BSIT-S22-003

03-135191-002 Anas Nawaz

03-135191-004 Hamza Iqbal

Classification of Ransomware Attacks Using Machine Learning

In partial fulfilment of the requirements for the degree of
Bachelor of Science in Information Technology

Supervisor: Muhammad Zunnurain Hussain

Department of Computer Sciences
Bahria University, Lahore Campus

January 2023

C e r t i f i c a t e

We accept the work contained in the report titled

Classification of Ransomware Attacks Using Machine Learning

written by

Anas Nawaz

Hamza Iqbal

as a confirmation to the required standard for the partial fulfilment of the degree of

Bachelor of Science in Information Technology.

Approved by:
Supervisor: Muhammad Zunnurain Hussain

_______________________________________ (Signature)

January 10, 2023

DECLARATION

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

Enrolment	Name	Signature
03-135191-002	Anas Nawaz	
03-135191-004	Hamza Iqbal	

Date : January 10, 2023

Specially dedicated to

my beloved grandfather, mother, and father

(Anas Nawaz)

my beloved grandmother, mother, and father

(Hamza Iqbal)

ACKNOWLEDGEMENTS

We would like to thank everyone who had contributed to the successful completion of this project. We would like to express our gratitude to my research supervisor, Mr. Zunnurain Hussain his invaluable advice, guidance and his enormous patience throughout the development of the research.

In addition, we would also like to express my gratitude to our loving parent and friends who had helped and given me encouragement.

Anas Nawaz

Hamza Iqbal

Classification of Ransomware Attacks Using Machine Learning

ABSTRACT

To make money, steal information, and harm computer systems, malware takes on a kind of dangerous presence in the online world. Ransomware is a unique form of virus that poses serious hazards to the entire planet. It has resulted in an immeasurable loss for the businesses, the government, and the people. The previous anti-malware technology employed signatures to detect malware when it came to creating a defense against it. However, once the ransomware has been installed on a victim's machine, further investigation is no longer feasible. The signature-based strategy has already started to lose its impact.

Machine learning research and advancements in ransomware detection and classification have led to effective and precise differentiation. By gathering and studying ransomware characteristics, machine learning algorithms have significantly improved the ransomware defense technology. To discover the dataset with the greatest representation of ransomware behavior, this research will proceed from basic feature collections to feature engineering. Iterative techniques are being used to construct this system.

Table of Contents

List of tables:

List of figures

Chapter 1

Introduction

1.1 Background:

Since the creation of computing and networking technologies, cybersecurity has been one of the main issues. Significant study has been done to build the defenses to protect people and organizations from such sabotage as criminals are becoming more sophisticated and posing new risks. Ransomware is one of the many varieties of malware that has been developing tremendously in recent years and has a particularly negative influence on the world.

1.2 Problem Statement:

The irreversible nature of a ransomware attack sets it apart from earlier computer assaults. After encryption is complete, the decryption key is the sole way to unlock the user files. To decrypt the data, the attackers demand payment in the anonymous currency known as bitcoin. Infections by ransomware damage both individuals and companies looking to boost sales. Data is disclosed in most situations. Approximately 70% of ransomware assaults, according to research, exposed the victim's data. Our goal is to review of ransomware attacks and detect the ransomware attack in dataset than visualize the attack using machine learning.

1.3 Aims and Objectives:

The aims and objectives are given below:

i. Machine learning-based visualization of ransomware attacks.
ii. Detection the Ransomware attacks in dataset.
iii. The Efficiency and Accuracy of the results.

1.4 Scope of Project:

What: Ransomware attacks have increased during the previous few years, many of them in the public eye. According to a report by Compatriotic, in 2020 alone, ransomware attacks will have cost the healthcare sector more than $20 billion in lost revenue, legal fees, and ransom payments. Over 600 hospitals, health centers, and other healthcare facilities were hit by 92 ransomware attacks throughout the year. Implementing the technology to identify the ransomware assault is thus the primary goal.

Why: Many hackers use unrest and chaos in times of crises in search of possible financial benefit. The COVID-19 problem, which started in 2020, brought further attention to cyberattacks in the healthcare industry. Cybersecurity is more crucial than ever before for protecting user or organizational data; thus, we must put this system in place.

How: First we will get the data from dataset of ransomware attack and then process the data in different virtual machine**.** We will visualize the result of different datasets and use different tools like RapidMiner, rattle, and Window OS.

i. Comprehensive review of Ransomware attacks.
ii. Detection the Ransomware attacks in dataset.
iii. Visualization of Ransomware attacks using machine learning.

Chapter 2

LITERATURE REVIEW

2.1 Technology:

Technology is always developing new ways to improve society living. Technology advancements enable industry to carry out current activities more effectively, having a direct impact on other industries and eventually society. In the 1980s, personal computers started to become widely used. These stand-alone, permanent terminal devices made computer ideas like operating systems, programming, early business applications, and games accessible to the public. For military use, ARPANET created protocols for establishing connections between distant computers in the late 1960s. A small number of nodes interconnected via the current communications infrastructure were used in this. Hypertext Markup Language (HTML) helped the World Wide Web (WWW) gain popularity in the late 1980s (HTML). Since that time, more people have access to the internet. Access to and connectivity with the Internet are inextricably related to the development of the WWW. Advancements in data and communications have served as catalysts for the development of everything from mobile and internet services to dial-up modems and the Integrated Services Digital Network (ISDN).

Since the beginning of the 1980s, mobile Internet services have also developed. First generation (1G) gave 2.4 kbps, and second generation offered 64 kbps (2G). 2G development was supervised by Global Systems for Mobile Communications. Global Systems for Mobile marked the start of new initiatives and collaborations, including Third

Generation Partnership Projects and the Universal Mobile Telecommunications Systems (UMTS) (3GPP). From 144kbps in transit to 2Mbps stationary, 3G delivered data services. increased call, video, and communication services with download speeds of up to 100Mbps, 4G is an LTE wireless broadband service created by the 3GPP. The foundation for the development and eventual implementation of 5G services has been set through these initiatives and alliances. Society will become more valued and pertinently connected as the Internet and WWW develop. Nowadays, most communication takes place on gadgets like smartphones, laptops, tablets, etc.

A router that follows this protocol is required a network must be reachable through the internet. Up to this point, a 4th-generation Internet Protocol operated rather well. However, the arrival of more Internet-connected devices increased the pressure placed on IPV4's addressing capabilities. In 1993, the Internet Engineering Task Force (IETF) began developing concepts for IPv4's replacement and making suggestions for its implementation. Smart things will have a variety of communications capabilities, from the most basic to the most complicated. The effectiveness of the security measures will rely on the computing power of the item, making certain things more vulnerable than others. The physical properties of the smart object affect amount of processed or stored data as well. Variety installed advance items is growing along with IoT initiatives. Access control, participation, and data security are essential elements of IoT installations. For smart city residents who respect their privacy, protecting this data is crucial to building confidence. Data about an individual or a whole society may be compromised. Social media and mobile technologies have transformed the conventional idea of a smart city from one centered on industry to one that emphasizes holistic living. The confidentiality and safety of data are crucial regardless of the individual, social, or industrial IoT principles.

2.2 Malware:

Reproducible computer code was first proposed in 1949. According to this notion, programs might self-replicate and transmit their code to new programs. When Fred Cohen created a computer software that could infect a computer, duplicate itself, and spread to other devices in 1987, he coined the word "virus". In the 1980s, the personal computer and ARPANET were establishing a connected world. The possibility of developing dangerous software was also expanding. There was now a way for the self-replicating programs envisaged in 1949 to spread to other locations. Infected ARPANET terminals of the Creeper worm posted a message and established new connections with additional terminals. This annoyance proved that computer code might infect linked devices via automation and repetition even when no harm was done. These techniques led to the emergence of more harmful software. Targets included the Brain virus and MS-DOS operating system, while the Morris worm took use of a link to the ARPANET. In the early 1990s, HTML aided in the construction and growth of the WWW. The proliferation of harmful software was made easier by the increase of networked computers. Evolutionary Web X.0 paradigms gave rise to new malware techniques and varieties. The list of the WWW and the viruses it has produced is shown below:

- 1991: With the introduction of Michelangelo in 1991, the public on the Internet became aware of the danger posed by viruses. This infection increased public awareness of the risks posed by viruses and helped pave the way for the development of anti-virus software.
- 1999: The broad infection of the Internet by Melissa, ILOVEYOU, and Anna Kournikova was made possible by the efficient delivery of harmful code via email.
- 2003: SQL flaws are made public after Slammer and Conficker were exploited to launch widespread DDoS assaults.
- 2005: Koobface is disseminated through social media channels.
- 2007: ZBot, which infects Windows workstations and is engineered to steal financial information, becomes the most successful botnet ever.
- 2010: The well-known Stuxnet malware, which aims to undermine the Iranian nuclear program, targets industrial control systems.
- 2013: Introduction of Crypto locker, which encrypts user data on computers, ransomware debuted as software that generates cash. Before the decryption key is

provided, a ransom, often in Bitcoin, must be paid. In 2017, WannaCry spread internationally, harming individuals, businesses, and governmental organizations in over 150 nations.

- 2016: Targeting insecure IoT devices led to the creation of the Mirai botnet in 2016. High-profile web services were paralyzed because of the DDoS assault.

2.2.1 Malware Method:

The first Creeper worm spread and infected terminals via automated and repeated methods. These traits are present in all the malware categories listed above. The malware is created, programmed, and released by humans. A person botmaster is also in charge of the command and control (C&C), may be contacted by infected computers. End users unknowingly permit material, which promotes to infection due to human naivety. Global infection, however, mostly on automation and repetition techniques. Botnets offer a method for the infection and control of cyberattacks on a worldwide scale. A botnet assault will typically include two sets of IP addresses. The hacked hosts are the first group of IP addresses. These are ordinary infected devices that unintentionally take part in an assault. The C&Cs make up the second group of IPs.

There are three ways for C&C and the compromised host to communicate:

- A concept based on Internet Relay Chat that uses take orders from C&C.
- An internet approach that uses host-based pull instructions.
- Peer-to-peer (P2P) paradigm, where bots communicate with one another.

2.3 Ransomware Attack:

Users are attacked by ransomware, a type of malware that encrypts data without their permission. Limits authorized access to user data. Users are not allowed to utilize their own assets because of this. The irreparable nature of a ransomware attack sets it apart from earlier computer assaults. The user files can only be unlocked using the decryption key after encryption is complete. To decrypt the data, the attackers want money in bitcoin, an undetectable money. Threats by ransomware damage both individuals and corporations by reducing income. Attackers benefit from the undetectable currency and long-lasting harm that ransomware assaults create. Threats are made against the victim, including that his data will be misused, destroyed, or revealed, as well as that private details like search history [1]. The data is disclosed in most situations. According to a survey, data from the victim was exposed in around 70% of ransomware assaults.

2.3.1 Ransomware Attack Variety:

There are a huge variety of ransomware malware strains. [2]

Locky

Locky allows for the encryption of 160 file types, mostly those used by designers, engineers, and testers. It was first launched in 2016. Hackers send e - mails inviting recipients to download malicious ZIP files or Word, Excel, or PowerPoint files from Microsoft Office. It is frequently spread by phishing or exploit kits.

WannaCry

WannaCry is a beginner-level malware that spreads itself among computers by exploiting a hole in the Windows SMB protocol. The WannaCry packager, a self-contained application, extracts the encryption/decryption software, files containing encryption keys, and the Tor communication software. It is not difficult to locate and remove, nor is it disguised. 2017 saw the rapid spread of WannaCry across 150 nations, resulting in $4 billion in damage to 230,000 machines.

Crypto locker

Nearly half a million computers were hacked by the 2017 version of Crypto locker. Typically, e - mail, file-sharing websites, and unprotected downloads are how malware spreads to PCs. In addition to files on the local workstation, it may also encrypt objects it has authority to write to and search mapped network devices. Current iterations of Crypto locker can evade firewalls and anti-virus software from the past [3].

Cerber

Cerber, a ransomware-as-a-service tool, may be used by cybercriminals to initiate attacks and distribute their loot alongside the malware author. Cerber runs covertly while encrypting data and may attempt to disable antivirus and Windows security features in order to prevent users from reinstalling the operating system. The desktop wallpaper changes to a ransom message after the computer's data has been successfully encrypted.

Petya

Using the Master File Table to encrypt the whole hard disc, Petya is a ransomware malware that seizes control of a computer (MFT). The entire disc is inaccessible even though the files are not encrypted. Petya spread mostly through a fake cover letter for a job that contained a link to an infected Dropbox file, which was how it was first identified in 2016. PCs with only Windows were affected.

Grand Crab

2018 saw the release of Grand Crab. The attackers threatened to reveal the victims' propensity for watching porn in ransomware-based extortion activities, and it encrypts data on the victim's computer and demands a payment. There are various variations, and they are all designed for Windows computers. Most Grand Crab versions may currently be decrypted for free.

Ryuk

To access computers, Ryuk utilizes drive-by downloads or email spam. It makes use of a dropper, which downloads a trojan and establishes a persistent network connection on the victim's machine. Attackers can use Ryuk as the foundation for an Advanced Persistent Threat (APT) and then add keylogging software, carry out privilege escalation, and engage in lateral movement. Ryuk is installed on every system to which the attacker gains access after that.

2.3.2 How Ransomware Works:

The ransomware attack continues as following when the infected code is found on a device. Ransomware may stay dormant on a system and wait to launch an assault at a time when it is least secure [4]. Seven-stage of Ransomware:

Execution- In order to carry out its harmful actions, ransomware seeks and register’s locations for specific documents, including locally stored files, mapped and unmapped network-accessible systems. Backup files and folders may potentially be lost or encrypted as a result of some ransomware attacks.

Infection- Ransomware is installed and stealthily downloaded onto the computer.

Encryption- During the encryption stage, ransomware trades keys with the command-and-control system. Then, during the Execution stage, the ransomware utilizes the encryption key to encrypt any files discovered. Similarly, the information's accessibility is restricted.

User Notification- Before displaying a ransom note to the victim, ransomware installs files with instructions explaining the compensation process.

Cleaning up-Ransomware often shuts down and deletes itself, just the files with the financial transactions remain.

Payment- The target hits an URL in the financial transactions, which takes them to a website with more instructions on how to transfer the required ransom. To prevent being detected by network traffic monitoring, these messages are commonly wrapped and disguised utilizing secret TOR facilities.

Decryption- After paying the ransom, the victim may get the decryption key via the attacker's Bitcoin address. However, there is no guarantee that the decryption key will be sent as promised [5].

2.3.3 Ransomware Protection:

The following recommended practices will assist you in preventing and guarding against Ransomware attacks in your business:

Endpoint Protection

The apparent first line of defense against ransomware is Endpoint Protection Antivirus, however outdated antivirus technologies can only offer limited protection. A component of modern endpoint security solutions, next-generation antivirus provides protection from signature - based attacks like WannaCry, zero-day malware, and ransomware whose signature is not yet available in malware databases. They also have device firewalls and endpoint detection and response capabilities, which help security teams recognize and block endpoint attacks fast.

Patch Management

Update the operating system, installed programs, and security patches on the device. Conduct vulnerability scans to find and swiftly fix known issues.

Data Backup

Data should be frequently backed up to an external drive using the 3-2-1 rule and versioning control create three backup copies on two different media with one backup stored in a separate location. If you can, unplug the hard drive from the computer to prevent the backup data from being encrypted [6].

Control and Application Spam filtering

Establish device restrictions to limit installed programs to a centrally controlled checklist. Users should boost their browser security settings, turn off Adobe Flash and other shoddy browser plugins, and use web filtering to prevent them from visiting hazardous websites. Turn off macros in word processors and other exposed programs.

Network Security

Use a firewall or web application firewall, intrusion prevention/intrusion detection systems, and other limitations to prevent ransomware from interacting with command-and-control centers.

Email Security

Tests should be given to evaluate if employees can recognize and avoid phishing emails, and they should be educated to recognize social engineering emails. Use spam prevention and endpoint protection software to automatically filter out dubious emails and to block the hazardous content if a user does happen to click on one of the links.

2.3.4 Ransomware Removal:

Here are the initial actions you should take to reduce the ransomware danger if you discovered an infestation in your network:

Isolate- Isolate the infected computers by locating them, cutting them off networks, then locking share discs to stop encode. Look into the backups that are accessible for encrypted data. Check to see what kind of ransomware you were exposed to and whether any decryptors are available. Determine whether the ransom is a feasible option [7].

Recover- Recover your data from a backup if no decryptor tools are available. In most nations, paying the ransom is not advised, although in some severe circumstances, it could be an alternative. Follow industry standards when erasing and reimaging affected systems to get rid of ransomware.

Reinforce- In order to comprehend how computer systems were compromised and prevent a recurrence, reinforce conduct a lesson learned workshop. Determine the critical flaws or inadequate security procedures that let the attacker's entry and fix them.

Evaluation- It's important to evaluate what happened and the lessons that might be used when the crisis is resolved. How did ransomware operate effectively? Which security flaws allowed for penetration? Why did email filtering and antivirus fail? How much of a spread did the illness have? Were infected computers able to be cleaned up and reinstalled, and was a backup restoration successful? To be better prepared for the next assault, address the areas where your security posture is lacking.

2.4 Machine Learning in Ransomware Attack:

Given the wide range of handwriting styles, it would be challenging to create a program that could recognize handwritten letters. Even if you could take into consideration these variations, creating the software itself could take too much time or be too difficult [8].

Such a challenge is not "game over" with today's machine learning technologies. Instead, by giving it instances to analyze, this technology may be used to address the same problem in novel circumstances. The examples act as a manual for correctly identifying letters. In essence, the goal is to teach the computer to solve issues using examples or recognize patterns, just like you could teach a young child to distinguish between a cat and a dog.

ML is a branch of AI research that creates statistical models using principles from computer science and statistics. These modules are used for two different things:

- Inferring from information requires finding connections.
- Use knowledge of the past to anticipate the future and make (very correct) assumptions about it.

ML focuses more on making predictions about the future than does AI, which frequently focuses on teaching computers to make judgments (predicated on Machine learning and logical sets of guidelines). This is where ML technology differs from AI.

2.4.1 How machine learning technologies can defend against ransomware

Ability to foresee is the hidden weapon that ML possesses. With more accurate information points available for it to learn from, this ability is enhanced. Imagine playing game repeatedly with the one same partner and then with different players. As you gain experience, you become more adept at predicting your adversaries' future moves. You have additional alternatives to think about by incorporating the lessons discovered from prior opponents, allowing you to modify your own approach accordingly. The foundation of the machine learning procedure in the context of data protection is stack trace analysis. Consider the fact that a program leaves a trail of what has occurred at various times in time. Normal activity is made evident by examining what transpires at each stage, and a reference model is established. In the case of a ransomware attack, new code will be inserted into this process, which would be evident.

The most effective software uses ML that ignores aberrations and just considers the most often used reference points. This method progressively develops the computer's understanding of legitimate vs harmful code, improving accuracy while also enhancing software speed due to the machine learning model's reduced data usage.

2.5 Visualization:

Graphs, charts, and plots can be used to display analysis of a dataset that has been captured. These do a better job of communicating the information. Additionally, it effectively conveys massive statistics to a target audience. To help people, comprehend and make sense of massive volumes of data, data visualization is a technique that makes use of a variety -of static and dynamic visualizations within a given context. To visualize patterns, trends, and connections that could otherwise go missing, the data is sometimes presented in a narrative style.

2.6 Systematic Literature Review (SLR):

TABLE 2.1:SLR

Year	Paper Title	Author	Objectives	Methodology	Contribution	Future Gap
2017	**STUDY ON RANSOMWARE ATTACK AND ITS PREVENTION**	**Ganesh Gupta**	**The objectives of this paper:** • Cybersecurity education can raise awareness among less experienced computer users. • Regular practice of preventative strategies can also be provided.	**The methodology used in this paper:** • Using various intrusion prevention system (IPS) technology, malicious traffic from exploit kit activity can be detected	**The contribution of the paper:** • The significance of having a traffic-filtering system that can offer proactive anti-ransomware defense.	**Future Work or Future Gap:** • While some commercial antivirus products come with an automatic update module and a real-time scanner, anti-ransomware security

				and blocked, preventing the ransomware installation process.		technologies can be a dependable alternative. • To provide more reliable antivirus product for future.
2019	**Prevention of Crypto-Ransomware Using a Pre-Encryption Detection Algorithm**	**S. H. Kok and Mahadevan Supramaniam**	**The objectives of this paper:** • Only focuses on crypto ransomware because it makes data unrecoverable and once suspect's documents were encrypted.	**The methodology used in this paper:** • The two-phased pre-encryption detection algorithm (PEDA). • A Windows application programming interface in PEDA-	**The contribution of the paper:** • In terms of test error, FPR, AUC, and detection rate, LA surpasses other learning algorithms like RF and NB. • LA or PEDA-Phase-l has	**Future Work or Future Gap:** • The PEDA concept attempts to save users from having to pay ransom by spotting ransomware before it encrypts data. It is possible to think of this restriction

			• Developed a pre-encryption detection method (PEDA) for ransomware cleanup with the least amount of danger.	Phase-I. (API). the PEDA-Phase-II signature repository.	successfully identified crypto-ransomware using only API data, demonstrating its effectiveness as a prediction model.	as a research goal for the future. • PEDA-Phase-II intends to create a Signature Repository by storing the signature of all discovered ransomware.
2016	**Automated Dynamic Analysis of Ransomware: Benefits, Limitations and use for Detection**	**Daniele Sgandurra**	**The objectives of this paper:** • Developed EldeRan, a framework to recognize the most important dynamic ransomware	**The methodology used in this paper:** • Elderan Sandboxed Training • Analysis Elderan Live Detection.	**The contribution of the paper:** • Demonstrated that ML is a workable and efficient method for identifying new	**Future Work or Future Gap:** • The authors predict how ransomware will develop in the future, including by focusing on the wearables

			traits and utilize them to identify ransomware. • EldeRan's capacity to identify new ransomware families, with an average detection rate of 93.3%.		ransomware families and variations for analysis and signature extraction in addition to AV. • It achieves substantially better outcomes than more simplistic methods and compares favorably with more complex algorithms in	sector (commonly known as "ransom wear"). By demonstrating that (prudent) statistics for Crypto Locker account for $3 million in revenue in 2013–2014, among other things, characterize the underground market for scareware and ransomware. In, the authors

					terms of output.	suggest a brand-new technique for identifying and learning about malware activity.
2020	**A Study of Ransomware Detection and Prevention at Organizations**	**Saurabh Kumar Sen**	**The objectives of this paper:** • This study lays the groundwork for future studies to address the issue of ransomware attacks in businesses. Chart representations	**The methodology used in this paper:** • Limit users' ability to install and utilize unsuitable software programs. You should prevent the attachment of file types of exe/url/tmp/p	**The contribution of the paper:** • The major objective of this research is to implement risk using machine learning and the Python programming language in accordance	**Future Work or Future Gap:** • With the advent of new technologies, it will become easier to detect malware in the future and to reduce business losses brought on by ransomware. Additionally, this paper inspires

			for cycles are known to IT security specialists and researchers.	if/vb/vbe/scr/reg/cer/pst/cmd/bat/dll/hlp/wsf/hta/js. • Set up the installation of host-level antiexploitation tools like the Enhanced Mitigation Experience Toolkit.	with organizational research. • Techniques to limit ransomware attack gaps in the network and methods to lessen harm from ransomware assaults.	new analysts and researchers for the decryption of contaminated files.
2019	**Situational Awareness of Ransomware Attacks—Detection and Prevention Parameters**	**Juan A. Herrera Silva and Lorena Isabel Barona López**	**The objectives of this paper:** • Offers a ransomware article classification based on	**The methodology used in this paper:** • Support vector machines, decision trees, and	**The contribution of the paper:** • The main goal of this page is to advance research in this field by	**Future Work or Future Gap:** • Ransomware attacks can seriously impact businesses of all sizes. They

			methods for detection and avoidance. • The ransomware life cycle and the threat detection model.	Bayesian networks (BN). • Prevention • Detection • • • • • • • • • Prediction.	the publication of updated papers that compile the most recent findings and offer a comprehensive analysis of ransomware.	safeguard systems from ransomware variations by employing customary procedures like antimalware. However, due to ransomware's intelligence and ongoing evolution, these techniques are insufficient to detect and prevent fresh attacks.

2022	**Ransomware Malware and Ransomware Detection Techniques**	**Sonal Yadav and Neha Soni**	**The objectives of this paper:** • The analysis of delivery product assault discovery strategies and ransomware network attacks are the main objectives of this research. There are numerous recognition techniques or methodologies that can be used to identify	**The methodology used in this paper:** • Detection By Signature • Detection By Behavior • Detection By Abnormal Traffic	**The contribution of the paper:** • Attacks should be stopped by individuals and organizations, and finding such attacks is a crucial step in developing a ransomware attack defense strategy.	**Future Work or Future Gap:** • In the future, individuals and organizations should stop attacking, and the detection of such an attack is a crucial step in the ransomware attack countermeasure to secure the systems.

			payment product assault. • Network security, malware, ransomware, and ransomware detection methods are some related terms.			
2020	**Android Ransomware and Its Detection Methods**	**Manish Kaushik and Leena Bhatia**	**The objectives of this paper:** • This paper's primary goal is the detection of: • Crypto Ransomware	**The methodology used in this paper:** • Static, dynamic, and hybrid approaches are used to	**The contribution of the paper:** • This study illustrates the detection of ransomware using static, dynamic, and	**Future Work or Future Gap:** • High accuracy detection techniques for mobile ransomware must be

			(File Encryptor Ransomware) • Locker Ransomware (Lock Screen Ransomware).	identify ransomware.	hybrid approaches. Compared to dynamic methods, static methods are more accurate. in order to find ransomware. However, they are useless. with infections during runtime. The shortcomings of static and dynamic approaches	developed using both static and dynamic techniques.

					are overcome through hybrid approaches.	
2019	**Systematic literature review and metadata analysis of ransomware attacks and detection mechanisms**	**Alhassan, Haruna Chiroma and Emmanuel Gbenga Dada**	**The objectives of this paper:** • We categorize ransomware assault strategies. • We looked at the criteria used to rate ransomware attack, protection, and detection systems. • For a future investigation of	**The methodology used in this paper:** • Search/data sources. • Search keywords. • Explicit inclusion and exclusion criteria. • Data collection and synthesis of results Study selections.	**The contribution of the paper:** • In this paper, we give a thorough analysis of ransomware attacks and countermeasures. The publications under consideration clarified several fundamental	**Future Work or Future Gap:** • For future research, we also provide proactive computational intelligent prediction models. Intelligent techniques published by Abdullahi and Ngadi, Abdulhamid et al., and others

			the anatomy of ransomware, we collated and summarized all research datasets that were available.		characteristics and signs of ransomware. • The evaluated articles focused a lot on the environment, particularly the Windows and Android platforms, which serve as a haven for ransomware activities due to their pervasive vulnerabilities.	can be used to predict a ransomware attack before it even happens.

