only A.

4.4 Multiple Feature Selection Using Ensemble Learning Technique

The notion of ensemble learning, which combines weak learners to generate a strong learner to address a specific computational intelligence problem, sets it apart from single-based learning in terms of performance. Ensemble learning promotes robustness while reducing variables like volatility and biases. A bagging strategy was utilized to combine the output of these many models, such as the text feature, behaviour feature, and misleading score feature. The stability the bagging approach offers the model and its ability to lessen overfitting in models are the driving forces for its use. The study displays predictions of deceit made using the Extra Trees Classifier. Unexpectedly, misleading opinion detection has made extensive use of text, behaviour, and deceptive score aspects. To enhance performance, the paper combines the results of classifiers trained on both text and behavioural data.

4.4.1 Extra Trees Classifier Technique

Extremely Randomized Trees Classifiers (also known as Extra Trees Classifiers) are an ensemble learning technique that combines the output of various de-correlated decision trees gathered in a "forest" to produce a single classification result. The only way it differs conceptually from a Random Forest Classifier is in how the decision trees in the forest are built. The Extra-Trees approach uses the traditional top-down construction method to generate an ensemble of decision trees or unpruned regression. By selecting cut spots entirely at random, it splits nodes, unlike other tree-based ensemble approaches, and instead of a bootstrap replica. The two parameters K, the number of characteristics selected at random for each node, and n_{min}, the minimal sample size for splitting a node, is used in the Extra-Trees splitting algorithm for numerical attributes. The (whole) original learning sample is utilised numerous times with it to create an ensemble model (the number of trees in this ensemble is indicated by M). It calculates the splitting value and each dimension's score as follows.

$$s_d = |L| f_{sc}(y_L) + |R| f_{score}(y_R) \quad (4)$$

The number of data points assigned to text features is $|L|(|R|)$, and the y values in the behaviour features are denoted by y_L (y_R). The Gini index is calculated as follows for the binary classification function f_{score}:

$$f_{sc}(y) = 1 - (p^2_{-1} + p^2_{1}) \quad (5)$$

Where, p_{-1} represents the percentage of samples where $y = -1$ and the percentage of $y = 1$ is p_1 respectively. The misleading score for regression is f_{score}, which is variance-negative.

$$f_{score}(y) = -\frac{1}{n}\sum^{n}_{i=1}(y_i - mean(y))^2 \quad (6)$$

From the above equation, n denotes the size of y and $m(y) = \frac{1}{n}\sum y_i$. It selects the dimension with the greatest score s_d, saves its split in the node, and then repeatedly executes these three stages on the two subtrees built of L and R, respectively. The number of features in a random sample of two features is determined by the value of k. Information Gain will be the deciding factor in this situation. The entropy of the data should first be calculated. Take note that the following formula is used to determine entropy:

$$Entropy(S) = \sum_{i=1}^{c} -p_i log_2(p_i) \quad (7)$$

Where p_i indicates the percentage of rows where the output label is , and c represents the number of distinct class labels. As a result, the entropy for the provided data is:

The system's entropy for a certain property is known as conditional entropy A

$$Entropy(S|A) = \sum_{a\epsilon A} p(a)Entropy(S|A = a)$$

(8)

The degree to which the information of attribute A reduces the uncertainty of the information of X is measured by the information gain.

$$G(S, A) = Ent(S) - Entropy(S|A) \quad (9)$$

The purpose of dehoaxing is to prevent possible future harm to the subject. Subjects may be given false pretest scores to test the effect of these scores on subsequent tests of motivation levels. The information acquisition rate is as follows because the information gain criterion favours more appealing features.

Where, $\underset{IV(S)}{inRatio(S, A)} = \frac{Gain(S,A)}{\qquad}$

$$IV(S) = -\sum_{a\epsilon A} p(a|A)logp(a|A)$$

(10)

(11)

$(a|A)$ denotes the percentage of a in A, and IV (intrinsic value) denotes a fixed value for an attribute. If A only contains one type of data, for example, if $(a|A) = 1$, then $(A) = 0$. As a result, the formula-based estimation of the information gain rate will be incorrect. The information gain rate in this instance is 0. Up until a predetermined minimum node size (N_{leaf}) is attained, and Extra Trees establishes each tree iteratively. Further, a leaf node is constructed, and it holds the output value that is produced most frequently (for classification) and the average value (for regression), respectively.

4.5 Deceptive Opinion Spam Classification

Opinion spam detection plays a vital role in ensuring accurate opinion mining, especially in the era of widely available online opinion reviews. These reviews serve as essential sources of information for both consumers and businesses. However, the rising prevalence of false opinions, driven by deceptive practices, poses a significant challenge. Crafted to appear authentic, these deceptive comments aim to mislead individuals into buying either low-quality products through positive deceptive opinions or high-quality products through negative deceptive opinions. Addressing this issue is crucial to maintain trust and reliability in online reviews, benefiting both consumers and businesses alike.

4.5.1 Bidirectional Long Short-Term Memory (Bi-LSTM)

Bi-LSTM is a variant of the Long Short-Term Memory (LSTM) model that processes input sequences in both forward and backward directions. This bidirectional processing enables Bi-LSTM to capture contextual information from both past and future words in a given review text. In Deceptive Opinion Spam Classification, this allows the model to understand the relationships between words and phrases and extract essential features from the text. Bi-LSTM helps in overcoming the challenge of long-distance dependencies in text, as it retains the memory of words encountered earlier in the sequence while processing subsequent words. Word representation is used by BiLSTM to compose each word; vectors are transformed into document vectors using document representation. BiLSTM has sequential, complete information about every word before and after it in a particular document, which helps to partially resolve the long-distance relationship. To transform the document vector into a real-valued vector whose length is the number of classes, a tan function is introduced in particular. Real vector values are converted into conditional probabilities by adding two linear layers. First-person pronouns and POS are two types of features that the proposed technique incorporates into the BiLSTM model.

Word embedding serves as a representation of a word's semantic and syntactic information, which converts each word into a real-valued vector. A hyperparameter called d_w size is typically set to 50 or 100 words. Additionally, other basic feature types are transformed into vector representations, or r^{fn}, where n denotes the nth feature type and d^{fn} stands for dimensions. The embedding function of the torch framework generates the vector's initial value at random. Each word is represented by concatenating all of the initial feature embeddings in the manner shown below: given a phrase $x = \{w_1, w_2, \cdots, w_n\}$.

$$x_i = \{r^w, r^{f1}, r^{f2} \cdots, r^{fn}\} \quad (12)$$

Where $\underset{i}{r^{fk}}$ denotes the kth type embedding feature and $\underset{i}{r^w}$ indicates the 300-dimension Glove embedding of words. The new embedding x_i is generated by joining the $\underset{i}{r^w}, \underset{i}{r^{f1}}, \underset{i}{r^{f2}} \cdots, \underset{i}{r^{fn}}$. There are forward transfer and backward transfer layers in BiLSTM. Contextual data is fully taken into account by the bidirectional structure while encoding review documents. The local feature vector and the previous layer are combined to create the global feature vector in the pooling layer. The study employs a one-dimensional maximum pool in this model because it performs better for the deceptive review detection work than other multiple-dimensional maximum pools. To extract the high-level characteristics, a $tanh$ layer transforms the obtained vector. Additionally, using

$tanh$ as the activation function will constantly increase the cycle's feature effect. All of the word vectors in each review are converted into a document vector via the document representation layer. By converting all word vectors in each review into a document vector through the document representation layer, the model can better capture the overall sentiment and context of the review text, contributing to more accurate and reliable detection of deceptive opinions.

