UNIVERSITY KASDI MERBAH OUARGLA
FACULTY OF APPLIED SCIENCES
DEPARTMENT OF ELECTRICAL ENGINEERING

Dissertation
ACADEMIC MASTER
Field: Science and Technology
Sector: Electrical Engineering
Specialty: Industrial Electrotechnics
Presented by:

BEKKOUCHE KHIREDINE KHELIF MOHAMMED MILOUD

Theme:

Power system fault detection and classification using machine learning

jury:

Dr.Boukaroura Abdelkader	MCB	President	UKM Ouargla
Dr.Bouhafs Ali	MCA	Supervisor	UKM Ouargla
Dr.Ayad Ahmed Nour El Islam	Professor	Examiner	UKM Ouargla
Dr.Rouibah Tahar	MCA	Co supervisor	UKM Ouargla

Academic year: 2023 – 2024

Acknowledgments

We express gratitude to the Almighty for granting us the strength and determination to finish this current task.

First of all, we sincerely thank Dr. BOUHAFS ALI for his supervision and especially for the precious help he gave us throughout the work.

We also thank Dr. ROUIBAH TAHAR

We would also like to thank Dr.TAIBI DJAMAL for the assistance and advice provided, which made it easier for us to complete this thesis.

We also extend our gratitude to each member of the jury for the privilege of agreeing to evaluate our work.

We extend our appreciation to all those who have contributed, whether directly or indirectly, to the accomplishment of this task.

Dedication

We stand on the threshold of graduation, feeling deep gratitude for the support and love you generously bestowed upon us throughout this journey. Your encouragement has fueled our ambitions, and your guidance has illuminated our path through the challenges and triumphs of academic life. To our families, you have continuously offered boundless love, understanding, and sacrifices to ensure our success. Your steadfast belief in us has been the strongest source of our strength. To our close friends and supporters, your laughter, social spirit, and unwavering support enriched our college experience like no other. You have been pillars of our growth, shaping our personalities as they are today. As we embark on this new chapter of our lives, we carry with us the memories we shared, the lessons we learned, the bonds we forged, and may we always find comfort and joy in each other.

Summary

Currently, the importance of artificial intelligence applications in the field of electrical energy, such as the power grid, has increased, especially in data analysis and fault detection. Therefore, this memo presents a multi-line electrical network system with fault detection, such as phase - to-phase, phase-to-ground, and double phase-to-ground faults, using machine learning algorithms. Simulation results using MATLAB software and implementing the nearest neighbor algorithm show high accuracy.

Résumé

Actuellement, l'importance des applications de l'intelligence artificielle dans le domaine de l'énergie électrique, telles que le réseau électrique, a augmenté, notamment dans l'analyse des données et la détection des pannes. Ainsi, ce mémoire présente un système de réseau électrique multi lignes avec détection des pannes comme phase à phase, phase à la terre et la double phase à la terre, en utilisant des algorithmes d'apprentissage automatique. Les résultats de la simulation utilisant le logiciel Matlab et l'application de l'algorithme des plus proches voisins montrent une grande précision.

ملخص

في الوقت الحالي، زادت أهمية تطبيقات الذكاء الاصطناعي في مجال الطاقة الكهربائية، مثل الشبكة الكهربائية، خاصة في تحليل البيانات وكشف الأعطال. لذلك، في هذه المذكرة يتم تقديم نظام شبكة كهربائية متعدد الخطوط مع كشف الأعطال مثل طور الى طور و طور الى الأرض ومزدوج الطور الى الأرض. باستخدام خوارزميات التعلم الآلي. تظهر نتائج المحاكاة باستخدام برنامج ماتلاب وتطبيق خوارزمية أقرب الجيران لأنها تتمتع بدقة عالية.

CONTENTS

CHAPTER I: General information on faults in electrical network

CHAPTER II: Machine learning

CHAPTER III: Implementation and results

LIST OF FIGURES

CHAPTER I: General information on faults in electrical network

CHAPTER II: Machine learning

CHAPTER III: Implementation and results

LIST OF TABLES

CHAPTER III: Implementation and results

GENERAL INTRODUCTION

General introduction

Failure in electric distribution feeders results in both customer inconvenience and economic losses due to power outages. The complexity of the issue is exacerbated by inaccurate fault information. Addressing such challenges necessitates the implementation of effective fault diagnosis techniques as integral components of power system protection.

Therefore, the development of efficient fault detection, classification, and localization schemes becomes crucial for streamlining the power restoration process. Accurate fault detection facilitates the isolation of the faulty segment from the healthy network through relay operations. This isolation protects the assets of the faulty segment, ensuring an uninterrupted energy supply to customers in the unaffected areas.

Moreover, the availability of fault class information provides essential insights into the fault location, expediting the restoration process by directing crews to the affected spots and minimizing customer downtime and revenue loss.

Despite the significant advancements in fault diagnosis approaches for transmission networks over the years, these methods cannot be directly applied to distribution feeders due to their inherent complexities. Additionally, the increasing integration of distributed generations (such as renewable and non-renewable energy resources and energy storage systems) aims at achieving techno-economic benefits, such as improving voltage profiles, system balance, reducing transmission power losses, greenhouse gases, and injecting reactive power into the network.

However, the integration of distributed generations into distribution feeders transforms them from passive to active networks, introducing complexities in operation and control techniques. The power flow shifts from unidirectional to multidirectional, further complicating the traditional operation of passive feeders. The intermittent nature of solar and wind power generation also impacts the regular operation of active feeders, affecting the accuracy of traditional protection techniques like protective relays and fault diagnosis methods.

In light of these challenges, it is imperative to redesign traditional protection schemes and fault diagnosis techniques, taking into account the trends associated with the incorporation of distributed generations into distribution feeders.

Chapter One, titled " General information on faults in electrical networks" provides a definition of the electrical grid and the faults that occur within it.

Chapter Two, titled "Machine Learning" provides a definition of machine learning, its types, and its applications. Additionally, we presented some algorithms and highlighted the importance of using them in various fields.

Chapter Three, titled "Implementation and results" explains the definition of our system and the use of machine learning in fault detection. We present the steps and tools used to implement our work, as well as the results obtained. Finally, we conclude this document with a summary that encapsulates the essence of our work and its future prospects

CHAPTER I

GENERAL INFORMATION ON FAULTS IN ELECTRICAL NETWORKS

I.1 Introduction

The intricate web of power generation, transmission, and distribution systems forms a complex electrical grid designed to provide a continuous and reliable supply of electricity. However, despite the advancements in technology and infrastructure, electrical failures and disruptions remain an inevitable challenge.

Electrical failures can stem from diverse factors, ranging from natural disasters such as storms or earthquakes to technical issues like equipment malfunctions or human errors. These disruptions have the potential to impact businesses, disrupt essential services, and create widespread inconvenience.

This introduction will delve into the significance of electrical power, exploring its various sources and the intricate network that delivers it to end-users. Additionally, it will touch upon the challenges and consequences associated with electrical failures, emphasizing the importance of proactive measures to enhance the resilience and stability of our electrical infrastructure.

I.2 Definition of the electrical network

A power distribution system is a network of cables designed to connect electrical power from its generation point to end users.[1] Electricity distribution systems are primarily designed to meet the demands of consumers for power. This is achieved by taking power from main stations and delivering it to various customer substations, either through underground cables or overhead lines. In densely populated areas, some cities rely on underground cables, while rural areas may use overhead lines due to space constraints. The choice between underground and overhead delivery methods depends on enhancing protection and control.

When it comes to selecting the best means for an area, three planning components should be considered: long-term planning, network planning, and construction planning. Long-term planning involves anticipating the future. These distribution systems help provide more electricity to homes, businesses, and other organizations to carry out their daily tasks. [2]

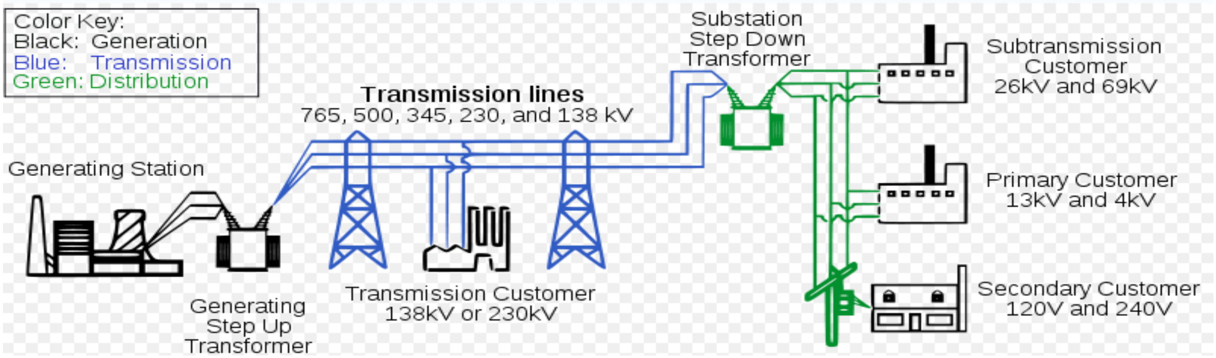


Figure I.1: Global view of the electrical network [3]

I.2.1 Electricity production

The production process involves utilizing turbo-alternators to convert mechanical energy from turbines into electrical energy sourced from various primary sources such as gas, oil, and hydraulics. The distribution of primary sources differs from country to country; for instance, in Algeria, natural gas accounts for over 70% of production, while in France, 75% of electricity is nuclear-generated. Typically, each power plant incorporates multiple turbo-alternator groups to ensure continuous operation during maintenance periods. For instance, the Jijel power station in Algeria comprises three 196 MW groups, and the Cap Djenet facility in Boumerdes has four 168 MW groups. Moreover, industrialized nations are installing power plants with increasingly higher capacities to meet rising electrical energy demands. For example, the Gravelines nuclear power plant in France has six units of 900 MW each, while the Three Gorges hydroelectric power plant in China has 34 units of 700 MW and 2 units of 50 MW, making it the world's largest power plant as of 2014.[4]

I.2.2 Electric power transmission

Electric power transmission involves the large-scale transfer of electrical energy from a power generation site, such as a power plant, to an electrical substation. The interconnected lines forming a transmission network facilitate this movement. This process is distinct from the local wiring connecting high-voltage substations to customers, which is typically known as electric power distribution. Together, the transmission and distribution network constitute the electricity delivery system, commonly referred to as the electrical grid.

