Text

Description automatically generated with medium confidence

UNIVERSITY OF HERTFORDSHIRE

Department of Computer Science

MSc Data Science and Analytics

Data Science and Analytics Masters Project

7COM1075-0509

**THE CONVERGENCE OF 5G, ARTIFICIAL INTELLEGENCE, DATA ANALYTICS, INTERNET OF THINGS AND END USER PREDICTION**

Name: Tohib Olalekan Owoade

Student ID: 20054509

Supervisor: Dr. Justin Mifsud

Date: 20/09/2022

# ABSTRACT

In this study, the researcher extensively examined how the rapidly developing fields of 5G networking, AI, Data Mining, and the Internet of Things all interact with one another (IOT). The Internet of Things (IoT) market encompasses products and services that integrate artificial intelligence (AI) into IoT infrastructure and provide IoT-based support for a wide variety of AI-enabled use cases. This document evaluates the primary players, strategies, remedies, and available assets. The researcher claims that AIoT is revolutionary in both contexts because AI enhances IoT with machine learning and decision making, while IoT enhances AI with communication and data exchange. With this in mind, it is important to note that factors some might interfere with 5G network signals, the researcher proposes a question of whether we can be able to know if our infrastructure still have the required 5G connection even when we are not able to detect this by network speed and other relatively obvious factors.

The researcher proposes to answer this pertinent question by using machine learning methods and models to predict 5G network usage when obvious characteristics are not in place.

Table of Contents

[ABSTRACT ii](#_Toc114862184)

[1 INTRODUCTION 6](#_Toc114862185)

[1.1 Background of the Study 6](#_Toc114862186)

[1.2 PROBLEM STATEMENT 8](#_Toc114862187)

[1.3 RESEARCH QUESTION 8](#_Toc114862188)

[1.4 AIM 8](#_Toc114862189)

[1.5 OBJECTIVES 9](#_Toc114862190)

[1.6 PURPOSE OF THE STUDY 9](#_Toc114862191)

[1.7 SCOPE OF THE STUDY 9](#_Toc114862192)

[1.8 LIMITATION OF THE STUDY 9](#_Toc114862193)

[1.9 PROJECT STRUCTURE 9](#_Toc114862194)

[1.10 DEFINITION OF TERMS 10](#_Toc114862195)

[2 LITERATURE REVIEW 12](#_Toc114862196)

[2.1 INTRODUCTION 12](#_Toc114862197)

[2.2 5G TECHNOLOGY AND APPLICATIONS 13](#_Toc114862198)

[2.2.1 Smart Mobility: 14](#_Toc114862199)

[2.2.2 Smart Energy: 14](#_Toc114862200)

[2.2.3 Smart Health: 15](#_Toc114862201)

[2.2.4 Industrial Applications: 15](#_Toc114862202)

[2.2.5 Consumer Applications: 16](#_Toc114862203)

[2.3 INTERNET OF THINGS AND ARTIFICIAL INTELLIGENCE REVOLUTION 16](#_Toc114862204)

[2.3.1 Sustainable development’s areas for AI and IoT. 16](#_Toc114862205)

[2.3.2 AI and IOT Technical Issues 17](#_Toc114862206)

[2.4 BLOCKCHAIN, 5G ROBOTICS AND ARTIFICIAL INTELLIGENCE 18](#_Toc114862207)

[2.5 BIG DATA IN 5G DISTRIBUTED APPLICATION 21](#_Toc114862208)

[2.5.1 Key Technology and Business Drivers 22](#_Toc114862209)

[2.5.2 Mobile Cloud/Edge Computing 22](#_Toc114862210)

[3 METHODOLOGY 28](#_Toc114862211)

[3.1 INTRODUCTION 28](#_Toc114862212)

[3.2 LEGAL, ETHICAL, PROFESSIONAL AND SOCIAL ISSUES CONSIDERATION 28](#_Toc114862213)

[3.2.1 Ethical Issues and Mitigation 28](#_Toc114862214)

[3.2.2 Legal Issues and Mitigation 29](#_Toc114862215)

[3.2.3 Professional Issues and Mitigation 29](#_Toc114862216)

[3.2.4 Social Issues and Mitigation 29](#_Toc114862217)

[3.3 TECHNICAL REQUIREMENT 30](#_Toc114862218)

[3.3.1 Python Language 30](#_Toc114862219)

[3.3.2 Google Colab Notebook 30](#_Toc114862220)

[3.4 DATASET 31](#_Toc114862221)

[3.5 DATA PREPARATION 31](#_Toc114862222)

[3.5.1 Importing Libraries and Dataset 31](#_Toc114862223)

[3.5.2 Cleaning of Dataset 32](#_Toc114862224)

[3.5.3 Pre-processing of Dataset 34](#_Toc114862225)

[3.5.4 Normalisation of Data 36](#_Toc114862226)

[3.5.5 Training and Testing Split 36](#_Toc114862227)

[3.6 MODEL TRAINING AND PREDICTION 37](#_Toc114862228)

[3.6.1 Naïve Bayes 38](#_Toc114862229)

[3.6.2 Logistic Regression 38](#_Toc114862230)

[3.6.3 Support Vector Machine (SVM) 39](#_Toc114862231)

[3.6.4 Decision Tree 39](#_Toc114862232)

[3.6.5 Artificial Neural Network (ANN) 40](#_Toc114862233)

[4 RESULT ANALYSIS 41](#_Toc114862234)

[4.1 INTRODUCTION 41](#_Toc114862235)

[4.2 PREDICTION PERFORMANCE EXPLANATION 41](#_Toc114862236)

[4.3 CONFUSION MATRIX EXPLANATION 42](#_Toc114862237)

[5 CONCLUSION, EVALUATION & SUGGESTION 45](#_Toc114862238)

[6 REFERENCES 46](#_Toc114862239)

[7 APPENDIX A: SOURCE CODE 47](#_Toc114862240)

[8 APPENDIX B: PROJECT PLAN TIMELINE 48](#_Toc114862241)

[Figure 1 Library import and data loading 32](#_Toc114856865)

[Figure 2 Data cleaning 32](#_Toc114856866)

[Figure 3 Discarded repetitive features 33](#_Toc114856867)

[Figure 4 Final dataset used for predictions 33](#_Toc114856868)

[Figure 5 Final dataset info 34](#_Toc114856869)

[Figure 6 Density population of 5g users based on city 35](#_Toc114856870)

[Figure 7 Correlation heatmap of final dataset features 35](#_Toc114856871)

[Figure 8 Data Normalisation 36](#_Toc114856872)

[Figure 9 Train Test Split 37](#_Toc114856873)

[Figure 10 Naive Bayes Model and performance 38](#_Toc114856874)

[Figure 11 Logistic regression model and performance 38](#_Toc114856875)

[Figure 12 SVM model and performance 39](#_Toc114856876)

[Figure 13 Decision Tree Model and performance 39](#_Toc114856877)

[Figure 14 ANN Model and performance 40](#_Toc114856878)

[Figure 15 5G prediction performance 42](#_Toc114856879)

[Figure 16 Models confusion matrix visualisation 44](#_Toc114856880)

# INTRODUCTION

## Background of the Study

There is constant development in the four primary technologies that will be necessary to navigate the global information and communication technology (ICT) landscape. Fifth-generation (5G) cellular networks, artificial intelligence (AI), information analytics, and the Internet of Things (IoT) are all examples of such technology. Any one of these developments in information and communication technology (ICT) and across all major business verticals that consider telecommunication and information technology services would have a substantial impact in its own right. However, the convergence of these developments is set to create opportunities for electronic messaging, applications, interactive entertainment, and commerce that will significantly improve consumer experiences and also in relevant professional and research areas like medicine and surgery, geology, aviation and space exploration. Due to the fact that they created previously unheard-of advantages and comforts for the functioning of society and the economy, information and communication technologies (ICT) are the most often encountered constituents (Yu et al., 2017).

To avoid confusing AIoT with a specific name, we'll refer to the "Analytics of Things" (AoT) as the analysis of information from IoT networks, processes, computers, and applications for decision-making tasks. The internet of things and the enormous advancement in technology have created a huge demand for mobile traffic which in turn have created new obstacles that the previous generations of mobile networks which are the fourth generation (4g) third generation (3G) and second generation (2G) must face.

Business insight into the Internet of Things is not innate; it requires careful planning. We can't let this go unseen. Software, chipsets, networks, and human-made things like appliances are all potential sources of AIoT market components. Like the human nervous system, an IoT network has both autonomous and cognitive components that intelligently monitor and run nervous endpoints like neural (detector and communication activation) terminals and nevus channels linking the whole system.

A key differentiator is that the IoT technology industry is supported by AI architecture and cognitive positioning in both consolidated and specialized computing. The term AIoT5G was coined by the tech industry to describe the integration of artificial intelligence, the internet of things, and the fifth generation of wireless technology. Integrating these breakthroughs will spark new ideas, leading to expanded use across a wide range of vertical markets and other technologies, such as robots and augmented and virtual reality. Sensors have been linked to the majority of IoT applications. Internet of Things applications in smart cities would monitor things like air quality and traffic congestion. One such sensor must be radio-connected and provide uplink data on a regular basis or immediately when an event occurs.(Condoluci et al., 2016)

Trade in the "Early Market" Although AIoT solutions tend to be all-encompassing, it is expected that industry and business sector convergence will eventually provide more nuanced inter- and cross-industry AIoT solutions. As a result of vastly improved investigation and decision-making procedures, these strategies will center on machine and network activities, as well as the extraction of value from corporate data. As IoT networks develop locally and globally, the volume of unstructured data generated by computers is expected to grow at an exponential rate. Opportunities for artificial intelligence to aid in the framework of data analytics solutions will grow substantially as the amount of data generated by humans and machines continues to rise. Data generated by IoT-enabled equipment may be invaluable for product life cycle management and other customer-facing internal operations.(Mzahm et al., 2013).

## PROBLEM STATEMENT

Every day, more and more people are logging into wireless networks. The global network is responsible for the development of the Internet of Things (IoT), as well as device-to-device (D2D) and machine-to-machine (M2M) communication, cloud-based software, and services. Improved uplink and downlink data speeds, improved Quality of Service (QoS), a more consistent quality of experience (QoE) across all channels, and higher quality streaming material are all still desired by the public. Thus, these standards cause very high levels of mobile network data traffic. This necessitates the development of a novel mobile device to meet these requirements. Security risks are present in big data, just as they are in most modern systems. This means that safeguarding customer or company information is an absolute need. Establishing a reliable and secure network architecture, from servers to endpoints, is a key component of 5G design and architecture.

With this in mind, it is important to note that factors some might interfere with 5G network signals, this raises the question of how we can be able to know whether our infrastructure still gave the required 5G connection even when we are not able to detect this by network speed and other relatively obvious factors? (Luini et al., 2020).

## RESEARCH QUESTION

This research wok is set out to find an answer to the question “Can 5G network usage be predicted using machine learning methodologies without the inclusion of network speed, network latency and device information?”

## AIM

The purpose of this thesis is to examine the interplay between promising new technologies that will revolutionize the ICT industry, such as 5G networks, AI, Big Data analytics, and the Internet of Things (ICT) and also predict 5G usage from non-obvious network factors.

## OBJECTIVES

To achieve the aim of this research work, the following:

1. Acquire relevant dataset pertaining to the scope of this research.
2. Pre-process and noramlise the dataset to be used for the prediction.
3. Train machine learning models on the dataset and perform prediction.
4. Collate and evaluate the results of these trained machine leaning models.
5. Interpret the results of the predictions and provide an answer to the research question.

## PURPOSE OF THE STUDY

The goal of this research is to analyze how different ICT’s have been used to improve cutting-edge technical programs. The primary points of convergence in this research are 5G, AI, and Big Data Analytics.

## SCOPE OF THE STUDY

The scope of this research is to determine potential uses of emerging technologies in 5G networks and provide a prediction whether 5G usage can be determine by generic network information. Modern society relies on a wide range of high-tech applications made possible by advances in areas like Artificial Intelligence (AI) and the Internet of Things (IoT). As part of an accountability system, blockchain technology was a forerunner in documenting the robot's actions throughout an operation and ensuring the accuracy of the models it provided. The application of smart contracts in cryptography is also implied. The storage, retrieval, and processing power needs of Big Data in 5G dispersed applications go much beyond those of conventional databases and data mining methods.