4.5.2 Capsule Neural Network Approach

Capsule Neural Network is a deep learning architecture designed to represent hierarchical relationships within data. It uses capsules, which are groups of neurons, to encode different parts of an object or feature. In the context of Deceptive Opinion Spam Classification, the capsule network can learn to represent complex linguistic patterns and dependencies in reviews. This allows the model to capture the "part-and-whole" relationships between words and phrases, enabling it to understand the context and semantic meaning in the text. It takes as input the text representation produced by the syntax and attention module to extract additional features. The module's layers each can extract various levels of characteristics. Eventually, a feature representation of the complete text for categorization is formed by further combining low-level features to produce higher-level features.

The text representations are stated as X_1 and X_2 produced by the attention and syntax module into a single-layer network in the first layer, which is called the fusion layer:

$$X_3 = W^{f1} \cdot X_1 + W^{f2} \cdot X_2 \tag{13}$$

Where, W^{f1}, $W^{f2} \epsilon \mathbb{R}^L$. This stage involves the linear transformation and combination of two-sentence matrices. The convolutional layer, which makes up the second layer, extracts N-gram phrase characteristics from various textual locations. To generate the N-gram feature matrix M, this layer convolutions the sentence matrix X_3 using k_1 convolution filters.

$$M = [M_1, M_2, \cdots, M_{k_1}] \epsilon \mathbb{R}^{(L-N+1) \times k1} \tag{14}$$

In this case, $M_{k_1} = [m_1, m_2, \cdots, m_{L-N+1}] \epsilon \mathbb{R}^{(L-N+1)}$ is the k_1-th column vector in M, and procedure (15) is used to acquire each element in this vector, m_i.

$$m_i = (W^{c1} \cdot x_{i:i+N-1} + b_1) \tag{15}$$

Where, $x_{i:i+N-1}$ indicates that N-word vectors in the sentence are connected in series, f is the nonlinear activation function, $W^{c1} \epsilon \mathbb{R}^{N \times d}$ denotes the k_1-th convolution filter and bias item are stated as b_i. The principal capsule layer, which is the third layer, combines the N-gram phrase features that were collected from the same area as the capsules. The feature matrix M is transformed into the principal capsule matrix P by this layer using k_2 transformation matrices.

$$P = [P_1, P, \cdots, P_{k_1}] \epsilon \mathbb{R}^{(L-N+1) \times k2 \times l} \tag{16}$$

Where, $P_{k_2} = [p_1, p_2, \cdots, p_{L-N+1}] \epsilon \mathbb{R}^{(L-N+1) \times l}$ is the k_2-th column capsule in P. The prediction vector is operated (17):

$$u_r = (\sum_i c_i \cdot \hat{p_i}) \tag{17}$$

Where c_i denotes the coupling coefficient and g indicates the nonlinear activation function, the dynamic routing algorithm updates c_i. Within a window, the similarity between the generated convolutional and primary capsules varies. A primary capsule with a lot of resemblances needs to weigh more. This idea is the foundation of the equation (17). The final layer is used to generate the fully connected capsule layer, which represents the capsule Y for the category

$$Y = [y_1, y_2, \cdots, y_j] \epsilon \mathbb{R}^{j \times l} \tag{18}$$

Where, the $j - th$ category's capsule is indicated by the notation $y_j \epsilon \mathbb{R}^l$. The process (17) is carried out to obtain y_j after the capsules in U are linearly transformed (16) to provide the prediction vector $u_{j|r}$. Figure 7 depicts the fully linked capsule layer, which may alternatively be thought of as the fourth layer of a convolution window operation.

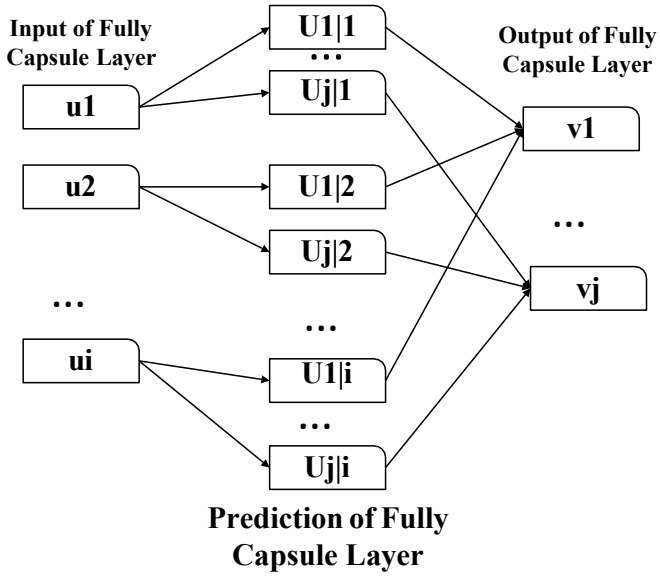


Figure 7: Fully Connected Capsule Layer

The modulus of the capsule vector ultimately determines the likelihood of belonging to the category that represents it in the completely connected capsule layer. Accordingly, the technique classifies the deceptive opinion based on positive and negative sentiment to detect the truthful and deceptive opinions among the collected reviews. Combining Bi-LSTM and capsule neural network, the Deceptive Opinion Spam Classification model can effectively capture both local and global features within the review text. The Bi-LSTM provides contextual information and sequential dependencies, while the capsule network focuses on the hierarchical structure of the text. This integration allows the model to represent the reviews comprehensively and capture nuanced patterns, resulting in accurate and reliable classification of deceptive opinions.