2017	**Ransomware-Prevention Technique Using Key Backup**	**Kyungroul Lee, Insu Oh, and Kangbin Yim**	**The objectives of this paper:** • This work offers a preventative technique for user PCs in addition to a range of systems, such as massive platforms, to provide security from cybercrime based on ransomware.	**The methodology used in this paper:** • Ransomware infiltrates the target system and activates the encryption feature to encrypt the target system's files; in this case, the prevention application passes the encryption feature rather	**The contribution of the paper:** • The suggested ransomware-prevention technique in this study uses a main-backup procedure to recover the encrypted files from a system that has been infected by ransomware. Because ransomware uses aberrant	**Future Work or Future Gap:** • In the future, having an extremely robust endpoint security solution will be crucial to preventing ransomware. Your endpoint devices are equipped with these solutions, which prevent malware from infecting your systems.

				than blocking it. • When the ransomware calls the key-generation and key-import methods of the CNG library, the hooking code gives the key to the prevention software and then transfers execution control to the ransomware.	behaviour, such as locking the victim's system or encrypting system or files, to interfere with a victim's system, this paper suggests a key-backup technique for which the encryption key is maintained in a safe repository.	

2021	**Android Ransomware Detection Based on a Hybrid Evolutionary Approach in the Context of Highly Imbalanced Data**	**Iman Almomani Raneem Qaddoura and Maria Habib**	**The objectives of this paper:** • The goal of this study is to detect malware with good performance using an extremely unbalanced dataset. • Similar to the Android Market, there aren't many ransomware programmers compared to other kinds of software.	**The methodology used in this paper:** • The suggested method is based on a combination of an oversampling methodology, a classification strategy, and an evolutionary process for optimizing unbalanced data.	**The contribution of the paper:** • Outline in detail the most recent advancements in ransomware detection technologies. • By taking into account the most recent Android release, provide a resent dataset of the Android OS (version 11, API level	**Future Work or Future Gap:** • Although more data and more advanced models to handle the acquired big data, such as deep learning algorithms that are more capable of inferring accurate patterns of relationships, could be used in this research study to expand it into further research, the results have

				A. Structure of the particle B. Fitness function (internal evaluation) C. How the algorithm is implemented	30). An unbalanced dataset of safe and malicious applications will be created in order to simulate the real-market situation.	shown the merits of the proposed approach's capacity to detect Ransomware efficiently (97.5% of g-mean).
2019	**Situational Awareness of Ransomware Attacks—Detection and Prevention Parameters**	**Juan A. Herrera Silva and Lorena Isabel Barona López**	**The objectives of this paper:** • Modernized methods and strategies for analyzing, preventing, and defending against ransomware	**The methodology used in this paper:** • Intelligent techniques, including Bayesian Networks (BN), decision trees, and	**The contribution of the paper:** • Cloud-based recovery solutions and distributed computing backup and recovery systems can	**Future Work or Future Gap:** • Predicting ransomware is one of the upcoming trends in order to spot the danger before encryption and

			attacks on Windows devices. It will act as a beginning point for further research. • The first proposal to recapitulate the criteria used in current investigations is this paper.	support vector machines (SVM), have been presented in recent study. The focus of each of these segments is the following action: • Detection Prediction • Prevention	both contribute to reducing result of ransomware assaults. Ransomware attacks are rendered useless, there will be less incentive in creating new dangers.	stop the attack in time. • Future research will also focus on establishing a database of information on the financial features of ransom payment systems.
2017	**CRYPTO RANSOMWARE ANALYSIS**	**ASHWINI BALKRUSHNA KARDILE**	**The objectives of this paper:** • This system's goal is to introduce new	**The methodology used in this paper:** • The file system access and	**The contribution of the paper:** • A traditional method for the analysis and	**Future Work or Future Gap:** • This paper offers a wealth of insightful

	AND DETECTION USING PROCESS MONITOR		methods for automated ransomware detection employing dynamic methodology, not merely dynamic analysis of malware. • This system tracks file access and I/O traces to find user-level malware.	I/O traces implemented using Process Monitor and how to setup Cuckoo Sandbox along with Virtual machine configuration. • Generating realistic user environment • File paths. • Valid Contents. • Monitoring file system	detection of the most recent ransomware was described in this paper. This technology can identify the typical actions of ransomware, such as the harmful encryption of user's data. • This study also demonstrates the	information and can act as a cornerstone for numerous upcoming efforts. The authors, like anybody else who creates malware-fighting methods, express concern about the possibility that malware writers will modify their programs once more to counter even hardware-based strategies like the one

				activities using Process Monito.	interactions a ransomware sample has when it attacks a machine, namely with the file system.	covered in their study.
2019	**Machine Learning Based File Entropy Analysis for Ransomware Detection in Backup Systems**	**KYUNGROUL LEE, SUN-YOUNG LEE, AND KANGBIN YIM2**	**The objectives of this paper:** • The key advantage of employing a backup solution is the user's ability to back up their files. If a user's files have been	**The methodology used in this paper:** • Entropy measurement methods • Machine learning MODELS 1. LINEAR MODEL 2. KNN	**The contribution of the paper:** • The ransomware detection methods now in use do not detect malware files within backup.	**Future Work or Future Gap:** • In the future, we'll get outcomes for a range of file types and investigate a method for artificially identifying

			encrypted by ransomware, they can restore their original contents by synchronizing or transferring data from backup systems, including cloud services like Dropbox and Google One Drive, USB storage, and external devices. However, if the ransomware-	3. DECISION TREE ENSEMBLE 4. DECISION TREE ENSEMBLE 5. KERNEL TRICK 6. NEURAL NETWORK (DEEP LEARNING) • Model validation	However, this paper effectively detects ransomware-infected files delivered to the backup system in real time using the reference value derived through Machine Learning Based File Entropy Analysis for Ransomware Detection in	ransomware by figuring out the ideal settings and parameters for every individual user on every user's backup files.

			infected files are synced to the backup system, the files cannot be restored using the backed-up files.		Backup Systems machine learning based on entropy according to different file formats.	
-2021	**A framework for supporting ransomware detection and prevention based on hybrid analysis**	**Francesco Mercaldo**	**The objectives of this paper:** • The system created in this research aims to reduce and prevent ransomware threats. It includes a top-level design or an evaluation	**The methodology used in this paper:** • Static analysis enables us to extract from the executable beneath scrutiny a list of such APIs and libraries	**The contribution of the paper:** • In this study, we proposed a hybrid solution to counter the ransomware threat that uses both API calls and commands	**Future Work or Future Gap:** • As part of our ongoing research, we intend to assess the suggested framework's performance on a wider range of applications,

			of the suggested framework.	being used using a reverse engineering technique.	(via static analysis) (by dynamic analysis). We tested the ability of using API calls and commands, separating malware from genuine programs using the Cuckoo framework, and we found positive findings.	both good and bad. • To improve accuracy for the tasks of ransomware detection and mitigation, we also intend to consider the adoption of formal approaches.

2019	**A Study of Ransomware Attacks: Evolution and Prevention**	**Aini Khalida Muslim1, Dzunnur Zaily Mohd Dzulkifli**	**The objectives of this paper:** • The evolution of ransomware assaults and methods for diagnosing ransomware were examined in this study. This study poses two questions, including "How have ransomware assaults evolved over time?" likewise, "How	**The methodology used in this paper:** • Ransomware attacks have been analyzed using qualitative research as a tool. The information gathered for this study's research is secondary information. Field research can be used to gather	**The contribution of the paper:** • This study will help new researchers find research gaps by providing summaries of previously published research publications.	**Future Work or Future Gap:** • The widespread usage of industrial robots in industry and the infrastructural sectors that link smart cities are examples of larger targets that attackers may choose to attack in the future. • Cybercriminals can invent, launch, and profit greatly from this threat of

			to handle the escalating ransomware attacks?"	secondary data. • Examples of secondary data for social science include information from organizations and scrutinized government agencies, in addition to data that were first obtained for various		cybercrime in the future.

				research purposes.		
-2021	**SDN-Based Detection of Self-Propagating Ransomware: The Case of BadRabbit**	**FAHAD M. ALOTAIBI AND VASSILIOS G. VASSILAKIS**	**The objectives of this paper:** • One of the main goals of our work is to understand the process through which this kind of targeted ransomware operates.BadRabbit Analysis	**The methodology used in this paper:** • THE WORM COMPONENT • THE ENCRYPTION COMPONENT • ENCRYPTION PROCESS • ENCRYPTION PROCESS PROPAGAT	**The contribution of the paper:** • BadRabbit underwent a thorough investigation, and it was discovered that this family of ransomware does not interact with other entities in order to exchange an encryption key. Instead, it	**Future Work or Future Gap:** • In upcoming work, a strategy to evaluate the IDPS's effectiveness and performance in a real network. The existence of various programmers, realistic background traffic, and the operation of additional security

				ION METHODS	makes use of a public key that is built into its data.	appliances and features will all be taken into account for validation reasons.
2014	**DNA-Droid: A Real-time Android Ransomware Detection Framework**	**Amirhossein Gharib**	**The objectives of this paper:** • Discovered novel traits with strong discriminative strength that enable the DNA-Droid to identify unidentified ransomware samples.	**The methodology used in this paper:** • Static Analysis • Text Classification Module (TCM) • Image Classification Module (ICM) • API calls and permissions	**The contribution of the paper:** • A freely accessible, fully automated Android sandbox that can report the order of API calls as a web service was made available.	**Future Work or Future Gap:** • Explore new sources of information. • Visualization of the DNAs. • Experiment on a larger dataset. • Experiment with real malware.

			• Examined how well Deep Auto Encoder reduced and picked up new features.	Module (APM) • Feature Learning and Reduction • Dynamic Analysis • Sandbox • API Calls Refining • Multiple Sequence Alignment (MSA) • Detection Module • Static Classification	• Used an extensive collection of various ransomware samples for experimental examination.	

				• DNA Matching		
2018	**The Ransomware Detection and Prevention tool design by using signature and anomaly-based detection methods**	**Baris CELIKTAS**	**The objectives of this paper:** • Demonstrate the hybrid process ransomware prevention and detection solution, which seeks to operate well on Windows OSs with a minimal amount of false positive alerts. • Consider the fact that this concept will act	**The methodology used in this paper:** • Static detection technique • Dynamic detection technique • Hybrid detection technique • Method of detection based on signatures Method for anomaly-	**The contribution of the paper:** • A thorough analysis of pertinent literature and expert reports indicates that relying solely on the signature-detection process to identify and stop malware is ineffective.	**Future Work or Future Gap:** • Users have a better understanding of the key traits of the Ransomware Prevention and Identification Tool that may be used as a remedy, software developers, and security managers as a result of this study.

			as a guide for scholarly investigation of ransomware.	based detection • Ransomware Detection Methods		• This work will serve as a guide for upcoming scholarly investigations of malware, including ransomware.
2022	**Ransomware Detection, Avoidance, and Mitigation Scheme:** **A Review and Future Directions**	**Adhirath Kapoor, Ankur Gupta, and Innocent E. Davidson**	**The objectives of this paper:** • Extremely risky Ransomware assaults have suddenly increased, crippling both individuals and most enterprises.	**The methodology used in this paper:** • Ransomware recognition, static, dynamic, hybrid, string extraction, PE file segments, static linking, stub analysis,	**The contribution of the paper:** • We present DAM, a conceptual framework for evaluating and classifying the tools, approaches, and mitigation	**Future Work or Future Gap:** • Future work will concentrate on developing a browser extension powered by artificial intelligence that will be used to monitor both

			Ransomware is a serious menace that requires an international response. • The best ransomware prevention methods require specialized mitigation and recovery efforts.	automated sandboxing, manual code reversing, manual debugging, malware reconstruction, machine learning classifiers, and memory dump evaluation.	methods for ransomware. • We proposed a continuum for preventing ransomware. Different enterprises, from small businesses to critical deployments, can use this continuum.	personal and corporate online safety.
2020	**Analysis, Detection, and Prevention of Cryptographic Ransomware**	**Ziya Alper Genç**	**The objectives of this paper:** • To develop a protection system that	**The methodology used in this paper:** • The steps that make up the detection	**The contribution of the paper:** • In this research, we investigated	**Future Work or Future Gap:** • This section serves to alert the scientific

			pushes the boundaries of technology by researching ransomware behavior, flaws, and cryptographic origins.	process we use in this chapter are as follows: First, we collect the traces by repeatedly running a malware sample in a sandbox. • Subsequently, we look to see if the sample engaged in any suspicious behavior	potential restrictions decoy tactics may run against when used to combat ransomware. We start by addressing the problem theoretically, and we then explain a real-world proof-of-concept that demonstrates how certain existing	community to potential ransomware threats. • Keeping anti-ransomware ideas in mind in advance could be a game-changing element because it is predicted that the ransomware threat will grow in sophistication rather than in quantity of attacks.

				during the initial run but then behaved maliciously during subsequent runs. If this is the case, it indicates that the malware has some evasive capabilities.	decoy-based solutions can be easily thwarted.	
2018	**Ransomware Activity Detection**	**Marvic Grima**	**The objectives of this paper:** • The main objective of this research is to determine whether	**The methodology used in this paper:** • Sandbox Environment • Prototype Ransomware	**The contribution of the paper:** • The results of this study suggest that monitoring file access by	**Future Work or Future Gap:** • Additional investigation using a larger sample size can reveal additional

			behavior of ransomware detection can enhance security prior to the delivery of new anti-malware signatures by the anti-malware solution provider.	Detection Application • Hardware • Software • Features • Activity Monitoring Mechanism • Process Information Collection • Watchdog • Detector • Protection Mechanism • Configuratio n • Controlled Ransomware Execution	active processes on a Windows computer is a useful method for spotting dangerous ransomware activities.	techniques to enhance and maximize the effectiveness of the detection systems as well as uncover fresh defenses against the execution of the destructive encryption process itself.

2019	**Analysis of Infection, Persistence, Recovery Prevention and Propagation Mechanisms**	**Maxat Akbanov and Vassilios G**	**The objectives of this paper:** • In addition to conventional security measures, new countermeasures are seen as a crucial and fashionable responsibility in this industry. • However, to create such a solution, a thorough examination of ransomware behavior and	**The methodology used in this paper:** • Static and dynamic are two major categories that apply to techniques. While dynamic analysis includes running the malicious binary in a controlled environment, static analysis is	**The contribution of the paper:** • This study concentrated on the first interactions and infection process of WannaCry, as well as its persistence mechanism, encryption process, recovery prevention, and communicatio	**Future Work or Future Gap:** • The results of this study could be applied to the development of efficient WannaCry and other ransomware families that display similar behavior mitigating measures. The work on this is postponed.

			functionality is necessary.	carried out without doing so.	n with C&C servers.	
2020	**RAPPER: Ransomware Prevention via Performance Counters**	**Manaar Alam1, Sayan Sinha**	**The objectives of this paper:** • Provide a two-step unsupervised detection tool that finds malicious process activity with the least number of traces possible when it thinks a process activity to be malicious.	**The methodology used in this paper:** • The RAPPER two-step detection framework employs Fast Fourier Transformation and Artificial Neural Network to create a highly accurate,	**The contribution of the paper:** • In this paper, we give a thorough explanation of how ransomware affects typical system operations. Using a two-step detection methodology, we enlist the help of an artificial	**Future Work or Future Gap:** • Recovering the AES key by focusing on the AES CBC process would be a difficult task. We will save that for a later scope of work, though.

				quick, and reliable ransomware detection method with a minimum number of trace points. • The two phases of the detection process are called the Offline Phase and the Online Phase.	neural network to find the presence of ransomware.	
2020	**A Three-Level Ransomware**	**Amos Loh Yee Ren**	**The objectives of this paper:**	**The methodology used in this paper:**	**The contribution of the paper:**	**Future Work or Future Gap:**

	Detection and Prevention Mechanism	**and Chong Tze Liang**	• Instead of enabling malware to infect the host system, the objective is to separate potentially hazardous items into a virtual environment and place it in quarantine.	• Vaccination from Not Petya • Updates and Patches • Anti-Malware & Anti-Ransomware software • Least Privilege Principle • Prudence, Self-awareness, and Logic	• The notion of Petya and Not Petya is covered in this study; whereas Petya encrypts only the MBR, Not Petya encrypts both files and the MBR. • With the ability to act like a worm and take advantage of open vulnerabilities, Not Petya arose.	• As technology advances, we anticipate being able to run more virtual computers on a single computer. Virtual machines may prove to be an efficient deterrent to malware, according to our research, and this is a positive development in the fight against malware.

2020	**Analysis of Ransomware and its prevention**	**A. D. C Navin Dhinnesh**	**The objectives of this paper:** • Users should be warned not to click on any dubious links they receive through email. For these kinds of attacks, companies must employ a few security measures. Their software needs to be updated properly.	**The methodology used in this paper:** • Keep your System Isolated • Avoid Paying Ransom • Do not click the unknown links • Never open unknown email attachments • Use proper filtering • Update software periodically	**The contribution of the paper:** • The history and development of ransomware are explained in this study. It also examines the decision to use encryption for ransomware attacks. • In this paper, the author discusses how to avoid ransomware	**Future Work or Future Gap:** • Ever since ransomware was first identified in 2000, it has caused extensive damage. Up until now, ransomware prevention has been described. Let's examine how to react to a ransomware attack now.

				• Periodic Data Back up	and how to deal with an attack.	

Chapter 3

DESIGN AND METHODOLOGY

3.1 Methodology:

Utilize exploratory data analysis to examine dataset columns and rows. We will then have a dataset with pre-selected columns and traits for further research. We will now produce visualizations or graphs of the data we collected after running the program.

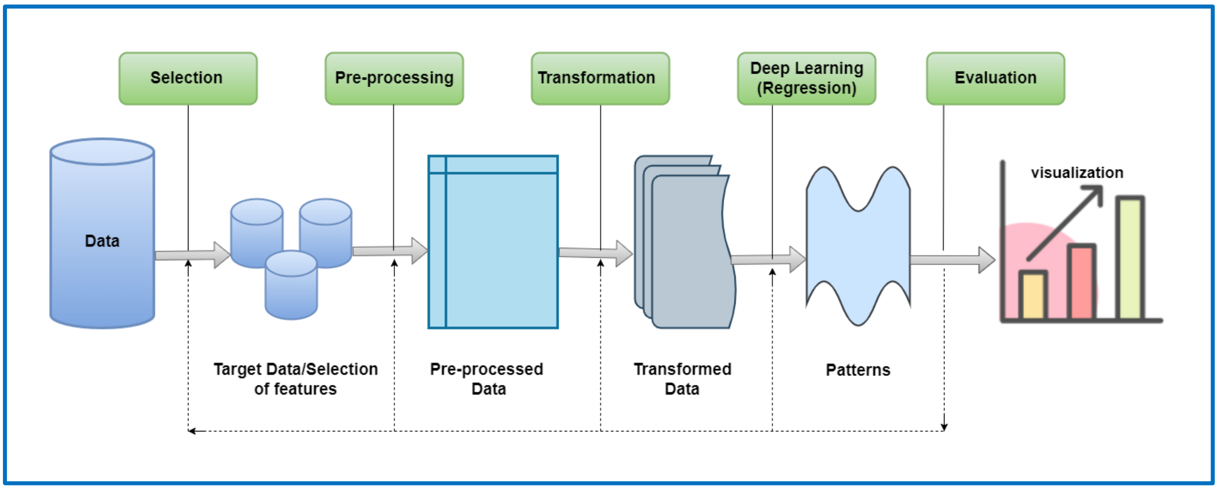

Figure 3.1: Methodology Diagram

Stages of progress throughout the project

i. Data collection

ii. Target Data

iii. Data preprocessing

iv. Transformed Data

v. Patterns

vi. Knowledge or Result

3.1.1 Data:

The first step is to gather data. Data can take the shape of text, comments, graphics, photographs, statistics, charts, and signs. The data might, for example, comprise specific dates, costs, sizes, addresses, ages, names, temperatures, or lengths. Data lacks significance and use on its own since it is an incomplete sort of knowledge. A data set is a collection of data or dataset. Within scenario of tabular data, a data set is associated with one or more database tables, where each row alludes to a particular record in the associated data set and each column to a particular parameter. Our dataset consists of 100,000 samples of ransomware attack.

3.1.2 Target Data (Dataset):

Our target data is Ransomware. While gathering the ransomware set of data was an essential component of our study, we go into considerable depth in this section regarding how we chose the ransomware samples. We gathered malware samples from many sources to create a complete ransomware data collection. The collected ransomware data set consists of 100,000 samples. Dataset is **Bitcoin Heist** that contain different amount of data in the form of table which has numbers of entities like row number, address, year, date, day, length, weight and count etc. Every row and column contain different of different person and of different year. The dataset needs some data mining like data cleaning or data pre-processing.

3.1.3 Data preprocessing

Data preprocessing, which is an important step in the data mining procedure, may be described as the modifying or deleting of data before to use in order to ensure or enhance performance. Rubbish in, garbage out is especially true for projects requiring data mining and machine learning. For the preprocessing of dataset number of attributes are label, year, count, date, length, day, and address etc. Then heat map is generated of the processed data [9].

3.1.4 Attributes and Type

Attributes	Type
Address	String
Income	Decimal
Looped	Integer
Length	Integer
Weight	Float
Count	Integer
Day	Integer
Neighbours	Integer
Year	Integer
Label	String

TABLE 3.1:ATTRIBUTES AND TYPES

3.1.5 Transformed Data:

Data transformation is the process of converting data from one format, such as an Excel spreadsheet database file, or XML document, into another. Transformations often entail cleaning, validating, and making useable a raw data source. Then heat map is generated of the processed data.

3.1.6 Transformed Data Attributes and type

Attributes	Type
Length	Integer
Weight	Float
Count	Integer
Neighbor	Integer
Income	Decimal
Label	String

TABLE 3.2: TRANSFORMED DATA ATTRIBUTES AND TYPE

3.1.7 Patterns:

Data analysts search for patterns in the present data by looking for sets of data that have a recognized pattern. Because each dataset is unique, it's important to recognize patterns and trends in the underlying data. If a company wants to provide accurate, trustworthy results, it must choose the algorithm and strategy that are most suited for the data and analysis.

3.1.8 Knowledge or Result:

See chapter five.

3.2 Deep Learning:

The core of H2O's deep learning system is a multi-layer feedforward artificial neural network that was trained using stochastic gradient descent via back-propagation. The network may have several hidden layers made up of neurons with the tanh, rectifier, and max out activation functions [10]. Let's think about how a neural network calculates a single unit. Y is the output, z is the weighted input, and (z) is the activation function that simulates the sigmoid function.

Nowadays, ReLU function is advised as an activation function instead of sigmoid function since it solves "The vanishing gradient problem." The function that mimics the neurons in a human brain, step function, was replaced by the sigmoid function, a basic differentiable activation function. In the tutorial that follows, we'll use the sigmoid function to help us better comprehend backpropagation. In this tutorial, the sigmoid function will be used.

$$z = x1w1 + x2w2 + b$$

$$y = \sigma(z) = 11 + exp(-z)$$

The inputs are x1 and x2. The coefficient weights for each input are w1, w2.

In essence, x1 and x2 represent data that have undergone normalization or standardization. Better performance is made possible by techniques like applying normalization or standardization to input data. For instance, when normalizing a picture with a 0–255 color range, we divide the picture by 255 to get a 0–1 color range. Gradients exist in the early learning state because the weights initialize in a small range. There are techniques for initializing weights, including using a Gaussian distribution. We will set the values in this tutorial from 0 to 1. Due to initial 0 bias producing improved learning accuracy, bias is set to 0. The bias will be updated as more is learned.

3.3 Gradient-boosted trees (GBM):

This algorithm's main idea is to create models sequentially while aiming to reduce the shortcomings of the previous model. However, how should we approach that? What can be done to reduce the error? This is achieved by building a new model on the residuals or mistakes of the previous one [11].

Progressive Boosting When the target column is continuous, a regressor is utilized; if classification is the issue, a gradient boosting classifier is. The "Loss function" is the sole difference between the two. Gradient descent will be used to increase weak learners and decrease this loss function. We will have a variety of loss functions for regression issues, such as for classification problems and Mean Squared Error, such as log-likelihood, since it is based on a loss function.

Formula:

$$Fm(x) = Fm\_1(x) + vmhm(x)$$

The number of decision trees created is m. Here, nu is the learning rate, which is typically chosen between 0-1, and Fm-1(x) is the prediction of the base model (prior prediction). Long-term accuracy is increased since it lessens the impact that each tree has on the outcome of the prediction. The most recent DT performed on the residuals is Hm(x).

3.4 Random Forest:

To address classification and regression problems, the Random Forest Algorithm, a very well-liked supervised machine learning method, is used. A forest is made up of several different species of trees, and the forest will be more vigorous the more trees there are. In this way, as the number of trees in a Random Forest Algorithm increase, so do its accuracy and ability to solve problems [12].

The steps listed below are how the Random Forest Algorithm functions:

- Step 1: Select randomly selected samples from a specified data collection or training set.
- In step 2, this algorithm will create a decision tree for each training batch of data.
- The third decision tree's average will be used to perform the voting.
- As the last prediction result in step 4, select the outcome that garnered the greatest support.

This combination of several models is referred to as an ensemble. Ensemble uses two methods:

Boosting: Is the process of transforming weak learners into strong ones through the development of subsequent models with the aim of reaching the highest level of accuracy. XG BOOST and ADA BOOST are two examples.

Bagging: Bagging is the method of replacing a sample training dataset with a different training subset. A majority vote is required to determine the result.

Chapter 4

DATA AND EXPERIMENTS

4.1 Download Dataset from Kaggle:

Data scientists and machine learning experts may connect online at Kaggle. Users of Kaggle may work together, access and share datasets, use notebooks with GPU integration, and compete with other data scientists to solve data science problems [13].

The dataset is Bitcoin Heist Ransomware Address.