## LIMITATION OF THE STUDY

This study is limited to only machine learning prediction for 5g usage and the literature exploration of Big Data Analytics, the Internet of Things (IoT), and Artificial Intelligence (AI) that are discussed in this thesis, among other technological developments or methods (AI).

## PROJECT STRUCTURE

Chapter 1: This section contains introduction about the research work consisting of sections like research question, aim, objectives, scope and limitations of the study.

Chapter 2: This section contains literature review into 5G technology, its applications and how it encompasses other fields of study such as IoT, Big data etc.

Chapter 3: This section contains the methodology used to achieve the technical aspects which contains tasks like acquiring and loading dataset, ethical, professional, legal and, social issues and mitigation, as well as implementing machine learning methods.

Chapter 4: This section primarily focuses on interpreting and describing the analysis and predictions generated in chapter 3 and also providing relevant visualization.

Chapter 5: This chapter includes summary of the research work, critical reflection, recommendation for further work and research and limitation encountered.

Appendix A: This section contains the source code used for the prediction of 5G network usage.

Appendix B: This section contains the project timeline gantt chart.

## DEFINITION OF TERMS

1. Information Communication Technology (ICT): Information technology (IT) encompasses not only the hardware and software required to display, process, distribute, and manage data for users, but also the centralized communications and convergence of phone lines and WLAN signals.
2. 5G Networks: In 2019, the international rollout of the fifth-generation cellular broadband infrastructure standard (5G) for mobile telecoms providers will take place. 5G is the intended successor to 4G networks, which currently provide access to the vast majority of mobile phones.
3. Artificial Intelligence (AI): Humans and animals may not have the same level of artificial intelligence as computers, but they do show signs of natural intelligence. The field of artificial intelligence (AI) is typically defined as the study of "intelligent agents," which may be thought of as any system that takes in data from its environment and makes adjustments to increase the likelihood that it will succeed in achieving its goals.
4. Big Data Analytics: Research strategies, such as the continuous compilation or other retrieval of data sets that are too vast or difficult to deal with for traditional data analysis, are essential to the field of big data.
5. Internet of Things (IoT): The term "Internet of Things" refers to a global network of items and systems that are interconnected and exchange and communicate data via the Internet through the use of embedded sensors, software, and other technologies.
6. Blockchain: In reality, blockchain technology is defined as a decentralized, distributed directory tracking the provenance of a digital asset. Blockchain is more simply described as a decentralized, distributed leader technology recording the origins of a digital commodity.

# LITERATURE REVIEW

## INTRODUCTION

Big data, the Internet of Things, and the fifth-generation cellular network (5G) are just a few of the new technologies that have transformed our society and made the impossible possible in the last decade. (Wang et al., 2018). There is actually no suggestion that the additional changes to come would be gradual or less detrimental in the future, but there is no denying that ICT has had a huge effect over recent years and has provided various advantages to all of our lives. ICT permeates every facet of our lives, from job to play to social interaction. Since its humble beginnings as a specialized discipline, information and communication technology (ICT) has grown to become an indispensable tool for bridging traditionally separate areas of study and industry. As a result, new areas of expertise and academic disciplines emerge at related interfaces, such as those between manufacturing and IT (Industry 4.0, smart grids), population issues and IT (AAL), automotive engineering and IT (autonomous driving), and medicine, biology, and IT (bioinformatics), to name just a few examples (Technology, 2015). Operators must plan for a network infrastructure that is 5G-ready in order to effectively support the next generation of devices and networks in the coming decade. The Next Generation Mobile Network was one of the earliest platforms to propose a variety of 5G use cases, business models, and infrastructure and design ideas (NGMN). (Ruffini & Member, 2017). Guidelines address some of the forthcoming advancements, but they are technology specific and lack the holistic perspective necessary for the 5G goal. Examples of the positive effects of ICT on growth may be seen in many parts of the world, including those where more access to education, increased agricultural output, improved health care, accountable leadership, and greater economic opportunity have all been achieved. They're crucial to economic and social progress since they allow for long-lasting change to be made. (Ibadan et al., 2013).

## 5G TECHNOLOGY AND APPLICATIONS

It's not always been clear what "5GI" means, and the topic has generated much debate. Current expectations of 5G technologies, including a hyper-connected future and cellular access technology for the next decade, are presented in a study issued by GSMA Intelligence. The first is a ground-breaking technology that combines the best features of previous generations like 2G, 3G, and 4G into a single, improved system that will have a greater impact on users than current technologies like 4G and LTE. Despite the fact that this technology outperforms all others in its field, we shouldn't expect a true "generational shift" to occur. Imagined here is a radio access system for the twenty-first century (Paudel & Bhattarai, 2018).

A cutting-edge technology, 5G will allow for the usage of virtually any desired application by connecting the world's current communication networks into a single unified system. The 5G terminals will be equipped with an upgradable multimode and cognitive radio. The program dictates the radio modulation methods that will be used. The next generation of mobile networks, 5G, will prioritize the development of consumer terminals that can simultaneously connect to several data streams from different technologies, such as the exclusive Wi-Fi technologies that are used today (Hossain, 2013).

Mobile networking improvements have spawned a variety of various technologies to better the quality of life of end-users, including intelligent mobile, remote banking, social networking, and healthcare. The Internet infrastructure that mobile devices are a part of has evolved rapidly over the last several decades (Ding & Janssen, 2018). We highlight five application categories that can take use of 5G's converged network and the cyber-physical infrastructure that will be in place over the course of the next decade.

### Smart Mobility:

There are a wide variety of applications for 5G mobility, from traditional highway planning to the development of autonomous driving systems (connected cars) and ecosystems of intelligent ridesharing. Balanced traffic, optimum routes, fewer accidents, less energy used, less expenses, and less pollution are only some of the advantages of intellectual mobility. These application developers have high hopes for 5G because of its potential for ubiquitous connectivity, low latency, high performance and dependability in transmission, security, and low battery consumption. The growth of cities is, in essence, due to planning that prioritizes vehicular traffic. This method resulted in a polycentric metropolis that was separate, dispersed, and isolated (Arce-ruiz, 2016).

### Smart Energy:

New energy-saving services for homes and businesses, energy markets, smart charging stations for electric cars, and monitoring and control of power plants are just a few examples. It is anticipated that the use of smart technology would enhance the efficiency, dependability, and smart distribution of green energy systems. 5G must meet stringent requirements for secure connections, private data transmission, and other critical features. The term "Smart Energy System" is sometimes used interchangeably with "Smart Grid," especially in the fields of control engineering and administration, or to imply that the applicability of this control technology would extend beyond the energy industry.(Lund et al., 2017).

### Smart Health:

Since health-related applications that focus on mobile users' wellbeing are becoming more widespread. Applications in this area included mobile condition monitoring and diagnostics, environmental quality assessment, and the creation of intelligent wearables. Applications More information from wearable tracker sensors is one way in which smart wellness might improve healthcare and medicine. One emerging use case in this area is augmented and virtual reality-enabled surgical procedures, which have specific needs in terms of low latency and high bandwidth on top of the more general needs of 5G in terms of low capacity, security, and data privacy. The term "smart health care system" (sometimes "smart healthcare") refers to a healthcare infrastructure built on top of the Internet of Things (IoT). The acronym for autonomous, enlightened, scalable, resource-rich, and technological (SMART). Instead of just being an intelligent health care system, this smart healthcare system represents a revolutionary shift in the medical paradigm for today's digital natives. The monitoring and monitoring of patients necessary for healthcare services in the 21st century is simplified by the Internet of Things (IoT) and cloud-based healthcare technology (Akpojaro et al., 2019).

### Industrial Applications:

Next-generation cyber-physical services include applications like Industry IoT that focus on development, M2 M connection, 3D printing, and AI-supported architecture (Internet of Things). These industrial applications will have far-reaching effects on society at large, not only in manufacturing and processing facilities. Extremely low latency, massive deployment, security, and privacy support are some of the most important features of 5G networks. Smart "Things" help consumers learn about their surroundings and save costs by automating routine tasks. The user of an Internet of Things device is a hybrid machine/human. The majority of consumer IoT connections may be classified as either machine-to-user or client-server exchanges (Sisinni & Saifullah, 2018).

### Consumer Applications:

A large number of user-created apps show the innovative possibilities of 5G mobile business and technology. When compared to the traditional mobile apps we are used to using on our smartphones and tablets, new applications such as mobile streaming, financial technology, overview games (like Pokemon Go), augmented reality and virtual reality platforms for unorthodox aircraft, and holography technologies like HoloLens are revolutionary. Computer programs are also making advantage of these applications. To facilitate wide communication, high bandwidth, low latency, low power depth, dependability, and stability, all of these cutting-edge networks need 5G. IoT consumer market growth and consumer response to IoT marketing operations are both expected to be affected by how quickly and enthusiastically this new technology is adopted by its users (Dergisi, 2019).

## INTERNET OF THINGS AND ARTIFICIAL INTELLIGENCE REVOLUTION

### Sustainable development’s areas for AI and IoT.

Finance, economics, marketing, management, R&D, transportation, energy, environment, ecology, and healthcare are just a few of the crucial fields that may be used to gauge sustainable development. Findings such as the number of works dedicated to the most popular list of topics and the most commonly used engineering-controlled terms and keywords map the AI and IoT implementation fields of sustainable development, including the power market, smart cities, architectural architecture, and smart houses, mobility, monitoring systems, traffic congestion, and smart roads, business operations, manufacturing and industrial control, product design, and lifec. The utilization of artificial intelligence (AI) methodologies and the reshaping of the Internet of Things (IoT) into novel approaches are on the rise in decision support systems. Many other types of IoT exist, such as IIoT, EIoT, GreenIoT, and SocialIoT. Computer science, mathematics, psychology, linguistics, philosophy, neurology, and other specialized fields like artificial psychology all come together in the field of artificial intelligence (AI) (Paper, 2016). Several of these relate to the use of AI-enabled technological advancements in nations with varying cultural norms and available resources. In order to develop in areas like artificial intelligence (AI), research, and product creation, it is necessary to house enormous processing centers. (Communications, 2020).

### AI and IOT Technical Issues

One of the most important technological components of sustainability objectives is likely to be AI and the Internet of Things if we can unlock their full potential. As intelligent, knowing, and autonomous technologies, they bring about a new era of process realization; but, they are not without their drawbacks. The advantages of Artificial Intelligence (AI) extend far beyond the aforementioned areas, and include, but are not limited to: strengthening medical practices by identifying nuanced trends; decreasing the amount of time needed for analysis; and fostering more proactive decision-making on the part of decision-makers. strength in dealing with noisy data and a superior grasp of time series analysis. One of the problems with AI is that it may make predictions without first fully comprehending the nature of the relationships between the variables. In econometrics, the non-stationary nature of time series analysis leads to the spurious regression issue.

Increase CPU, memory use, and power consumption; Improve automation, asset utilization, employee competitiveness, supply chain administration, customer service, and cost efficiency; Shorten the time it takes to mitigate marketing risks. Some of the advantages of the Internet of Things include improved complete responsibility throughout the value chain for knowledgeable environmental management (IoT). One potential drawback is that it lacks completely specified IoT architectural standards; there is a dearth of evidence for inexpensive IoT-based services; and power requirements for an appropriate system discovery protocol remain elusive. (Jabłońska & Zajdel, 2018).

## BLOCKCHAIN, 5G ROBOTICS AND ARTIFICIAL INTELLIGENCE

5G might have far-reaching consequences beyond just networking. The present trend towards more hierarchical structures in Computation is shared by the fields of joint networking, robotics, artificial intelligence (AI), the Internet of Things (IoT), and blockchain. Together, this and the concept of network slicing (and its related term, edge computing) are the bedrock of 5G and beyond in terms of facilitating low-latency and network load management (Pe, 2020).