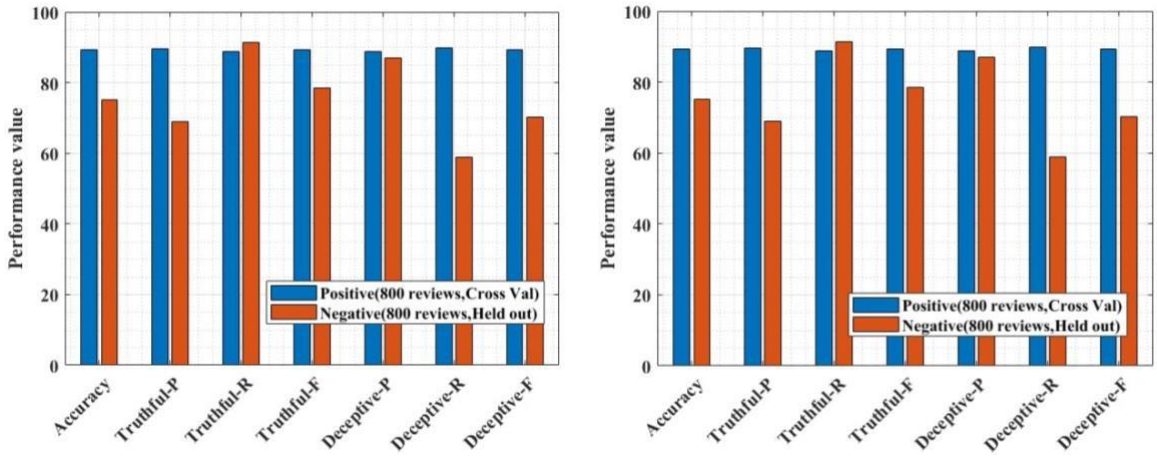
5. EXPERIMENTATION AND RESULTS DISCUSSION

The research utilized the Benchmark dataset of the Deceptive Opinion Spam Corpus, comprising 1600 reviews of hotels and restaurants, to investigate and evaluate various techniques. The hotel review dataset was sourced from https://www.kaggle.com/datasets/rtatman/deceptive-opinion-spam-corpus, while the restaurant data was obtained from http://odds.cs.stonybrook.edu/yelpzip-dataset/. This Yelp dataset encompasses both recommended and filtered reviews, originating from 201 hotels and restaurants and authored by 38,063 reviewers. The reviews were categorized as truthful or deceptive. To conduct the evaluation, the data was split into test and train sets, with 80% used for training the model and the remaining 20% for result computation. Pre-processing methods were employed for any new test case, generating a feature vector with a class label. Evaluation measures, such as accuracy, precision, F1-score, recall, and specificity, were utilized to assess the proposed system's performance. The experiments were implemented using Python programming language.

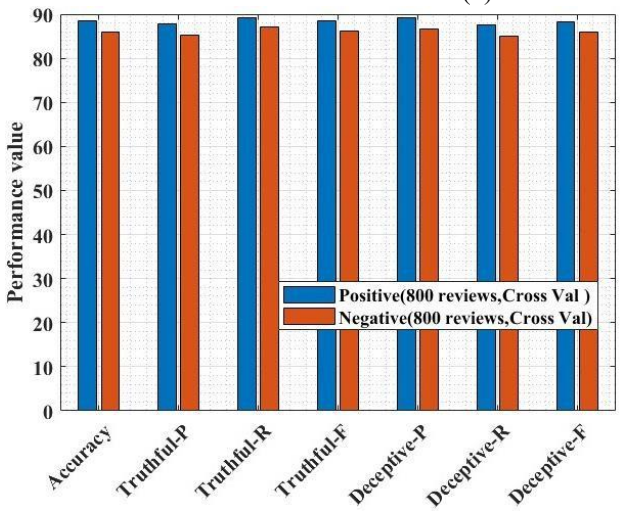
Table 1: System Configuration

Python Jupiter	Version 3.8.0
Operation System	Windows 10 Home
Memory Capacity	6GB DDR3
Processor	Intel Core i5 @ 3.5GHz

Python Software has implemented the recommended algorithm. This approach has been tested and assessed by using Python Jupiter. Table 1 depicts the system settings for simulation. Windows 10 Home and its 6GB DDR3 memory capacity. The Python 3.8.0 processor is an Intel Core i5 @ 3.5GHz.



(a) Train Sentiment Positive (b)Tarin Sentiment Negative



(c)Train Sentiment Combined

Figure 8: Performance Value of Deceptive and Truthful Reviews

Performance for the three trained sentiments appears in figure 8. Deception detection performance includes Precision, Recall, and (F)1-score, for three trained sentiments on a set of 800 negative, positive and combined reviews are depicted in figure 8(a), (b) and (c). This shows the deception detection performance of two positive and negative that aggregate the assessments with the 800 reviews: (1) the majority predicts deceptively two out of others predict deceptive (and truthful otherwise), and (2) the in combined strategy the accuracy is about 88.62 and the performance of f1-score, precision, and recall of truthful is (85.9, 88.0, 87.6) and deceptive (87.8, 87.0, 89.9) respectively.

5.1 Performance Metrics

F1 score, accuracy, recall, and precision were the model evaluation metrics used in the experiment to assess the performance of several prediction models. According to the study, FP and FN stand for sample numbers of false-positive and false-negative instances, respectively, whereas TP and TN stand for sample numbers of true-positive and true-negative cases, respectively. The following definitions apply to the computation parameters:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (19)$$

$$Precsion = \frac{TP}{TP+FP} \quad (20)$$

$$Recall = \frac{TP}{TP+FN} \quad (21)$$

F1 Score: This measurement, which has the parameter value of $\beta = 1$, is the most popular parametric family member of the F-measures. The harmonic mean of recall and precision is what makes up the F1 score, and as a function of M, it has the following properties:

$$F1 = \frac{2*precision*recall}{Precision+Recall} \quad (22)$$

F1 has a range of [0, 1], with the minimum value for TP = 0, or incorrect classification of all positive samples, and the maximum value for FN = FP = 0, indicating accurate classification.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}} \quad (23)$$

The majority of positive data instances and the majority of negative data instances must have been properly predicted by the binary predictor for MCC to produce a high score. The extreme values, 1 and +1, are attained in the situation of perfect misclassification respectively. MCC = 0 is the anticipated outcome for the coin-tossing classifier. The study utilized the false positive rate (FPR), true positive rate (TPR), precision, F1-score, and G-mean to evaluate the effectiveness of the suggested spam detection approach in class imbalance problems because the accuracy rate does not adequately reflect the whole situation. These indicators are computed using a confusion matrix.

Table 2: Confusion Matrix

	Predicted Spam	**Predicted Non-Spam**
Actual Spam	True Positive (TP)	False negative (FN)
Deceptive	False positive (FP)	True negative (TN)

In Table 2, each column and row represents the anticipated and deceptive class. The minority class is used as the majority and the positive class is used as the negative class in this paper.

True Positive (TP): Number of successfully predicted spam reviews, indicating that both the actual and predicted values of the class are spam.

True Negative (TN): Number of reviews that were accurately predicted as being true, indicating that both the value of the actual and predicted class is accurate.

False Positive (FP): The number of spam reviews that were erroneously forecasted as spam, meaning that the value of the actual class is spam while the predicted class is accurate.

False Negative (FN): Number of true reviews wrongly predicted as being truthful, indicating that the actual class is truthful while that of the anticipated class is spam.

G-mean: Using the ratio of positive accuracy to negative precision, the G-mean measure assesses the level of inductive bias. The classifier performs better at classifying both majority and minority classes, as indicated by a larger G-mean.

$$G - mean = \sqrt{\frac{TP}{TP+FN} * \frac{TN}{TN+FP}} \quad (24)$$

Deep learning is used to determine all the ideal parameters for all the base classifiers based on 10-fold cross-validation to increase performance and produce superior classification results. According to the experimental protocol, the classification performance is best when three hidden layers of the BiLSTM are used, with 64 hidden units in the first layer, 32 in the second, and 16 in the third. The hidden layer is activated using a sigmoid function in this method. To prevent overfitting, the model also employs dropout.