Efficient long-distance transmission of electric power necessitates the use of high voltages to minimize losses caused by intense currents. Transmission lines utilize either alternating current (AC) or direct current (DC), with voltage levels adjusted using transformers. Voltage is increased for transmission purposes and decreased for local distribution.

A broad area synchronous grid, referred to as an interconnection in North America, directly links power generators supplying AC power with numerous consumers at the same frequency. North America features four major interconnections: Western, Eastern, Quebec, and Texas. Similarly, most of continental Europe is connected through a single grid.

Traditionally, transmission and distribution lines were commonly owned by the same entity. However, since the 1990s, many countries have liberalized electricity market regulations, leading to separate entities handling transmission and distribution.[5]

Figure I.2: Electric power transmission [6]

I.2.3 Interconnection stations

The distribution network is used to provide large industrial customers under high or medium voltage and to distribute electricity in various rural and urban areas. It originates in the transmission network from the interconnection stations THT/HT. Typically, this kind of network uses 60 and 30 kV.

I.2.4 Distribution

Electric power distribution represents the ultimate phase in the supply chain of electricity. It involves the transportation of electricity from the transmission network to individual users. Distribution substations serve as the link between the transmission infrastructure, where voltage levels are high, and consumers, by reducing the voltage to a medium range typically between 2 kV and 33 kV via transformers.[7]

Primary distribution lines transport medium voltage electricity to distribution transformers situated close to customers' locations. These transformers then reduce the voltage to a level suitable for use by lighting, industrial machinery, and household appliances. Frequently, multiple customers receive power from a single transformer via secondary distribution lines. Commercial and residential clients are linked to these secondary lines through service drops.

Customers with higher power demands may be directly connected to either the primary distribution level or the sub transmission level.[8]

The transition from transmission to distribution happens in a power substation, which has the following functions:[8]

- Circuit breakers and switches allow for the substation to be disconnected from the transmission grid, or for distribution lines to be disconnected.
- Transformers reduce high transmission voltages, typically 35 kV or higher, to lower primary distribution voltages, which typically range from 600 to 35,000 V. [7]
- Electricity from the transformer is directed to the busbar, which serves as a hub for dividing the power into various directions for distribution. The busbar then channels the power to distribution lines, which branch out to supply electricity to customers.

Urban distribution primarily occurs underground, often within shared utility ducts, while rural distribution predominantly utilizes above-ground infrastructure such as utility poles. Suburban distribution, on the other hand, encompasses a combination of both methods. As the power gets closer to the end consumer, a distribution transformer reduces the primary distribution power to a lower voltage secondary circuit, typically around 120/240 V in the United States for residential users. This electricity is delivered to the customer through a service drop and an electricity meter. In urban areas, the final circuit length may be under 15 meters (50 feet), whereas it could extend beyond 91 meters (300 feet) for rural customers.[7]

I.2.5 Load balancing

The transmission system provides for base load and peak load capability, with margins for safety and fault tolerance. Peak load times vary by region largely due to the industry mix. In hot and cold climates home air conditioning and heating loads affect the overall load. They are typically highest in the late afternoon in the hottest part of the year and in mid-mornings and mid-evenings in the coldest part of the year. Power requirements vary by season and time of day. Distribution system designs always take the base load and the peak load into consideration.

The transmission system usually does not have a large buffering capability to match loads with generation. Thus, generation has to be kept matched to the load, to prevent overloading generation equipment.

Multiple sources and loads can be connected to the transmission system and they must be controlled to provide orderly transfer of power. In centralized power generation, only local control of generation is necessary. This involves synchronization of the generation units.

In distributed power generation the generators are geographically distributed and the process to bring them online and offline must be carefully controlled. The load control signals can either be sent on separate lines or on the power lines themselves. Voltage and frequency can be used as signaling mechanisms to balance the loads.

In voltage signaling, voltage is varied to increase generation. The power added by any system increases as the line voltage decreases. This arrangement is stable in principle. Voltage-based regulation is complex to use in mesh networks, since the individual components and setpoints would need to be reconfigured every time a new generator is added to the mesh.

In frequency signaling, the generating units match the frequency of the power transmission system. In droop speed control, if the frequency decreases, the power is increased. (The drop in line frequency is an indication that the increased load is causing the generators to slow down.)

Wind turbines, vehicle-to-grid, virtual power plants, and other locally distributed storage and generation systems can interact with the grid to improve system operation. Internationally, a slow move from a centralized to decentralized power system has taken place. The main draw of locally distributed generation systems is that they reduce transmission losses by leading to consumption of electricity closer to where it was produced.[9]

I.3 Organization of the power system

The purpose of Power Systems is to link production centers, like thermal power plants or hydraulic power plants, with consumer centers, like cities or factories. Power is transported at high voltage or very high voltage to minimize Joule effect losses, which are proportional to the square of the electric current intensity. The voltage is then gradually lowered to the level of the end user. The Power System is made up of all the devices intended for the production, transmission, distribution, and use of electricity from Power Plants to the farthest-flung country houses.[10]

Figure I.3: Organization of the power syste[11]

For the electrical energy to be functional, the System needs to fulfill certain conditions:

- To ensure the customer the power he needs
- To provide a stable voltage whose variations do not exceed ±10% of the nominal voltage
- To provide a stable frequency whose variations do not exceed ±0.1% Hz
- To provide energy at an acceptable price
- To maintain rigorous safety standards
- Please pay attention to environmental protection

I.4 Power stations

There are five primary categories of power stations: Traditional thermal power plants fueled by fossil fuels like coal, oil, and natural gas, nuclear power stations, which also fall under the thermal power plant classification, hydroelectric power stations, solar or photovoltaic power stations, and wind power stations.

I.4.1 Thermal power plants

Thermal power plants represent the predominant form of power generation worldwide, constituting approximately 60% of the global electricity production. These facilities utilize diverse fuel sources such as coal, natural gas, oil, and nuclear energy to generate electricity.

Boilers are essential components within Thermal Power Plants, serving a vital function in the energy production process. Specifically engineered, these vessels efficiently transform water into steam, which then propels turbines linked to generators. Through the transfer of heat generated by burning fossil fuels or alternative heat sources, water within the boiler is converted into high-pressure steam. This steam, in essence, drives turbines, ultimately generating electricity. Consequently, the presence of boilers significantly influences the overall efficiency and effectiveness of Thermal Power Plants, establishing them as indispensable elements in the realm of energy production.[12]

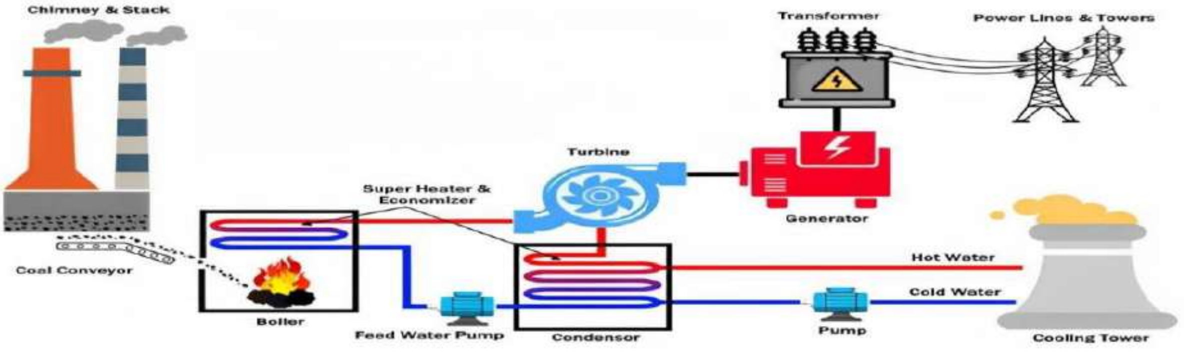


Figure I.4: Thermal power plants [13]

I.4.2 Nuclear power plant

A nuclear power plant (NPP) [14] operates as a thermal power station, utilizing a nuclear reactor as its heat source. Following the conventional model of thermal power stations, the heat produced is utilized to create steam, which in turn propels a steam turbine connected to a generator, ultimately generating electricity. According to the International Atomic Energy Agency's report as of September 2023, there were 410 operational nuclear power reactors across 32 countries globally, with an additional 57 reactors under construction.[15][16]