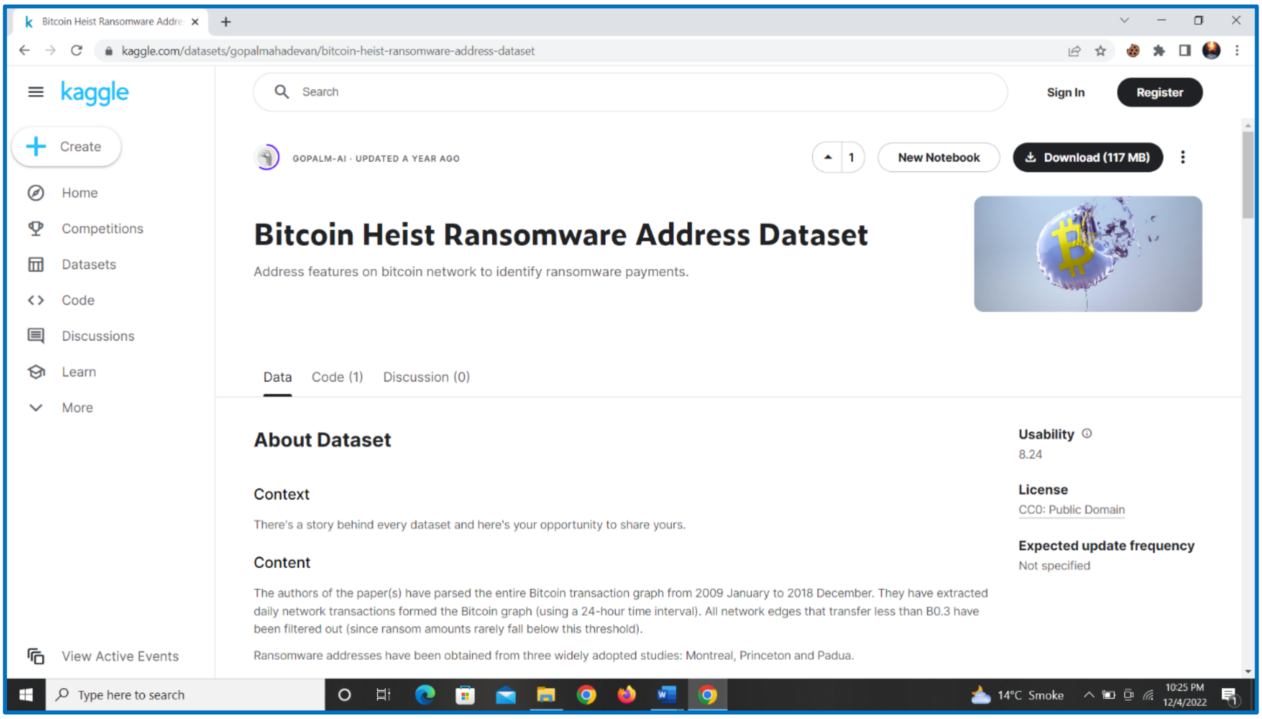

FIGURE 4.1: KAGGLE DATASET RANSOMWARE

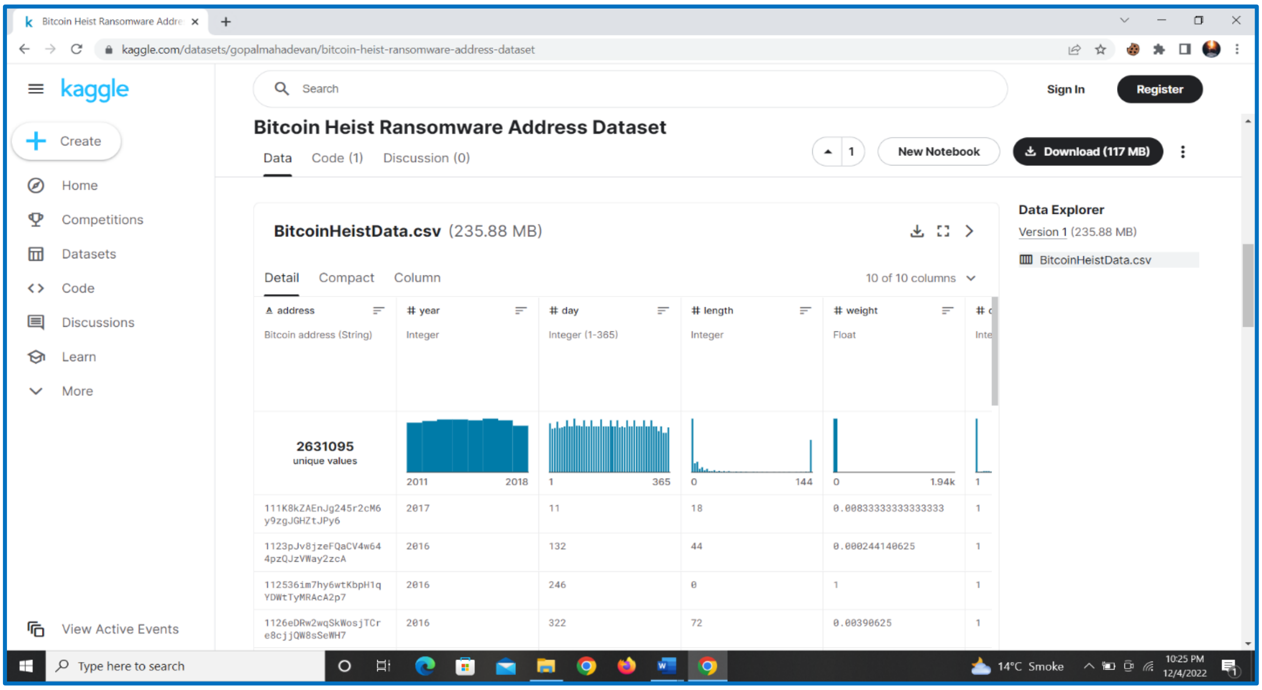

FIGURE 4.2: BITCOIN HEIST RANSOMWARE ADDRESS DATASET

4.2 Download and Install RapidMiner:

RapidMiner is a potent data mining program that supports model deployment, model operations, and data mining. All the data preparation and machine learning skills required to make a significant effect throughout your business are provided by our end-to-end data science platform [14].

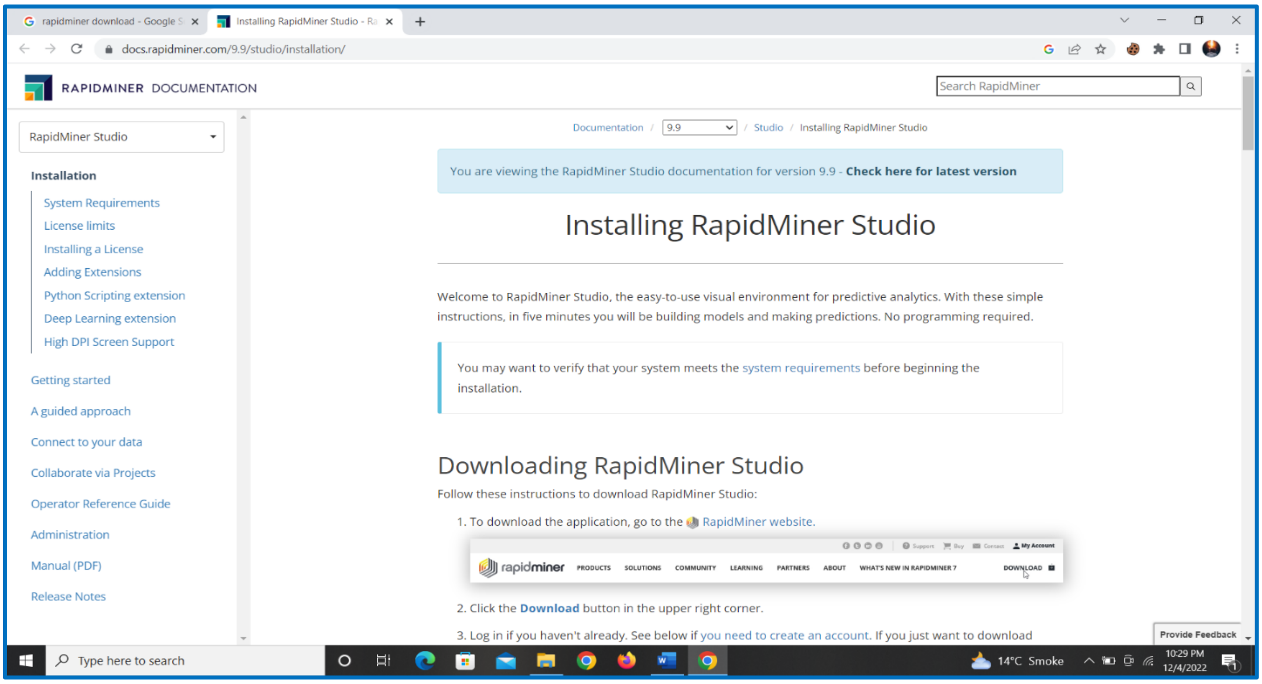

FIGURE 4.3: RAPIDMINER INSTALLATION

4.3 Heat Map:

In a two-dimensional heatmap, a graphical representation of data, the individual values included in a matrix are shown as colors. A matrix of the variables that is colored according to the intensity of the value is called a heatmap. As a result, it provides an excellent visual tool for contrasting numerous objects. This heat map show there is no null values in dataset [15].

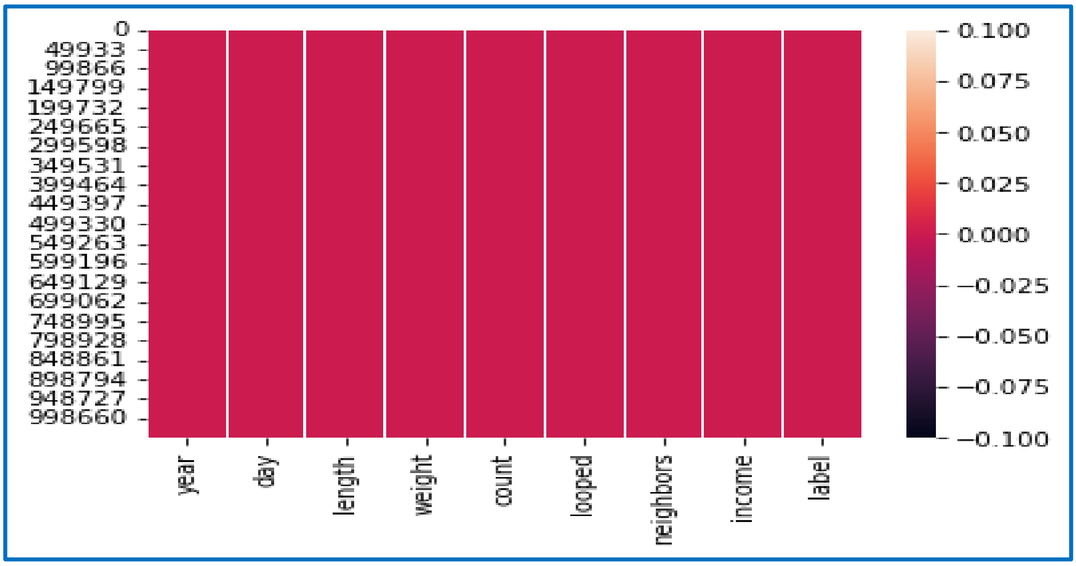

FIGURE 4.4: HEAT MAP

4.4 Import the Dataset:

Utilizing the drag and drop functionality is all that is required to import data into your repository.

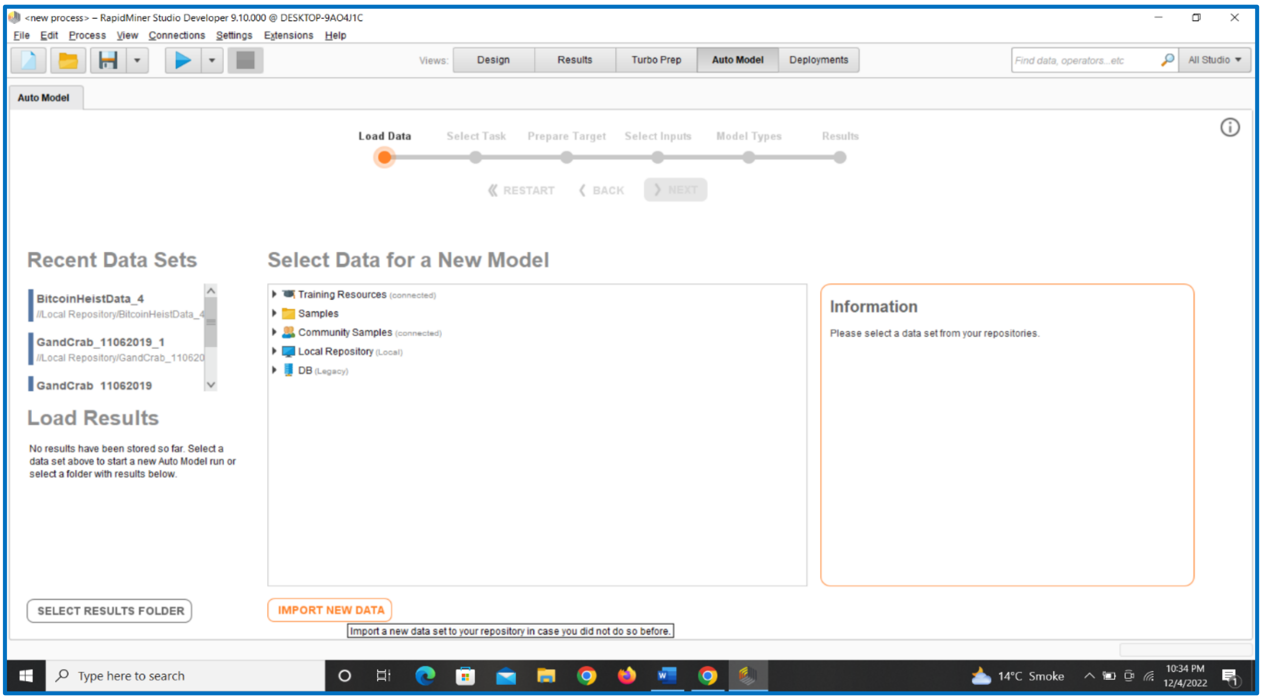

FIGURE 4.5: IMPORT DATA

4.4.1 Select the location:

Simply drag the file into the canvas from your file browser and continue.

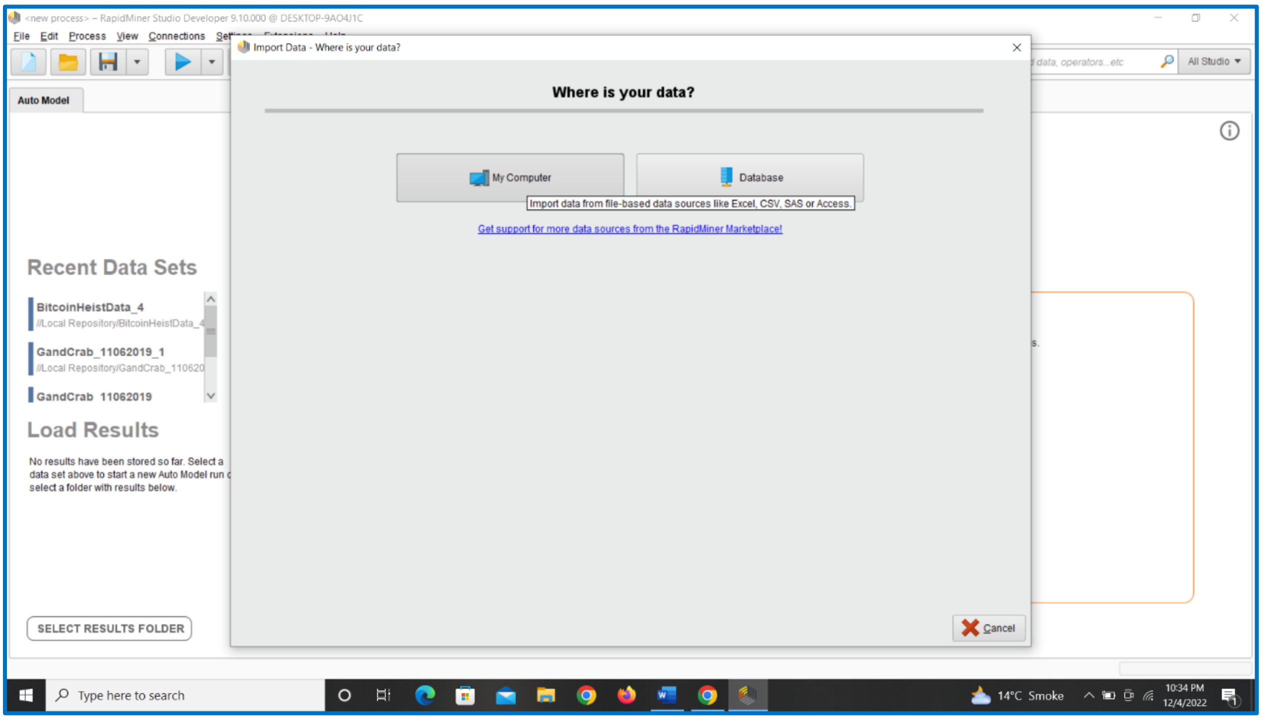

FIGURE 4.6: SELECT LOCATION

4.4.2 Select the Data:

Verify that the target or label is correctly tagged and that the data types are accurate. This method of opening data differs significantly from other methods in that it does not constantly read the original source file from scratch. Therefore, you must overwrite the stored data if you want to update.

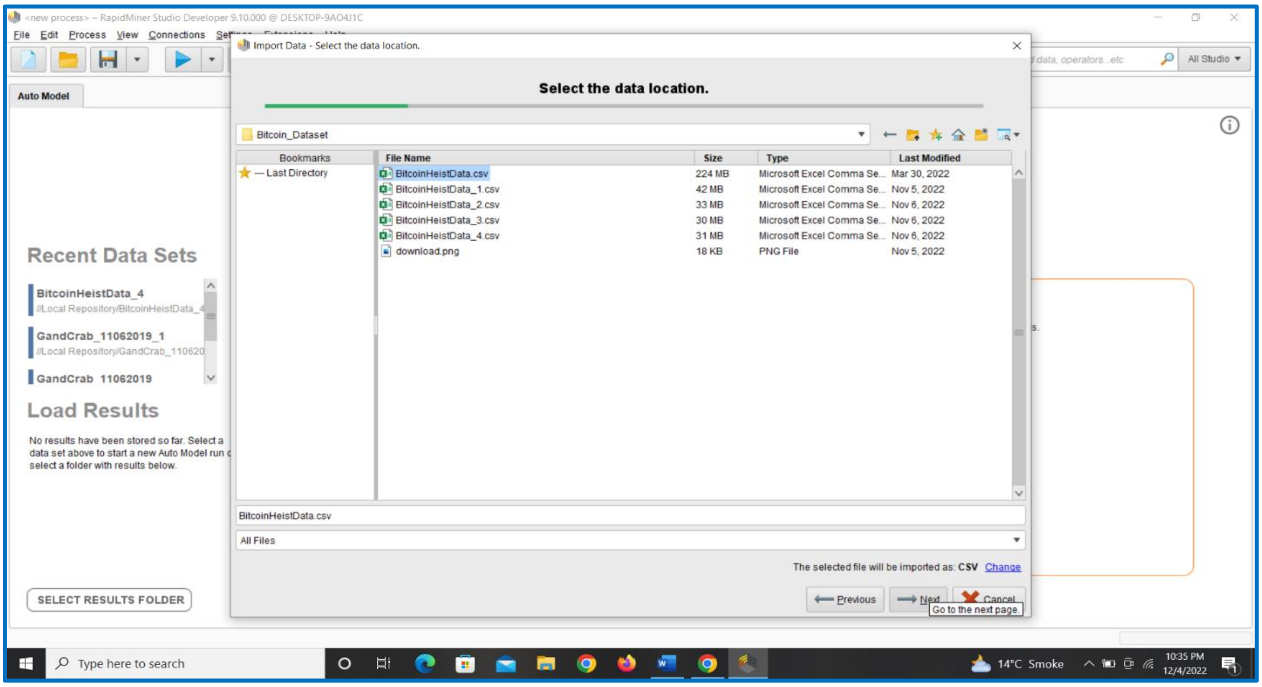

FIGURE 4.7: SELECT DATA

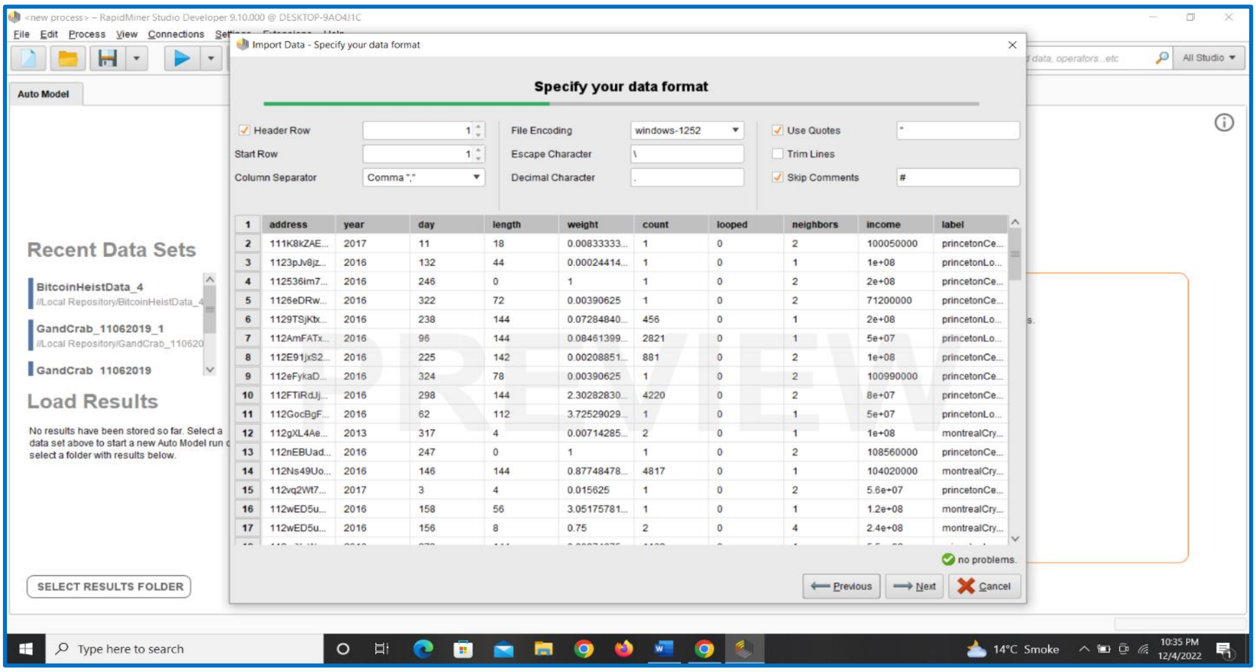

FIGURE 4.8: SELECT DATA FORMAT

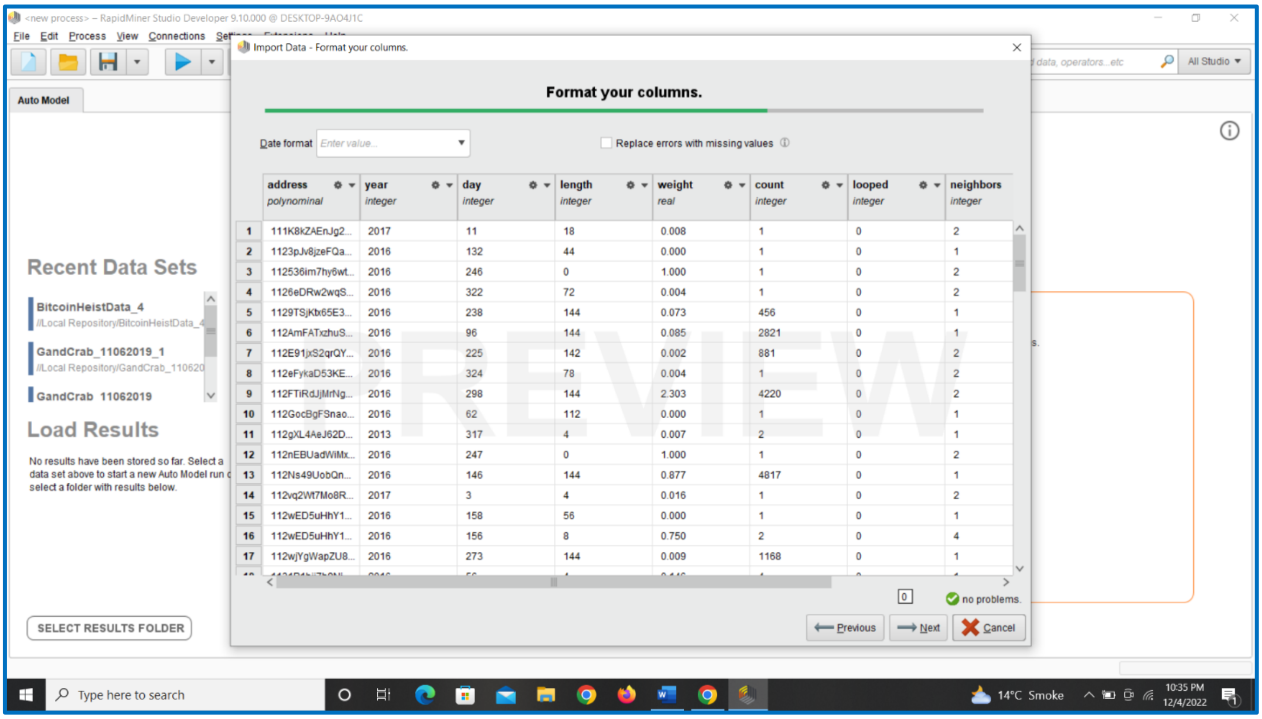

FIGURE 4.9: FORMAT COLUMNS

4.4.3 Save the Data:

However, once the import is complete, a local copy is maintained in RapidMiner's repository, allowing you to choose to delete the original source file if you so desire.