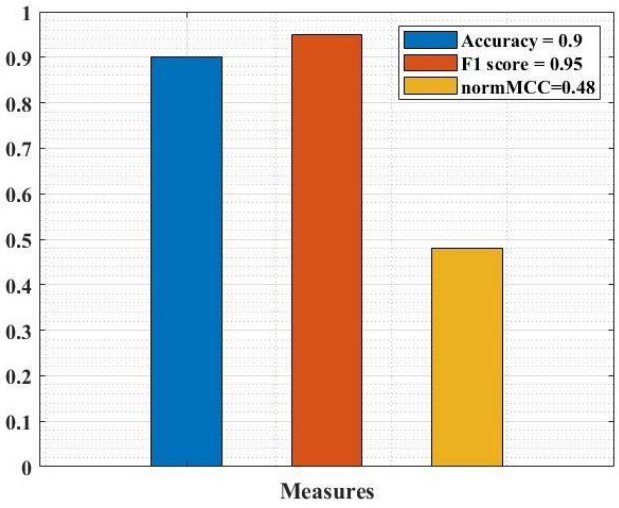


Figure 9: Performance Analysis of Imbalanced Dataset

Figure 9 portrays the performance measures of the imbalanced dataset. Over all other deep learning models, the suggested Bi-LSTM with capsule Network model recorded the greatest assessment metrics, followed by the BiLSTM. The proposed model has a higher F1-score than the machine learning model, at 95%. This finding emphasises how superior deep learning models are to machine learning at detecting aggressiveness. Moreover, the proposed technique accuracy is about 0.90 and the value of normMCC for the suggested technique is obtained at about 0.48 respectively. Accuracy and F1 scored highly in this instance, though, with accuracy at 0.90 and F1 at 0.95, both near the maximum allowable value of 1.00 in the [0, 1] interval. At this moment, if one selects to assess the efficacy of this classifier just based on accuracy and F1 score, one would overly believe that the computational method produced superior predictions.

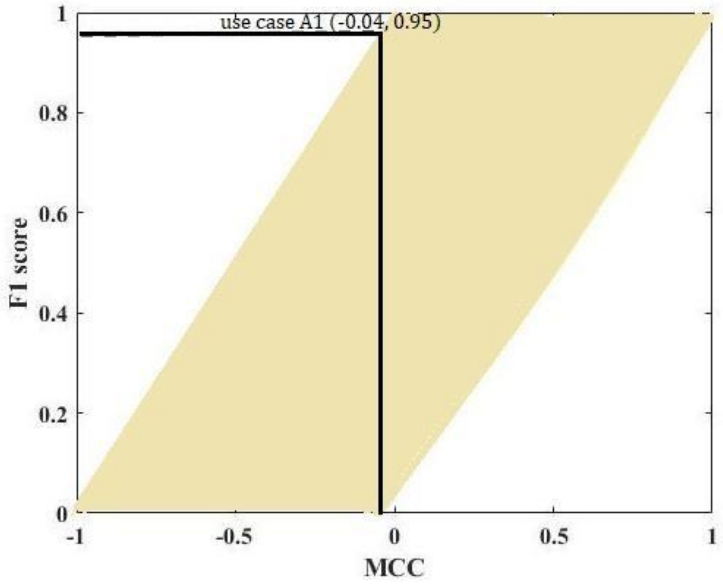


Figure 10: F1 Score and MCC Relationship in Use Case Analysis

For a dataset with 500 samples, Figure 10 shows the scatterplot of the MCCs and F1 scores for each potential confusion matrix. This allows for the utilisation of the scatterplot to show the relationship between the F1 score and MCC. The scatterplot cloud indicates that there is a large range of MCC values for each value of the F1 score and vice versa, however with different widths, even if the two measurements are relatively concordant. The Pearson correlation value between F1 and MCC climbs to 0.956585 when the analysis solely takes into account specific confusion matrices with datasets of size 500 when TP=TN.

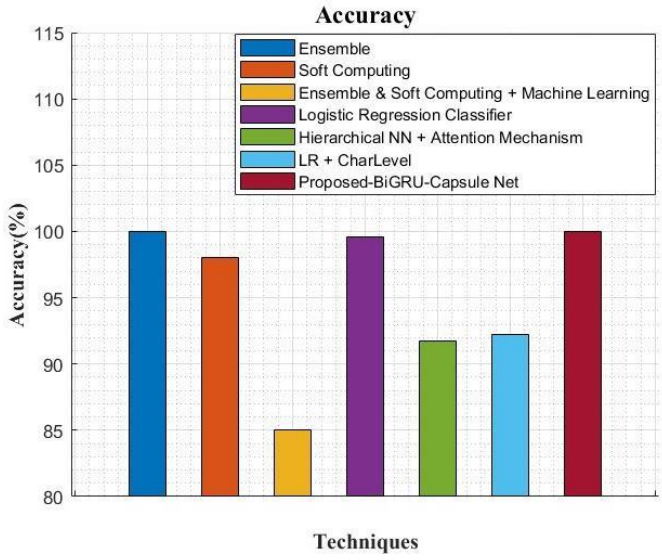


Figure 11: Accuracy Comparison Analysis

Figure 11 provides a comparison of different techniques used for deceptive opinion spam detection along with their corresponding accuracy scores. Each technique is evaluated based on its performance in correctly classifying deceptive and truthful reviews. In the evaluation of different techniques for deceptive opinion spam detection, several approaches were considered. The ensemble method demonstrated exceptional performance with an accuracy of 99.98%. On the other hand, the logistic regression classifier, a linear model for binary classification, proved surprisingly effective with an accuracy of 99.55%. The hierarchical neural network model with an attention mechanism achieved an accuracy of 91.7%. Designed to capture hierarchical relationships in the data, with the assistance of the attention mechanism, this technique can effectively focus on relevant parts of the input. Additionally, LR + CharLevel, which involves using logistic regression with character-level features, exhibited a respectable accuracy of 92.19%. Moreover, the highest accuracy was obtained in the proposed model of 99.99%, which combines Bidirectional Long Short-Term Memory (BiGRU) with a Capsule Neural Network. The model demonstrates outstanding performance, achieving almost perfect accuracy in detecting deceptive opinion spam.

5.2 Comparative Analysis with Existing Techniques

The paper contrasts the suggested approach with four established machine learning algorithms in this part. All traditional algorithms operate with their default settings for simplicity. The model uses a 10:1 class imbalance ratio as an illustration. The results of the categorization are expressed as TPR, FPR, precision, and accuracy of the test data.