Nuclear plants are frequently employed for base load electricity generation due to their relatively low operational, maintenance, and fuel costs compared to other sources of power generation.[17] Constructing a nuclear power plant typically entails a timeframe of five to ten years, which can result in substantial financial expenditures, contingent upon how the initial investments are funded.[18]

Nuclear power plants possess a carbon footprint similar to that of renewable energy sources like solar and wind farms, [19][20] and notably lower than fossil fuels such as natural gas and coal. Additionally, nuclear power plants rank among the safest forms of electricity generation, [21] comparable to the safety standards of solar and wind power plants.[22]

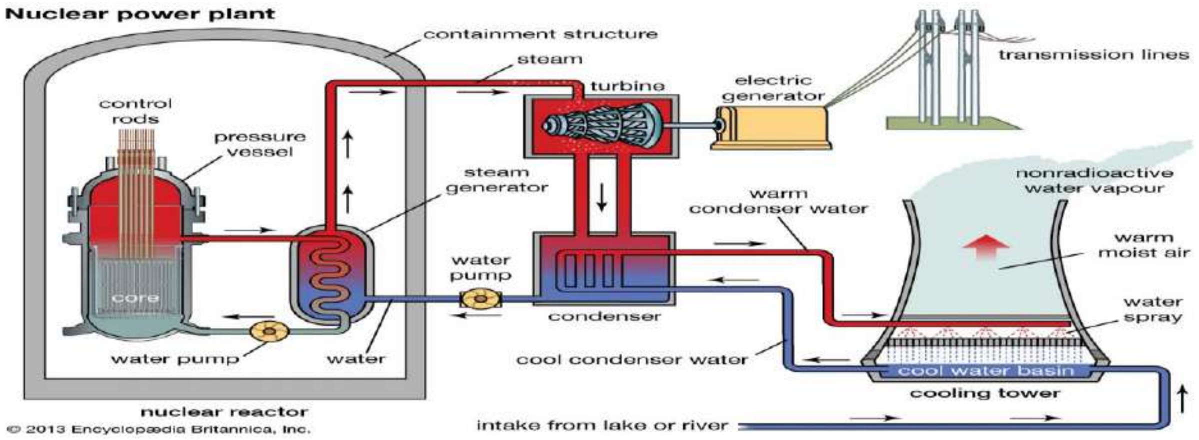


Figure I.5: Nuclear power plant [23]

I.4.3 Hydroelectric power plant

Hydroelectricity, or hydroelectric power, is the generation of electricity through hydropower, harnessing the energy of water. It accounts for approximately one-sixth of global electricity production, reaching nearly 4,500 TWh in 2020. This output surpasses all other renewable sources combined, as well as nuclear power.[24] Hydropower offers the advantage of providing substantial amounts of low-carbon electricity when needed, contributing significantly to the establishment of reliable and clean electricity supply systems.[24] Hydroelectric power stations equipped with dams and reservoirs are flexible sources, capable of adjusting electricity production rapidly to meet fluctuating demand within seconds or minutes. Once constructed, hydroelectric complexes generate no direct waste and typically emit far fewer greenhouse gases than fossil fuel-powered energy plants.[25] However, if built in lowland rainforest areas where parts of the forest are submerged, substantial greenhouse gas emissions may occur.[26]

As of 2021, the worldwide installed capacity for hydropower electricity approached nearly 1,400 GW, marking it as the most extensive among all renewable energy technologies.[27] Countries such as Brazil, Norway, and China [28] prominently feature hydroelectricity in their energy landscapes. However, there are constraints imposed by geographical factors and environmental considerations.[29] In coastal regions, tidal power presents an alternative renewable energy source that can be harnessed.

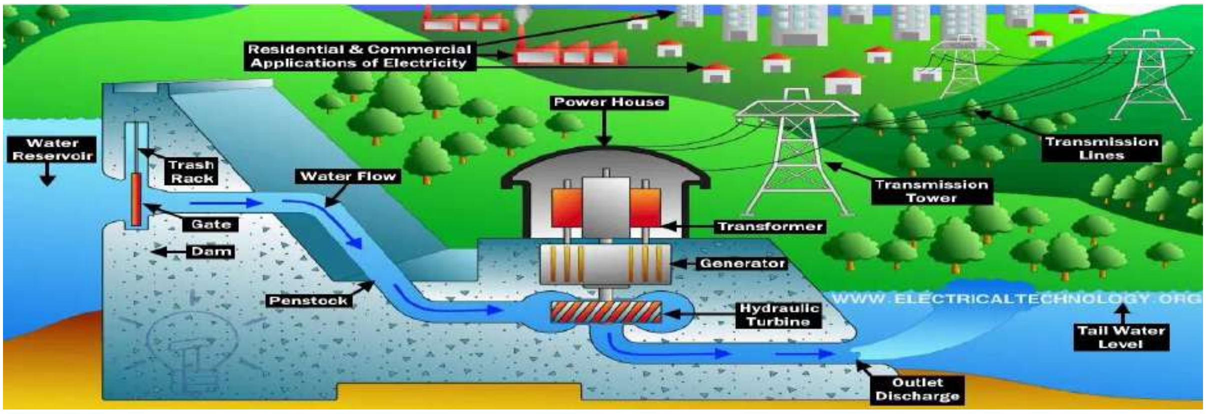


Figure I.6: Hydroelectric power plant [30]

I.4.4 Photovoltaic power station

A photovoltaic power station, also referred to as a solar park, solar farm, or solar power plant, constitutes a large-scale grid-connected photovoltaic power system (PV system) designed to supply merchant power. Unlike most building-mounted and decentralized solar power setups, these stations deliver power at the utility level rather than to individual local users. This distinction is sometimes labeled as utility-scale solar.

This approach contrasts with concentrated solar power, another significant large-scale solar generation technology, which utilizes heat to drive various conventional generator systems. While both approaches have their respective advantages and drawbacks, photovoltaic technology has seen broader adoption due to several factors. As of 2019, approximately 97% of utility-scale solar power capacity utilized PV technology.[31][32]

In certain countries, the nameplate capacity of photovoltaic power stations is denoted in megawatt-peak (MWp), representing the theoretical maximum DC power output of the solar array. Conversely, in other nations, manufacturers provide information regarding the surface area and efficiency of the solar panels. However, in Canada, Japan, Spain, and the United States, the standard specification is often the converted lower nominal power output in MWAC, which enables a more direct comparison with other forms of power generation. Typically, most solar parks are developed with a capacity of at least 1 MWp.

As of 2018, the world's largest operating photovoltaic power stations have exceeded 1 gigawatt. By the end of 2019, there were over 9,000 solar farms larger than 4 MWAC (utility scale), collectively boasting a capacity of over 220 GWAC. [31]

Figure I.7: Photovoltaic power station [33]

I.4.5 Wind power plant

A wind farm, also known as a wind park, wind power station, or wind power plant [34], is a collection of wind turbines situated in the same area to generate electricity. These farms can range in size from a few turbines to several hundred, covering a significant area. They can be established either onshore or offshore.

Some of the largest operational onshore wind farms are situated in China, India, and the United States. For instance, the Gansu Wind Farm in China, recognized as the world's largest, boasted a capacity exceeding 6,000 MW by 2012, [35] with aspirations to reach 20,000 MW [36] by 2020.[37] As of December 2020, the Hornsea Wind Farm in the UK, with a capacity of 1,218 MW, holds the title of the largest offshore wind farm globally.[38] Advancements in individual wind turbine designs have led to increased power output, resulting in fewer turbines being required to achieve the same total output.

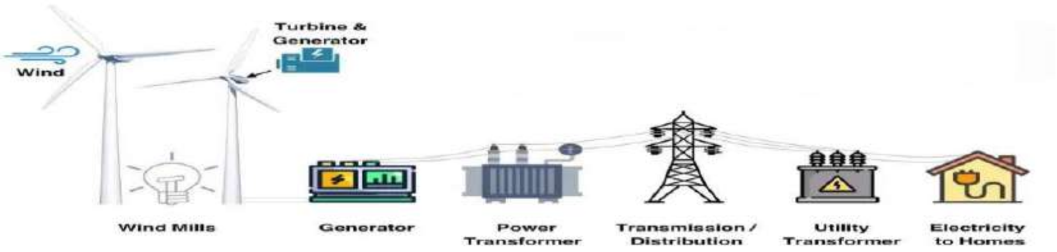


Figure I.8: Wind power plant [39]