Satoshi Nakamoto pioneered blockchain technology for use in Bitcoin. Blockchain's notoriety, worth, and utility have all risen as of late, and not only because of its ability to transport crypto-currency but also because of its ability to facilitate the decentralized and stable inclusion of a wide variety of applications on a shared network. Smart contracts and other enhancements to blockchain technology introduced in 2013 with the release of Ethereum4 have made the underlying blockchain protocol more flexible and useful across a wide range of industries and academic disciplines. Energy use and the speed at which a block must be authenticated are two more crucial factors that aren't being addressed by these projects. Each year, thousands of new ideas and services are offered that use blockchain technology to address these problems, but there is no one solution that works for every scenario. This lack of uniformity inspires more innovation. Alignment with blockchain is an area where robots might benefit from more strategies that have been shown to successfully integrate all relevant technology. Many of the problems that now exist with robotic systems might be solved by using blockchain technology. Since many networks have trust and information security vulnerabilities, blockchain can provide effective peer-to-peer connection over a secure network, which is the first obstacle to solve. The capacity to make decentralized choices is another advantage of blockchain's introduction, as it will ensure that all participants in a network have the same worldview.

Show how robotic events may be stored in a blockchain on the level of Bruno Degardin and Luis A. Alexandre. This idea allows for the development of intelligent contracts that make use of data collected by autonomous robots in the wild (potentially from different manufacturers) and triggers based on the contracts themselves, all of which are processed and checked by the blockchain. The time it takes to do things like replenish the screws for a robot that shows on the blockchain that it needs more screwdrivers before work can begin will be reduced, which means the factory as a whole will be able to crank out more stuff. Following on from previous efforts, the authors propose using Tezos's technology to create a blockchain for the registration of robotic events, which would then be protected more effectively thanks to Tezos's use of structural verification. Ongoing work like this will benefit smart contracts on the blockchain by ensuring that the appropriate artificial intelligence (AI) programming is used (i.e. to do exactly what their specification defines). The idea also includes modifying the blockchain so the device can process significantly more events per second and manage a far larger number of communicating robots than is currently possible with the existing standard.

To put it simply, Aitheon is a blockchain-powered platform built on the ERC-20 standard. One of its goals is to create a comprehensive forum that helps developers save time on tasks like document organization. Their aim is to provide a unified interface for AI and robotics-based enterprise processes. Their own platform was comprised of five separate pieces. One is an AI component designed to automatically gather and automate data on commonly used procedures. Digibots, the second module, is similar to the first but focuses on the automation of programming activities such backend solutions and data-driven difficulties. The third component, Mechbots, enables businesses to use robotic automation to boost efficiency and effectiveness. The remaining two modules, Aitheon Technicians and Pilots, consist of human professionals who perform non-automated tasks and oversee the robots. Aitheon provides a discussion board for the use of AI and robotics to automate tasks over extended periods of time.

The benefits of combining blockchain technology with robotics, and notably swarm robotics and robotics, are highlighted in the research of Eduardo Castell'o Ferrer. Scalability at a rapid rate and resistance to breakdowns are two of the many advantages of using robotic swarms. These advantages result from the dispersion of these swarms' members. This field will also grow in the manufacturing sector, helping firms become more efficient via examples like AmazonRobotics' deployment of a horde of robots to manage its warehouses. In most robotic swarms, each robot only has access to information about itself and its immediate surroundings; however, by incorporating blockchain technology into such systems, global data may be made available to the robots, which can be put to use in a number of ways. Because of the global nature of information stored in blockchains, the system as a whole will be able to adapt its demeanor more quickly to meet the demands of individual robots. A controller robot may do this by querying the blockchain for the current state of the device and then making the appropriate commits. These updates, together with the system's global data, will make operation and upkeep easier and more efficient.

To address data security concerns raised by robots' usage of personal data in human-robot interactions, the authors of RoboChain, a computational mechanism for safe sharing of sensitive data between robots, suggested a method. This approach employs MIT's OPAL to provide an extra degree of security and ensure that the information robots depict remains hidden. In order to verify the accuracy of the supplied models and keep track of the robot's actions throughout an operation, this accountability system incorporates blockchain technology. Additionally, the authors advocate for a consensus procedure wherein each node uses block chain’s smart contract technology to vote on which model to adopt (Lopes, 2019).

## BIG DATA IN 5G DISTRIBUTED APPLICATION

Big Data are data sets so large and growing so quickly that traditional tools and techniques can't keep up. The term "Big Data" is often used to describe datasets that exceed the capacity of conventional databases and data mining methods to store, retrieve, and process (Riahi & Riahi, 2018). Data amount (volume), semi-structured and unstructured data kinds (variation), the rate at which data is updated or created (velocity), data relevance (value), and knowledge content (confidence, reputation, and honesty) are the five dimensions (so-called five 'five Vs'). Diverging from traditional methods of data analysis (veracity). All the use scenarios discussed are centered on collecting information from the vast quantity of heterogeneous data created by connected devices to aid in the choice, and this suggests that Big Data tactics may play a key role by the time 5G is utilized commercially. Specific criteria for their application growth are outlined, along with a discussion of some potential problems to resolve positioning issues and solutions based on semantics, and it is shown how these can all benefit from the advent of new 5G network capabilities in tandem with Big Data technologies. Such a scenario is indicative of a future setting in which we will use synergies between blockchain and other revolutionary technologies and paradigms such as the Internet of Things (IoT), Big Data, and next-generation networking. The academic world and service providers will benefit from this (business stakeholders). It does two things: it sheds light on the state of 5G R&D, and it investigates how the results of that research may be used to better future services that would meet the needs of consumers (Nejkovic et al., 2019).

### Key Technology and Business Drivers

Big Data Analytics has the ability to integrate well with other advancements in 5G technology, like SDN/NFV and MEC (Multi-Access Edge Computing). Among the most influential trends and commercial imperatives shaping the adoption of data analytics in 5G networks are the following:

### Mobile Cloud/Edge Computing

The Advanced and Intelligent Setting includes the Mobile Cloud Sensing, Big Data, and 5 G Network. Mission-critical applications like public safety and healthcare will need analytics in real time. The future 5G air interface promises latency advancements and slicing-based traffic prioritization, both of which pave the way for the widespread adoption of mission-critical edge analytics and haptic internet applications. All data travels easily from the cloud to a plethora of end points and vice versa; with 5G, it is feasible to not only detect and assess at the edge, but also to activate actuators within a fraction of a second to start response behavior (Ahmed et al., 2018).

#### IoT over 5G (Industrial IoT)

The massive amounts of data generated by the Internet of Things have the potential to dramatically improve the efficiency and profitability of many different industries, from manufacturing to healthcare to the development of smart communities. One fleet management company, for example, found it could cut the cost of maintaining its 180,000-truck fleet from 15 cents per mile to only 3 cents. Industrial IoT has its own set of communication requirements, such as high dependability, low latency, adaptability, and security. The 5G mobile infrastructure naturally provides them, therefore it might be used to serve Industrial IoT applications (IIoT) (Varga et al., 2020).

#### Data Monetization

Telco mostly employed data up till 4G/LTE to improve the quality of operations and the customer experience. Nonetheless, they will explore other ways to monetize their 5G network services by integrating IoT and AI, such as with intelligent corporate application solutions. Telco's future success hinges on the importance of the device and network intelligence layers it offers to businesses, in addition to the monetization of data. The ever-increasing data quantities accessible motivate businesses to find ways to profit from them. Today, information is increasingly valued as a trade-able asset that can be purchased and sold digitally. It's true that in the current business world, organizations may gain significant competitive advantage by properly using and monetizing their data. Recent advances in technology in the realm of big data have sparked a surge of interest in the study and practice of data monetization (Baecker et al., 2020).

#### Cognitive Analytics

Instead of the usual useful analytics with basic BI reporting, 5G analytics may morph into computer/deep learning. Analytics in 5G have the potential to grow to the point where they can learn from their surroundings, predict what will happen next, prescribe the next appropriate move or moves, learn from past behavioral trends to make the next effective choice, and automate the next action for fully autonomous use cases. Decisions will be guided more and more by analytics-generated insights, and the speed of 5G will allow for the collection and analysis of even more data, resulting in cognitive intelligence applications more quickly than ever before. The two learning systems are cognitive computers. They bring analytics and management of data closer to storage, memory, switching, and processing by using data-centric architectures and embedded analytics. They process massive amounts of data in a way that is neither linear nor predictable (Noor, 2015).

**2.6 RELATED WORKS**

|  |  |  |  |
| --- | --- | --- | --- |
| S/N | Author | Year | Work Focus |
| 1. | Cengiz | 2018 | Describe some of the key breakthroughs that enable 5G networks to provide such significant improvements in bandwidth, data rate, network overhead signals, and energy-spectral performance. However, there are typically significant difficulties associated with these breakthroughs when attempting to implement them in a 5G network. |
| 2. | Varga et al. | 2020 | Draws attention to the state of the art analysis and policy concerns concerning 5G-enabled IoT Industrial based on the first criteria and the potential of both industries. |
| 3 | Condoluci et al. | 2016 | Describe that the Internet of Things will soon be seeing a significant change from a sensor driven platform to a model augmented by actuators, drones and robots (IoT) (IoT). The uplink is responsible for relaying sensed data in real time, as well as perceiving and controlling data, to the edge cloud. Given that all these applications are important of the missions, the cellular connection family is the most suited, because of the shortage of licensed spectrum able to secure the communication service. |
| 4. | Lin et al.. | 2016 | Technology like 5G mobile broadband, the IoT, big data analytics, cloud networking, and software-defined networking (SDN) have advanced at a dizzying pace, fostering deep dependencies and fueling fast development. IoT networks that generate low-volume, high-speed data, for example, will need 5G's high data rate and low latency to transmit this information quickly and cheaply. SDN must be elastically coupled to the network and the Cloud must analyze and store this data for efficient data transfer. |
| 5. | Riahi & Riahi | 2018 | In his research, the author hopes to explain Big Data, its underlying ideas and concepts, the challenges it presents in practice, and the role Big Data Analytics plays in this field. |
| 6. | Dergisi | 2019 | One of the most influential forces in shaping the 21st century so far has been the Internet of Things. The term "Internet of Things" (IoT) refers to a network of permanently connected, individually identifiable items. |
| 7. | Ruffini & Member | 2017 | In this article, we will explore a fresh perspective on network integration by proposing a multi-dimensional, end-to-end strategy for network design to support the arrival of 5G services in the future. We also offered a tutorial on the infrastructure, frameworks, and concepts around network convergence, and we highlighted a wide variety of high-level notions that we believe are of fundamental importance in the 5G vision. |
| 8. | Yu et al. | 2017 | Users, not network operators, give the perspective for evaluating the performance of the emerging 5G mobile networks. Telecom companies are hard at work constructing 5G networks, with a focus on studying the future of mobile services. |
| 9. | Sisinni &Saifullah | 2018 | In this piece, we provide examples to demonstrate the ideas behind the Internet of Things, Industrial Internet of Things for manufacturing, and Industry 4.0. We provide examples of the potential outcomes of this paradigm shift and the challenges it faces in becoming a reality. In particular, we examine the difficulties posed by issues like power consumption, time lag between data collection and analysis, compatibility issues between different systems, and the need of privacy and security. |
| 10. | Jabłońska & Zajdel | 2018 | By lowering the unsustainability that disrupts economies, ecosystems, and natural resources, AI and the IoT hold great promise for the future of sustainable development. Due to the relative originality of the idea and the lack of literature on its role in attaining goals in terms of sustainable development, there is need for future study into the rapid multi-disciplinary growth of AI and IoT. |
| 11. | Ghosh et al | 2018 | Data science and artificial intelligence (AI) studies have attempted to address this problem. So, the combination of the Internet of Things with artificial intelligence will lead to significant progress. This isn't about just decreasing costs, using innovative goods, minimizing manual labor, or following the latest trend. More than that, this improves people's quality of life. |
| 12. | Singh et al., | 2016 | The gadgets' ability to talk to one another depends heavily on their established social norms and dynamics. The new 5G network also seeks to interconnect all network technologies through a unified system, including but not limited to the Internet of Vehicles (IoV), Big Data, Mobile Cloud Computing, the Internet of Things (IoT), and Device-to-Device communications (D2D). |
| 13 | Dergisi, | 2019 | The Internet of Things Customer Perspective in the Technology Adoption Paradigm (TAM). Perceived ease of use and perceived utility were added into TAM as significant factors of a potential consumer's behavioral willingness to utilize a contemporary technology. The data has been evaluated using structural equation modeling (SEM) |

# METHODOLOGY

## INTRODUCTION

In this chapter, we look at how 5G popularity projection come together, as well as the processes involved, datasets, pre-processing techniques, and machine learning algorithms that are used for the prediction of the usage of 5g network. In this step, we establish the criteria for success and the means by which information will be gathered. Algorithms for analyzing data was established, and metrics for calculating and ranking the model performances were established.