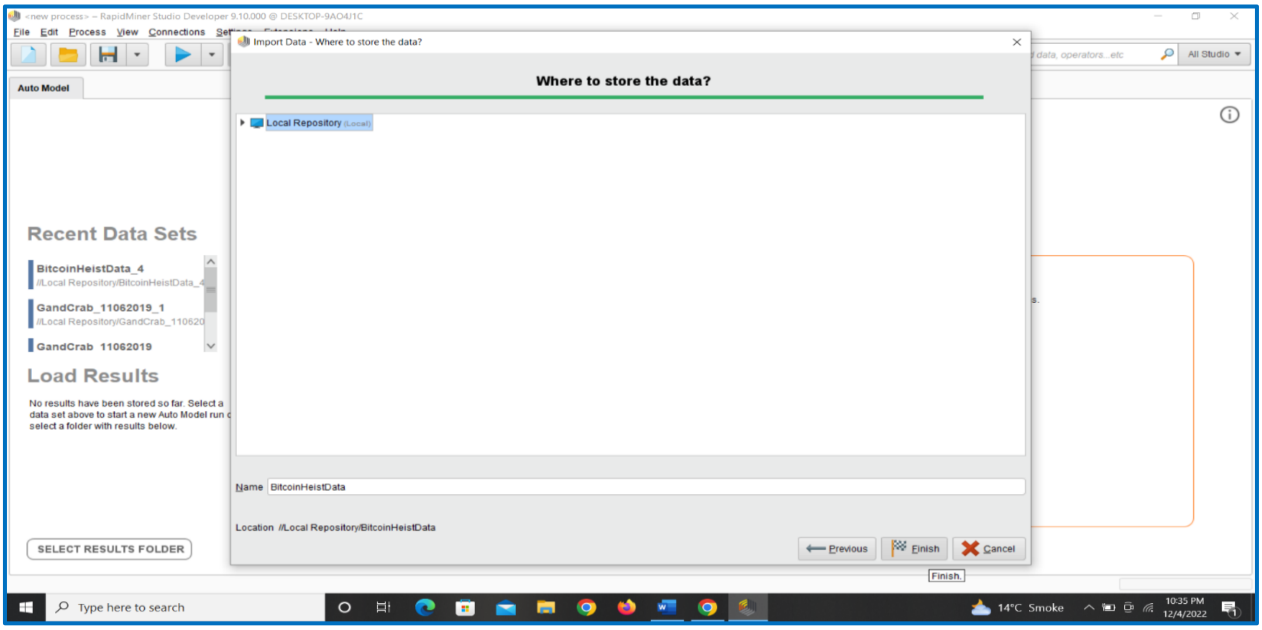

FIGURE 4.10: SAVE THE DATA

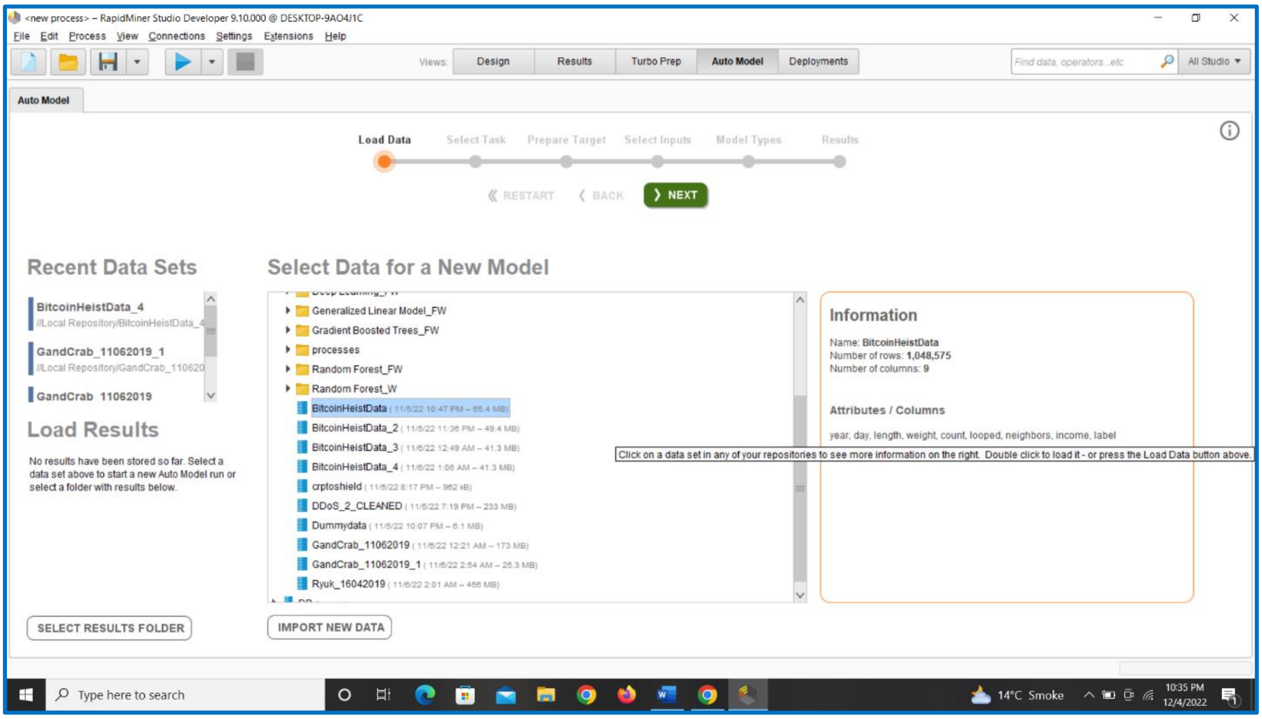

FIGURE 4.11: SELECT DATA OF NEW MODEL

4.4.4 Set the target Class:

After choosing a data set, you must determine the kind of issue you wish to address. Three different tasks are identified by Auto Model:

- Predict
- Clusters
- Outliers

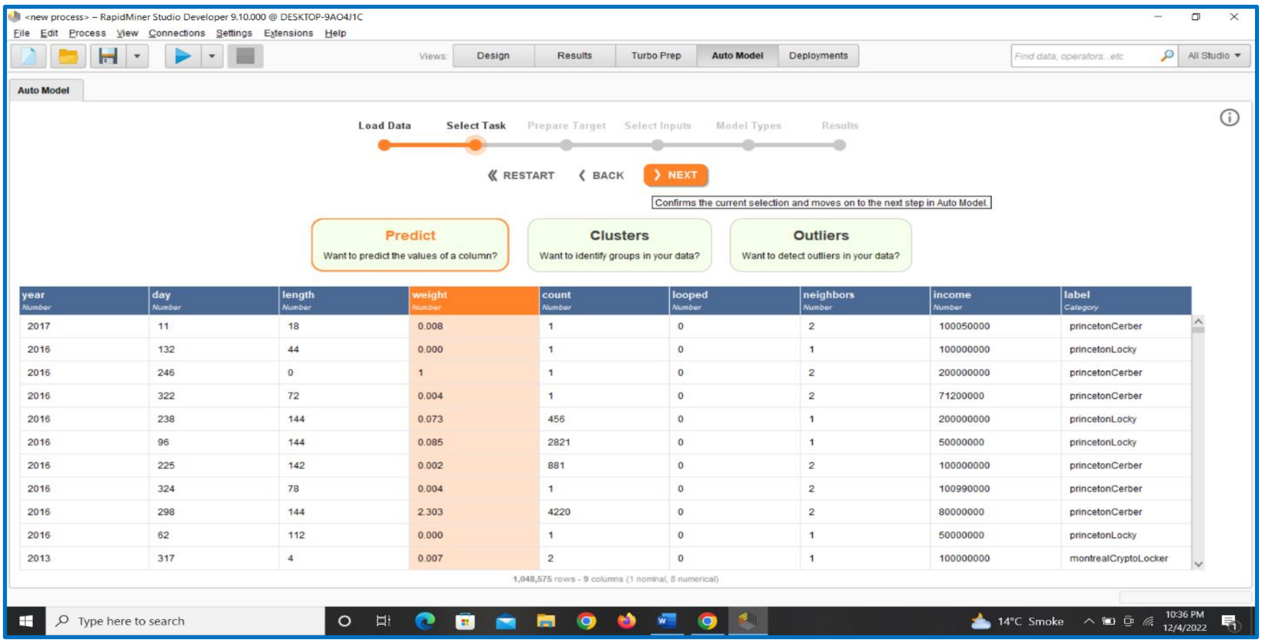

FIGURE 4.12: SELECT TARGET CLASS

4.4.5 Prepare Target:

The issue is a classification issue because there are only two possible answers for "Survived," "Yes" or "No." Auto Model will typically show a bar chart with the data points in each class categorization issues. Only ten classes with the greatest number of data points are shown when there are more than ten classes.

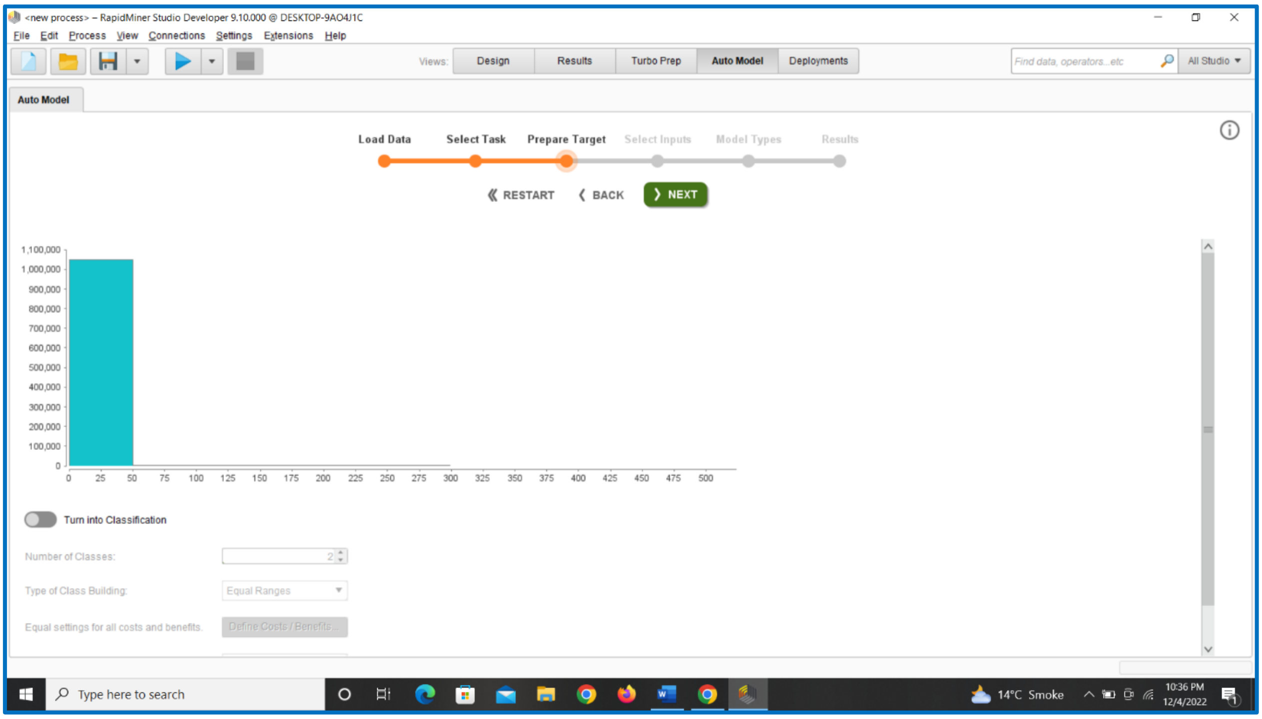

FIGURE 4.13: PREPARE TARGET

4.4.6 Select Inputs Fields:

Not all the data columns in your table will be useful for prediction. You could speed up and/or enhance the performance of your model by removing some of the data columns. However, how do you decide that? The fact that you're searching for patterns is important. The data is unlikely to be meaningful without some variance and some clearly visible patterns.

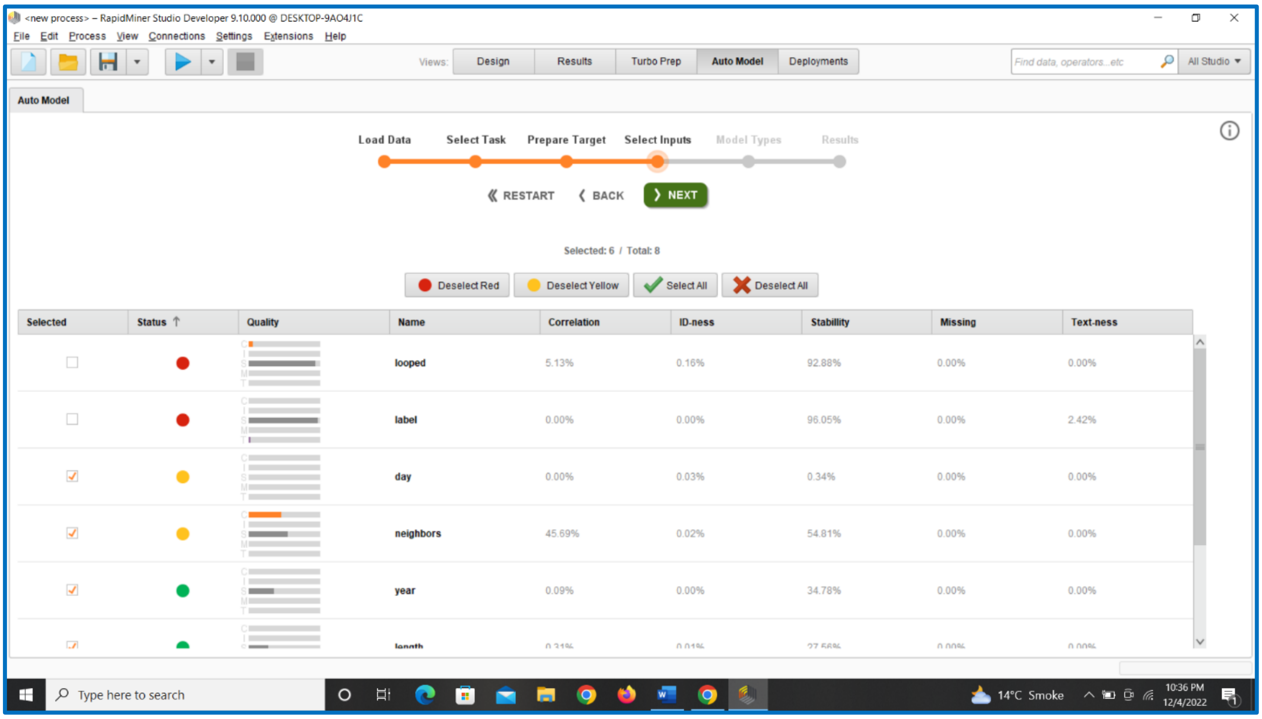

FIGURE 4.14: SELECT INPUT FIELDS

4.4.7 Select the Model:

You can choose from several models that Auto Model suggests are pertinent to your issue. The ideal choice, if there is no time limit, is probably to build every one of them, then evaluate how they operate once they are all complete.

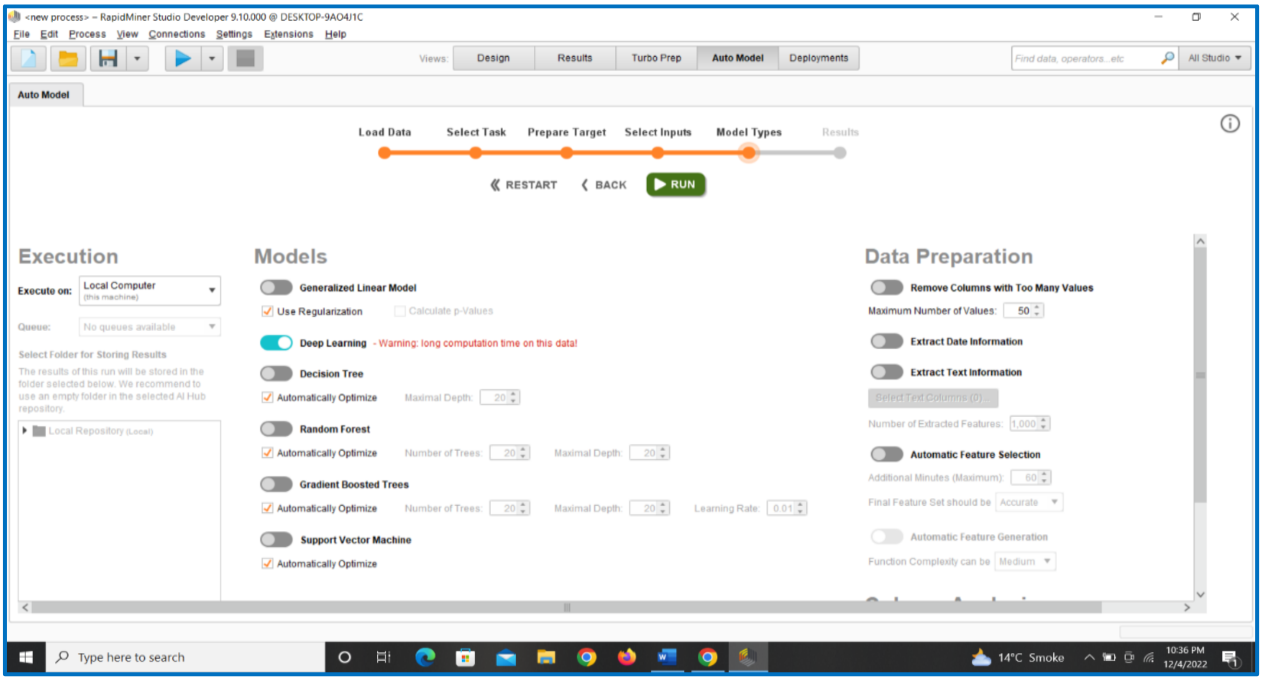

FIGURE 4.15: SELECT THE MODEL

4.5 Intrusion Detection System (IDS)

An intrusion detection system is a hardware or software program that keeps an eye out for malicious activities or policy breaches on a network or in a system. Any intrusion activity or violation is often recorded centrally using a security information and event management system, alerted to an administrator, or both.

4.5.1 SolarWinds

One pane of glass IT administration for on-premises, hybrid, and software as a service (SaaS) environment is made easier with the SolarWinds Orion Platform, a robust, scalable infrastructure monitoring and management platform.

The Orion Platform consolidates the entire set of monitoring capabilities into one platform with cross-stack integrated functionality, eliminating the need to deal with numerous incompatible point monitoring products.

NPM

With the help of the robust and reasonably priced SolarWinds Network Performance Monitor (NPM), you can easily identify, analyze, and fix network performance issues and outages.

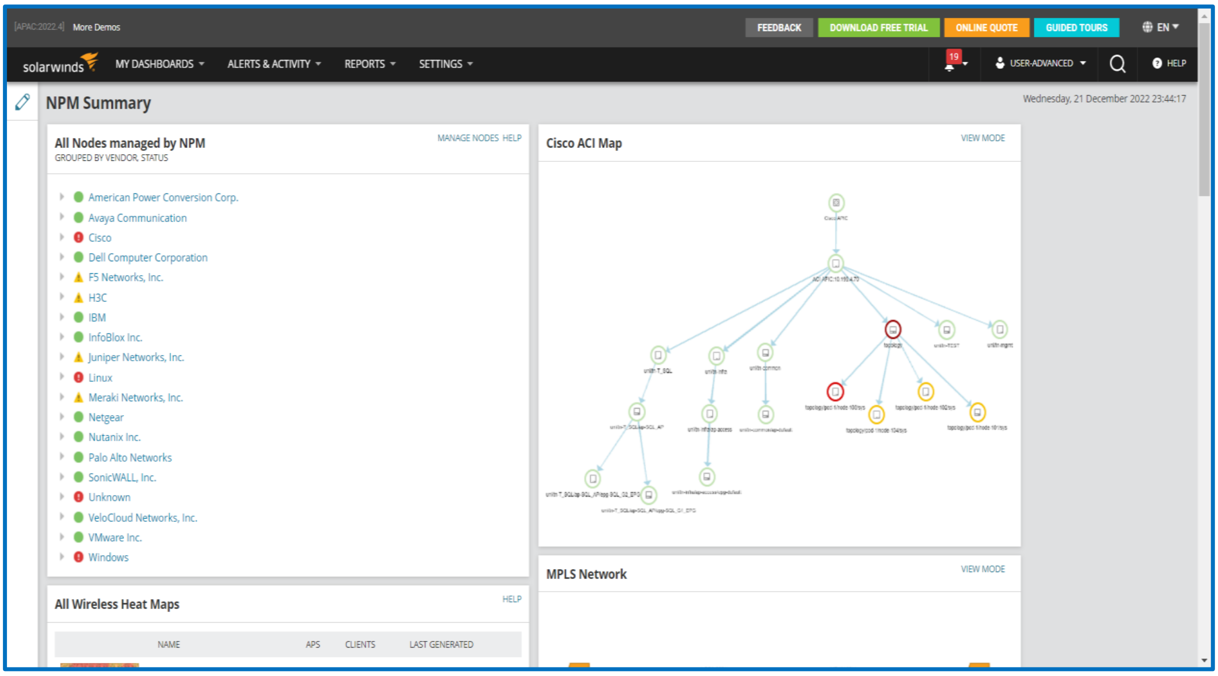

FIGURE 4.16: NPM SUMMARY

Wireless Network

Finding the devices on your wireless network is the first step in effective wireless network monitoring. As soon as the network discovery process is finished, wireless access points (APs) and controllers can be identified as wireless devices using SolarWinds Network Performance Monitor (NPM).

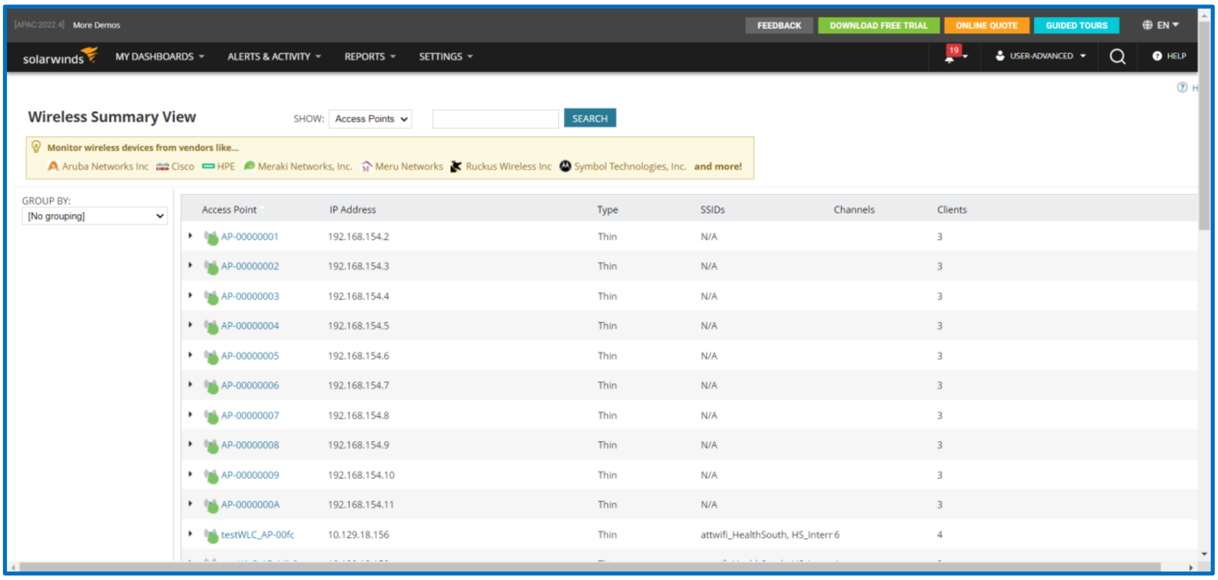

FIGURE 4.17: WIRELESS NETWORK SUMMARY

Load Balancer

A load-balanced service is made up of numerous cooperating parts. You can browse each of these elements, their connections, and their current states in the Balancing Environment widget.

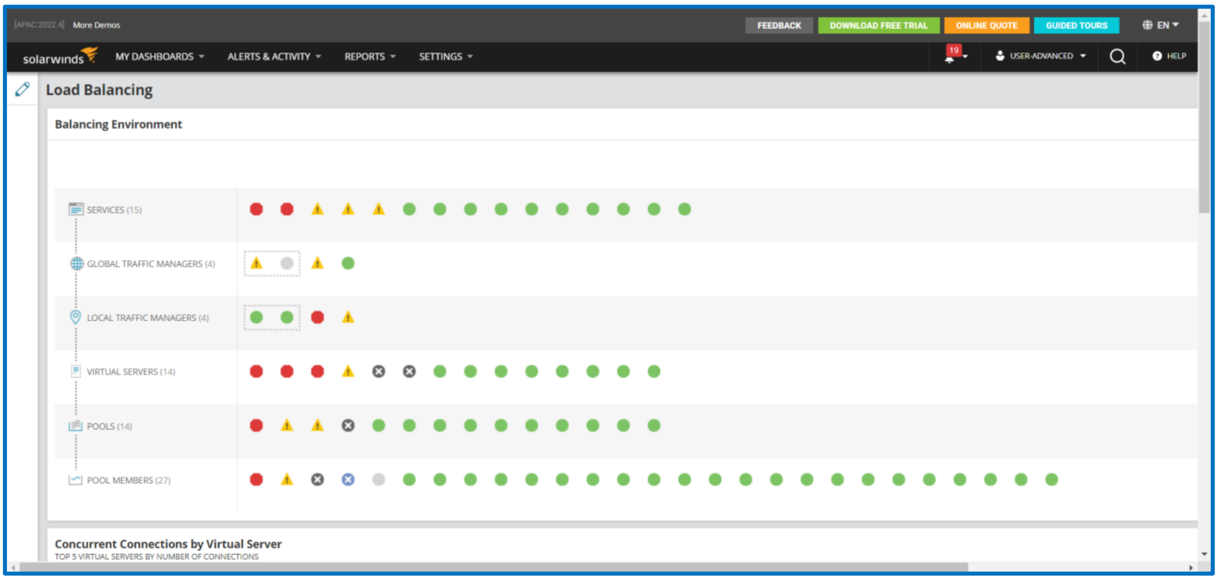

FIGURE 4.18: BALANCING ENVIRONMENT

NOC View

The Network Operations Center (NOC) view offers vital statistics for each device in your network that is being monitored. This view can be used to fill a mobile device or a wall-mounted monitor in a technical support facility used by network administrators and IT specialists who administer a network around-the-clock.

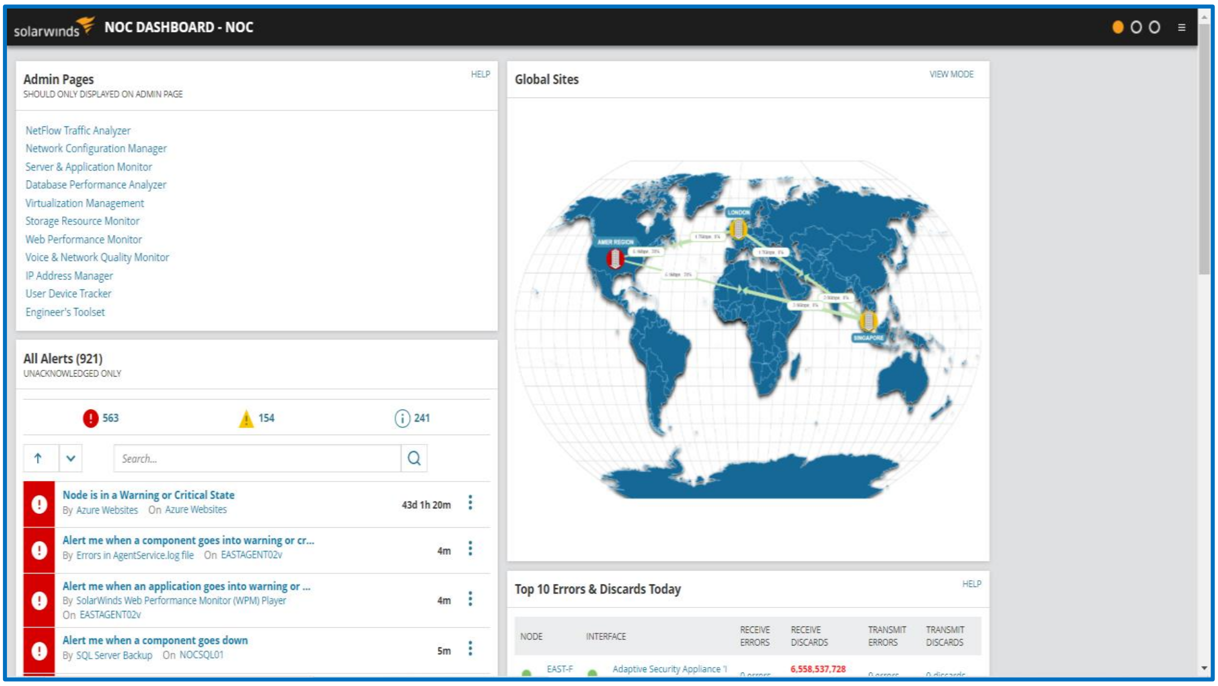

FIGURE 4.19: NOC DASHBOARD

Energy Wise

You can manage your energy costs with the help of Energy Wise. You can remotely set recurring policies and plan power usage with NCM, which can help you spend less energy. Additionally, SolarWinds NPM enables you to keep an eye on your power and energy usage.