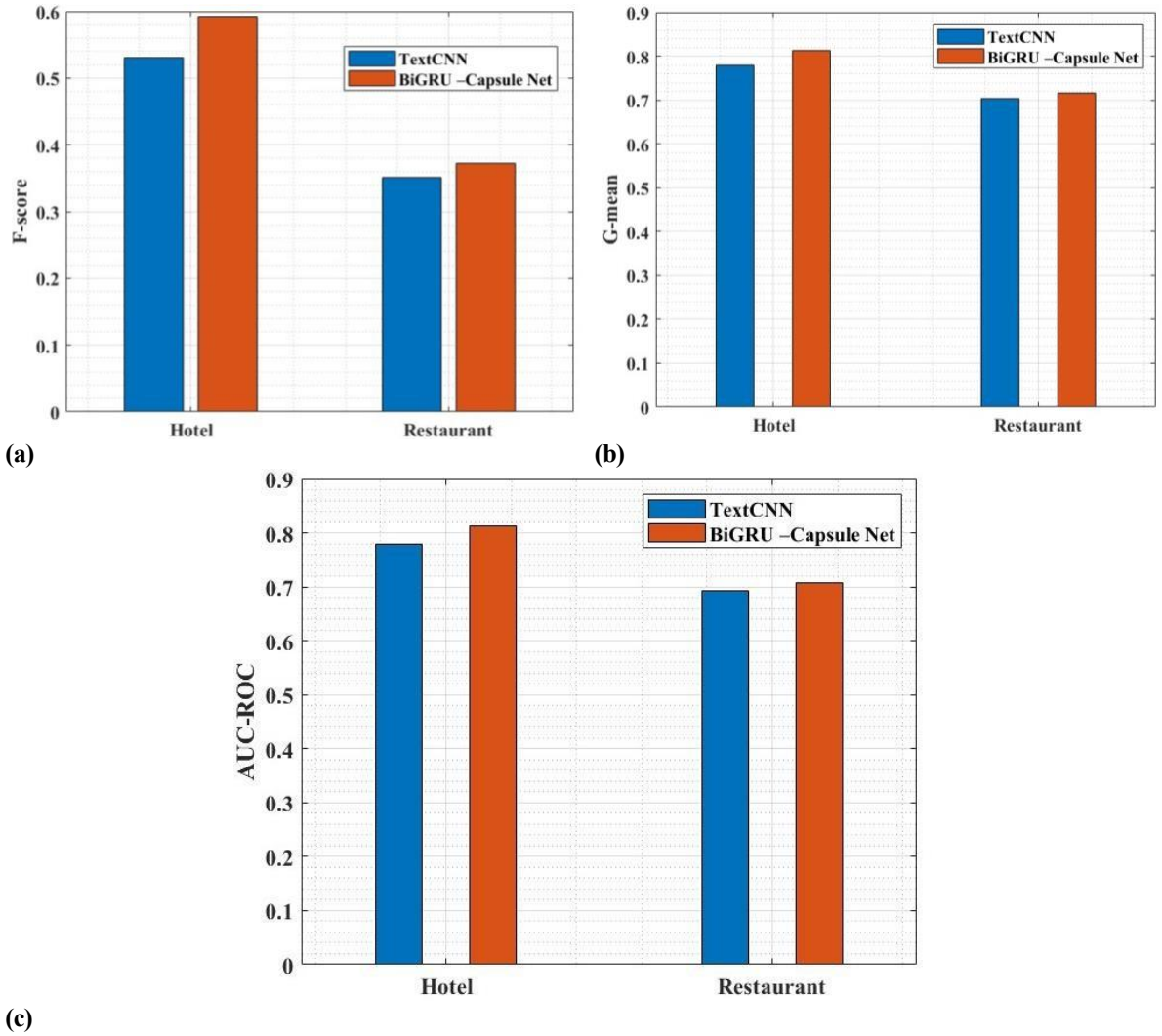


Figure 12: Performance Metrics Comparison Analysis with Dataset

Figure 12 depicts the results of different performance metrics with classifiers on the hotel and restaurant dataset. Figure 12(a) illustrates the f-score comparison analysis of the different datasets. Figure 12(b) exhibits the G-mean performance comparison analysis, and in figure 12(c) the AUC-ROC performance of the dataset is presented. It is worth noting that for the hotel dataset, the suggested BiLSTM-Capsule Net performed consistently

well for other techniques. F-score performed slightly well for the hotel dataset for opinion spam detection, whereas the proposed outperformed others for G-mean and AUC-ROC with a value of 0.82. In figure 12(a), the f-score is slightly lower for the restaurant dataset when compared to the hotel. Figures 12(b) and 12(c) show the metrics performance of base classifiers of the hotel dataset with 0.712, where BiLSTM-Capsule Net outperformed TextCNN classifiers with greater performance.

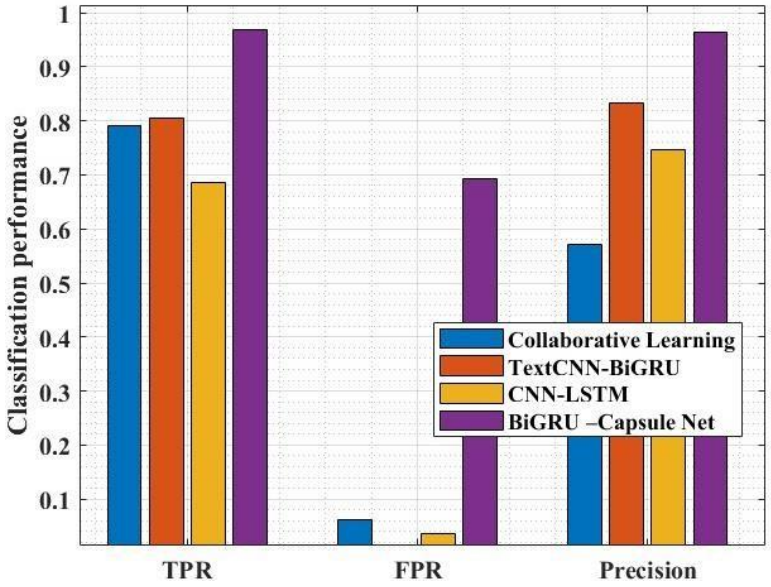


Figure 13: Classifier Performance of Metrics

With an unbalanced data rate of 10, Figure 13 shows the classification performance of each technique in terms of TPR, FPR, and accuracy. The bar graph demonstrates that only the TextCNN-BiGRU method and the proposed method have stable metrics, whereas the other methods have metrics that fluctuate greatly. The precision of BiLSTM-Capsule Net, for instance, is 0.985, but its TPR and FPR are 0.99 and 0.7, respectively. This shows that the categorization results in a tilt towards non-spam and that the technique is strongly influenced by the dominant class. The TextCNN-BiGRU approach's TPR and accuracy demonstrate consistent performance in the 0.8-0.9 range, however, their false positive rates are lower than those of the suggested method.

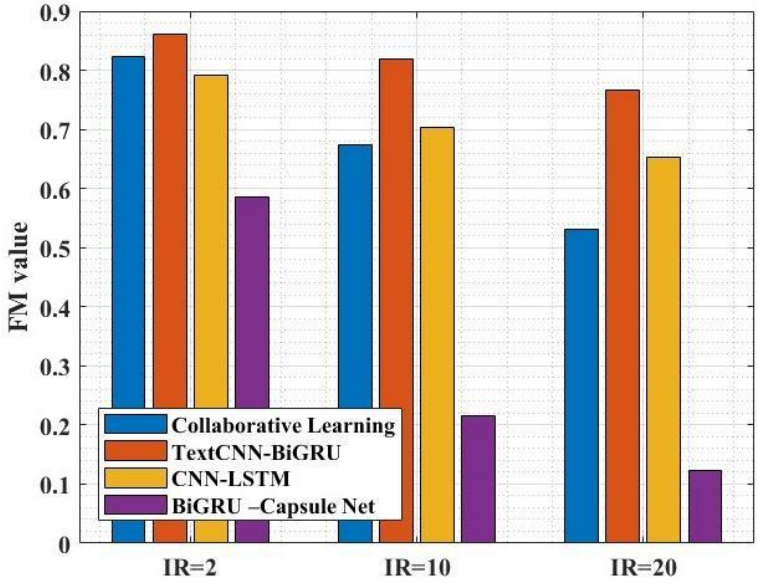


Figure 14: F-Measure Comparison with Imbalanced Data Rate

The effectiveness of each strategy is shown in Figure 14 in terms of F-measure (FM). According to the model, the trend of the F-measure metric declines as the rate of class imbalance rises by 0.91%. For instance, the performance of CNN-LSTM, which comes in second place, declines from roughly 0.76 with IR = 2 to 0.63 with IR = 20. With a rise in imbalance ratio, TextCNN-BiGRU performs the worst, yielding only roughly 0.12 FM value; as a result, the classification is useless from a practical standpoint. Despite also exhibiting a downward tendency, the collaborative strategy nevertheless gets the highest FM value among all the categorization strategies. For instance, when IR = 20, it exceeds the second-best strategy, the C4.5 decision tree, by more than 0.11, respectively.