## LEGAL, ETHICAL, PROFESSIONAL AND SOCIAL ISSUES CONSIDERATION

Before proceeding to perform any form analysis, the researcher assessed the dataset with scrutiny to verify if any form of issue could arise as a result of this research work or the data being implemented for the predictions in this work and to ensure measures to mitigate any of such issues.

### Ethical Issues and Mitigation

Using the Ethical OS Toolkit as a basis of valuation, I believe in terms of ethical concerns, the scope of this project in its entirety falls within the Risk Zone 4: Machine Ethics & Algorithmic Biases. This is because the dataset that will be used in this research work contains sensitive attributes which are Age and Sex and these two features if left unattended has the potential to influence the entire result and leave room for bias or segregation which is outrightly unethical from the trained model and I, the author of the models.

To mitigate this issue, I removed the age and sex columns from the dataset before beginning my analysis because ideally, I do not need age and sex to have an influence in the prediction of 5g network from user activity.

Furthermore, based on the university’s rule for the application of ethics approval when making use of data relating to human information or expression of thought either by survey or online download of human related data, I do not require an ethics approval from the university because my research and dataset in no way contains personal human data. The only attributes in the original data that has human connections are user\_id which is encoded, age and sex which were not eventually used in this analysis and report.

### Legal Issues and Mitigation

Legal issues could arise as a result of using making use of an intellectual property without proper acknowledgment of the author. In an attempt to mitigate this issue, to the best of my knowledge, there are no legal issues that could arise as a result of this research work as every source of material ranging from articles, journals, books, web pages, and source codes used in the writing of this project was duly cited and authors were acknowledged. There is no form of intellectual theft in this project.

### Professional Issues and Mitigation

BSC code of conduct serves as a guideline for recognising and mitigating professional issues. This research work was done in accordance to the code of conduct to mitigate and prevent professional issues from arising as a result of this research work.

### Social Issues and Mitigation

Generally, social issues arises from situations that collectively affect a group of persons who share a common factor. This factor could be gender, age group, location, race or even the kind of car they drive. Social issues mostly shares the same ideology with ethical issues, key difference is that it relatively affects people on a much larger scale. However, this project work does pose any social issues and to mitigate any form of unforeseen issues, attributes that could pose issues like discrimination and segregation are kept at the barest minimal.

## TECHNICAL REQUIREMENT

In order for the analysis of the dataset for the prediction of 5g usage to occur, technical analytical tools, languages and libraries needs to be in place. Below are the platforms on which the machine learning prediction was done.

### Python Language

Python programming language is basically the most common language in the field of machine learning, artificial intelligence, big data and, data science and analytics. The researcher chose to make use of this language because of how well it interacts with data frames, its support for multiple platforms, low latency in computation and finally because of the abundance of libraries for relatively any analytical or preprocessing or visualisation tasks. Moreover, Python programming language is a high level language which makes it very easy to write, read, and, understand. It also has a very faster execution time compared to other readable high level languages.

### Google Colab Notebook

This is a cloud based notebook environment owned and hosted by Google Inc. for free in order to enable users and execute python codes on the cloud. I chose to use this platform because it allows for ubiquitous access to the codes used in this analysis, as well as allowing smaller computational power required to run complex machine learning algorithms like the artificial neural network (ANN). Furthermore, this platform provides all the relevant frameworks and python libraries that might be required withing having to worry about any installation on a local computer. The only requirement to use this platform is a reliable internet connection. Therefore, all the codes used to create the predictions generated as a result of this research work was executed on the Google Colab platform.

## DATASET

The dataset that was used in training and testing the models used in this research work is obtained from Kaggle.com ([5g-user-prediction | Kaggle](https://www.kaggle.com/datasets/liukunxin/dataset?select=train.csv)). From the repository, there are three datasets which are “train\_data.csv”, “test\_data.csv”, and “sample.csv”. For this research purpose, I only made use of the “train\_data.csv” because it is far more detailed and comprehensive compared to the other two datasets. The data appears to be already cleaned with clearly defined features and non-empty cells. However, the dataset is enormously ambiguous in size, containing 60 columns and 70,000 rows and consuming a storage space of about 250Megabytes. To reduce ambiguity, redundancy and excess use of computational power, I limit the analysis and predictions of this research project to the first one thousand (1000) rows of the dataset before loading the data for analysis.

## DATA PREPARATION

This stage of the research work comprises of the technical and analytical tasks that was performed on the dataset in order to achieve the desired predictions.

### Importing Libraries and Dataset

Before a dataset file can be imported into a notebook environment, the required libraries needed to read the file and structure it into a dataset variable needs to be imported. This section contains the loading of the most basic python libraries to read the dataset file, structure it as a data frame, assign the tata frame to a given variable.Graphical user interface, text, application, table

Description automatically generated

Figure 1 Library import and data loading

### Cleaning of Dataset

It should be noted that before loading the data, I have used a third party application to reduce the number of rows to exactly one thousand. I couldn’t perform this operation on python because of I was encountering problems uploading the dataset due to the gigantic file size. Moreover, despite the reduction of rows, the dataset still needs some cleaning and preprocessing.

Graphical user interface, text, application

Description automatically generated

Figure 2 Data cleaning

As shown above, the dataset information shows that there are no empty cells but also shows that some file types are not figures and would not be used in this prediction. It also reveals that some columns contains values that are repetitive and utterly useless and would only cause more model complexity due to higher dimensionality and increase variance in the prediction. Some of those columns are identified and deleted from the data frame. The deleted columns are shown below.

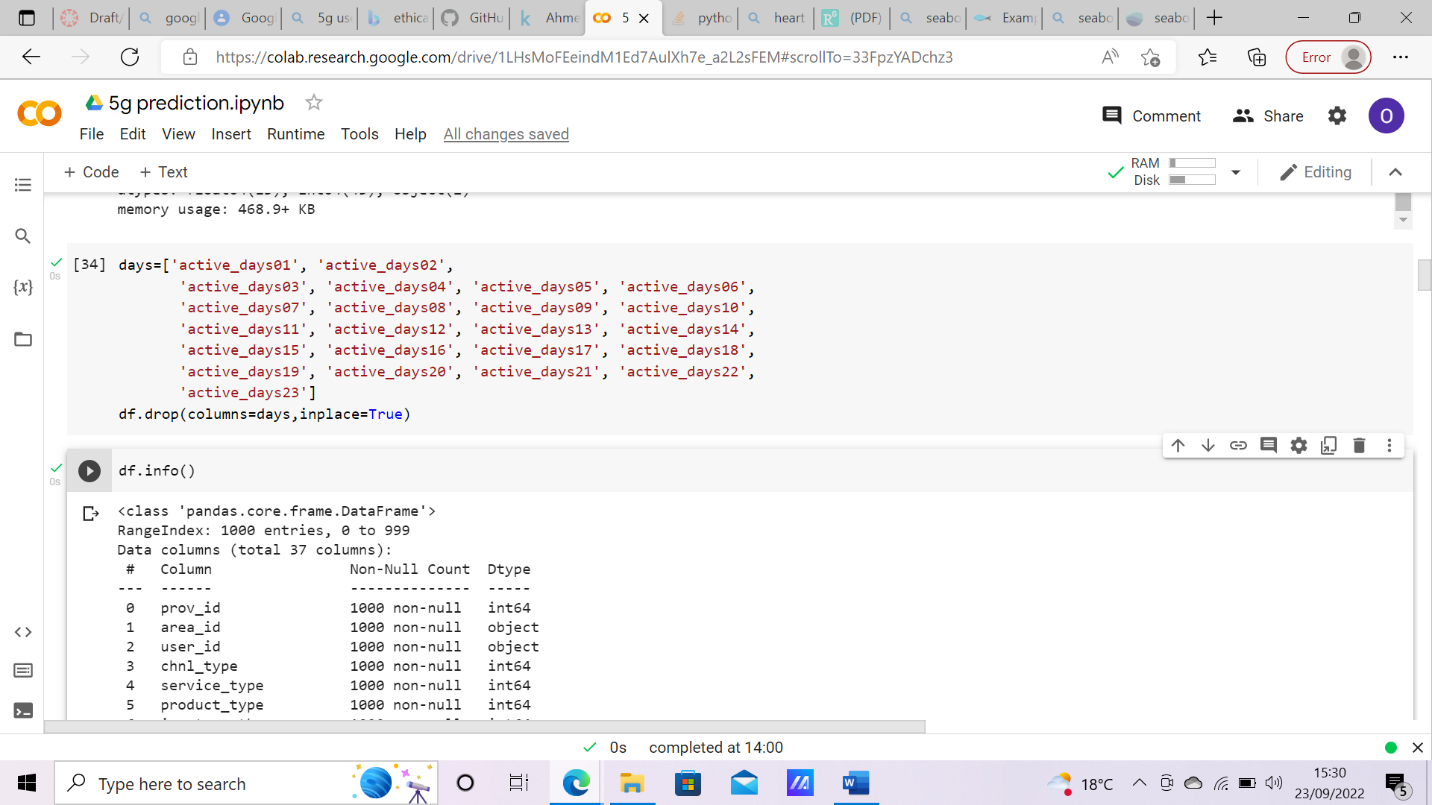


Figure 3 Discarded repetitive features

After deleting repetitive columns, the researcher identifies columns that are pertinent to the cause of the 5g network usage prediction and dropped every other columns that do not fall in this category from the dataset. This procedure is shown below.

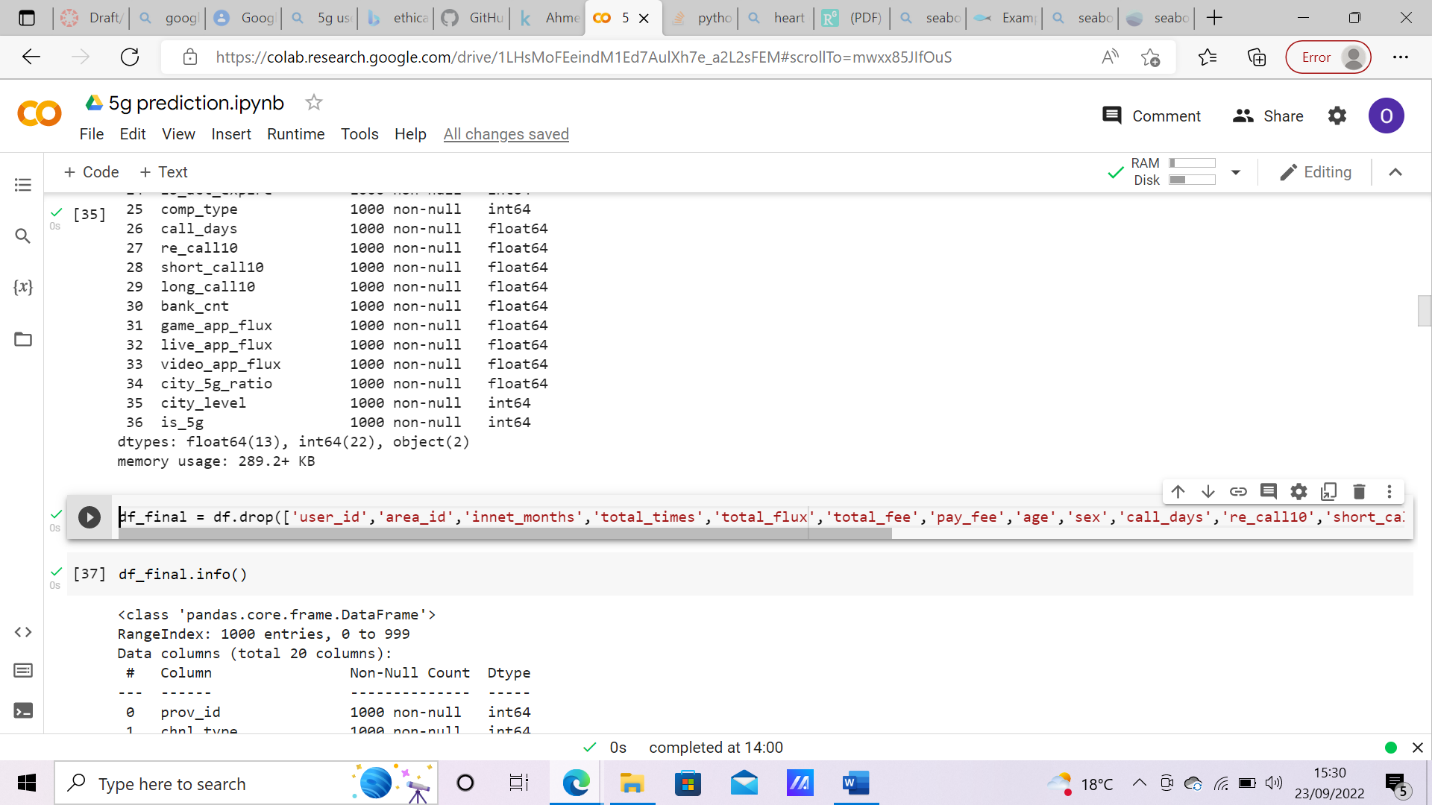


Figure 4 Final dataset used for predictions

At this stage, the final data frame that will be used for prediction has been cleaned from all irrelevant features, and the researcher is satisfied with the features to be used for prediction. These features are shown below:

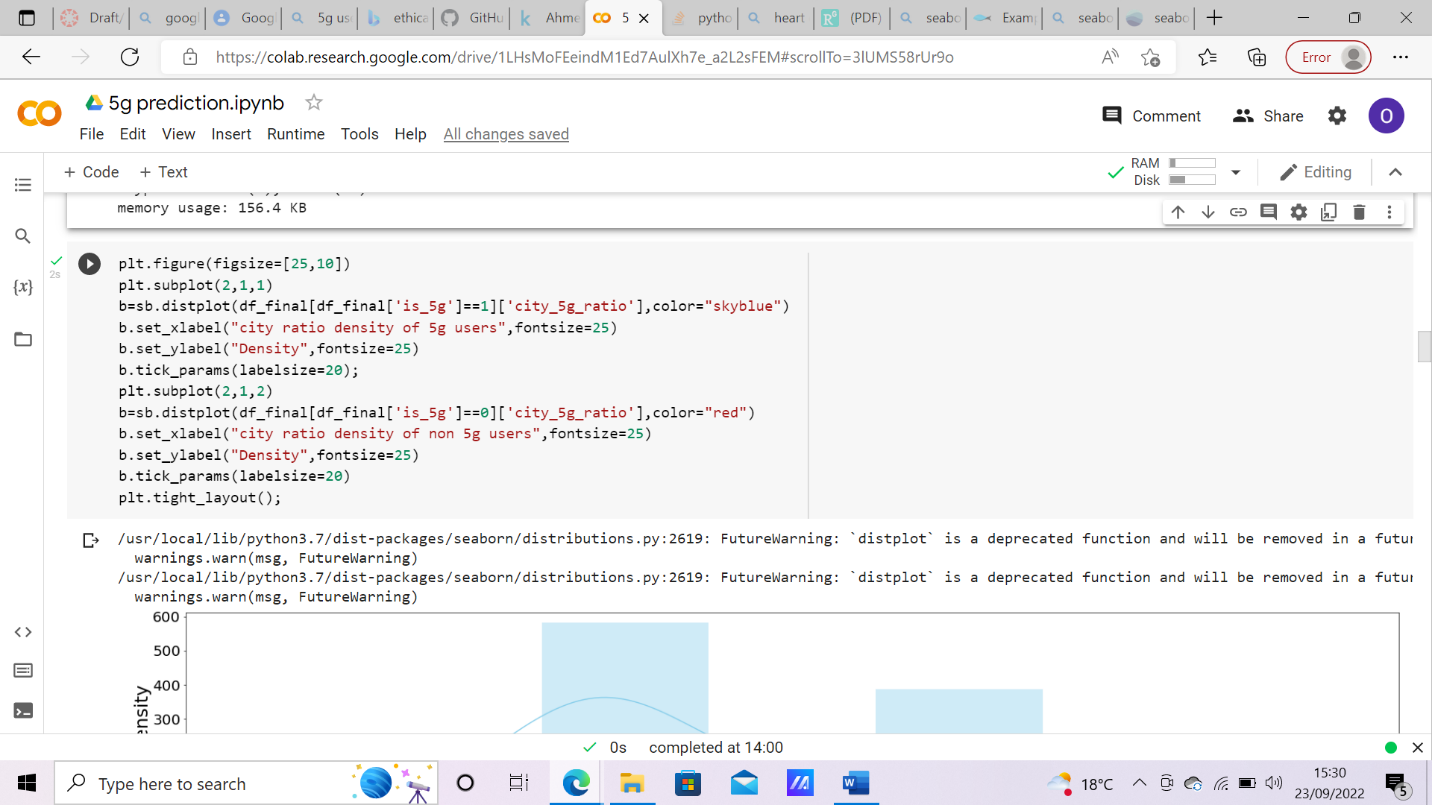
Graphical user interface, text, application

Description automatically generated

Figure 5 Final dataset info

### Pre-processing of Dataset

Before pre-processing, it is important to have a visual overview of how the dataset is, this lead to visualising the density ratio of cities and 5g and city density ratio and non 5g users as well as the heatmap of how the features correlates interacts with each other of which is shown below:



Chart, histogram

Description automatically generated

Figure 6 Density population of 5g users based on city

**A picture containing qr code

Description automatically generated**

Figure 7 Correlation heatmap of final dataset features

After visualising the dataset as a whole, the next task that was performed was distinguishing and separating the target feature from the main dataset and assigning it to a new variable as this would be our prediction evaluation.

### Normalisation of Data

It is not uncommon in data analysis that a dataset would contain enormous variance and/or bias between them, whenever prediction or analysis is done without correcting this error, it is mostly going to give false results there would be bias to favour features with larger numerical values as opposed to the ones with smaller numerical value even if they both have equal level of importance. Data normalization helps to curb this problem by making the average mean in the dataset equal zero and all features are balanced to have and have a fair amount of feature importance irrespective of how large or small are the numerical values of its content. The normalization process is shown below:

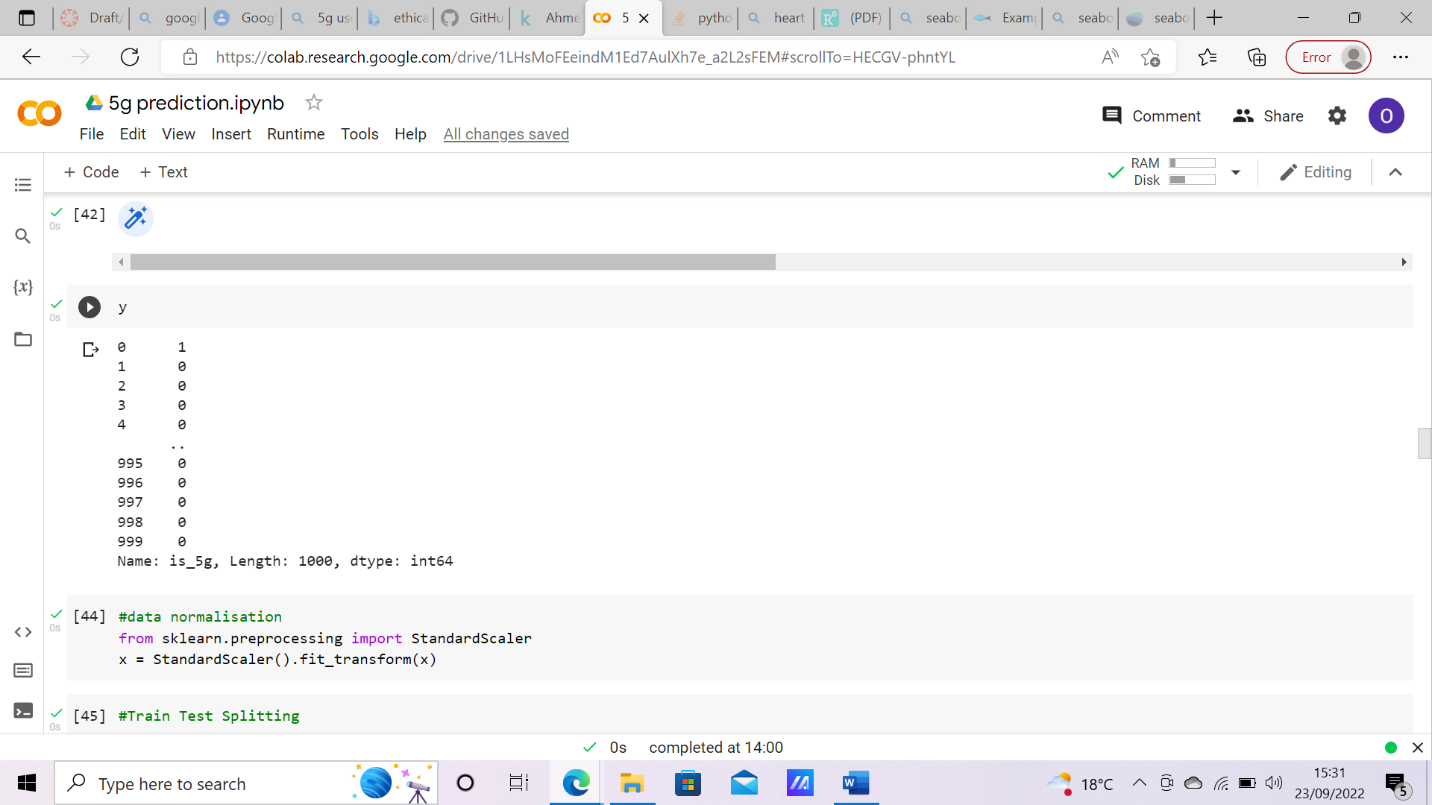


Figure 8 Data Normalisation

### Training and Testing Split

This process involves preforming a separation on the dataset to use a certain percentage of the dataset as training for the models and preserve the rest as test data on which the actual prediction accuracy will be generated. For this research project, I decided to use test\_size=0.2, which translates to 80% of the dataset to be used as test data and 20% to be used as test data. There is no validation split in this research project. Also, I used attribute random\_state = 40 to randomize the data selection and shuffle 40 times. The code used to implement this feature is shown below.

Graphical user interface, text, application, email

Description automatically generated

Figure 9 Train Test Split

## MODEL TRAINING AND PREDICTION

This stage is where the machine learning models that are used for the prediction of 5g usage are being trained on the test data and are being executed. I performed training on five machine learning algorithms which are Naïve Bayes, Logistic Regression, Support Vector Machine(SVM), Decision Tree and, Artificial Neural Network and then tested these trained models on the test data to predict the usage of 5g Networks. Below is the diagrammatic representation of the codes to train these models, performance metrics and confusion matrix.