I.4 Types of faults in electrical power systems

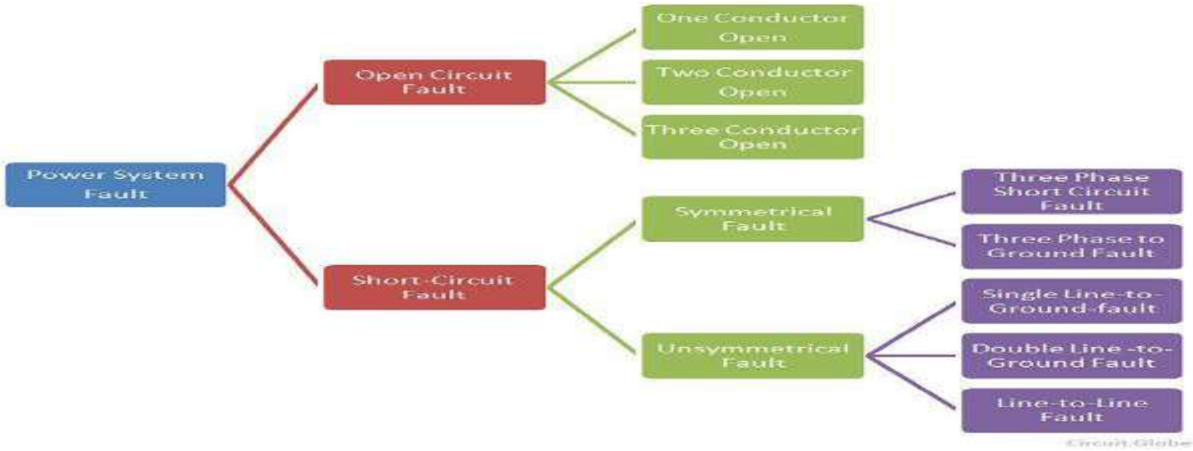
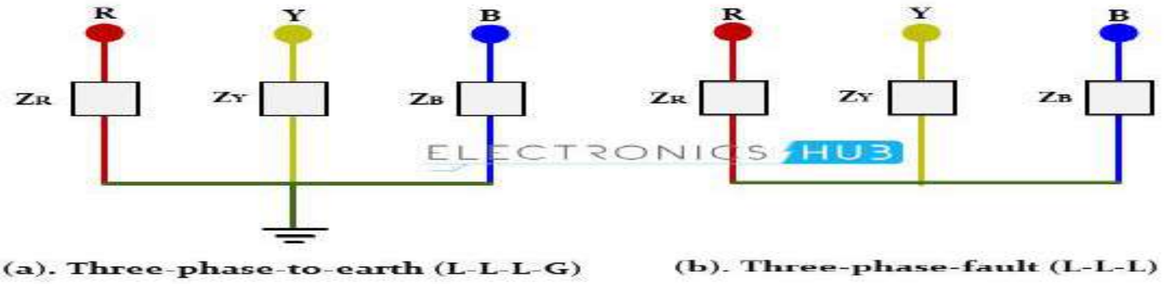


Figure I.9: Types of faults in electrical power systems [40]

I.3.1 Symmetrical faults

Symmetrical fault currents that are displaced by 1200 are produced by a symmetrical fault. A balanced fault is another name for a symmetrical fault. When all three phases are shorted out at the same time, a fault occurs.

Compared to asymmetrical defects, these faults are uncommon in real-world situations. Line to line to line (L-L-L) and line to line to line to ground (L-L-L-G) are two types of symmetrical faults, as seen in the image below.[41]



(a). Three-phase-to-earth (L-L-L-G) (b). Three-phase-fault (L-L-L)

Figure I.10: Symmetrical faults [41]

In general, symmetrical defects account for two to five percent of all system faults. Even if the system is still in a balanced state, if certain errors happen, the equipment suffers quite serious damage.[41]

To choose the rupturing capacity of the circuit breakers, set-phase relays, and other protective switchgear, these faults must be analyzed. Using the Thevenins' theorem or the bus impedance matrix, these defects are examined phase by phase.[41]

The symmetrical faults are classified into two types:

I.3.1.1 L – L – L Fault

These kinds of faults are balanced which means the system remains balanced after the fault occurs. So, this fault rarely occurs, although it is the harsh kind of fault that holds the largest current. So, this current is used to determine the rating of the CB. [42]

I.3.1.2 L – L – L – G Fault

The 3-phase L – G fault mainly comprises all the 3- phase of the system. This fault mainly occurs among the 3-phases as well as the ground terminal of the system. So, there is a 2 to 3% of probability to occur the fault. [42]

I.3.2 Unsymmetrical faults

Unsymmetrical faults are the faults that happen in the power system network most frequently. Unsymmetrical fault currents (varying in size with uneven phase displacement) result from this type of fault. Because they result in unbalanced currents in the system, these faults are also known as imbalanced faults.

With the exception of L-L-L-G and L-L-L, unsymmetrical faults comprise both open circuit faults (single- and two-phase open state) and short circuit faults.

The three symmetrical fault types—phase or line to ground (L-G) fault, phase to phase (L-L) fault, and double line to ground (L-L-G) fault—caused by short circuit conditions are depicted in the picture below.[41]

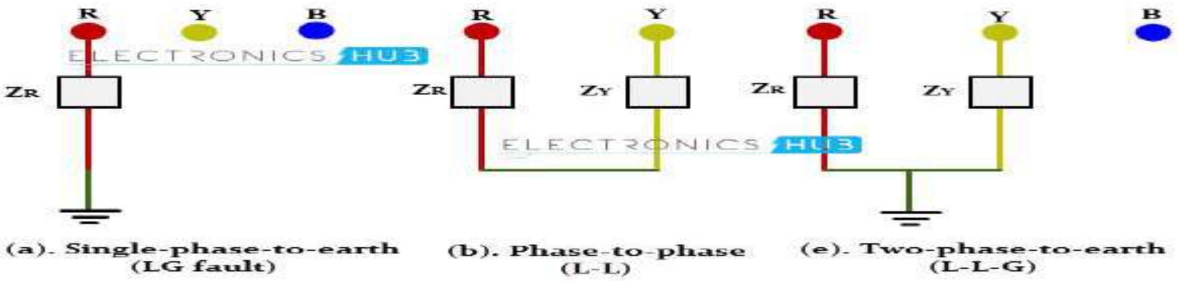


Figure I.11: Unsymmetrical faults [41]

One of the most prevalent types of failures is a single line-to-ground (LG) fault, which accounts for 70–80% of power system malfunctions. This creates a path of short circuit between ground and the line. When compared to other defects, these are far less serious faults.[41]

When two live conductors come into touch, a line-to-line fault happens. The primary cause of this problem, which allows overhead conductors to swing together, is strong winds. These are less serious errors that may occur in the range of 15% to 20%.

Double line to ground faults occur when two lines come into touch with the ground and each other. When compared to all system defects, the frequency of these severe faults isapproximately10%.

To ascertain the voltage and currents in every component of the system, unsymmetrical faults are examined through the use of unsymmetrical component techniques. Compared to symmetrical defects, these faults require more work in the analytical process.[41]
The unsymmetrical faults are classified into three types:

I.3.2.1 Single L – G Fault

The primary cause of this single L-G fault is when a single conductor starts to descend toward the ground termination. Thus, the single L-G problem accounts for between 70 and 80 percent of all power system faults.[42]

I.3.2.2 L – L Fault

This L-L fault generally happens when two conductors short circuit and when there is a lot of wind. As a result, strong winds have the ability to displace line conductors, which increases the risk of their coming into contact and shorting out. Thus, roughly 15% to 20% of the problems may develop. [42]

I.3.2.3 Double L – G Fault

This type of fault occurs when two lines come into contact with one another through the earth. Therefore, the probability of errors is 10%. [42]

I.3.3 Open circuit faults

The open-circuit faults mainly occur because of the malfunction of one otherwise more conductors used in the power system. The open-circuit faults diagram is shown below. This circuit is for 1-phase, 2- phases, and 3-phases open condition.

These faults mainly occur because of common issues like failure of joints in overhead lines, cables, failure in the phase of a circuit breaker, melting of conductor or fuse within one phase or more phases.

These faults are also known as series faults which are unbalanced types otherwise unsymmetrical types apart from 3-phase open fault.

For instance, a transmission line works through a balanced load before an open fault circuit occurs. In the transmission line, if any one of the phases gets dissolved then an alternator’s actual loading can be decreased & increases the acceleration of the alternator, so it works at a speed somewhat higher than the synchronous speed. In other transmission cables, this over speed can cause over voltages. Therefore, 1-phase and 2-phase open conditions can generate currents and voltages of the power system that causes huge damage to the apparatus. [42]

These faults are categorized into three types like following:

- Open Conductor Fault
- Two conductors Open Fault
- Three Conductors Open Fault

I.3.4 Short circuit faults

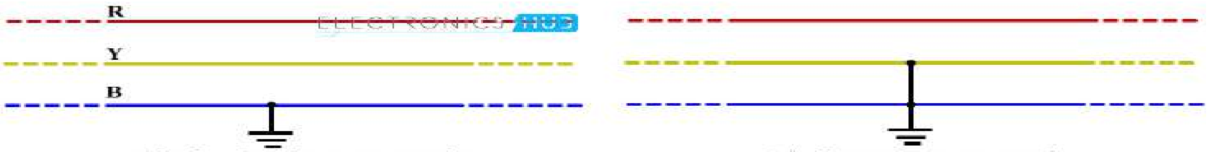
Short circuit faults mainly occur because of failure within insulation among phase conductors and earth. An insulation failure can cause a short-circuit path formation that activates short-circuit conditions within the circuit.

The definition of a short circuit is, an abnormal connection of extremely less impedance among two points of dissimilar potential, whether completed by chance or purposely. These faults are the most common types which result in the abnormal high current flow throughout the transmission lines or equipment.