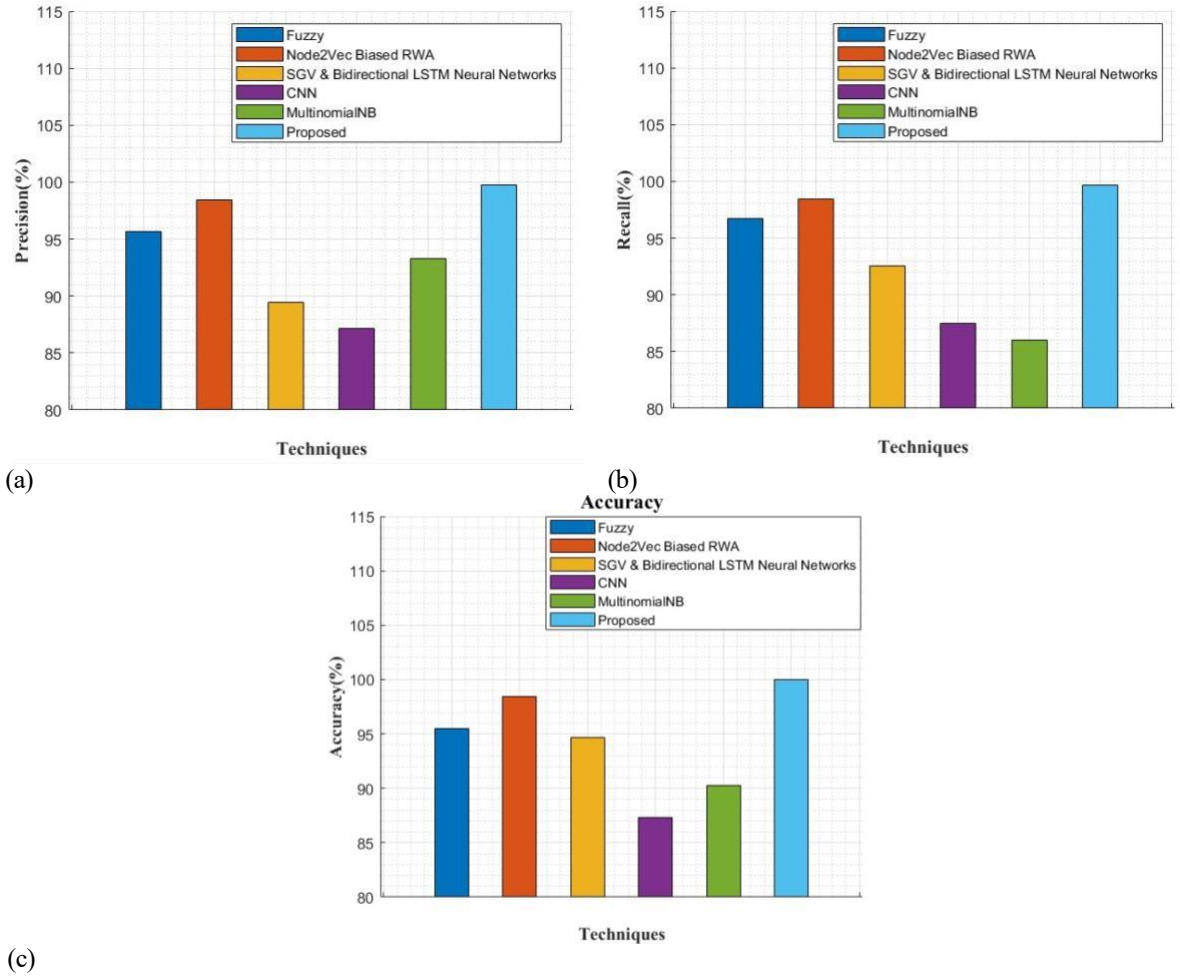


Figure 15: Evaluation Metrics Comparison Analysis with Existing Works

Figure 15 presents the performance evaluation of various methods for deceptive opinion spam detection, with each method being assessed based on precision (a), recall scores (b) and accuracy (c),. The fuzzy logic-based classification demonstrated strong results with an accuracy of 95.5%, precision of 95.7%, and recall of 96.7%, showcasing its effectiveness in handling uncertainty and imprecision in data to classify deceptive and truthful reviews. The Node2Vec biased random walk algorithm achieved an impressive accuracy of 98.44%, precision of 98.44%, and recall of 98.44%. Leveraging graph embedding techniques, Node2Vec learns representations of nodes in a network and effectively detects deceptive opinion spam with its biased random walk approach. MultinomialNB, a probabilistic model well-suited for text data, achieved an accuracy of 90.25%, precision of 93.25%, and recall of 86.01%, indicating its effectiveness in detecting deceptive opinions. Impressively, the deep neural network outperformed others with an accuracy of 99.9%, precision of 99%, and recall of 99%. Its ability to learn complex patterns makes it highly suitable for deceptive opinion spam detection. Finally, the proposed method surpassed all others with the highest accuracy of 99.99%, precision of 99.73%, and recall of 99.67%. Representing the model suggested in the study, it likely incorporates multiple techniques or innovations, contributing to its superior performance in detecting deceptive reviews.

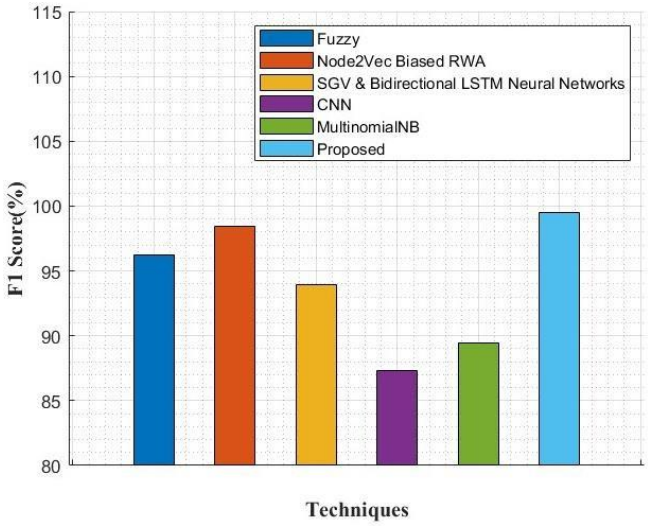


Figure 16: F1-Score Comparison Analysis

In this comparison analysis of different methods for deceptive opinion spam detection, figure 16 evaluated their performance based on the F1-score metric, which is a harmonic mean of precision and recall, offering a balanced assessment of a model's effectiveness. The fuzzy logic-based classification method achieved an F1-score of 0.962, indicating its ability to strike a good balance between precision and recall, and hence correctly classify both deceptive and truthful reviews with high accuracy. Further, the deep neural network exhibited outstanding performance with an F1-score of 0.99. The proposed method outperformed others with an impressive F1-score of 99.50%. Representing the model suggested in the study, it likely incorporates innovative techniques, leading to exceptional performance in detecting deceptive reviews.