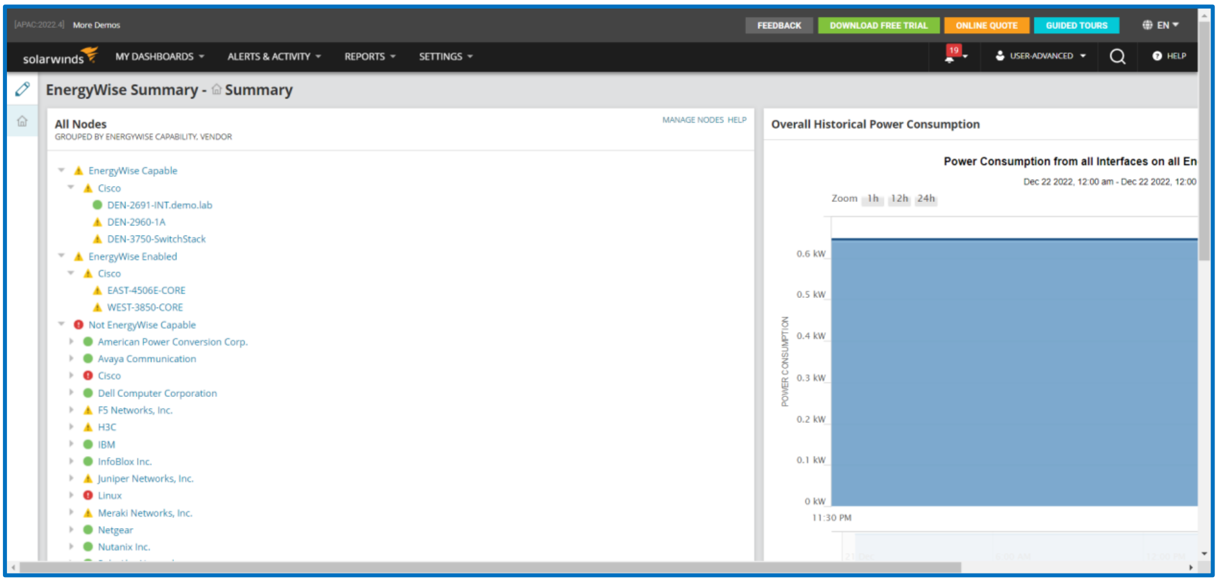

FIGURE 4.20: ENERGY WISE SUMMARY

Capacity Summary

The following metrics of monitored nodes, interfaces, and volumes are available for capacity forecasting:

- CPU uses across nodes
- Node memory usage
- Volumes' use of space
- Interface receive (in) utilization
- Interface transmit (out) utilization

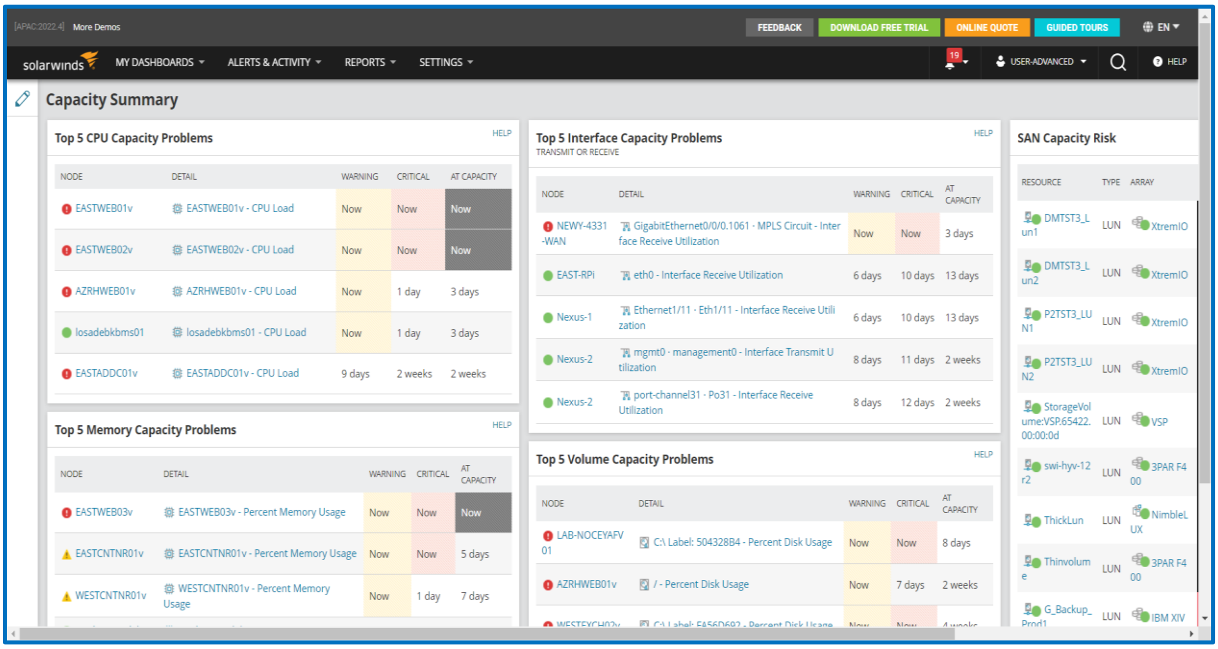

FIGURE 4.21: CAPACITY SUMMARY

VSAN

A piece of software called Star Wind Virtual SAN (VSAN) merely "mirrors" internal hard disks and flash between hypervisor servers, negating the requirement for real shared storage.

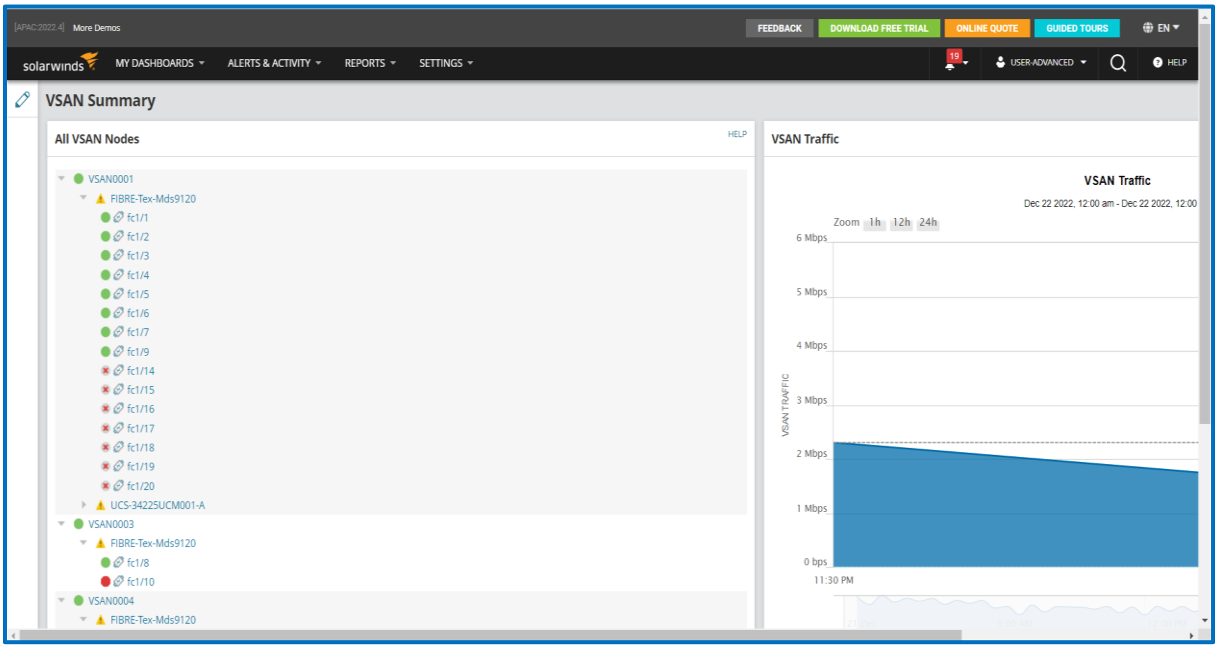

FIGURE 4.22: VSAN SUMMARY

NetPath

Slowdowns are simple to identify thanks to NetPath, which evaluates the performance characteristics of each network node and link. NetPath tracks the connectivity between your users and the services they use, identifies the infrastructure in the way, and pinpoints the locations of traffic snarls.

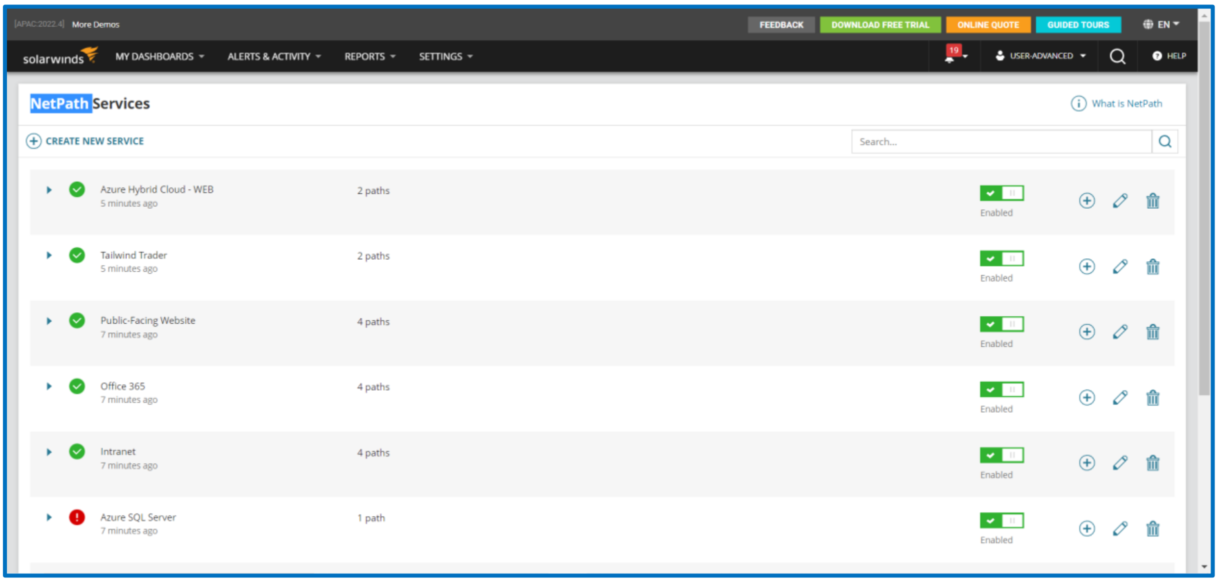

FIGURE 4.23: NETPATH SERVICES

Network Top 10

The Top 10 Interfaces by Percent Utilization, Top 10 Wireless Clients by Traffic, Top 10 Wireless APs by Clients Count, Top 10 Interfaces by Traffic, and Top 10 Interfaces by Traffic are all displayed by this service in SolarWinds. Top 10 Errors & Discards Today, Top 10 Nodes by Current Response Time Top 10 Nodes by Memory Usage, Top 10 Nodes by Average CPU Load, Top 10 Nodes by Percent Packet Loss, and Top 10 Volumes by Disk Space Usage. This window aids in our analysis of all metrics pertaining to network communication components.

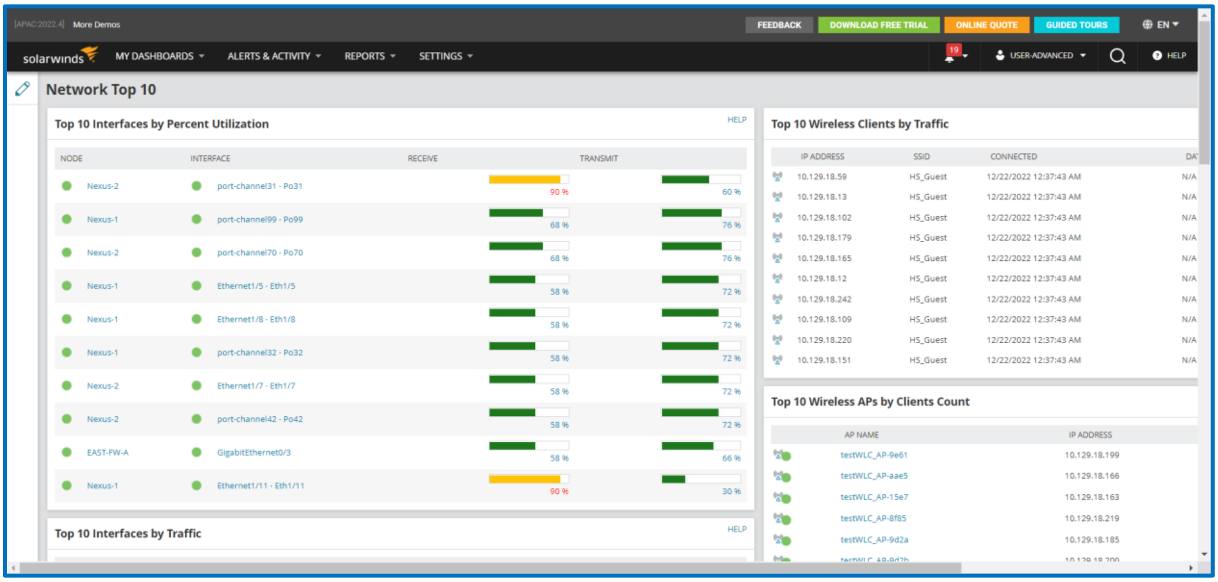

FIGURE 4.24: NETWORK TOP 10 VIEW

Alerts

An alert is a computerized notification that a network event, such a server becoming offline, has occurred. The conditions you specify when configuring an alert define the network event that initiates an alert. You may establish alerts that notify various people depending on how long the alert has been activated and schedule alerts to monitor your network during a specified time period.

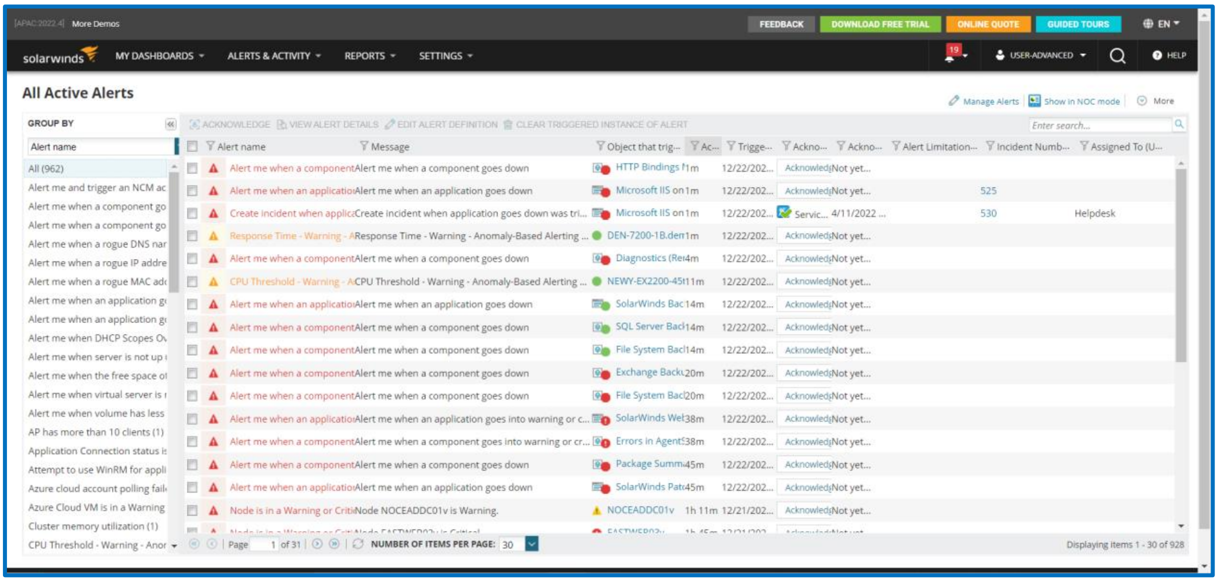

FIGURE 4.25: ALL ACTIVE ALERTS

Events

Any change in the state of a monitored object or an action taken in response to a state change is referred to as an event. To better understand the kinds of events you can anticipate, look over the list below. The range of potential outcomes is not covered by this list.

TABLE 4.1: EVENTS

Node events	**Down, Up, Warning, Deleted, Added, Unmanaged, Manage, Rebooted, and Changed.**
Interface events	Down, Up, Shutdown, Enabled, Unknown, Added, Deleted, Remapped, and Changed.
Volume events	Remapped, Changed, Added, Deleted, Disappeared, and Reappeared.
Monitoring	Started and stopped.
Failover	Failover and Failback.
Alert	Triggered and reset.

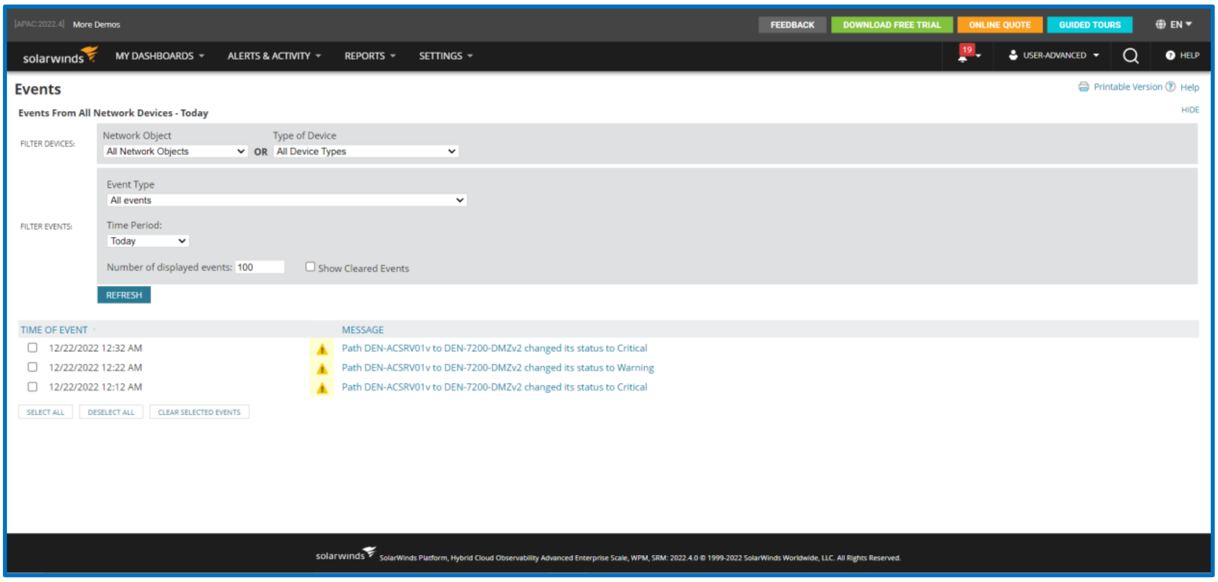

FIGURE 4.26: EVENTS FROM ALL NETWORKS

Syslogs

System Logging Protocol enables the transmission of data in a specific message format from network devices to a central server, also known as a syslog server. By making log message handling simpler, this logging protocol is an essential component of network monitoring since it enables you to monitor the general health of network devices.

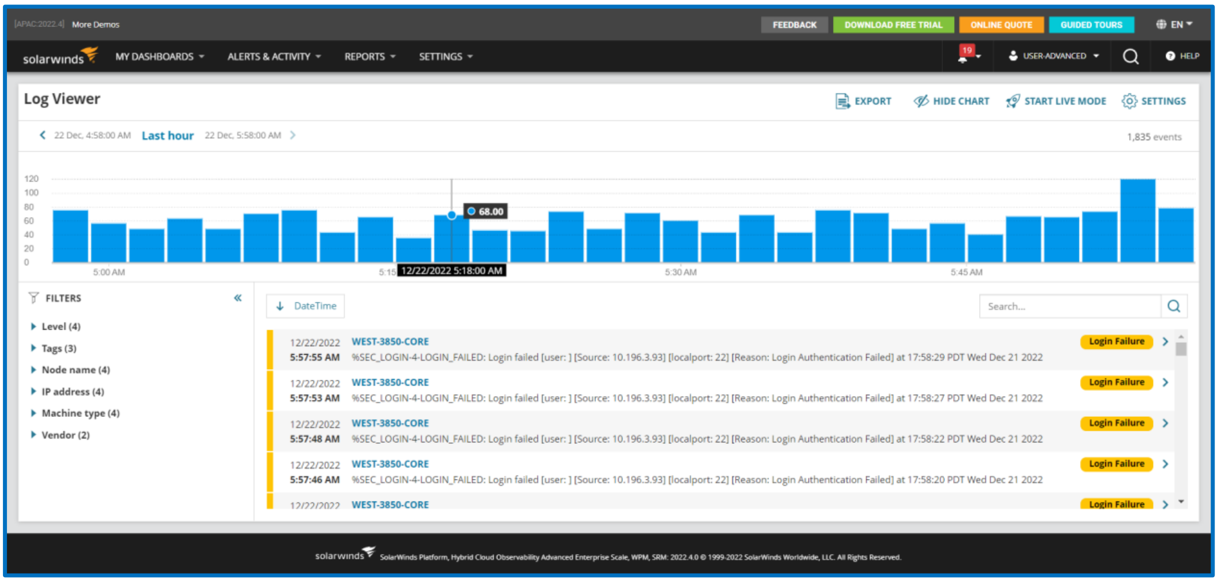

FIGURE 4.27: LOG VIEWER

SNMP Trap

An SNMP trap is sent to the designated SNMP manager by the SNMP Trap alert. Its purpose is to deliver the alert text to an SNMP manager for analysis using string pattern matching criteria, after which your current network management software reports and keeps track of it.

Any SNMP monitoring program receives an SNMP trap when you receive an SNMP trap alert. In addition to enterprise-specific and generic trap kinds like Cold Start, Warm Start, Link Down, and others, the alert supports Ip Monitor alert tokens.

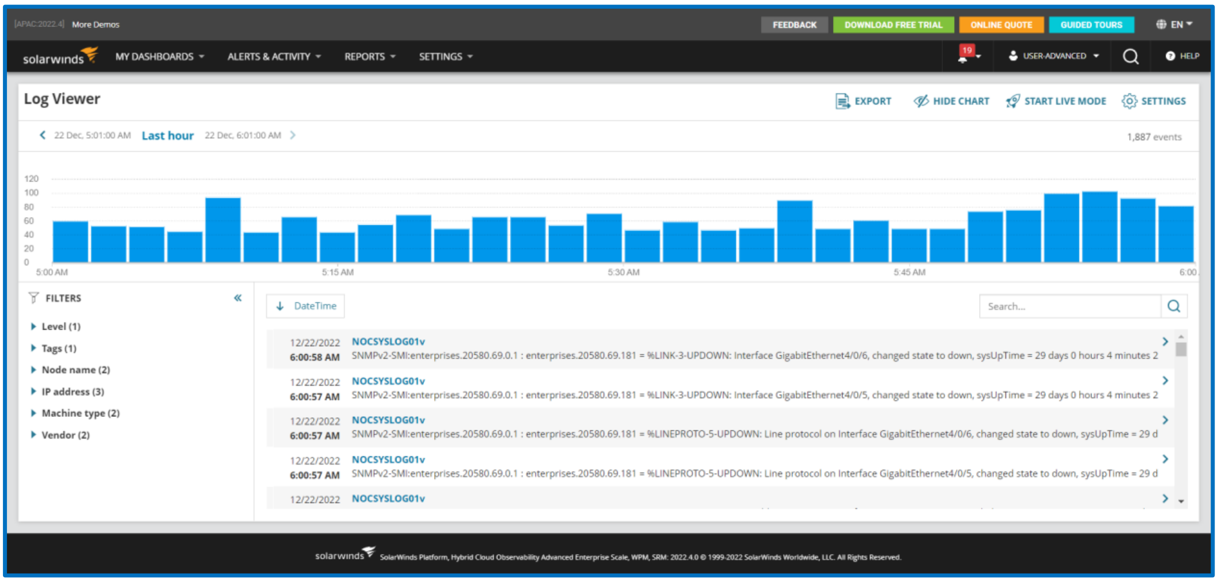

FIGURE 4.28: TRAPS

Message Center

You can view all network events, alerts, traps, and Syslog messages in the Message Center's view.

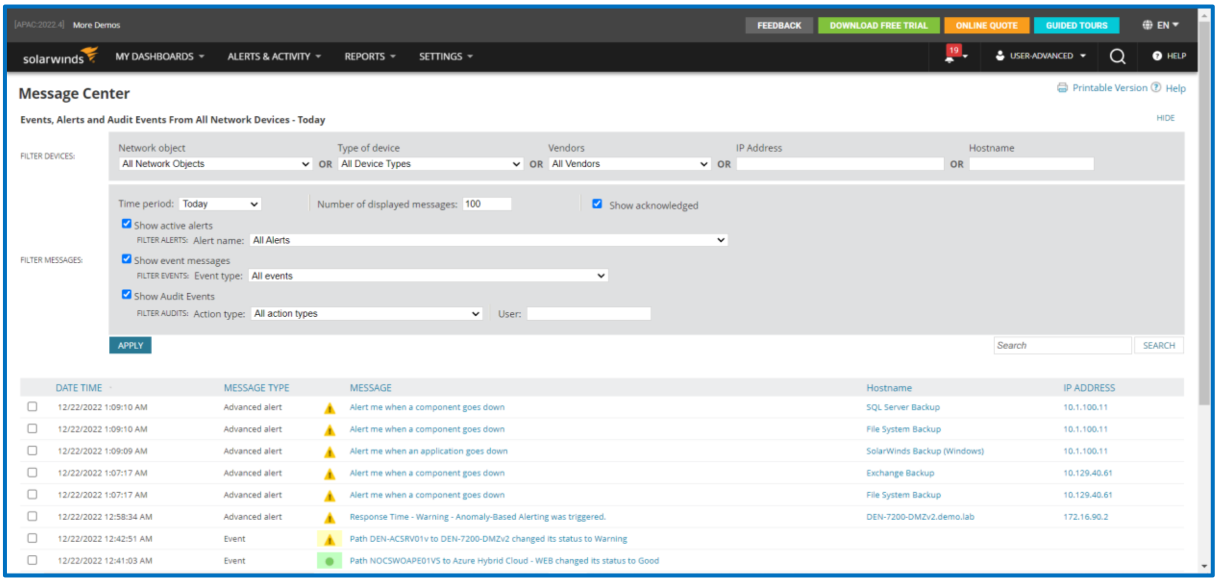

FIGURE 4.29: MESSAGE CENTER

Anomaly-Based Alerts

To enhance regular alerts, anomaly-based alerts combine Hybrid Cloud Observability (HCO) Alerting with anomaly detection as composed alerts. In order to increase the accuracy of warnings, HCO Anomaly Detection uses machine learning to identify outliers. The alert is only triggered when both the metric condition and the anomaly are present.