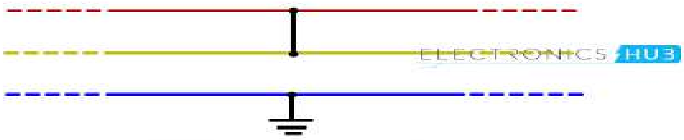
If short circuit faults are allowed to continue even for a small-time, then it leads to wide harm to the apparatus. Short circuit faults are also known as shunt faults because these faults mainly occur because of the failure in insulation among phase conductors otherwise among phase conductors and earth

The different achievable short circuit fault conditions mainly comprise 3-phases to earth, 3-phase clear of the earth, 1- phase to earth, phase to phase, 2- phase to earth, phase to phase and single-phase to earth.

Both the 3-phase fault clear of the earth, as well as the 3-phase fault toward earth, can be symmetrical or balanced while other faults are unsymmetrical faults. [42]



(d). Single-phase-to-earth (e). Two-phase-to-earth



(f). Phase-to-phase plus single-phase-to-earth

Figure I.12: Short circuit faults [41]

I.4 Causes of faults

The main reasons to cause electrical faults include the following:

I.4.1 Weather conditions

It includes lightning strikes, strong winds, rainstorms, and accumulation of snow and ice on transmission lines. It also involves salt deposition on overhead wires and conductors. Not only do these climatic conditions harm electrical infrastructure, they also interrupt the power supply. [42]

I.4.2 Equipment failures

Short circuit failures can occur in a variety of electrical equipment, including generators, motors, transformers, reactors, switching devices, and others, as a result of malfunctions, aging, cable insulation failure, and winding. High current flows through the devices or equipment as a result of these failures, severely damaging it. [42]

I.4.3 Human errors

Human mistake is another source of electrical failures. Examples of this include choosing equipment or devices with the incorrect rating, forgetting metallic or electrical conducting portions after maintenance or servicing, flipping the circuit while it is being serviced, etc. [42]

I.4.4 Smoke of fires

Sparks occur between overhead lines or between conductors to the insulator when smoke particles ionize the air around them. High voltages from this flashover cause insulators to lose their insulating ability. [42]

I.5 Effects of faults

The following factors are the main causes of electrical fault effects:

I.5.1 Over current flow

A very low impedance channel is created for the current flow when the fault occurs. As a result, the supply is drawn at an extremely high current, tripping relays and destroying insulation and equipment parts. [42]

I.5.2 Danger to operating personnel

People can potentially be shocked by a fault incidence. Depending on the current and voltage at the fault site, the shock's intensity could possibly be fatal. [42]

I.5.3 Loss of equipment

Excessive current resulting from short circuit problems causes the components to burn entirely, which affects how well the equipment or gadget functions. Equipment can occasionally become completely burned out due to intense fires. [42]

I.5.4 Disturbs interconnected active circuits

Faults disrupt the active connected circuits to the faulted line in addition to the point at which they occur. [42]

I.5.5 Electrical fires

When there is a short circuit, the air between two conducting channels ionizes, causing flashovers and sparks that ultimately result in fire, as we frequently see in the news when there are fires in shopping centers and buildings. [42]

I.6 Conclusion

Eelectrical power is the backbone of our modern world, driving progress and sustaining our daily activities. Despite its critical role, electrical failures pose significant challenges, stemming from natural disasters, technical glitches, or human factors. These disruptions can have far-reaching consequences, affecting businesses, communication, and essential services. Enhancing the resilience of our electrical infrastructure through proactive measures, investing in smart technologies, and promoting energy efficiency is essential. By addressing these challenges head-on, we can strive for a more reliable and sustainable electrical future, ensuring uninterrupted power supply for generations to come.

CHAPTER II

MACHINE LEARNING

II.1 Introduction

Machine learning (ML) and artificial intelligence (AI) are cutting-edge fields in computer science that are revolutionizing the way we interact with technology and the world around us.

Artificial intelligence refers to the development of computer systems that can perform tasks that typically require human intelligence. These tasks include problem-solving, understanding natural language, recognizing patterns, and learning from experience. AI aims to create systems that can mimic human cognitive abilities, enabling them to make decisions, solve problems, and interact with their environment in a manner similar to humans.

Machine learning is a subset of AI that focuses on the development of algorithms and statistical models that allow computers to perform specific tasks without being explicitly programmed. In other words, machine learning algorithms learn from data, identifying patterns and making predictions or decisions based on that data. This approach allows computers to improve their performance on a task over time as they are exposed to more data.

There are several types of machine learning, including supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, algorithms are trained on labeled data, where the correct answers are provided, allowing the algorithm to learn the relationship between inputs and outputs. Unsupervised learning involves training algorithms on unlabeled data, where the algorithm must discover patterns and structures on its own. Reinforcement learning involves training algorithms to make sequences of decisions by rewarding desired behaviors and punishing undesired ones.

II.2 Artificial intelligence

Machine learning's origins lie in the pursuit of artificial intelligence (AI). In the early stages of AI's development as an academic field, researchers sought ways for machines to learn from data. They explored different symbolic methods and what were then called "neural networks," mainly focusing on perceptrons and similar models, which were later recognized as variations of the generalized linear models used in statistics.[43]

Yet, a growing emphasis on the logical and knowledge-based approach created a division between AI and machine learning. Probabilistic systems encountered both theoretical and practical challenges in terms of data acquisition and representation. [44] By 1980, expert systems had risen to prominence within the field of AI, while statistics had fallen out of favor. In the 1990s, machine learning (ML) underwent reorganization and gained recognition as an independent field, experiencing significant growth. Its objective transitioned from aiming for artificial intelligence to addressing practical problems with attainable solutions. This shift moved the field away from the symbolic approaches inherited from AI, instead embracing methods and models drawn from statistics, fuzzy logic, and probability theory. [45]

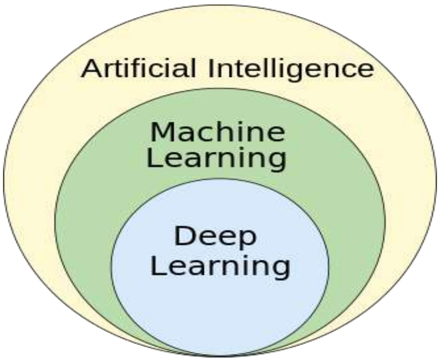


Figure II.1: Machine learning as subfield of AI [46]

II.3 Machine learning definition

Machine learning (ML) is a branch of artificial intelligence focused on creating and analyzing statistical algorithms. These algorithms are designed to learn from data, enabling them to generalize to new, unseen data and execute tasks without needing explicit instructions.[47] Lately, artificial neural networks have demonstrated the capability to outperform numerous previous methodologies in terms of performance. [48] [49]

Machine learning techniques have found application across various domains, encompassing natural language processing, computer vision, speech recognition, email filtering, agriculture, and medicine.[50][51] Machine learning is recognized in business settings as predictive analytics, where it's extensively used to tackle various challenges. While not all machine learning techniques rely on statistics, computational statistics serves as a significant source of methods within the field.

Mathematical optimization methods, also known as mathematical programming, form the mathematical basis of machine learning (ML). Data mining, which runs parallel to ML, is a related field concentrating on exploratory data analysis (EDA) through unsupervised learning techniques.[52][53]

II.4 The importance of machine learning

The vast availability of data, coupled with affordable storage solutions and the advancement of increasingly powerful processing capabilities, has driven the expansion of machine learning (ML). As a result, numerous industries are now creating more resilient models capable of analyzing larger and more intricate datasets, while also delivering quicker and more precise results on a massive scale. ML tools empower organizations to swiftly pinpoint profitable opportunities and potential risks.

The practical implementation of machine learning has a significant impact on business outcomes, often influencing a company's bottom line. With the rapid evolution of new techniques in the field, the application of ML has expanded to nearly limitless possibilities. Industries reliant on extensive datasets, requiring efficient and accurate analysis systems, have enthusiastically embraced ML as the optimal approach for model building, strategizing, and planning. [54]

II.5 Types of machine learning techniques

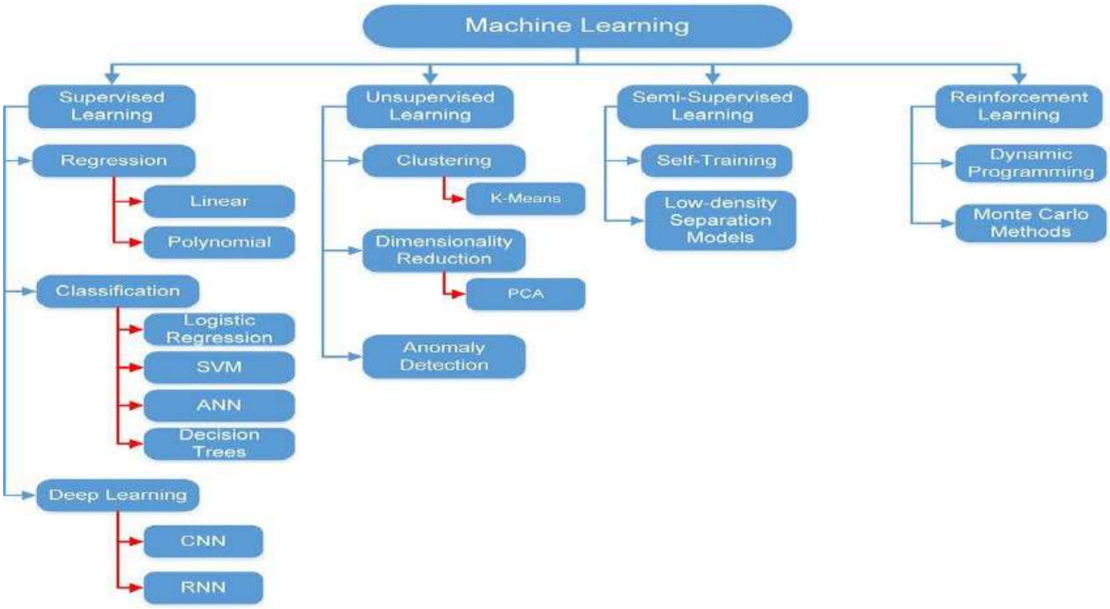


Figure II.2: Various types of machine learning techniques [55]

II.5.1 Supervised learning

Supervised learning involves overseeing or directing a specific activity to ensure it is carried out correctly. In this type of learning, the machine learns under guidance. In supervised learning, machines learn by using labeled data, where the output is already known. The model simply needs to map the inputs to the output. [56]

II.5.1.1 Supervised learning types

Supervised learning can be separated into two types: classification and regression.