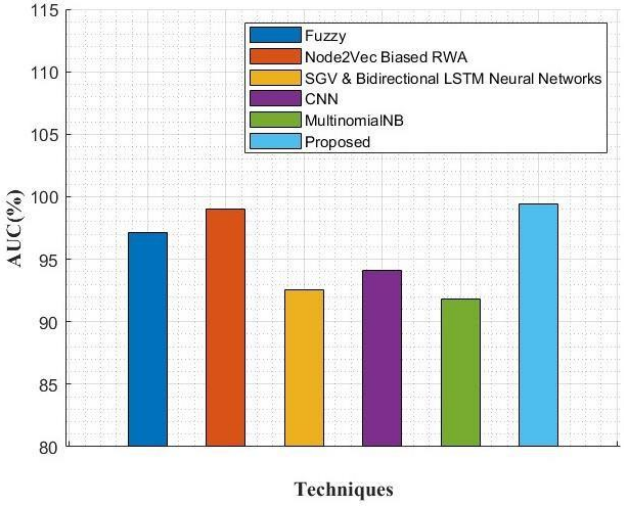


Figure 17: AUC Comparison Analysis with Existing Works

In the comparison analysis of figure 17 various techniques for detecting deceptive opinion spam, the study evaluated their performance based on the Area Under the Curve (AUC) values. The AUC values serve as a reliable metric to assess the models' ability to distinguish between deceptive and truthful reviews. The Fuzzy Logic-Based Classification method achieves an AUC of 0.971, indicating its effectiveness in discriminating between deceptive and truthful reviews. Similarly, the Node2Vec Biased Random Walk Algorithm demonstrates exceptional performance with an impressive AUC of 99%, and the Deep Neural Network stands out with an outstanding AUC of 0.99, reflecting its exceptional capability in discerning deceptive opinions from truthful ones. However, the Proposed Method overcomes all other techniques, achieving a remarkable AUC of 99.42%, which is very close to a perfect AUC score of 100%. This indicates its exceptional performance in detecting deceptive reviews.

Table 3: Comparison Analysis of Performance Metrics with State-of-the-Art Techniques

Method	Accuracy	Precision	Recall	F1-score	AUC
Fuzzy Logic-Based Classification	95.5%	95.7%	96.7%	96.2%	97.1%
Node2Vec biased random walk algorithm	98.44%	98.44%	98.44%	98.44%	99%
Stanford Global Vectors and Bidirectional-LSTM NN	94.66%	89.43%	92.53%	93.96%	92.53%
CNN	87.29%	87.13%	87.50%	87.31%	94%
Multinomial-NB	90.25%	93.25%	86.01%	89.48%	91.8%
Deep Neural Network	99.9%	99%	99%	99%	99%
Proposed	99.99%	99.73%	99.67%	99.50%	99.42%

Table 3 presents a comprehensive comparison analysis of various performance metrics for deceptive opinion spam detection using different methods, including fuzzy logic-based classification, Node2Vec biased random walk algorithm, Stanford Global Vectors with Bidirectional-LSTM Neural Networks, CNN (Convolutional Neural Network), Multinomial-NB (Multinomial Naive Bayes), Deep Neural Network, and the Proposed method. The Fuzzy Logic-Based Classification method achieved commendable accuracy (95.5%), precision (95.7%), recall (96.7%), and F1-score (96.2%). Its AUC of 97.1% indicates its effectiveness in correctly classifying both deceptive and truthful reviews. The approach combining Stanford Global Vectors with Bidirectional-LSTM Neural Networks displayed good performance with an accuracy of 94.66%, precision of 89.43%, recall of 92.53%, and F1-score of 93.96%. Although it showed slightly lower precision and recall compared to the Node2Vec algorithm, it still exhibits competence in detecting deceptive opinions, as indicated by its AUC of 92.53%. Despite being primarily designed for image recognition tasks, the CNN method provided competitive results in detecting deceptive opinion spam, with an accuracy of 87.29%, precision of 87.13%, recall of 87.50%, and F1-score of 87.31%. Its AUC of 94% further supports its ability to distinguish between deceptive and genuine reviews. The Deep Neural Network achieved outstanding performance, nearly perfect in classifying deceptive opinions, with accuracy, precision, recall, and an F1-score of 99%. Its AUC of 99% reflects its effectiveness in distinguishing between deceptive and truthful reviews. The Proposed Method surpassed all other techniques, achieving the highest accuracy of 99.99%, precision of 99.73%, recall of 99.67%, and F1-score of 99.50%. Its AUC of 99.42% demonstrates its exceptional performance in detecting deceptive reviews. The Proposed method represents the model suggested in the study, likely combining multiple techniques or innovations to achieve such superior performance. The comparison analysis showcases the varying levels of performance exhibited by different techniques in detecting deceptive opinions.

6. RESEARCH CONCLUSION

The article focused on enhancing the misleading spam opinion spam detection task. The model was tested using the misleading reviews gathered from several online communities. Consequently, the research proposed an ensemble feature selection technique which is extracted based on multiple features like text feature, behaviour feature, and deceptive score feature. In addition, a data resampling approach is used that integrates the Borderline- SMOTE algorithm to lessen the effects of the high dimensional imbalanced class category distribution. This research study developed a hybrid technique of Bi-LSTM with a Capsule Neural Network to detect the positive and negative false opinions spam. The experiment was conducted in two different datasets using the Yelp platform. The results on numerous benchmark datasets of fraudulent reviews outperformed the state-of-the-art at the time. The study assessed how well the suggested method performed on additional review-related tasks, such as review sentiment recognition, and it did so with state-of-the-art accuracy on two benchmark datasets. The suggested technique performs well in the task of detecting fake reviews, according to the performance analysis findings of the area under the curve (AUC), macro average precision, and weighted F1-score, with 87.8, 87.0, and 89.9%, respectively. The accuracy of the suggested technique is also higher than the accuracy of the current technique. The experimental results using the benchmark dataset show that both the suggested feature-based hybrid approach and the baseline achieve better performance.

Statements and Declarations

Conflict of Interest

The authors declare that they have no conflict of interest to disclose.

Funding

Not Applicable.

Author Contribution Statement

All authors contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

REFERENCES

[1] Mewada A, Dewang RK (2021) Deceptive reviewer detection by analyzing web data using HMM and similarity measures. Materials Today: Proceedings.

[2] Zhong M, Li Z, Liu S, Yang B, Tan R, Qu X (2021) Fast Detection of Deceptive Reviews by Combining the Time Series and Machine Learning. Complexity 2021.