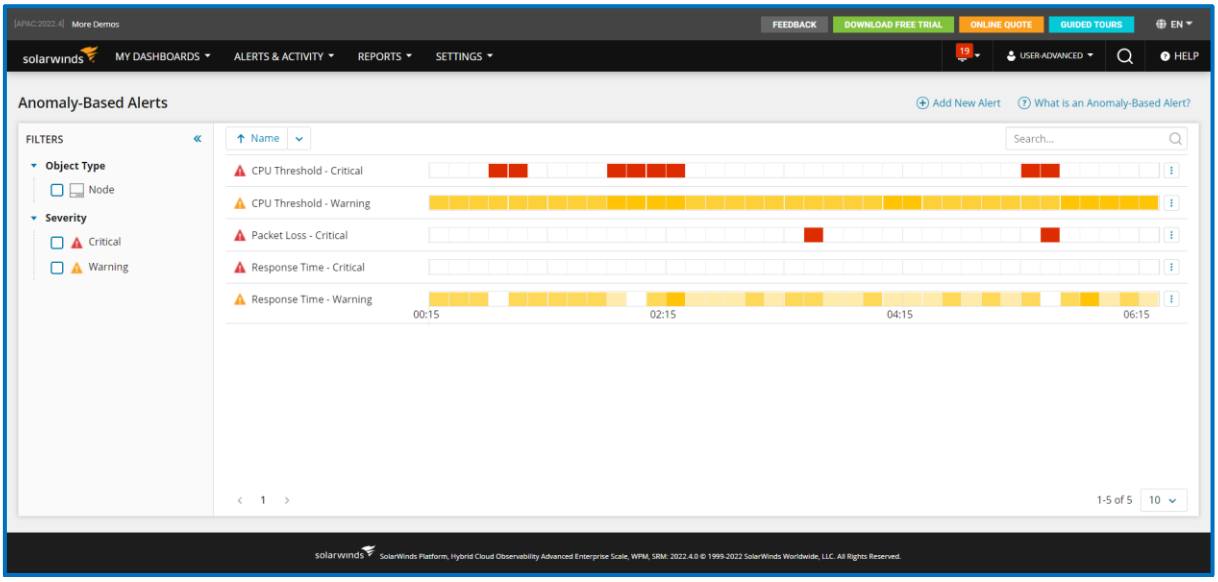

FIGURE 4.30: ANOMALY BASED ALERTS

Reports

For each SolarWinds Platform product, SolarWinds offers preconfigured reports. You can edit these predefined reports and make your own reports using the web-based interface.

By selecting Reports > All Reports from the menu bar, you can view a collection of predefined reports.

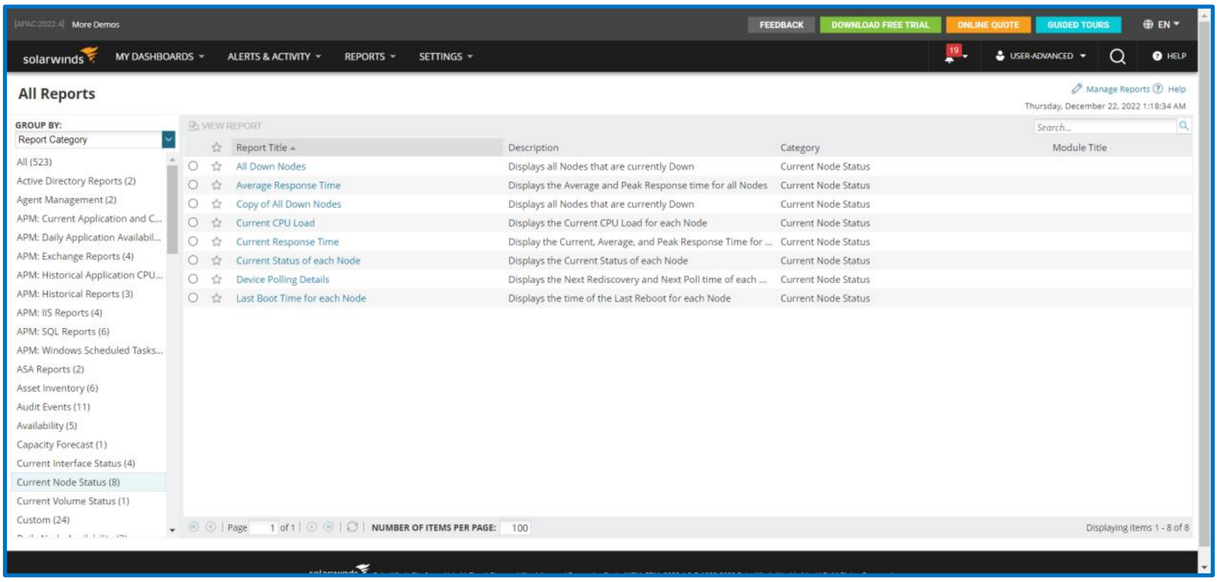

FIGURE 4.31: ALL REPORT

Network Sonar Discovery

The list of all the findings you've set up for your network is available in the Network Sonar view.

Consult the Status column on the Network Sonar Discovery tab to learn whether a discovery was successful.

- Completed: The finding was effective in its goal and did not need to be repeated.
- Scheduled: At least one more run of the finding will occur.

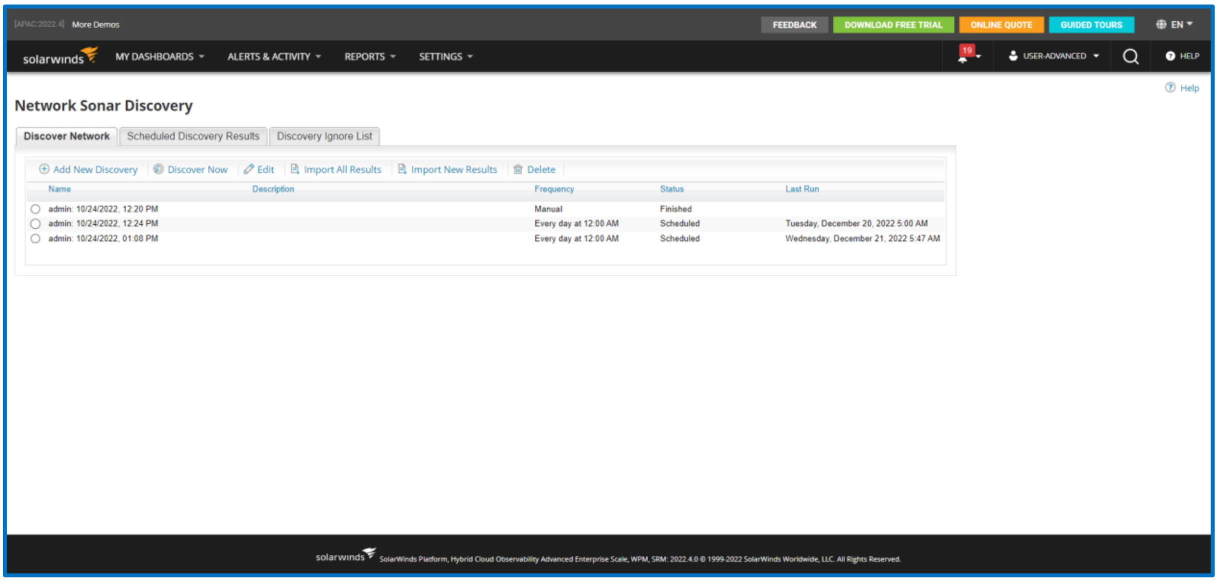

FIGURE 4.32: NETWORK SONAR DISCOVERY

Manage Nodes

In the SolarWinds Platform Web Console, the Manage Nodes view is the main view for managing devices. The terms "entities" can also be used to describe nodes and interfaces.

- Use the management actions offered in the toolbar after selecting the node or interface to manage.
- Select the devices to handle many devices at once.
- Select the box to the left of the Name column to access all monitored devices' management options.

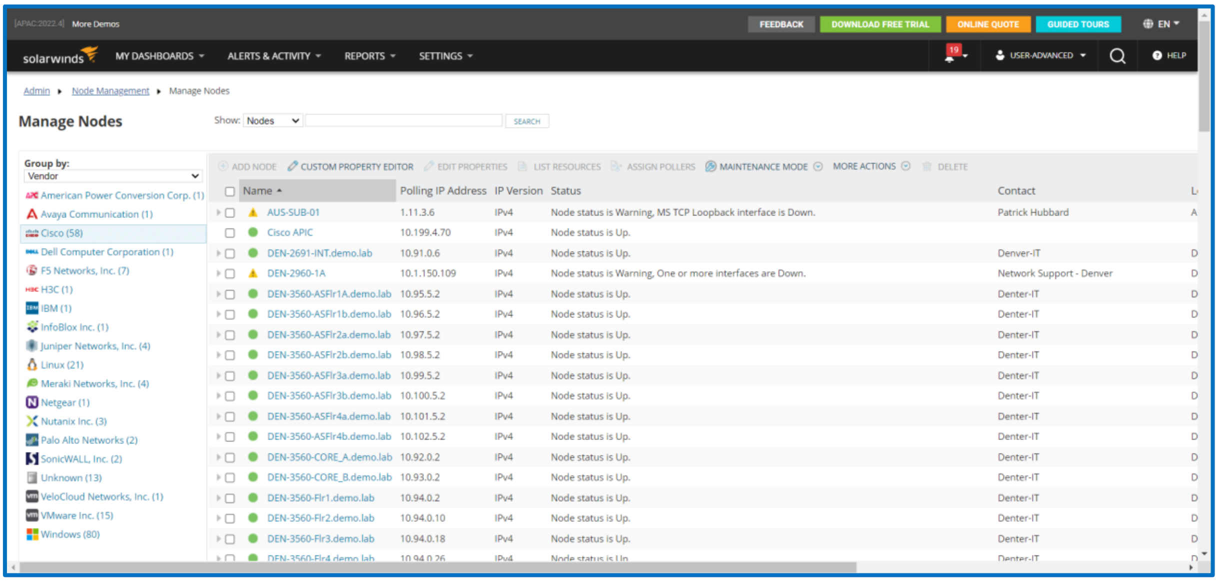

FIGURE 4.33: MANAGE NODES

Manage Dashboards

You can use a new, data-driven dashboard framework as in Orion Platform 2020.2.

These dashboards enhance the functionality of websites. They allow you to resize and arrange widgets in any desired position. These dashboards automatically update their data, so there is no need to force your browser to reload the page.

To customize contemporary dashboards, you require Administrator Rights or Manage Dashboard Rights.

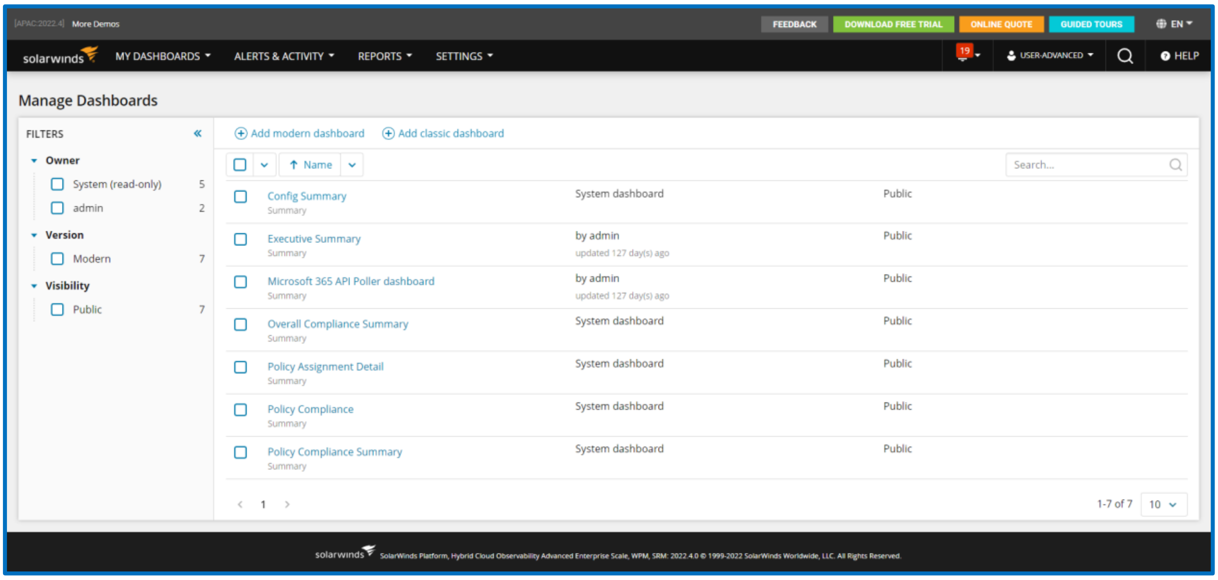

FIGURE 4.34: MANAGE DASHBOARDS

Chapter 5

RESULTS AND DISCUSSIONS

5.1 Result:

Models	Accuracy
Deep learning	95.4%
Gradient-boosted trees (GBM)	76.5%
Random Forest	54.7%

TABLE 5.1: RESULT

5.2 Deep learning:

A larger family of machine learning techniques built on artificial neural networks and representation learning includes deep learning. Unsupervised, semi-supervised, and supervised learning are all possible.

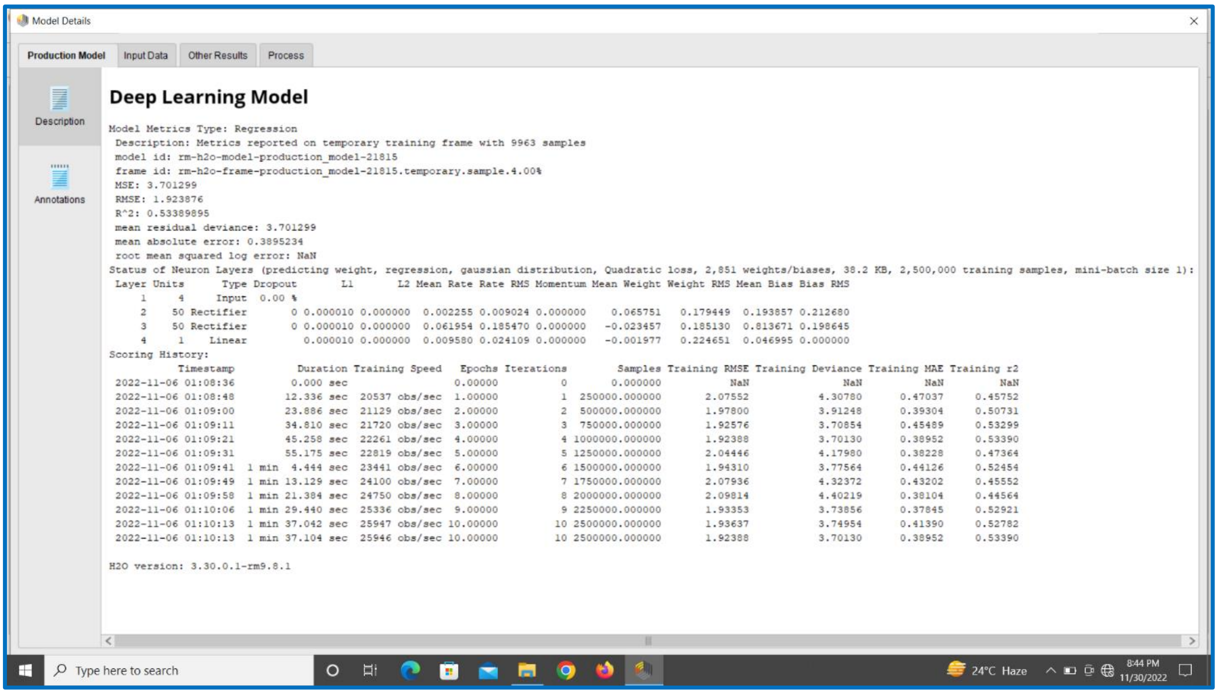

FIGURE 5.1: DEEP LEARNING MODEL

Select Simulator (DL) to gain further understanding. On the left are sliders and dropdown menus, while on the right are bar charts in this user interface. The Model Simulator selects average data values for its initial state.

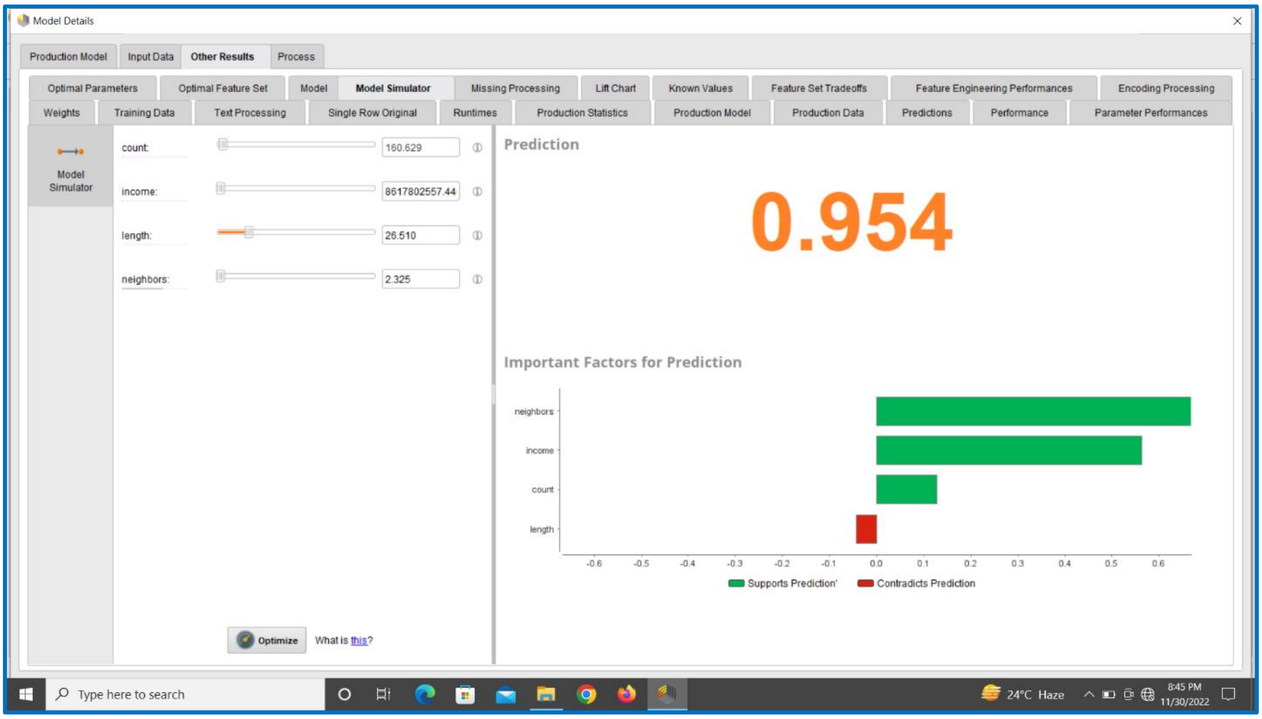

FIGURE 5.2: SELECT SIMULATOR DL

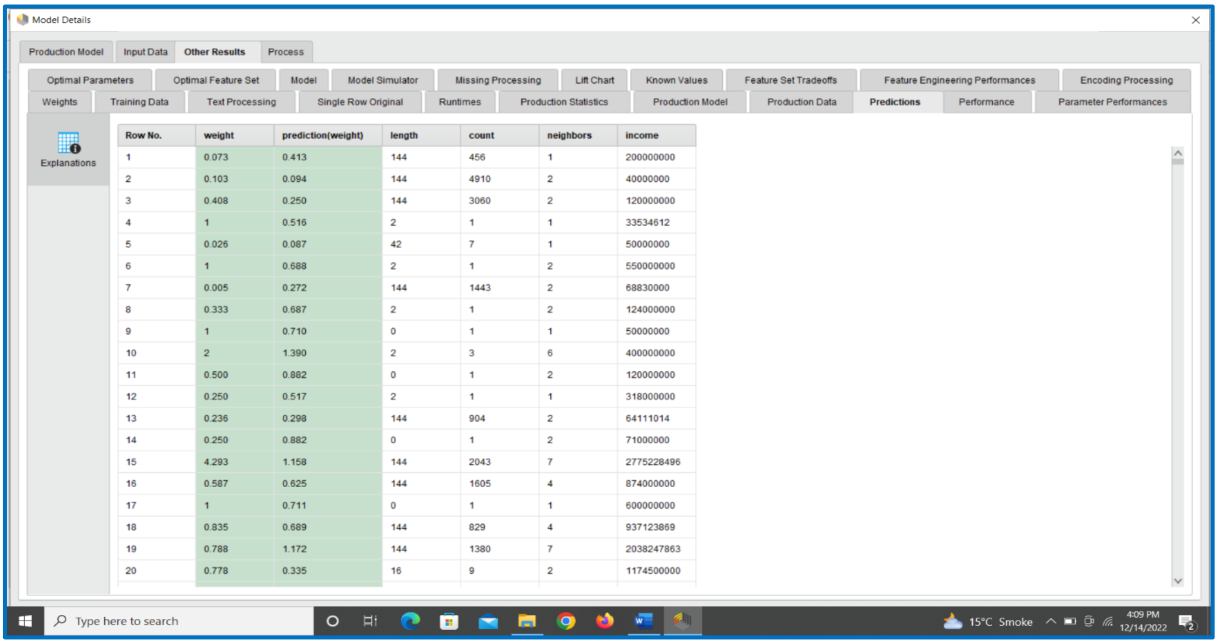

FIGURE 5.3: MODEL DETAILS

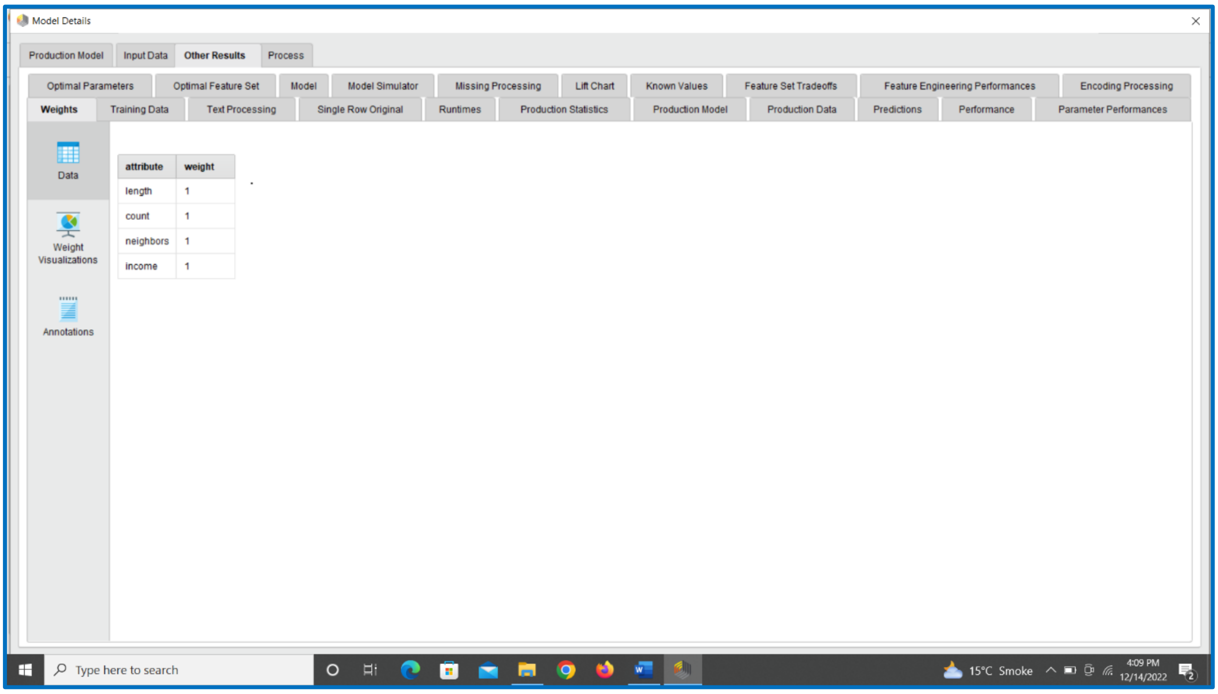

FIGURE 5.4: WEIGHT DATA

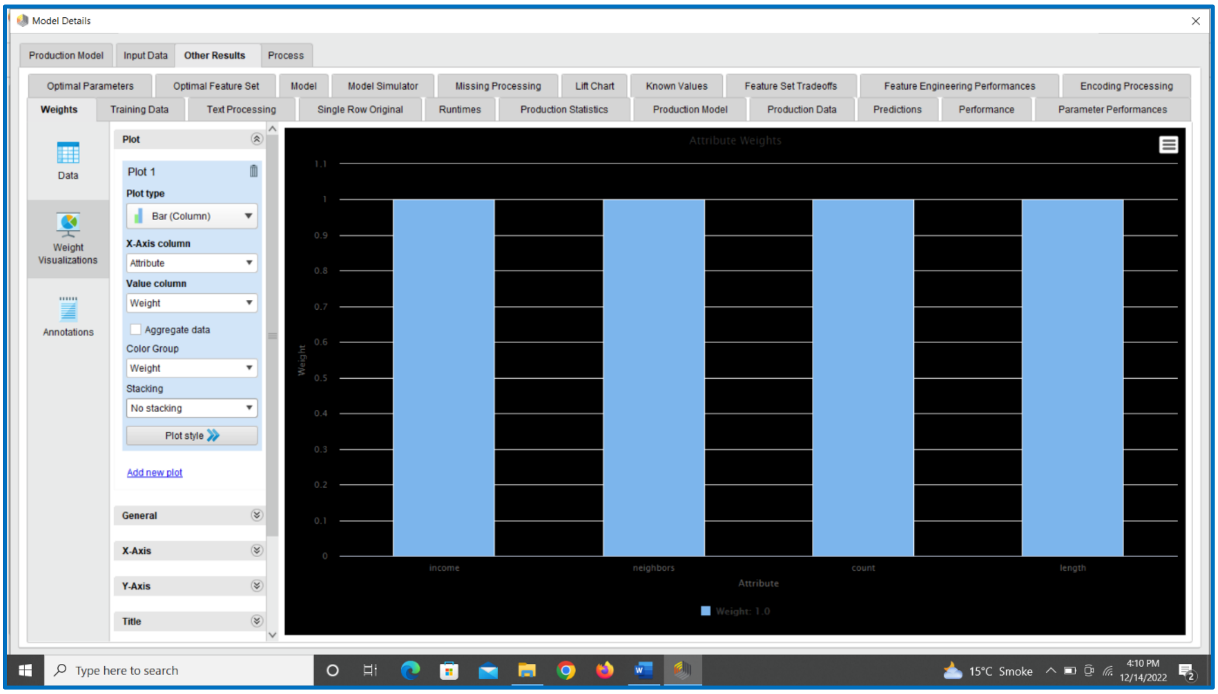

FIGURE 5.5: BAR CHART

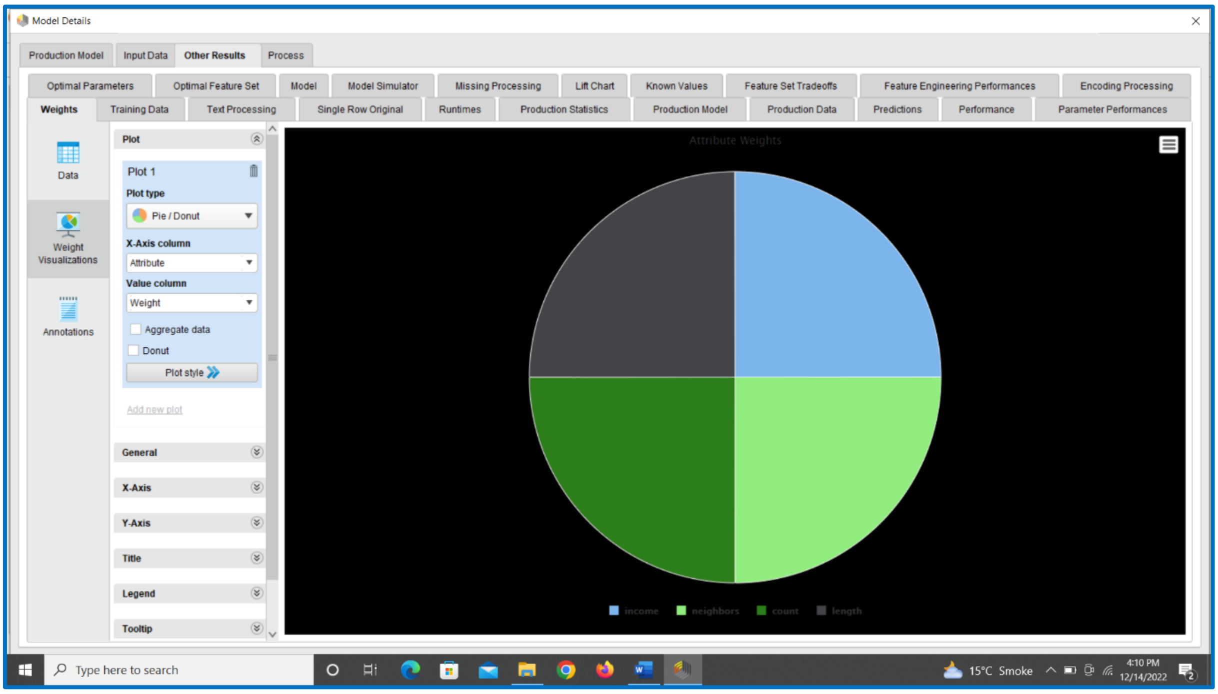

FIGURE 5.6: PIE CHART

Root Mean Squared Error class. the inaccuracy of the root-mean-square. The most widely used metric for evaluating the accuracy of numerical predictions is root mean-squared error and mean-squared error, error is the same size as predicted values themselves.