### Naïve Bayes

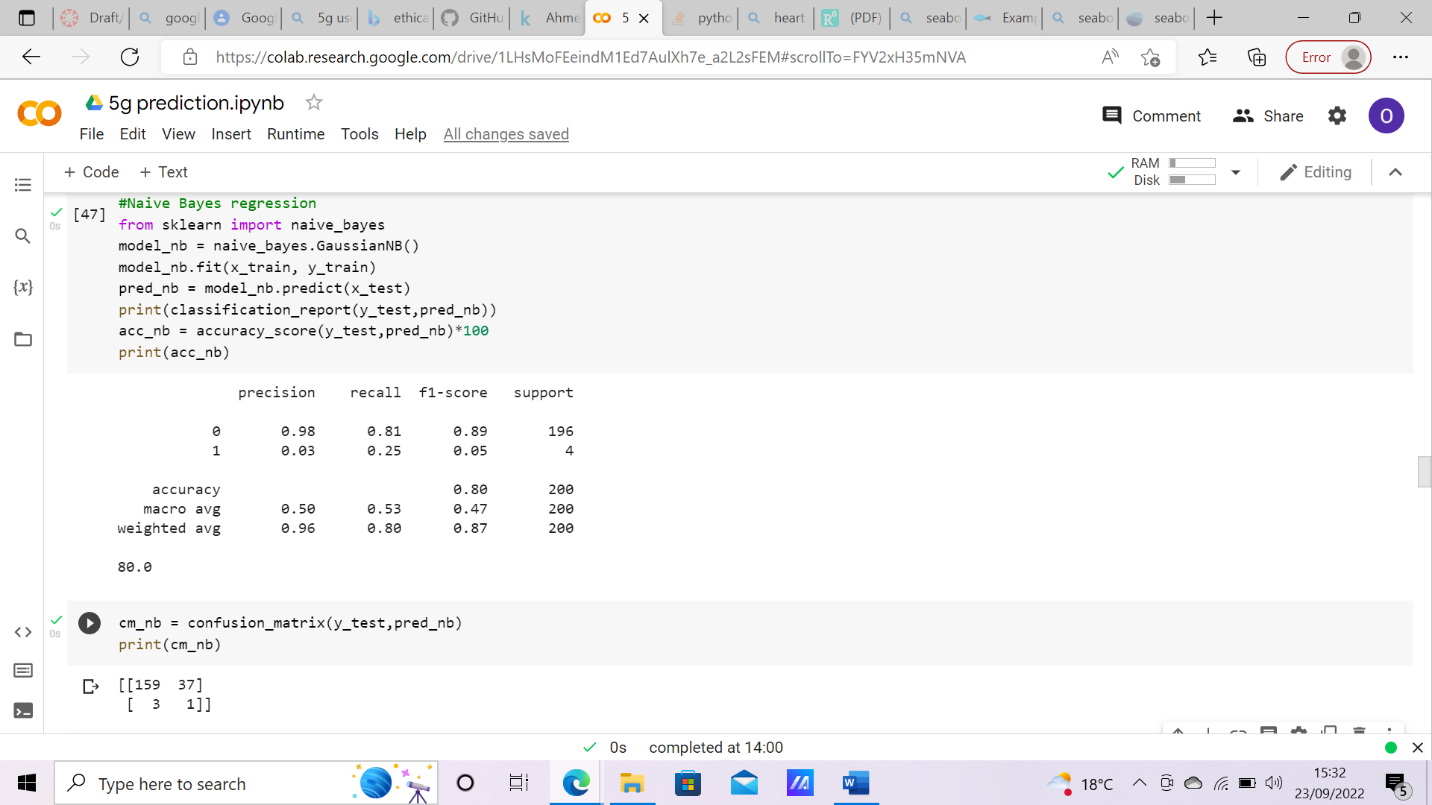


Figure 10 Naive Bayes Model and performance

### Logistic Regression

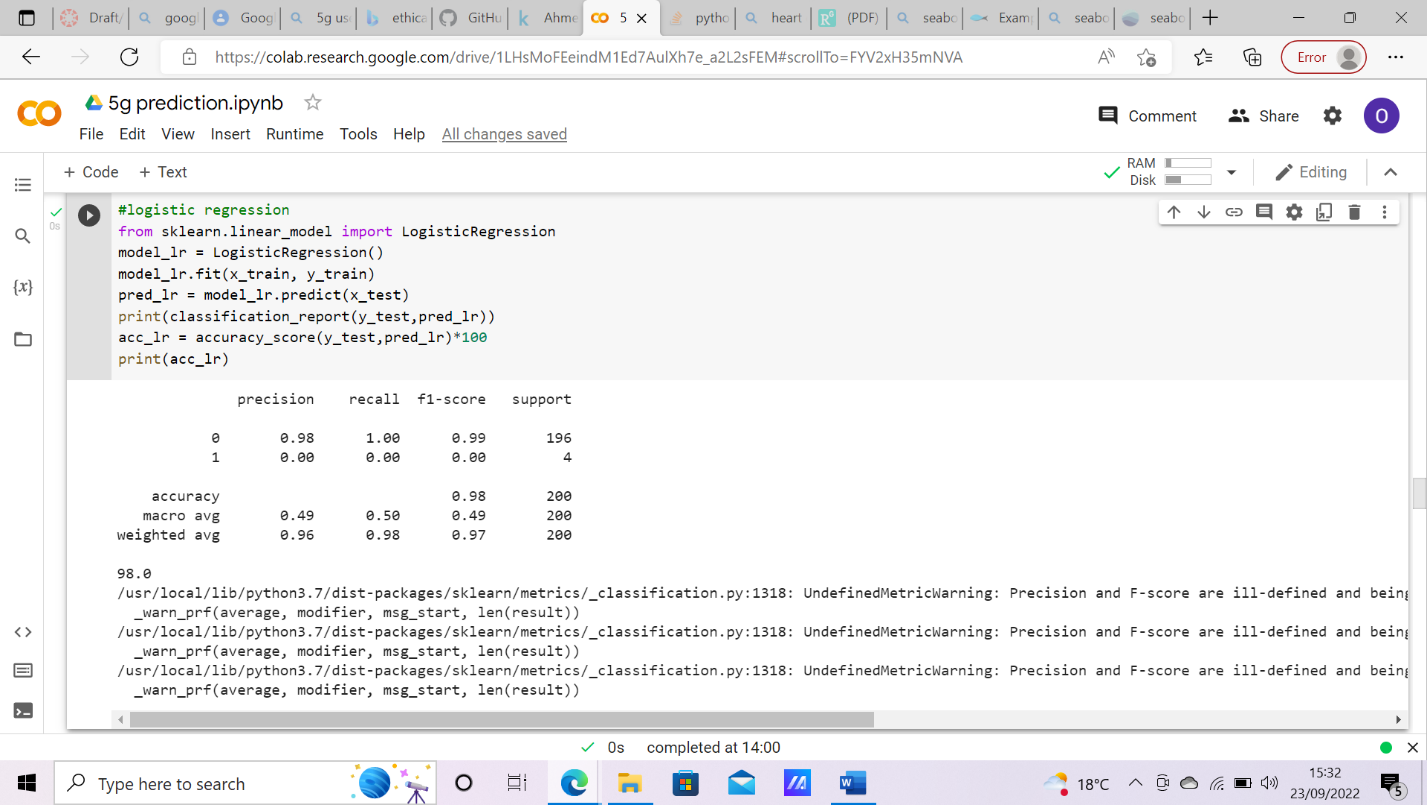


Figure 11 Logistic regression model and performance

Graphical user interface, text, application, email

Description automatically generated

### Support Vector Machine (SVM)

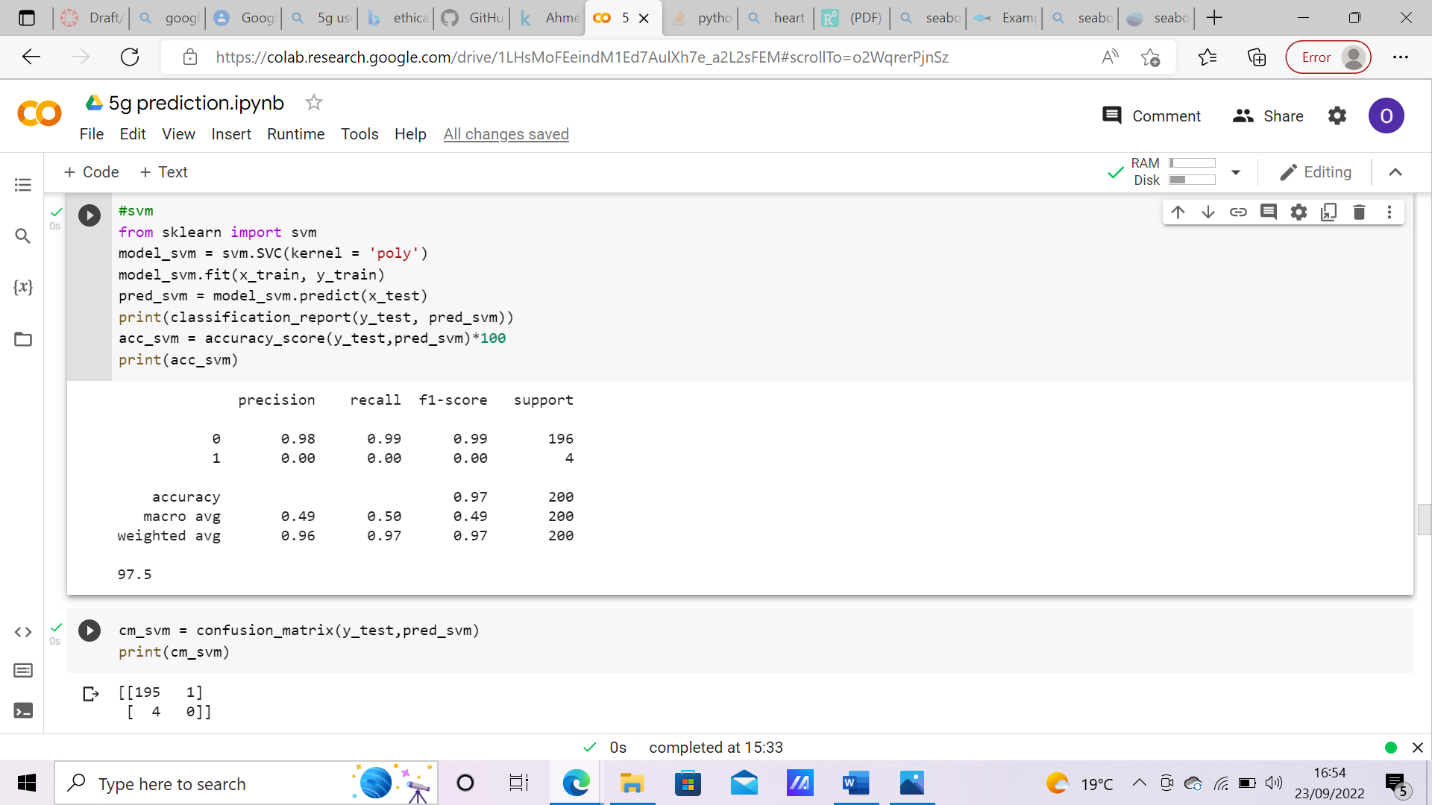
****

Figure 12 SVM model and performance

### Decision Tree

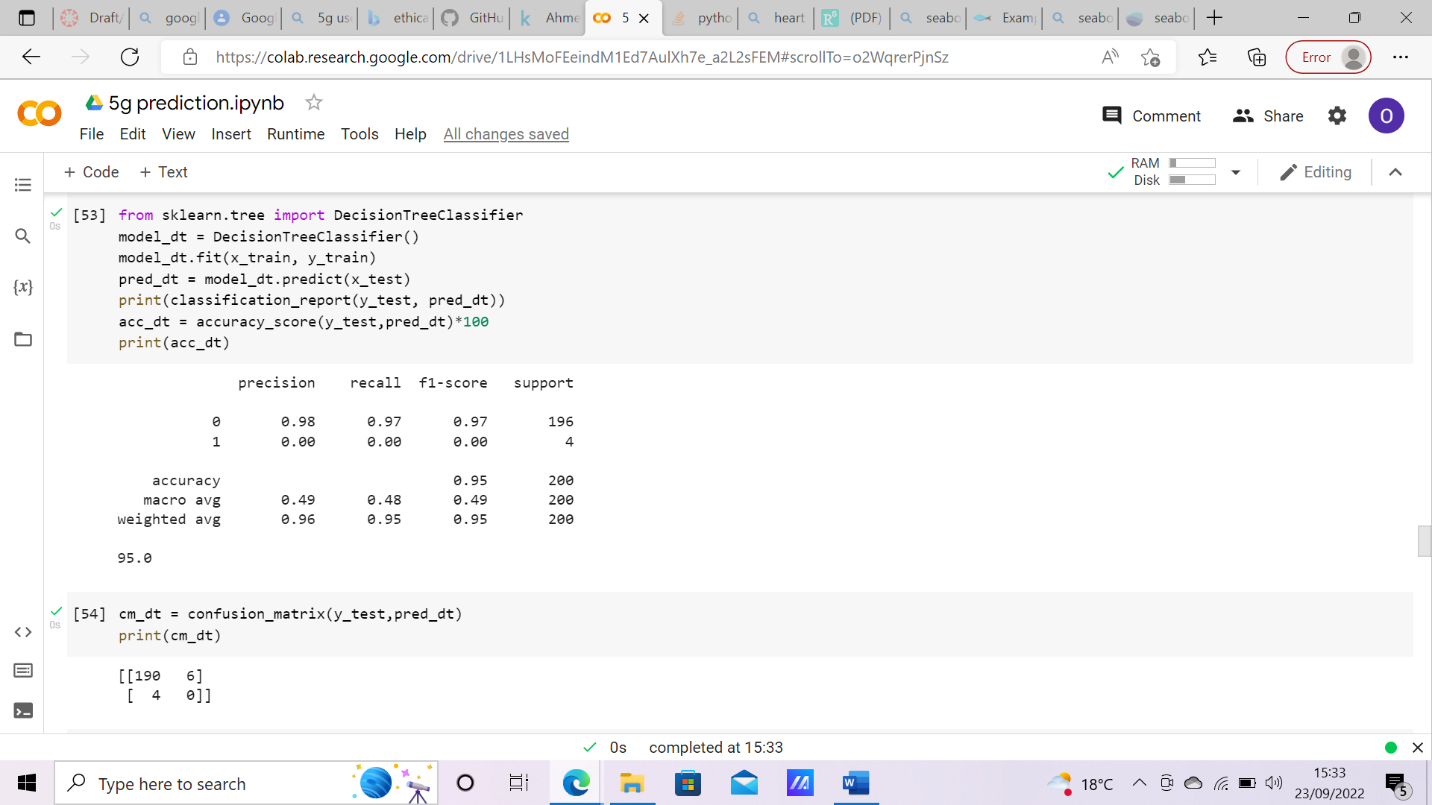
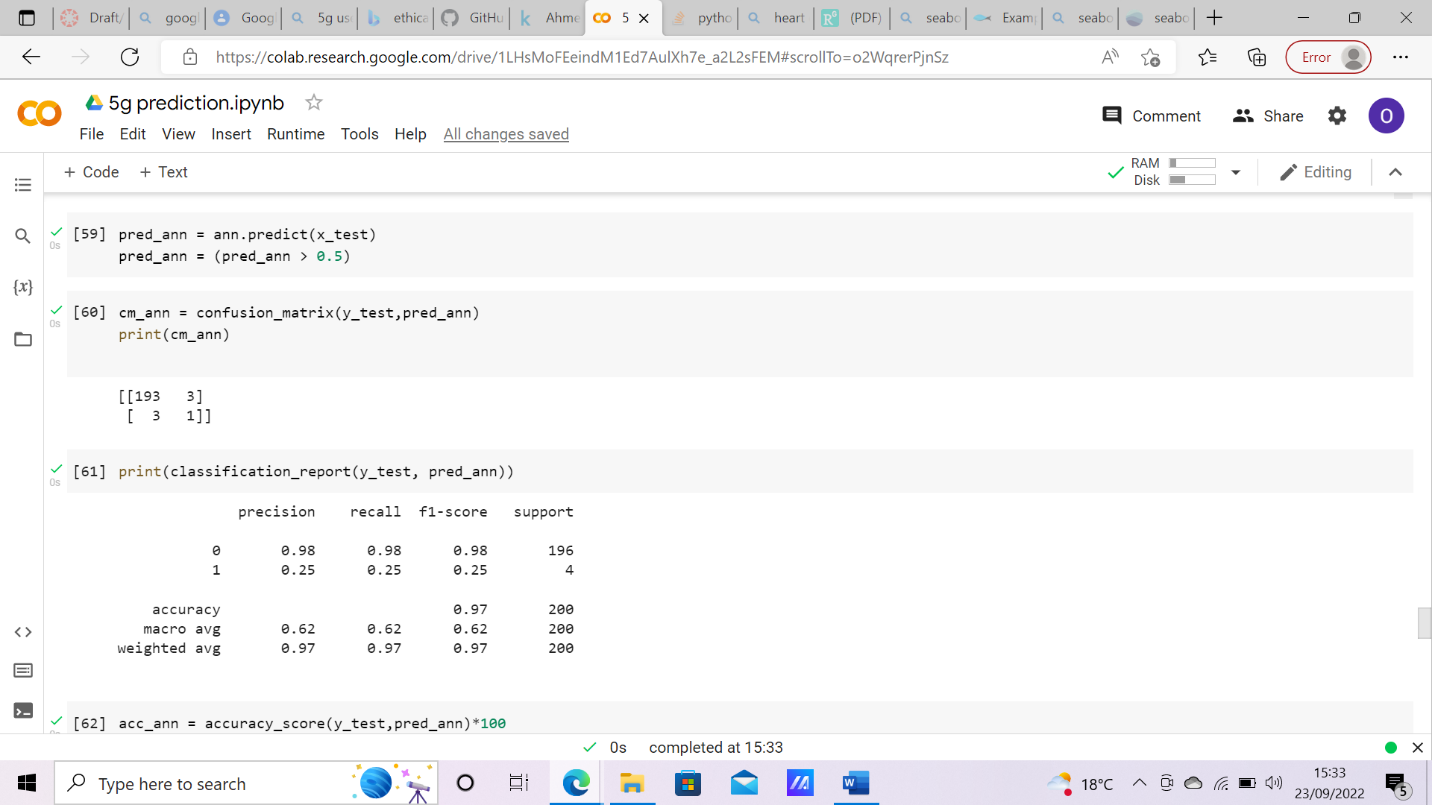
****

Figure 13 Decision Tree Model and performance

### Artificial Neural Network (ANN)

**Graphical user interface, text, application, email

Description automatically generated**



Graphical user interface, text, application

Description automatically generated

Figure 14 ANN Model and performance

# RESULT ANALYSIS

## INTRODUCTION

This chapter of this project analyses and visualize the findings of the analytical and predictive tasks done in the previous chapter to evaluate the predictive performance of the trained machine learning models. The findings that will be extensively discussed in this chapter will give a decisive answer to the research question aforementioned in this research work.

## PREDICTION PERFORMANCE EXPLANATION

This section sets out to show the prediction performance of each of the trained machine learning models. The prediction accuracy results for the models are highlighted below:

1. Artificial Neural Networks (ANN) : This is the most performing and most complex of all the trained models which is quite expected with an outstanding accuracy of 98%.
2. Logistic Regression: This model has a shared top spot with artificial neural network, I find the performance of this model most surprising because it is generally not as complex an thorough as artificial neural network but despite that, it still holds a top spot alongside it.
3. Support Vector Machine (SVM): This machine learning model also fits really well with this dataset to achieve a model predictive accuracy of 97.5%.
4. Decision Tree Classifier: The performance of this model is exceptionally impressive and for me, it was unexpected. Decision tree is a relatively simple algorithm although this is only valid when dealing with datasets with very low dimensionality and complexity, hence, a performance of 96% is a very good score for this model.
5. Naïve Bayes: This is the worst performing machine leaning model with an accuracy of 80%. This translates that 8 out of 10 times, this model will predict the right answer. This might appear like a good score, but when much more ambiguous dataset is being applied, this performance will only tend to get worse, and also in critical support systems, this performance rating is entirely unacceptable.

The image below shows the collective chart of the machine learning model prediction performance.

Graphical user interface, text, application

Description automatically generated

Figure 15 5G prediction performance

## CONFUSION MATRIX EXPLANATION

The confusion matrix is a very simple yet powerful accuracy measurement technique in machine learning. It is a basis on which many machine learning model performance metrics like accuracy and recall are built on. It helps to understand classification performance of a machine learning model on a test or validation dataset where correct and incorrect values are known. The keywords used in confusion matrix are:

TP – True Positive: Actual positive values that are predicted as positive.

FP - False Positive: Actual negative values that are predicted as positive.

FN – False Negative: Actual positive values that are predicted as negative.

TN – True Negative: Actual negative values that are predicted as negative.

In this research project and analysis it is important to note that the term positive refers to actually making use of the 5g network while the term negative refers to not making use of the 5g network.