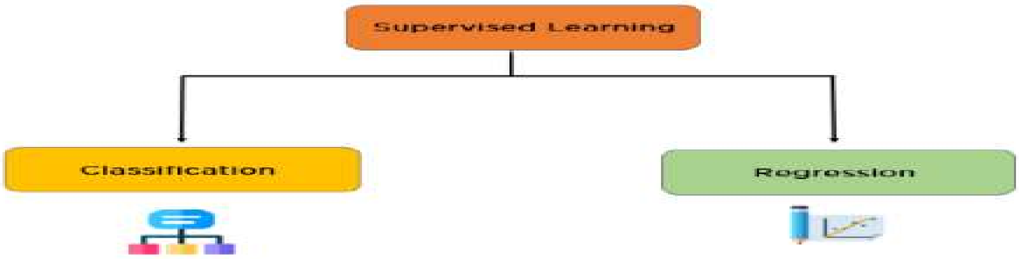


Figure II.3: Supervised learning types [57]

- **Classification:** It involves categorizing a specific dataset, regardless of whether it's structured or unstructured, into distinct categories. Typical classification algorithms include [58]
- linear classifiers.
- support vector machines (SVM).
- k-nearest neighbors (KNN)
- and random forest which are described in more detail below.
- **Regression:** Regression is a predictive modeling method employed to ascertain relationships among two or more variables. It is chiefly utilized for both prediction and causal inference purposes. Popular regression algorithms include linear regression and logistic regression.[59]

II.5.1.2 Supervised learning algorithms

Supervised machine learning encompasses numerous algorithms and computational approaches. Below is a concise overview of some of the most frequently utilized learning techniques:

1. **Neural networks:** A neural network comprises neurons organized into layers, tasked with converting input vectors into output. The procedure entails each neuron receiving input, applying a function typically nonlinear and then transmitting the output to the next layer.[60]
2. **Naive Bayes:** It is a classification algorithm rooted in Bayes' theorem, presuming independence among predictors. Essentially, it assumes that the presence of a feature in a class is independent of the presence of any other feature.[61]

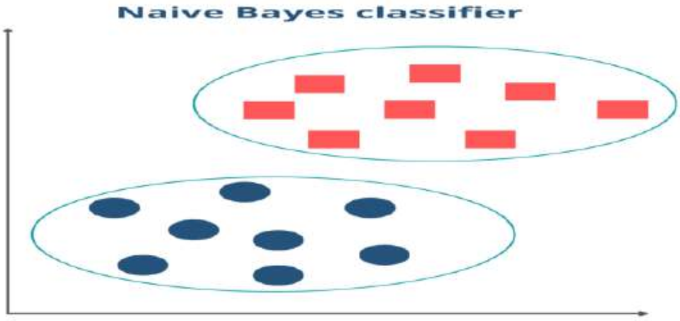


Figure II.4: Naive Bayes [62]

3. **Linear regression:** It's a machine learning classification algorithm that determines a result by utilizing one or more independent variables. Since the result is measured using a dichotomous variable, there are only two possible results [63]

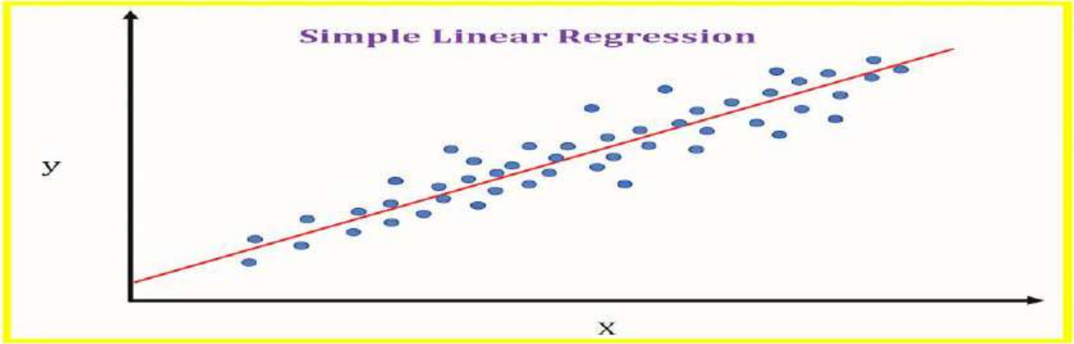


Figure II.5: Example of linear regression.[64]

4. **Logistic regression:** A classification process called logistic regression is used to determine the likelihood of altering a variable. The anticipated variable is of a dichotomous type, classified as either 1 (success/yes) or 0 (fails/no) depending on the encoded data.[65]
5. **Support vector machine:** It is one of the classification algorithms that is renowned for both its ease of use and results-guaranteed, though data extraction is still quite important. Its guiding idea is to divide data into groups as much as feasible using "simple" limits, like the maximum distance between various data groups.[66]

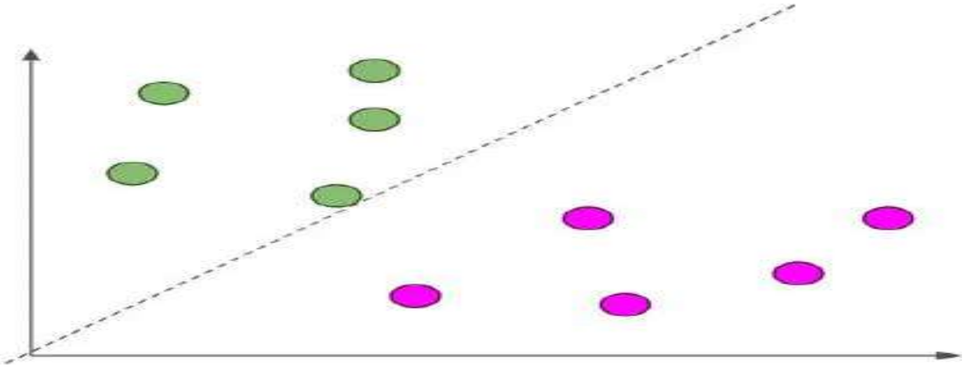


Figure II.6: Example of SVM [67]

6. **K-nearest neighbor:[**68] An approach for classification is the K-Nearest Neighbor (k-NN) algorithm. It is a straightforward algorithm that is simple to use. A set of data labeled with matching output values is given to the algorithm so that it can be trained and chosen as the prediction model. Next, fresh data can be fed into this algorithm to forecast the appropriate output values, by:
   - Select K number of neighbors.
   - Determine the distance between the unclassified point and the other points (using the techniques for estimating the distance as stated in [69]). The new point's classification among its closest neighbors.

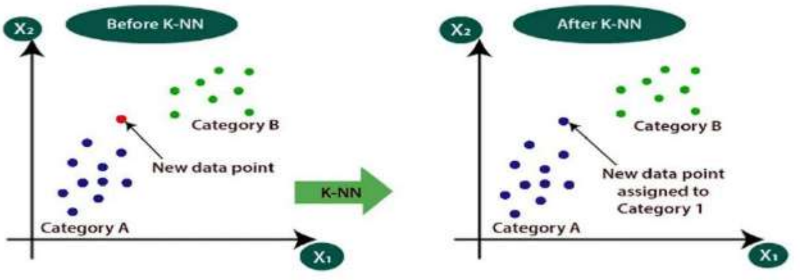


Figure II.7: Example of KNN [70]

7. **Random Forest:** They go by the name of random decision forests as well. One of the classification algorithms that works on the basis of building numerous decision trees from smaller data samples in order to identify and forecast the best tree and, ultimately, the right answer. [71]

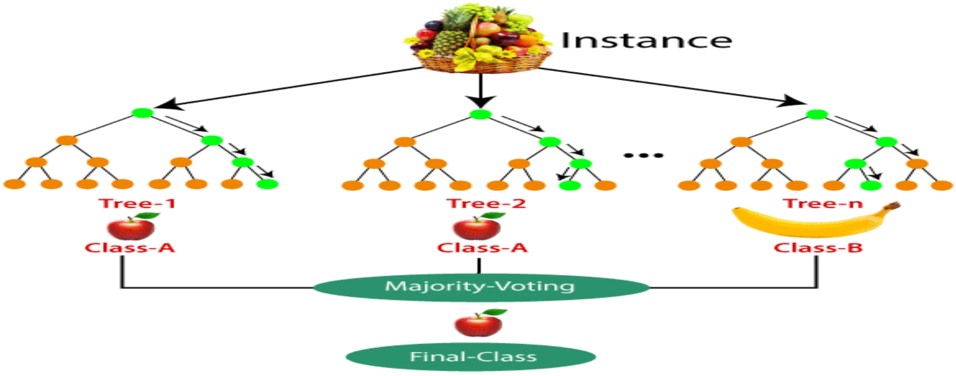


Figure II.8: Example of Random Forest algorithm.[72]