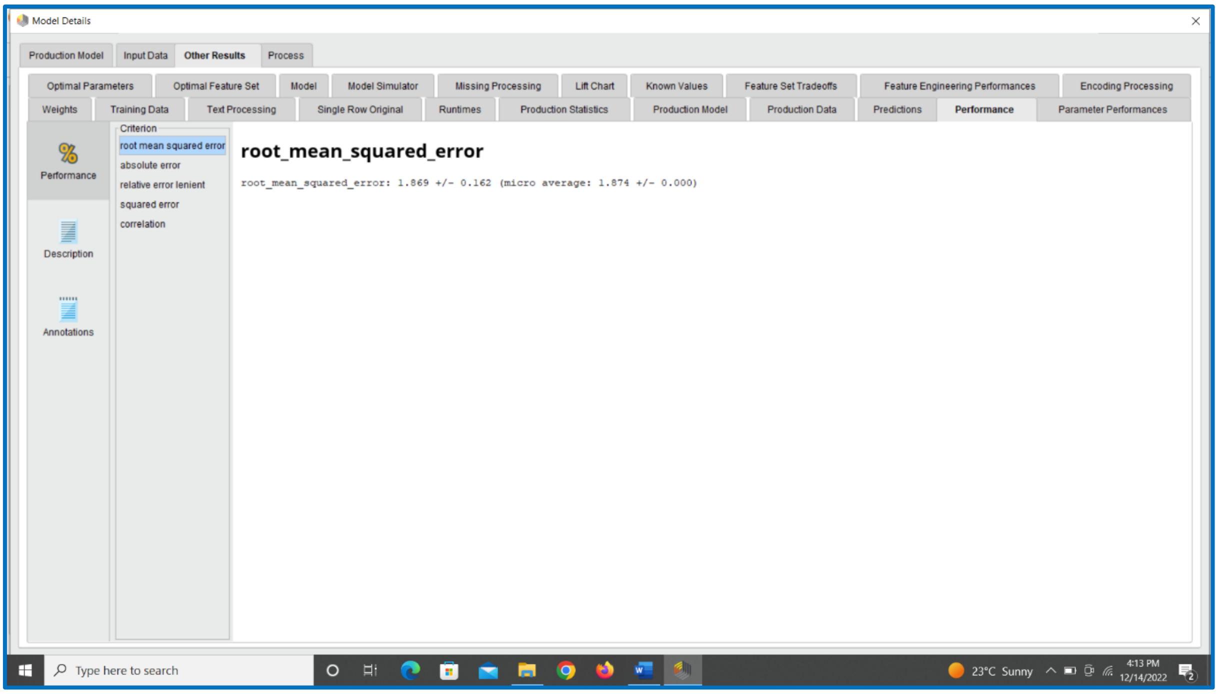

FIGURE 5.7: ROOT MEAN SQUARE ERROR

The absolute error is determined by summing the differences between all the label attribute's predicted values and actual values, Afterwards, divide the outcome by the total number of forecasts. To calculate prediction average, the actual label values are added together, and the total is divided by the overall number of instances.

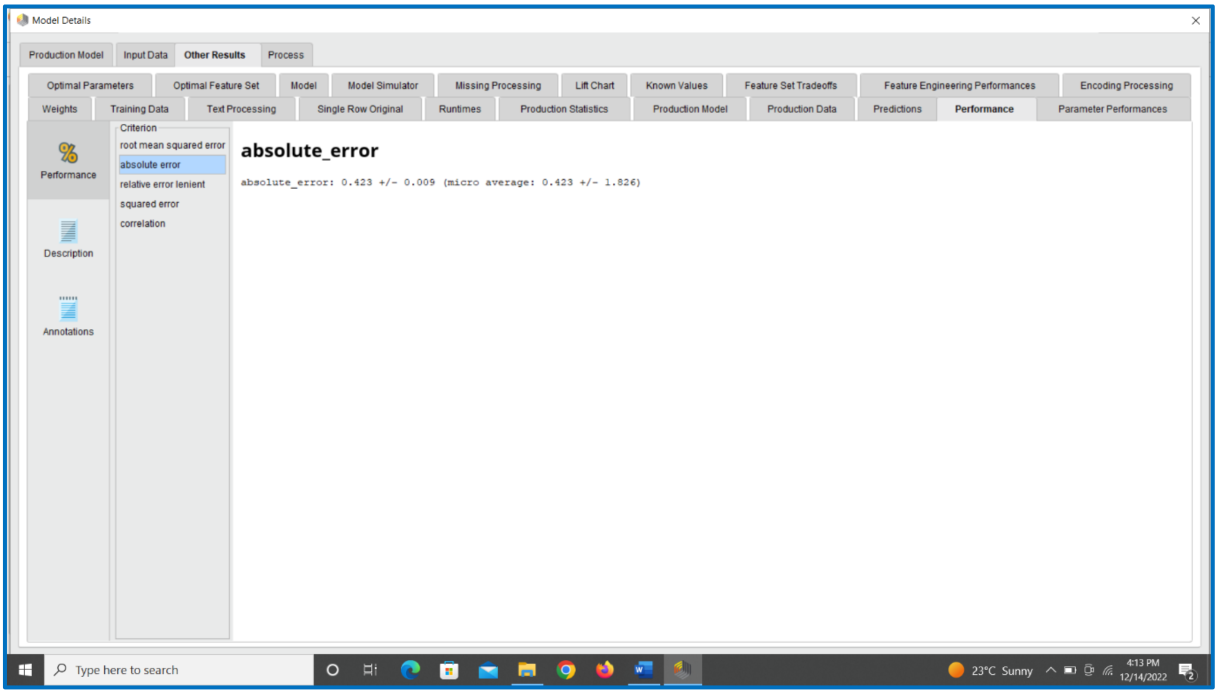

FIGURE 5.8: ABSOLUTE ERROR

The average lenient relative error is calculated by dividing the maximum of the actual value and the prediction by the average absolute deviation of the forecast from the actual value. The values of the label property correspond to the actual values.

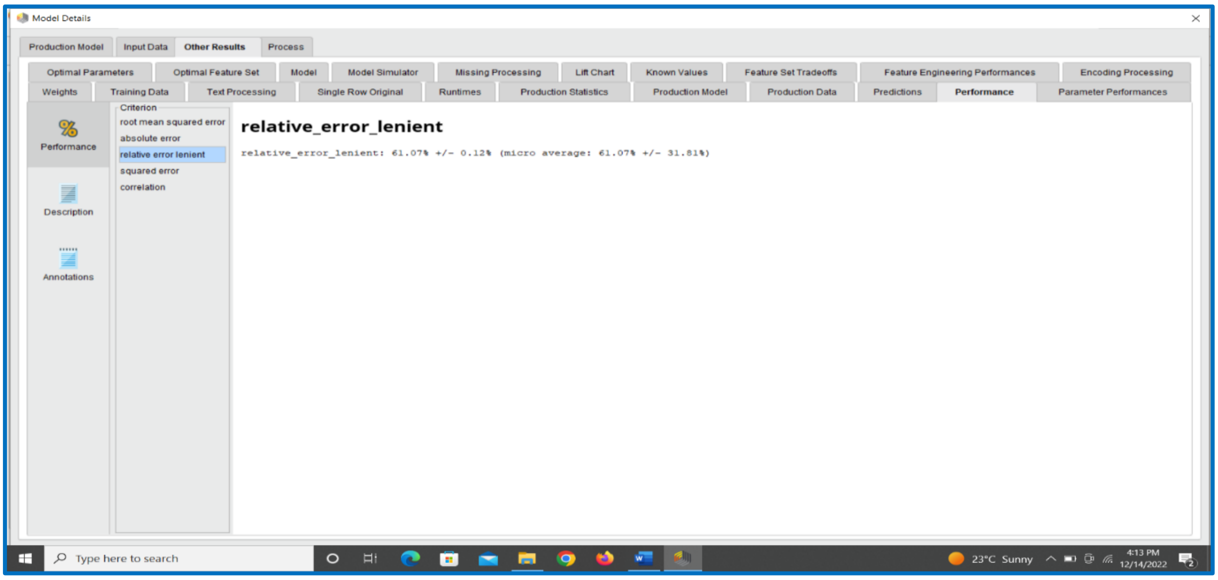

FIGURE 5.9: RELATIVE ERROR

How closely a regression line resembles a set of data points is determined by the Squared Error. It is a risk function that corresponds to the squared error loss's expected value.

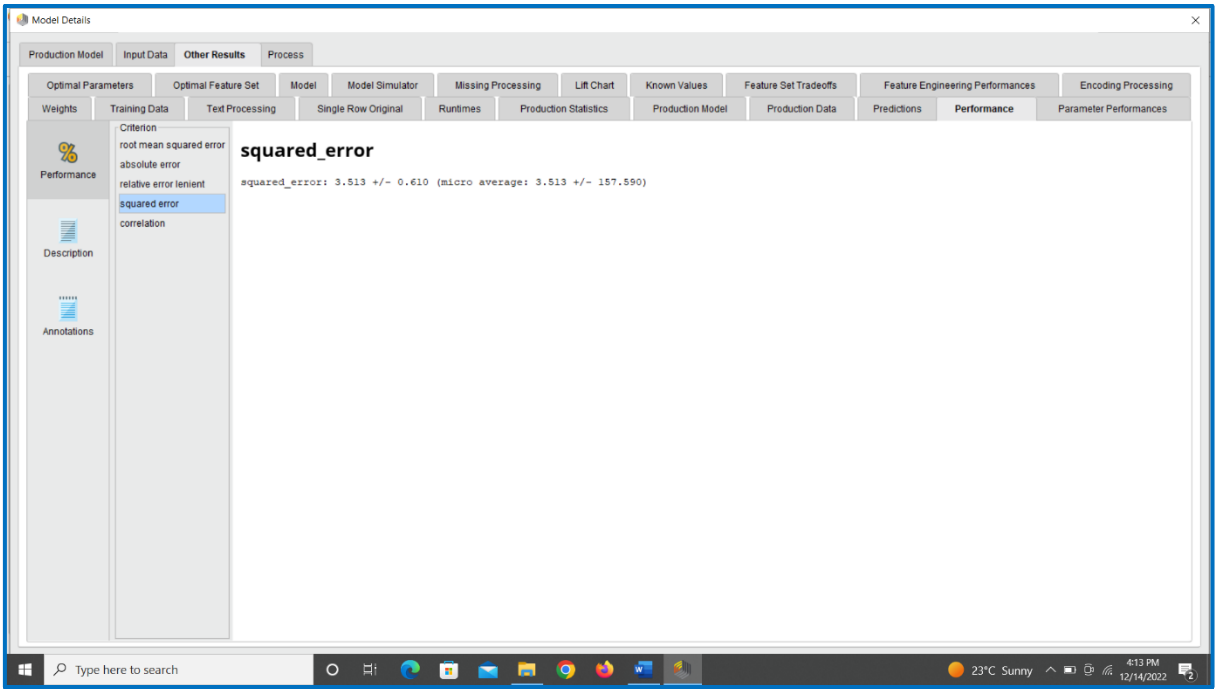

FIGURE 5.10: SQUARED ERROR

A correlation is a number that ranges from -1 to +1 and expresses how closely two attributes are related. A favorable connection is implied by a positive correlation value.

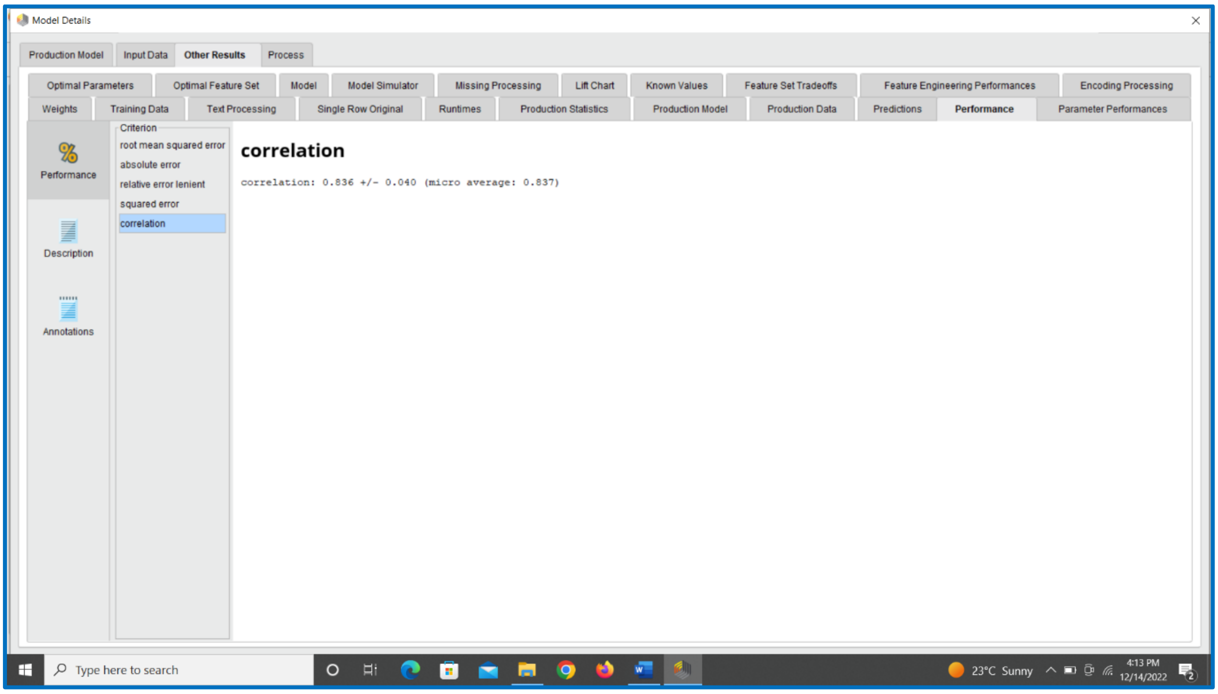

FIGURE 5.11: CORRELATION

This shows the missing values percentage with infinite, stability, and valid values. Show statistics name and value of minimum, maximum, average and standard deviation.

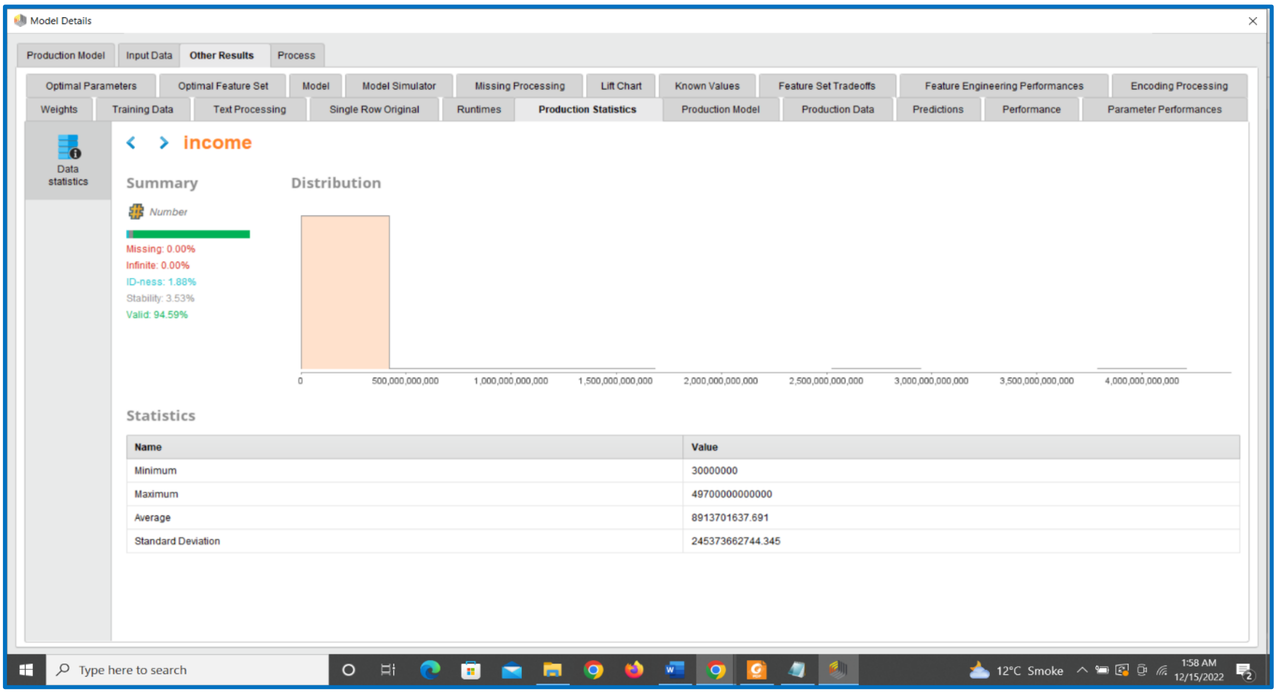

FIGURE 5.12: INCOME SUMMARY

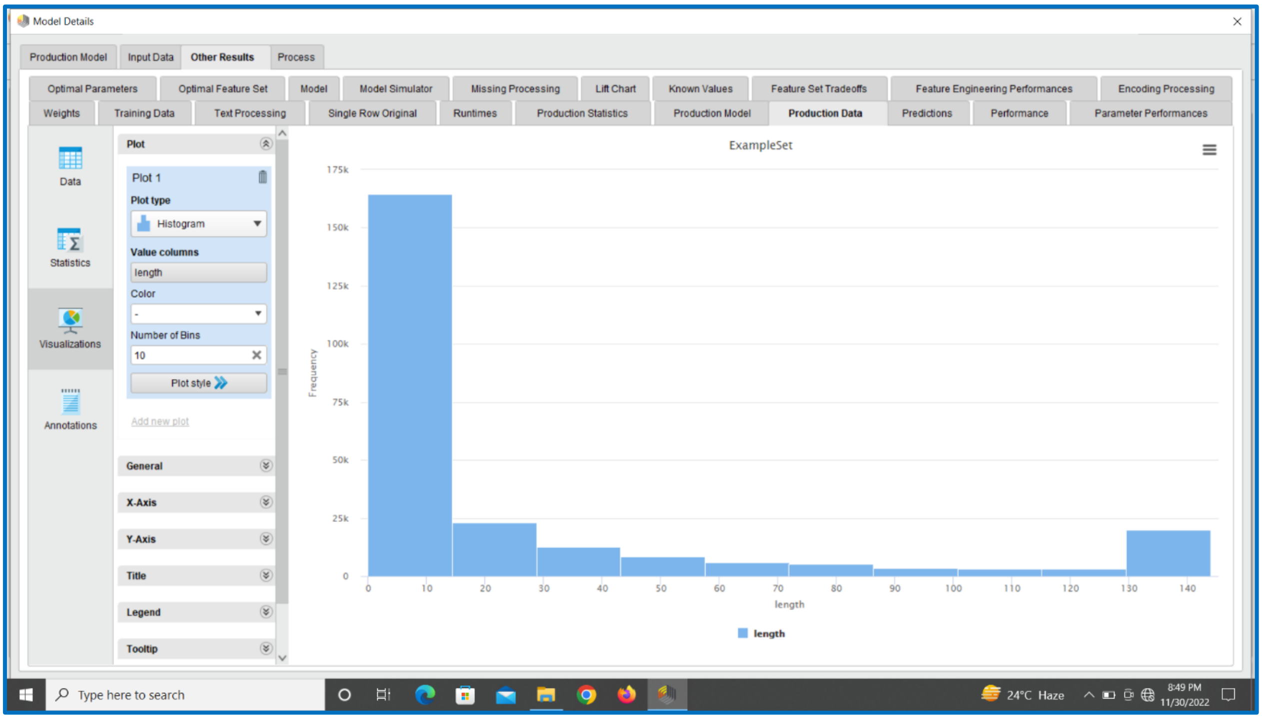

FIGURE 5.13: HISTOGRAM CHART

5.3 Gradient-boosted trees (GBM):

Gradient-boosted decision trees are a popular method for resolving prediction challenges in both the classification and regression domains. By simplifying the aim and needing fewer iterations to reach an appropriately optimum solution, the technique enhances the learning process [16].

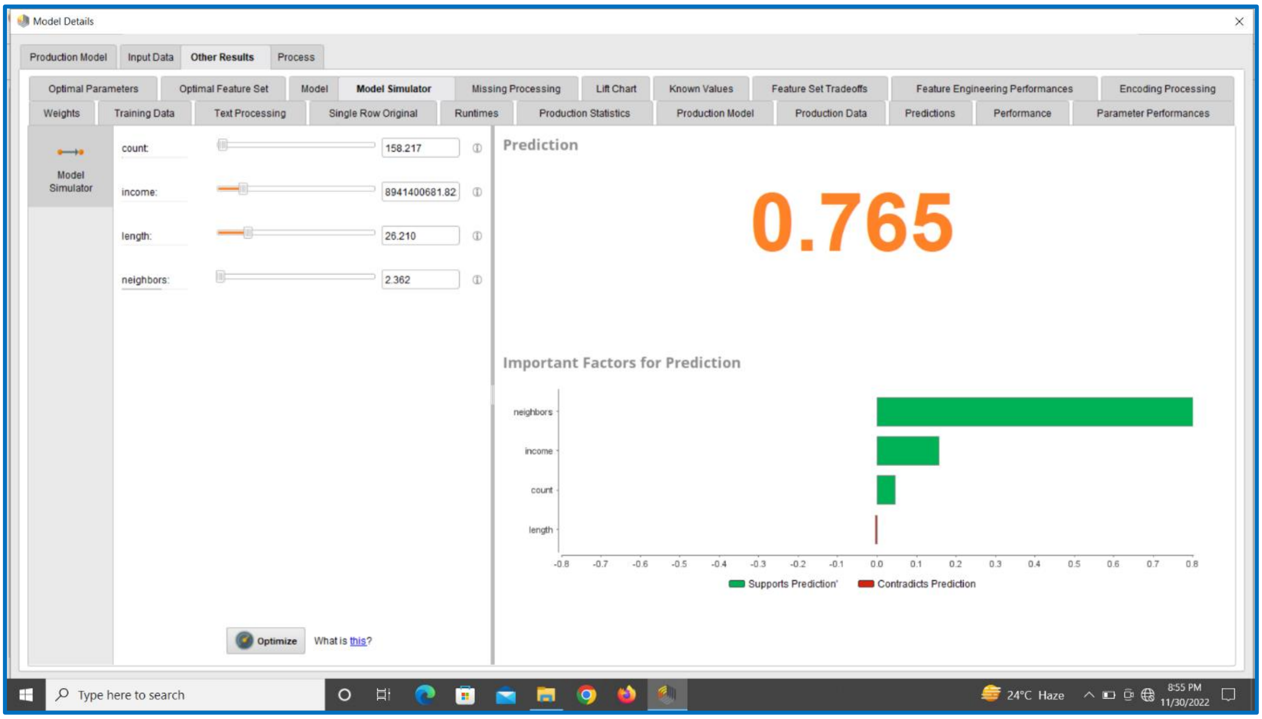

FIGURE 5.14: GRADIENT-BOOSTED TREES

A step line chart is a type of line graph in which points are linked by both horizontal and vertical line segments that resemble the steps of a staircase. When it's required to draw attention to the irregularity of changes, step line charts are utilized.