[3] Sultana N, Palaniappan S (2020) Deceptive Opinion Detection Using Machine Learning Techniques. International

Journal of Information Engineering & Electronic Business 12(1).
[4] Toplu A, Liu H (2021) Designing a Deceptive Comment Detection Platform with a Rule-based Artificial Intelligent Architecture. In 2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE 1442-1445.
[5] Jayathunga DP, Ranasinghe RM, Murugiah R (2021) A Comparative Study of Supervised Machine Learning Techniques for Deceptive Review Identification Using Linguistic Inquiry and Word Count. In International Conference on Computational Intelligence in Information System, Springer, Cham 97-105.
[6] Soldner F, Kleinberg B, Johnson S (2021) Confounds and Overestimations in Fake Review Detection: Experimentally Controlling for Product-Ownership and Data-Origin. arXiv preprint arXiv:2110.15130.
[7] Deepika DS, Sowmya A, Sravani M, Priyanka C, Ashesh K (2021) Identifying Deceptive Reviews: Using Linguistic and Spammer Behaviour. In International Conference on Image Processing and Capsule Networks, Springer, Cham 581-588.
[8] Vidanagama D, Silva T, Karunananda A (2021) Hybrid Filter-Wrapper Approach for Feature Selection in Deceptive Consumer Review Classification. In 2021 5th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI), IEEE 1-6.
[9] Ceballos Delgado AA, Glisson W, Shashidhar N, Mcdonald J, Grispos G, Benton R (2021) Deception Detection Using Machine Learning.
[10] Zhong M, Qu X, Chen Y, Liao S, Xiao Q (2021) Impact of Factors of Online Deceptive Reviews on Customer Purchase Decision Based on Machine Learning. Journal of Healthcare Engineering 2021.
[11] Du X, Zhao F, Zhu Z, Han P DRDF (2021) A Deceptive Review Detection Framework of Combining Word- Level, Chunk-Level, And Sentence-Level Topic-Sentiment Models. In 2021 International Joint Conference on Neural Networks (IJCNN), IEEE 1-7.
[12] Khan W, Crockett K, O'Shea J, Hussain A, Khan BM (2021) Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection. Expert Systems with Applications 169: 114341.
[13] Cao N, Ji S, Chiu DK, He M, Sun X (2020) A deceptive review detection framework: Combination of coarse and fine-grained features. Expert Systems with Applications 156: 113465.
[14] Catelli R, Fujita H, De Pietro G, Esposito M (2022) Deceptive reviews and sentiment polarity: Effective link by exploiting BERT. Expert Systems with Applications 209: 118290.
[15] Fahfouh A, Riffi J, Mahraz MA, Yahyaouy A, Tairi H (2020) PV-DAE: A hybrid model for deceptive opinion spam based on neural network architectures. Expert Systems with Applications 157: 113517.
[16] Fahfouh A, Riffi J, Mahraz MA, Yahyaouy A, Tairi H (2022) A Contextual Relationship Model for Deceptive Opinion Spam Detection. IEEE Transactions on Neural Networks and Learning Systems.
[17] Ren Y, Yan M, Ji D (2022) A hierarchical neural network model with user and product attention for deceptive reviews detection. Information Sciences 604: 1-10.
[18] Cao N, Ji S, Chiu DK, Gong M (2022) A deceptive reviews detection model: Separated training of multi- feature learning and classification. Expert Systems with Applications 187: 115977.
[19] Sharmila MG, Abinesh S, Dhanesh A, Annamalai SN Fake Review Detection Using Fuzzy Logic and Machine Learning.
[20] Kotriwal S, Raguru JK, Saxena S, Prasad Sharma D (2022) Deceptive Reviews Detection in E-Commerce Websites Using Machine Learning. In Data Engineering for Smart Systems, Springer, Singapore 489-495.
[21] Duma RA, Niu Z, Nyamawe AS, Tchaye-Kondi J, Yusuf AA (2023) A Deep Hybrid Model for fake review detection by jointly leveraging review text, overall ratings, and aspect ratings. Soft Computing 27(10): 6281- 6296.
[22] Jacob MS, Selvi Rajendran P (2022) Deceptive Product Review Identification Framework Using Opinion Mining and Machine Learning. In Mobile Radio Communications and 5G Networks, Springer, Singapore 57- 72.
[23] Liu Y, Wang L, Shi T, Li J (2022) Detection of spam reviews through a hierarchical attention architecture with N-gram CNN and Bi-LSTM. Information Systems 103: 101865.
[24] Rao S, Verma AK, Bhatia T (2023) Hybrid ensemble framework with self-attention mechanism for social spam detection on imbalanced data. Expert Systems with Applications 217: 119594.
[25] Rout JK, Sahoo KS, Dalmia A, Bakshi S, Bilal M, Song H (2023) Understanding Large-Scale Network Effects in Detecting Review Spammers. IEEE Transactions on Computational Social Systems.
[26] Velutharambath A, Klinger R (2023) UNIDECOR: A Unified Deception Corpus for Cross-Corpus Deception Detection. arXiv preprint arXiv:2306.02827.
[27] Tamimi M, Salehi M, Najari S (2023, January) Deceptive review detection using GAN enhanced by GPT structure and score of reviews. In 2023 28th International Computer Conference, Computer Society of Iran (CSICC) (pp. 1-7). IEEE.
[28] Dixit DK, Bhagat A, Dangi D (2023) An accurate fake news detection approach based on a Levy flight honey badger optimized convolutional neural network model. Concurrency and Computation: Practice and Experience 35(1): .e7382.
[29] Parte SA, Ratmele A, Dhanare R (2023) An Efficient and Accurate Detection of Fake News using Capsule Transient Auto Encoder. ACM Transactions on Knowledge Discovery from Data.
[30] Dangi D, Chandel ST, Dixit DK, Sharma S, Bhagat A (2023) An efficient model for sentiment analysis using artificial rabbits optimized vector functional link network. Expert Systems with Applications 225: 119849.
[31] Dangi D, Dixit DK, Bhagat A (2022) Sentiment analysis of COVID-19 social media data through machine learning. Multimedia Tools and Applications 81(29): 42261-42283.
[32] Amin I, Dubey MK (2022) Hybrid ensemble and soft computing approaches for review spam detection on different spam datasets. Materials Today: Proceedings 62; 4779-4787.
[33] Kotriwal S, Raguru JK, Saxena S, Prasad Sharma D (2022) Deceptive Reviews Detection in E-Commerce Websites Using Machine Learning. In Data Engineering for Smart Systems: Proceedings of SSIC 2021 (pp. 489-495). Springer Singapore.
[34] Zaki N, Krishnan A, Turaev S, Rustamov Z, Rustamov J, Almusalami A, Ayyad F, Regasa T, Iriho BB (2023) Node Embedding Approach for Accurate Detection of Fake Reviews: A Graph-Based Machine Learning Approach with Explainable AI.
[35] Ala'M AZ, Mora AM, Faris H (2023) A Multilingual Spam Reviews Detection based on pre-trained Word Embedding and Weighted Swarm Support Vector Machines. IEEE Access.

FACULTY OF INFORMATION TECHNOLOGY