II.5.2 Unsupervised learning

As the name implies, this is one of the machine learning techniques where you don't have a supervisor to guide you or are under any form of supervision. When using this kind of machine learning, the input is unlabeled, meaning there isn't a static output variable. Instead, the model learns from the data by looking for patterns and features before producing an output.[56]

II.5.2.1 Unsupervised learning type

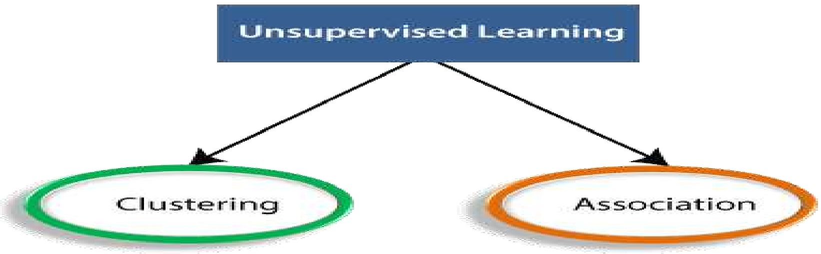


Figure II.9: Unsupervised learning type [73]

- **Clustering:** Clustering, an unsupervised machine learning method, gathers data samples and organizes them based on their proximity and similarities. Its significance lies in its ability to utilize assumptions about data similarities to segment it into various groups.[74]

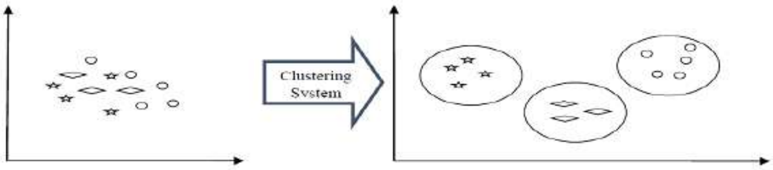
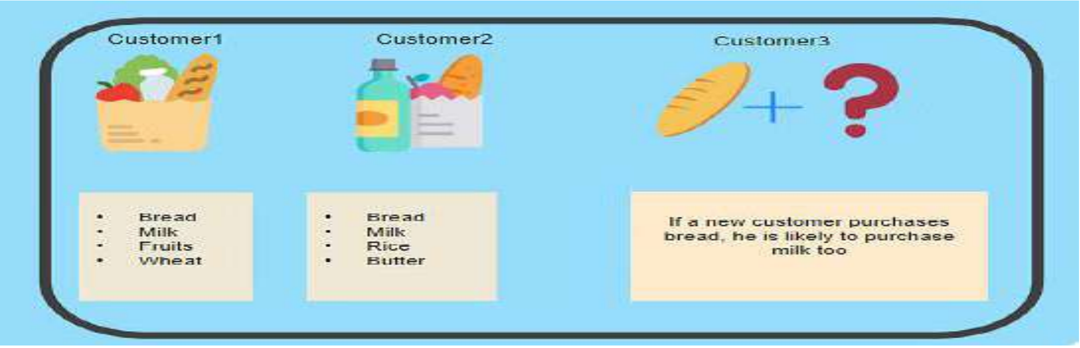


Figure II.10: Example of clustering algorithm.[75]

- **Association:** Association is a rule-based machine learning technique used to ascertain the likelihood of items co-occurring within a dataset. For instance, it can identify which products are commonly bought together. Consider a scenario in a supermarket where one customer purchases bread, milk, fruits, and wheat, while another buys bread, milk, rice, and butter. If a subsequent customer buys bread, it's probable they'll also purchase milk. This insight into customer behavior allows for the establishment of relationships and informs recommendations.[76]



If a new customer purchases bread, he is likely to purchase milk too

Figure II.11: Example of association algorithm.[76]

II.5.3 Semi-supervised learning

Semi-supervised learning, within machine learning, merges aspects of both supervised and unsupervised techniques by leveraging both labeled and unlabeled datasets to train AI models for classification and regression tasks.

Semi-supervised learning, typically applied in scenarios where supervised learning methods are used, stands out due to its utilization of techniques that integrate unlabeled data alongside the labeled data essential for traditional supervised learning.

Semi-supervised learning becomes particularly pertinent when acquiring a substantial volume of labeled data proves challenging or costly, while unlabeled data is readily available in abundance. In such contexts, neither fully supervised nor unsupervised learning methods suffice to offer satisfactory solutions.[77]

II.5.3.1 Semi-supervised learning type

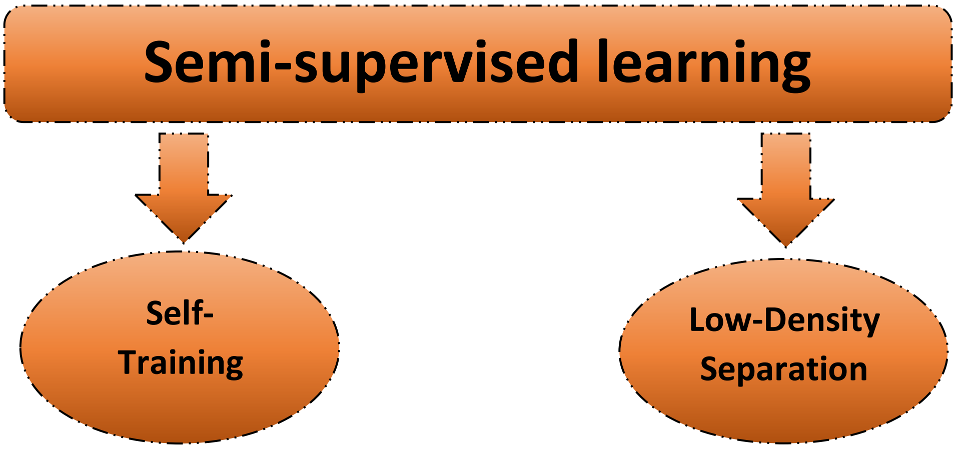


Figure II.12: Semi-supervised learning type

- **Self-Training:** Self-training is a machine learning technique that involves iteratively training a model using a combination of labeled and unlabeled data. Initially, the model is trained on a small set of labeled data. Subsequently, it predicts labels for unlabeled data, treating these predictions as pseudo-labels. These pseudo-labels are then combined with the original labeled data to create a larger labeled dataset. The model is retrained using this expanded dataset, incorporating both the original labeled data and the newly labeled data. This process is repeated across multiple iterations, gradually enhancing the model's performance by leveraging additional unlabeled data to augment the labeled dataset. Self-training is commonly employed in scenarios where acquiring significant amounts of labeled data is challenging or time-consuming, as it utilizes the model's predictions to enrich the training data.[78]

- **Low-Density Separation:** In situations where there's a limited number of labeled data points and a vast pool of unlabeled data, semi-supervised learning (SSL) aims to utilize the unlabeled data's arrangement to enhance classifier performance beyond what supervised methods achieve with labeled data alone. Successful SSL relies on assumptions about the data's structure, such as the likelihood of neighboring points sharing a classification or the positioning of decision boundaries in low-density regions. When dealing with intricate and multi-dimensional data, neural networks can acquire feature embeddings. These embeddings can then be subjected to traditional SSL techniques, resulting in what we refer to as hybrid methods.[79]

II.5.4 Reinforcement learning

Reinforcement learning (RL) is a branch of machine learning (ML) that focuses on training software to make decisions aimed at achieving optimal outcomes. It mirrors the trial-and-error learning approach employed by humans to reach their objectives. In RL, actions taken by the software that contribute to achieving the desired goal are strengthened or reinforced, whereas actions that hinder progress towards the goal are disregarded or ignored.