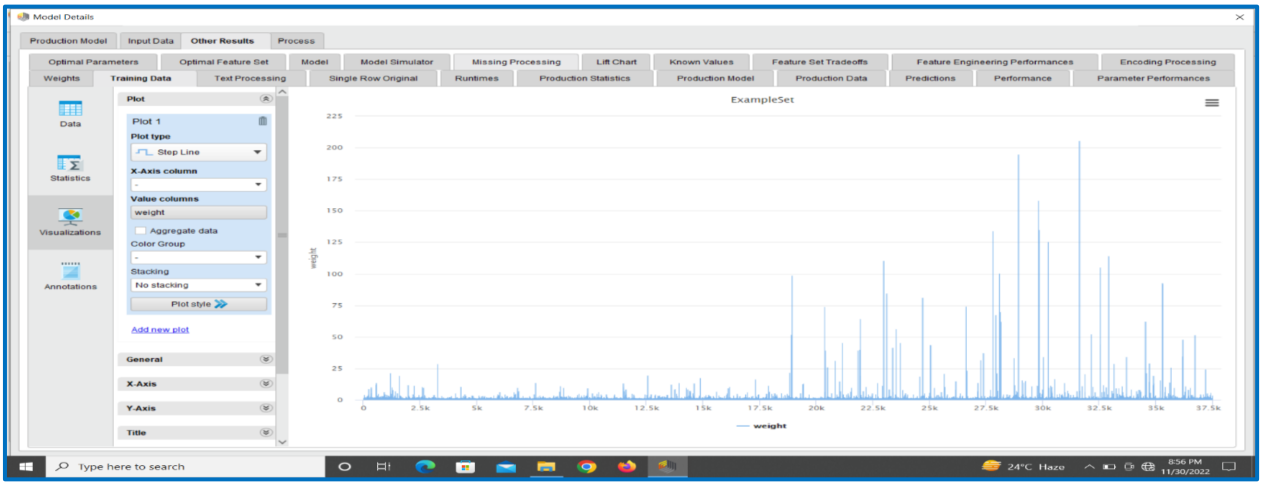

FIGURE 5.15: STEP LINE GRAPH

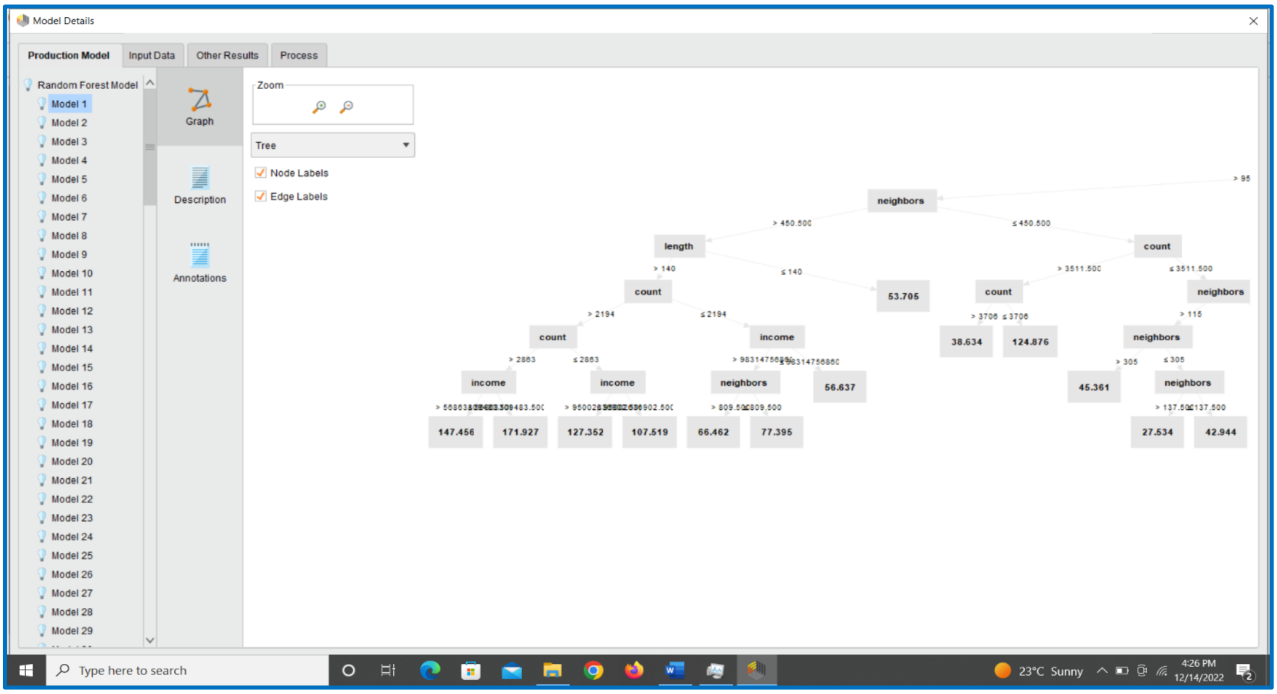

FIGURE 5.16: TREE DIAGRAM 1

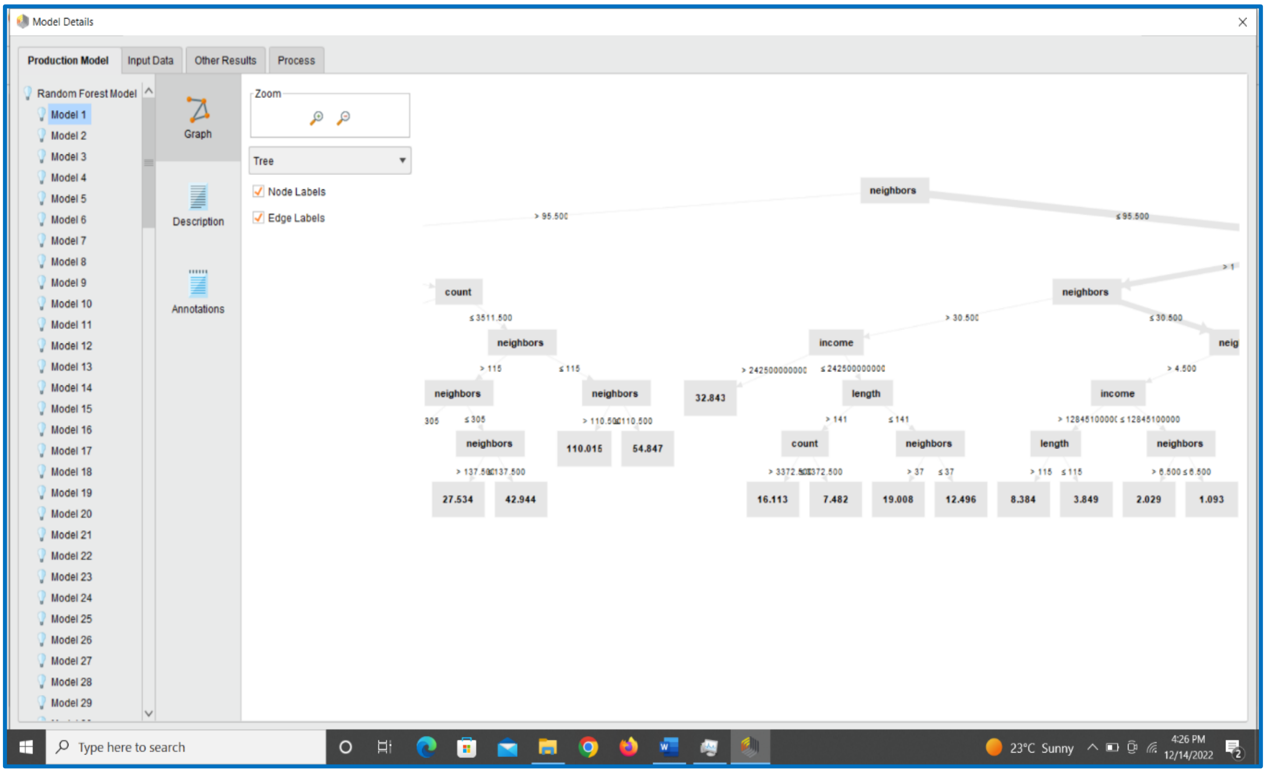

FIGURE 5.17: TREE DIAGRAM 2

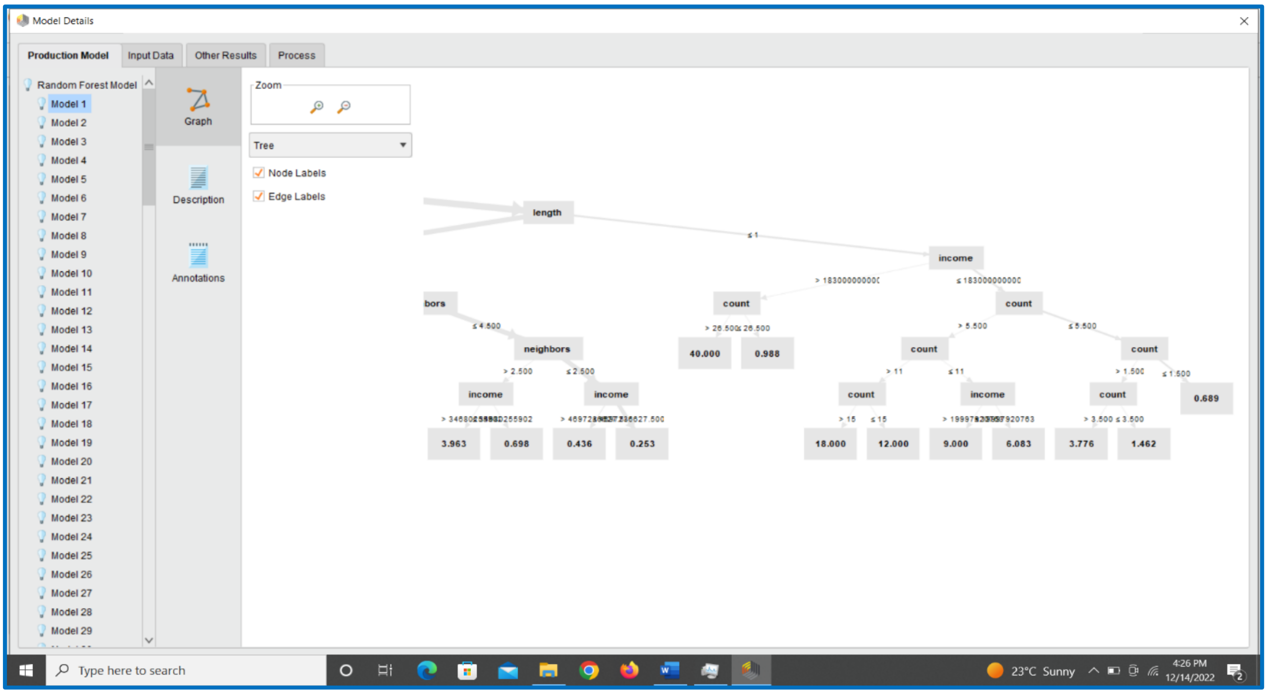

FIGURE 5.18: TREE DIAGRAM 3

5.4 Random Forest:

The ensemble learning strategy is used for regression in a supervised learning method known as Random Forest Regression. In order to provide predictions that are more accurate than those from a single model, the ensemble learning technique integrates predictions from several machine learning algorithms [17].

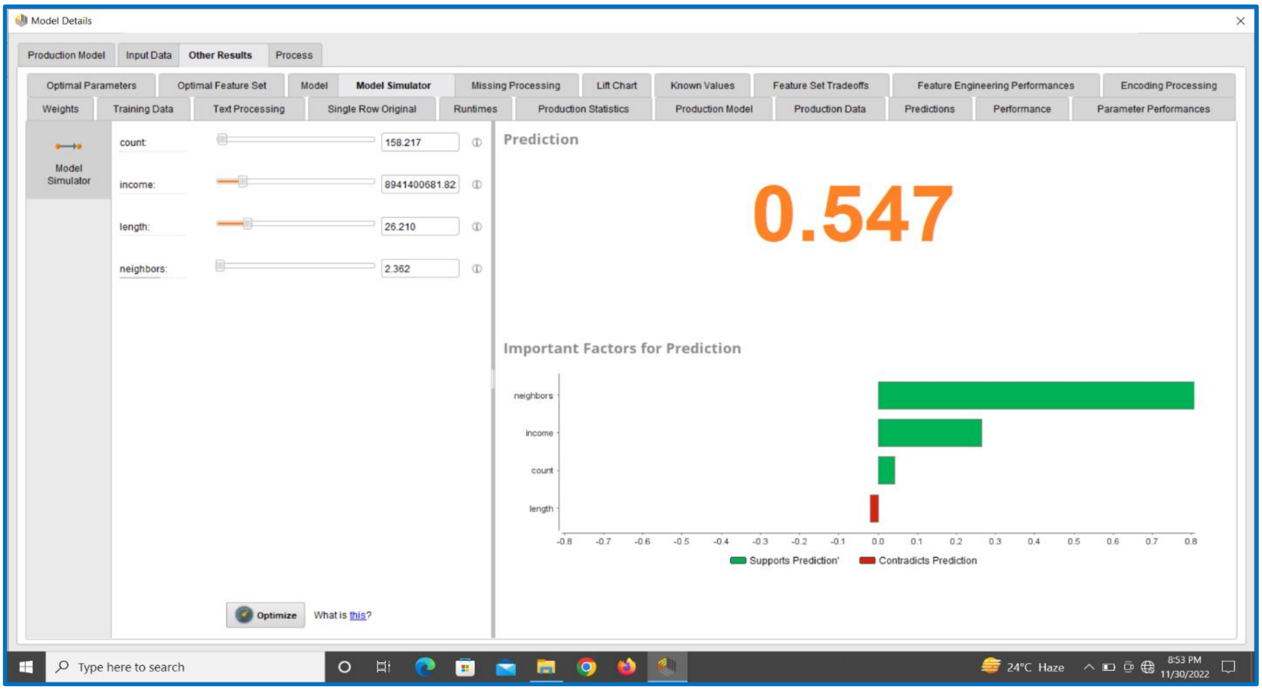

FIGURE 5.19: RANDOM FOREST

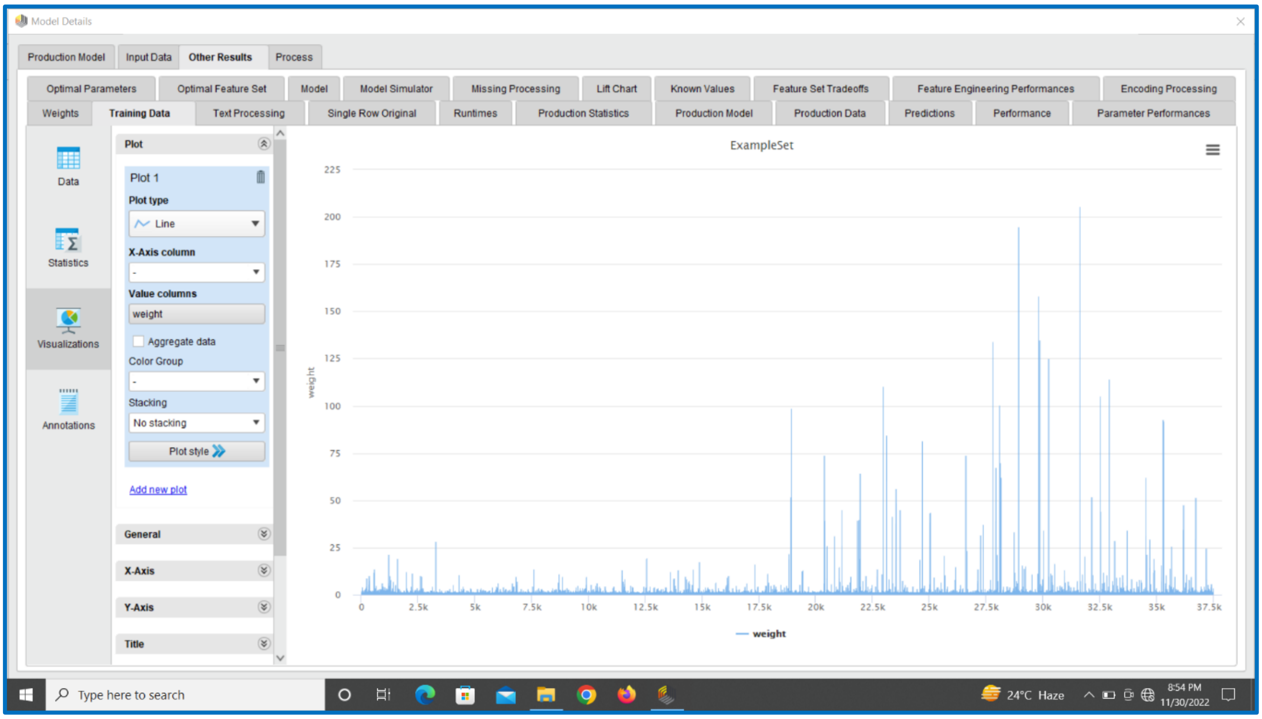

FIGURE 5.20: LINE GRAPH

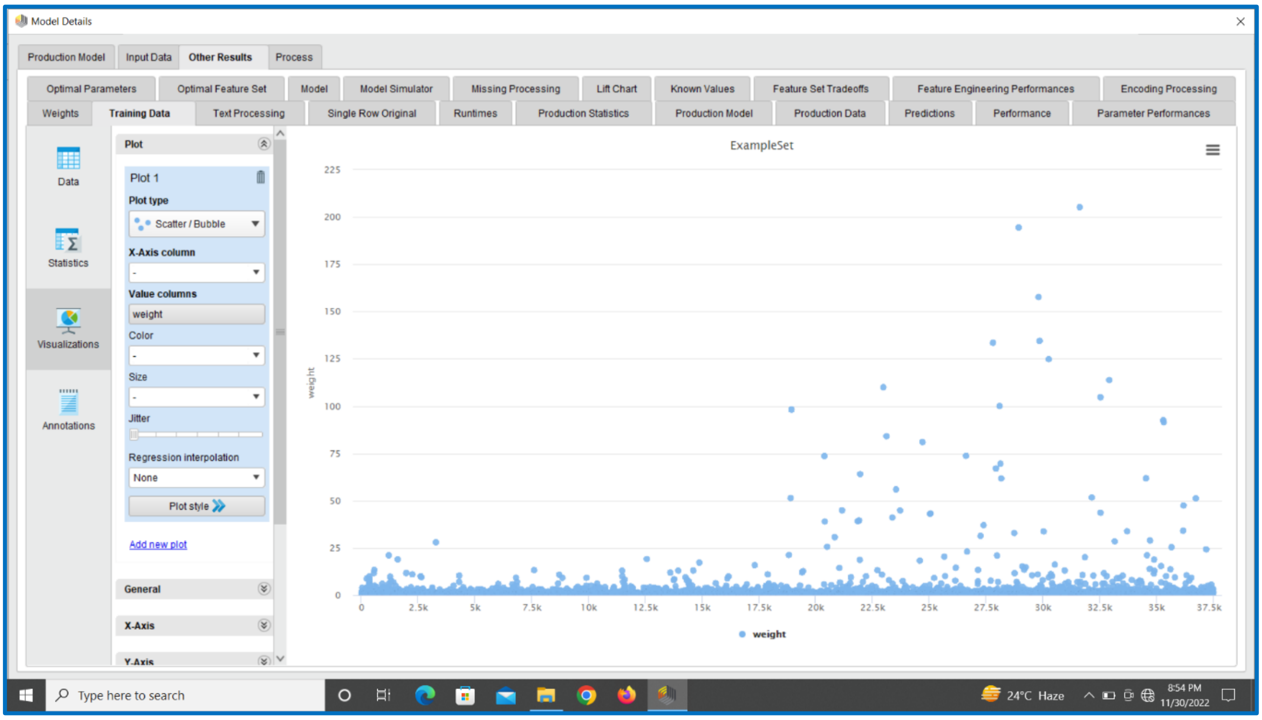

FIGURE 5.21: SCATTER / BUBBLES

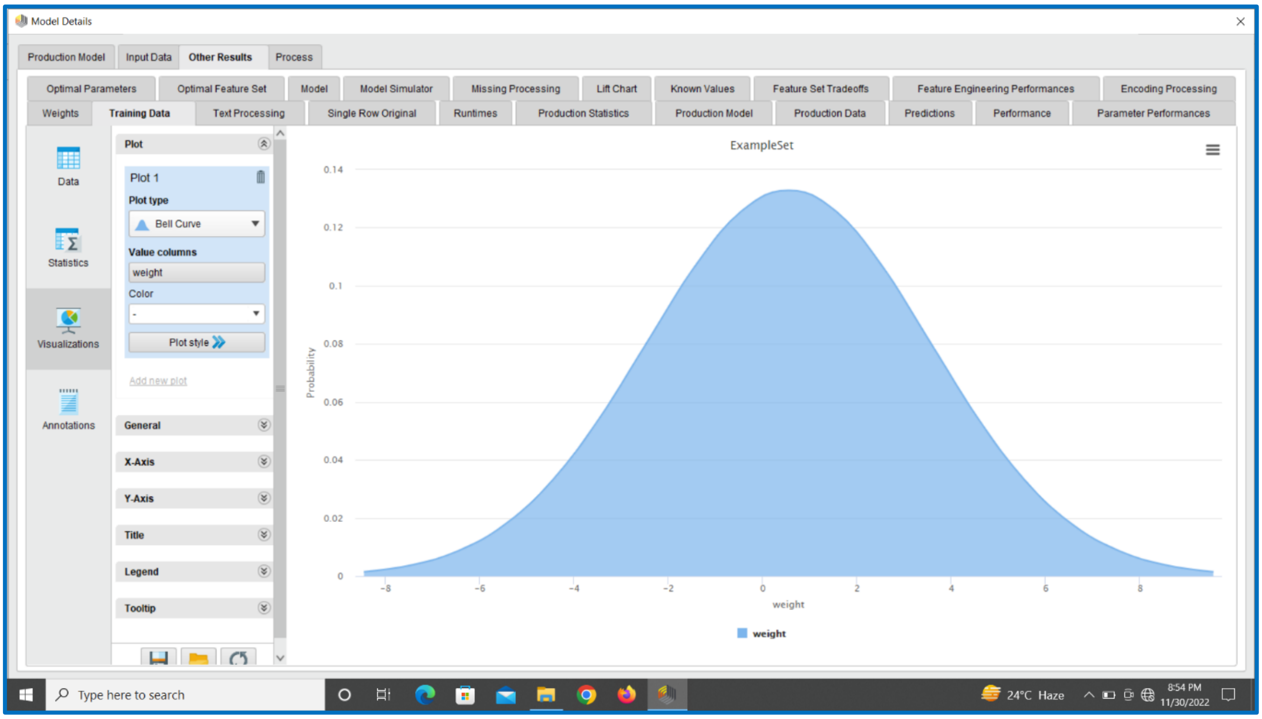

FIGURE 5.22: BELL CURVE GRAPH

Chapter 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion:

Deep learning techniques have been presented to analyze the dataset and produce effective findings (95.6% correct) by analyzing theories to find the presence of a ransomware attack. High-performance deep learning architectures are used and compared in this proposed work. Other methods are also used in this work Gradient-boosted decision trees (GBM) and random forest. Data pre-processing plays very important role in increasing the efficiency and accuracy of the results. We use a technique called backpropagation, also known as backward propagation of mistakes, is created to check for errors as they travel backward from input nodes to output nodes. For data mining and machine learning to increase the precision of predictions, it is a crucial mathematical tool. In fields like deep learning, backpropagation algorithms are frequently employed to train feedforward neural networks. The gradient of the loss function with respect to the network weights is easily computed. It enables the training of multilayer networks and the updating of weights to minimize loss using gradient methods, such as gradient descent or stochastic gradient descent.

Since the data is collected by Kaggle contain few datasets because of the privacy of data.

6.2 Future Work:

We have created a model where we can detect Ransomware Attack from the datasets. Secondly, we won't be restricted with only to Ransomware Attacks. We can create a model that will help to detect other attacks. The only thing that will be needed for achieving these goals is dataset. We can create a platform where individuals from all over the world can interact and check their work from our model.

REFERENCES

[1] M. Grima, "Ransomware Activity Detection," *Ransomware Activity Detection,* 2018.

[2] D. K. Tripathi, vol. 3, no. 5, 2017.

[3] kaspersky, "ransomware-attacks-and-types," [Online]. Available: https://www.kaspersky.com/resource-center/threats/ransomware-attacks-and-types.

[4] A. D. C. N. Dhinnesh, "Analysis of Ransomware and its prevention," *Analysis of Ransomware and its prevention,* vol. 5, p. 4, 2020.

[5] checkpoint, "What is Ransomware?," checkpoint, [Online]. Available: https://www.checkpoint.com/cyber-hub/threat-prevention/ransomware/.

[6] cisecurity.org, "steps-to-help-prevent-limit-the-impact-of-ransomware," [Online]. Available: https://www.cisecurity.org/insights/blog/7-steps-to-help-prevent-limit-the-impact-of-ransomware.

[7] imperva, "Ransomware Protection," imperva, [Online]. Available: https://www.imperva.com/learn/application-security/ransomware/.

[8] securityboulevard, "machine-learning-tackles-ransomwar," [Online]. Available: https://securityboulevard.com/2022/06/machine-learning-tackles-ransomware-attacks/.

[9] researchgate, "Data-preprocessing," [Online]. Available: https://www.researchgate.net/figure/Data-preprocessing-steps_fig3_228630212.

[10] docs.h2o.ai, "Deep Learning (Neural Networks)," docs.h2o.ai, [Online]. Available: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/deep-learning.html.

[11] "datascience," gradient-boosted-trees, [Online]. Available: https://towardsdatascience.com/a-visual-guide-to-gradient-boosted-trees-8d9ed578b33.

[12] "analyticsvidhy," understanding-random-forest, [Online]. Available: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/.

[13] "bitcoinheist-ransomware-dataset," .kaggle, [Online]. Available: https://www.kaggle.com/datasets/sapere0/bitcoinheist-ransomware-dataset.

[14] "development-platform-for-data-minin," analyticsvidhya, [Online]. Available: https://www.analyticsvidhya.com/blog/2021/10/intro-to-rapidminer-a-no-code-development-platform-for-data-mining-with-case-study/.

[15] N. Chourey, "A Study of Ransomware Detection and Prevention at Organizations," 2020.

[16] "neptune," boosted-decision-trees-guide, [Online]. Available: https://neptune.ai/blog/gradient-boosted-decision-trees-guide.

[17] "towardsdatascience," visualize-individual-decision-trees-in-a-random-forest, [Online]. Available: https://towardsdatascience.com/4-ways-to-visualize-individual-decision-trees-in-a-random-forest-7a9beda1d1b7.

[18] "neptune," gradient-boosted-decision-trees-guide, [Online]. Available: https://neptune.ai/blog/gradient-boosted-decision-trees-guide.