The confusion matrix do justice to show the actions of each machine learning model performed to arrive at the generated accuracy.

The table below shows the collated confusion matrix and accuracy to compare each model’s performance.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Algorithms** | **Accuracy** | **TN** | **FN** | **FP** | **TP** |
| **Naïve Bayes** | **80%** | **159** | **3** | **37** | **1** |
| **Logistic Regression** | **98%** | **196** | **4** | **0** | **0** |
| **SVM** | **97.5%** | **195** | **4** | **1** | **0** |
| **Decision Tree** | **96%** | **192** | **4** | **4** | **0** |
| **ANN** | **98%** | **195** | **3** | **1** | **1** |

This table easily depicts what metrics the accuracy way calculated on. In the model performance explanation in the previous module, you might have observed I rated Artificial Neural Network (ANN) higher above Logistic regression despite both algorithms boasting the same score, well, the reason is in the confusion matrix. The Artificial Neural Network (ANN) model managed to predict a true positive as opposed to the Logistic Regression model that predicted no true positive outcome. Although this is a personal point of view.

The visualization for the confusion matrix is demonstrated below for each of the machine learning algorithm

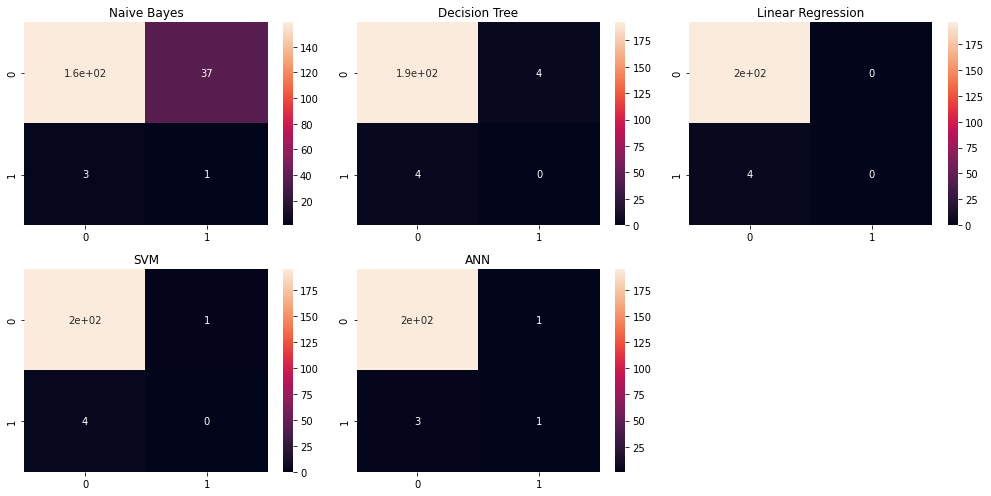


Figure 16 Models confusion matrix visualisation

# CONCLUSION, EVALUATION & SUGGESTION

The advent of the fifth generation (5G) network technology opened a massive gateway for advancement of many existing fields of technology like cloud computing, Internet of Things (IOT), ubiquitous computing, big data and data warehousing. Although these fields were thriving, but the 5G network technology made them better especially in terms of latency and data loss in communication. Due to this development other fields of study such as medicine and art have been able to advance their modes of operation. For example, doctors can now successfully perform surgery on a patient halfway across the world without any hiccup, or a geologist testing a rock sample on the moon from earth, all these are made possible by the advent of 5g network technology.

Although, 5g network can easily be recognised especially by the network speed, hence this research primarily set out to uncover if it can be predicted whether a user is on 5g network or not based on other generic network information and not including obvious attributes like network speed.

Based on the dataset and prediction analysis performed in this research work, I strongly opine that 5g network usage can be determined using non obvious network parameters.

This research truly is far from a complete research to unlock a brand new spectrum of unexplored field of knowledge, and as such, future work and improvement can still be done either to further make this research better and even more comprehensive or to use it as a guide or stating point to explore a whole new spectrum of research gaps. Personally, I think further wok could be done on implementing this prediction technique on a whole new dataset that will be much more comprehensive to verify research results. Also, more analytical research can be done on interaction between 5G and special case Internet of things (IOT) like driverless cars to determine if a faster network can enable these devices interact faster with their environment and/o make better decisions.

This research work was largely limited by the lack of workable dataset to use for 5G predictions. Many of the reasonable data that could be used for good prediction analysis are for sale and the authors are requesting ridiculous amounts of money in exchange for the dataset. Ultimately, I was eventually able to find and make use of the dataset I have in this project and the final result was, Yes, 5G network usage prediction is not dependent on network speed and other obvious attributes.

# REFERENCES

Adio-Moses, D. and Asaolu, O.S., 2016, February. Artificial intelligence for sustainable development of intelligent buildings. In Proceedings of the 9th CIDB Postgraduate Conference, At University of Cape Town, South Africa.

Ahmed, E., Rehmani, M.H., 2017. Mobile Edge Computing: Opportunities, solutions, and challenges. Future Generation Computer Systems 70, 59–63. <https://doi.org/10.1016/j.future.2016.09.015>

Bumi̇N Doyduk, H.B., Bayarçeli̇K, E.B., 2019. Nesnelerin İnterneti Teknolojisinin Tüketiciler Tarafından Kabulü. İstanbul Gelişim Üniversitesi Sosyal Bilimler Dergisi 6, 351–371. <https://doi.org/10.17336/igusbd.538182>

Condoluci, M., Araniti, G., Mahmoodi, T., Dohler, M., 2016. Enabling the IoT Machine Age With 5G: Machine-Type Multicast Services for Innovative Real-Time Applications. IEEE Access 4, 5555–5569. <https://doi.org/10.1109/ACCESS.2016.2573678>

fortiss - Forschungsinstitut des Freistaats Bayern für softwareintensive Systeme und Services, Munich, Germany, Baecker, J., Engert, M., Pfaff, M., Krcmar, H., Technical University of Munich, Chair for Information Systems, Munich, Germany, 2020. Business Strategies for Data Monetization: Deriving Insights from Practice, in: WI2020 Zentrale Tracks. GITO Verlag, pp. 972–987. <https://doi.org/10.30844/wi_2020_j3-baecker>

Lopes, V., Alexandre, L.A., 2019. An Overview of Blockchain Integration with Robotics and Artificial Intelligence. ledger. <https://doi.org/10.5195/ledger.2019.171>

Luini, L., Roveda, G., Zaffaroni, M., Costa, M., Riva, C.G., 2020. The Impact of Rain on Short ${E}$ -Band Radio Links for 5G Mobile Systems: Experimental Results and Prediction Models. IEEE Trans. Antennas Propagat. 68, 3124–3134. <https://doi.org/10.1109/TAP.2019.2957116>

Lund, H., Østergaard, P.A., Connolly, D., Mathiesen, B.V., 2017. Smart energy and smart energy systems. Energy 137, 556–565. <https://doi.org/10.1016/j.energy.2017.05.123>

Menditto, A., Patriarca, M., Magnusson, B., 2007. Understanding the meaning of accuracy, trueness and precision. Accred Qual Assur 12, 45–47. <https://doi.org/10.1007/s00769-006-0191-z>

Mzahm, A.M., Ahmad, M.S., Tang, A.Y.C., 2013. Agents of Things (AoT): An intelligent operational concept of the Internet of Things (IoT), in: 2013 13th International Conference on Intellient Systems Design and Applications. Presented at the 2013 13th International Conference on Intelligent Systems Design and Applications (ISDA), IEEE, Salangor, Malaysia, pp. 159–164. <https://doi.org/10.1109/ISDA.2013.6920728>

Nejkovic, V., Visa, A., Tosic, M., Petrovic, N., Valkama, M., Koivisto, M., Talvitie, J., Rancic, S., Grzonka, D., Tchorzewski, J., Kuonen, P., Gortazar, F., 2019. Big Data in 5G Distributed Applications, in: Kołodziej, J., González-Vélez, H. (Eds.), High-Performance Modelling and Simulation for Big Data Applications, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 138–162. <https://doi.org/10.1007/978-3-030-16272-6_5>

Noor, A.K., 2014. Potential of Cognitive Computing and Cognitive Systems. Open Engineering 5. <https://doi.org/10.1515/eng-2015-0008>

Queralta, J.P., Qingqing, L., Zou, Z., Westerlund, T., 2020. Enhancing Autonomy with Blockchain and Multi-Access Edge Computing in Distributed Robotic Systems.

Paudel, P. and Bhattarai, A., 2018. 5G telecommunication technology: history, overview, requirements and use case scenario in context of Nepal. In Conference: IT4D.

Ruffini, M., 2017. Multidimensional convergence in future 5G networks. Journal of Lightwave Technology, 35(3), pp.535-549.

Sisinni, E., Saifullah, A., Han, S., Jennehag, U., Gidlund, M., 2018. Industrial Internet of Things: Challenges, Opportunities, and Directions. IEEE Trans. Ind. Inf. 14, 4724–4734. <https://doi.org/10.1109/TII.2018.2852491>

Smail, G., Weijia, J., 2017. Techno-economic analysis and prediction for the deployment of 5G mobile network, in: 2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN). Presented at the 2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN), IEEE, Paris, pp. 9–16. <https://doi.org/10.1109/ICIN.2017.7899243>

Varga, P., Peto, J., Franko, A., Balla, D., Haja, D., Janky, F., Soos, G., Ficzere, D., Maliosz, M., Toka, L., 2020. 5G support for Industrial IoT Applications— Challenges, Solutions, and Research gaps. Sensors 20, 828. <https://doi.org/10.3390/s20030828>

Wang, D., Chen, D., Song, B., Guizani, N., Yu, X., Du, X., 2018. From IoT to 5G I-IoT: The Next Generation IoT-Based Intelligent Algorithms and 5G Technologies. IEEE Commun. Mag. 56, 114–120. <https://doi.org/10.1109/MCOM.2018.1701310>

Dataset: <https://www.kaggle.com/datasets/liukunxin/dataset>

# APPENDIX A: SOURCE CODE

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sb

%matplotlib inline

df = pd.read\_csv('train\_data.csv')

df.head()

df.info()

days=['active\_days01', 'active\_days02',

       'active\_days03', 'active\_days04', 'active\_days05', 'active\_days06',

       'active\_days07', 'active\_days08', 'active\_days09', 'active\_days10',

       'active\_days11', 'active\_days12', 'active\_days13', 'active\_days14',

       'active\_days15', 'active\_days16', 'active\_days17', 'active\_days18',

       'active\_days19', 'active\_days20', 'active\_days21', 'active\_days22',

       'active\_days23']

df.drop(columns=days,inplace=True)

df.info()

df\_final = df.drop(['user\_id','area\_id','innet\_months','total\_times','total\_flux','total\_fee','pay\_fee','age','sex','call\_days','re\_call10','short\_call10','long\_call10','bank\_cnt','game\_app\_flux','live\_app\_flux','video\_app\_flux'], axis = 1)

df\_final.info()

plt.figure(figsize=[25,10])

plt.subplot(2,1,1)

b=sb.distplot(df\_final[df\_final['is\_5g']==1]['city\_5g\_ratio'],color="skyblue")

b.set\_xlabel("city ratio density of 5g users",fontsize=25)

b.set\_ylabel("Density",fontsize=25)

b.tick\_params(labelsize=20);

plt.subplot(2,1,2)

b=sb.distplot(df\_final[df\_final['is\_5g']==0]['city\_5g\_ratio'],color="red")

b.set\_xlabel("city ratio density of non 5g users",fontsize=25)

b.set\_ylabel("Density",fontsize=25)

b.tick\_params(labelsize=20)

plt.tight\_layout();

cor = df\_final.corr()

plt.subplots(figsize=(15,10))

sb.heatmap(cor, xticklabels=cor.columns, yticklabels=cor.columns, annot = False, cbar= True, cmap = sb.diverging\_palette(150, 275, s=80, l=55, n=9))

x = df\_final.drop(['is\_5g'], axis = 1)

y = df['is\_5g']

x

y

#data normalisation

from sklearn.preprocessing import StandardScaler

x = StandardScaler().fit\_transform(x)

#Train Test Splitting

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import classification\_report

from sklearn.metrics import accuracy\_score

from sklearn.metrics import confusion\_matrix

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size = .20, random\_state= 40)

#Naive Bayes regression

from sklearn import naive\_bayes

model\_nb = naive\_bayes.GaussianNB()

model\_nb.fit(x\_train, y\_train)

pred\_nb = model\_nb.predict(x\_test)

print(classification\_report(y\_test,pred\_nb))

acc\_nb = accuracy\_score(y\_test,pred\_nb)\*100

print(acc\_nb)

cm\_nb = confusion\_matrix(y\_test,pred\_nb)

print(cm\_nb)

#logistic regression

from sklearn.linear\_model import LogisticRegression

model\_lr = LogisticRegression()

model\_lr.fit(x\_train, y\_train)

pred\_lr = model\_lr.predict(x\_test)

print(classification\_report(y\_test,pred\_lr))

acc\_lr = accuracy\_score(y\_test,pred\_lr)\*100

print(acc\_lr)

cm\_lr = confusion\_matrix(y\_test,pred\_lr)

print(cm\_lr)

#svm

from sklearn import svm

model\_svm = svm.SVC(kernel = 'poly')

model\_svm.fit(x\_train, y\_train)

pred\_svm = model\_svm.predict(x\_test)

print(classification\_report(y\_test, pred\_svm))

acc\_svm = accuracy\_score(y\_test,pred\_svm)\*100

print(acc\_svm)

cm\_svm = confusion\_matrix(y\_test,pred\_svm)

print(cm\_svm)

from sklearn.tree import DecisionTreeClassifier

model\_dt = DecisionTreeClassifier()

model\_dt.fit(x\_train, y\_train)

pred\_dt = model\_dt.predict(x\_test)

print(classification\_report(y\_test, pred\_dt))

acc\_dt = accuracy\_score(y\_test,pred\_dt)\*100

print(acc\_dt)

cm\_dt = confusion\_matrix(y\_test,pred\_dt)

print(cm\_dt)

from tensorflow import keras

from keras.models import Sequential

from keras.layers import Dense

#Artificial Neural Network

ann = Sequential()

ann.add(Dense(activation = "relu", input\_dim = 19,

          units = 8, kernel\_initializer = "uniform"))

ann.add(Dense(activation = "relu", units = 10,

          kernel\_initializer = "uniform"))

ann.add(Dense(activation = "sigmoid", units = 1,

          kernel\_initializer = "uniform"))

ann.compile(optimizer = 'adam' , loss = 'binary\_crossentropy',

        metrics = ['accuracy'] )

ann.fit(x\_train , y\_train , batch\_size = 8 ,epochs = 100 )

pred\_ann = ann.predict(x\_test)

pred\_ann = (pred\_ann > 0.5)

cm\_ann = confusion\_matrix(y\_test,pred\_ann)

print(cm\_ann)

print(classification\_report(y\_test, pred\_ann))

acc\_ann = accuracy\_score(y\_test,pred\_ann)\*100

print(acc\_ann)

plt.figure(figsize=(14,7))

plt.subplot(2,3,1)

sb.heatmap(cm\_nb, annot=True).set(title = 'Naive Bayes')

plt.subplot(2,3,2)

sb.heatmap(cm\_dt, annot=True).set(title = 'Decision Tree')

plt.subplot(2,3,3)

sb.heatmap(cm\_lr, annot=True).set(title = 'Linear Regression')

plt.subplot(2,3,4)

sb.heatmap(cm\_svm, annot=True).set(title = 'SVM')

plt.subplot(2,3,5)

sb.heatmap(cm\_ann, annot=True).set(title = 'ANN')

plt.tight\_layout();

total\_acc = {'Naive Bayes':acc\_nb, 'Decision Tree':acc\_dt, 'Logistic Regresion':acc\_lr, 'SVM': acc\_svm, 'ANN': acc\_ann}

acc=sorted(total\_acc.items(),key=lambda x: x[1], reverse=True)

acc=dict(acc)

x\_acc = list(acc.keys())

y\_acc = list(acc.values())

plt.figure(figsize=[25,10])

plt.plot(x\_acc, y\_acc)

plt.ylabel('Accuracy percentage',fontsize=30)

plt.xlabel('ML Classifiers',fontsize=30)

plt.xticks(fontsize=20)

plt.yticks(fontsize=20)

plt.title('5G prediction performance',fontsize=50)

plt.grid()

plt.tight\_layout()

#inspiration for this analysis was obtained from https://github.com/metwallusion/5g-user-prediction. Therefore there are similar ascpects of the code, However the analysis and results were totally different.

# APPENDIX B: PROJECT PLAN TIMELINE