RL algorithms operate within a reward-and-punishment framework as they analyze data. They iteratively learn from the feedback received for each action, gradually uncovering the most effective processing pathways to reach desired end results. These algorithms exhibit a capacity for delayed gratification, understanding that the best long-term strategy may necessitate short-term sacrifices. Hence, the optimal approach they uncover might entail enduring some penalties or revisiting earlier steps. RL stands as a potent technique for enabling artificial intelligence (AI) systems to attain optimal outcomes in unfamiliar or dynamic environments.[80]

II.5.4.1 Reinforcement learning type

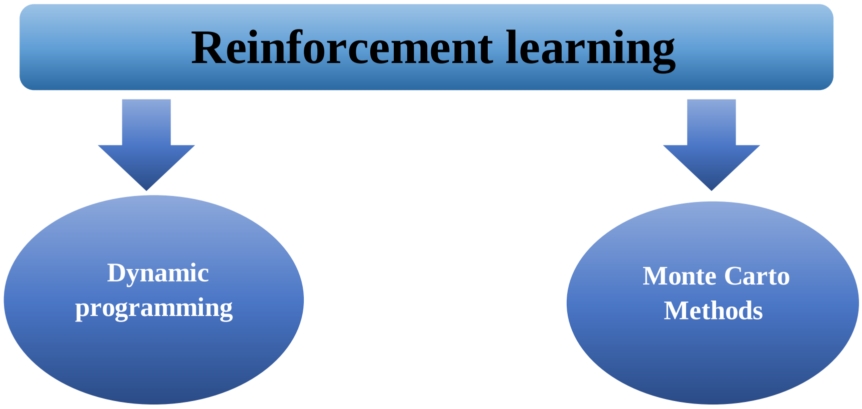


Figure II.13: Reinforcement learning type

- **Dynamic programming:** One machine learning technique that's commonly utilized to solve optimization challenges is dynamic programming. It entails decomposing a complicated problem into smaller, more manageable subproblems, solving each one just once, saving the solutions in a table (a process known as memorization), then employing those solutions again when needed to tackle bigger issues. Dynamic programming can greatly minimize the temporal complexity of solving problems, making it especially helpful in situations where there is an overlapping subproblem structure.[81]
- **Monte Carto Methods:** Monte Carlo (MC) techniques represent a category of computational algorithms employing repeated random sampling to estimate numerical values of unfamiliar parameters. They enable the simulation of intricate scenarios incorporating numerous random variables, facilitating risk assessment. MC finds diverse applications, yielding significant breakthroughs in physics, game theory, and finance. While encompassing a variety of approaches, all Monte Carlo methods fundamentally utilize random number generation to address deterministic problems.[82]

II.6 Application of machine learning

Machine learning technology has seen remarkable advancements and is increasingly integrated into everyday life through various applications. Below are examples of some of the most popular applications of machine learning:

II.6.1 Fault detection and predictive maintenance

ML models can analyze data from sensors installed in electrical equipment to detect anomalies or predict potential failures before they occur. This allows for timely maintenance, reducing downtime and preventing costly breakdowns.

II.6.2 Transportation

The transportation sector harnesses the power of machine learning in numerous ways, particularly with the emergence of autonomous vehicles. Machine learning algorithms play a pivotal role in empowering self-driving cars to maneuver through intricate surroundings by analyzing data from sensors and promptly making decisions to guarantee safe journeys. [83]

II.6.3 Robotics

Deep Learning plays a pivotal role in creating robots capable of mimicking human tasks. These robots leverage Deep Learning algorithms to continuously sense and react to obstacles in real-time, enabling them to instantly strategize their routes. This technology finds application across various domains, including transportation of goods in hospitals, factories, warehouses, as well as tasks like inventory management and manufacturing. [84]

II.6.4 Regulating healthcare efficiency and medical services

The medical sector is currently exploring the application of machine learning techniques, particularly in predicting wait times for patients across various hospital departments' emergency waiting areas. These models incorporate crucial parameters such as staff availability throughout the day, patient data, comprehensive records of department interactions, and emergency department setups. Additionally, machine learning techniques are instrumental in disease identification, treatment scheduling, and health condition analysis. This area stands as one of the most significant applications of machine learning in healthcare. [85]

II.7 Conclusion

This chapter provides an overview of machine learning and its various types, along with the algorithms used within the field. The following chapter will delve into the system we are focusing on, as well as explore the connection between machine learning and fault detection.

CHAPTER III

IMPLEMENTATION AND RESULTS

III.1 Introduction

When a malfunction arises within electrical networks, it is essential to promptly identify and resolve it to prevent harm to power system machinery and mitigate possible financial repercussions for the community.

In this chapter, we aim to simulate an electrical grid system and apply some faults to it, such as line-to-line fault (L-L), line-to-ground fault (L-G), and three-phase fault (L-L-L). We will also establish a database and then apply the machine learning algorithm (KNN) to predict and classify these faults.

III.2 Power system model

Simulation of the electrical power system model on MATLAB Simulink includes multiple lines and loads. In each experiment, we will apply different faults to the system to extract fault data. Based on this data, machine learning can predict the type of fault.

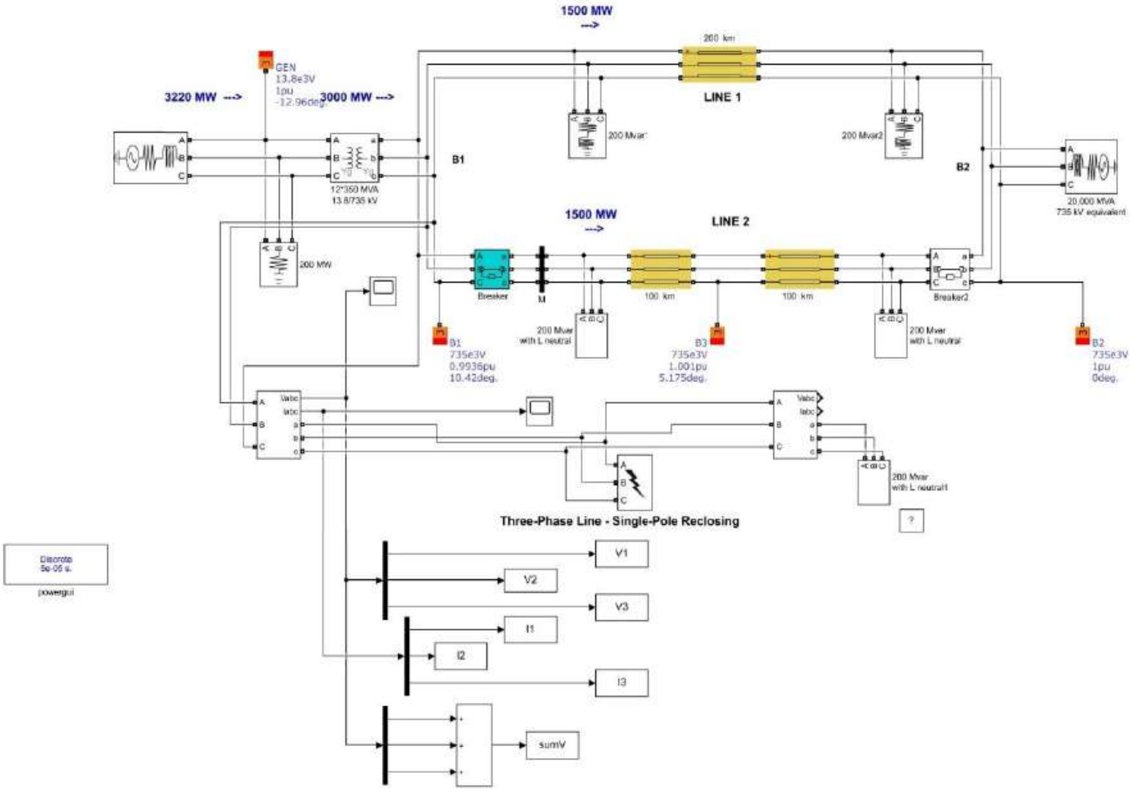


Figure III.1: Power system simulation model

III.3 Simulation results

III.3.1 Healthy system

All phases of voltage and current depicted in the graph exhibit a normal condition without any errors.

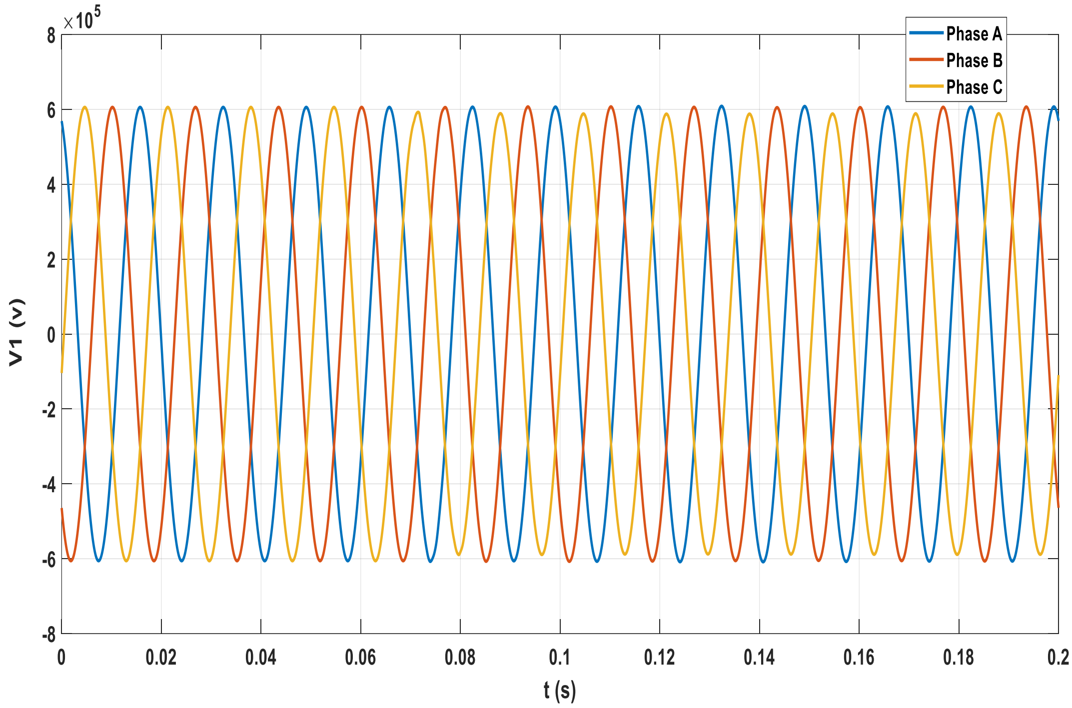


Figure III.2: Voltage of healthy system

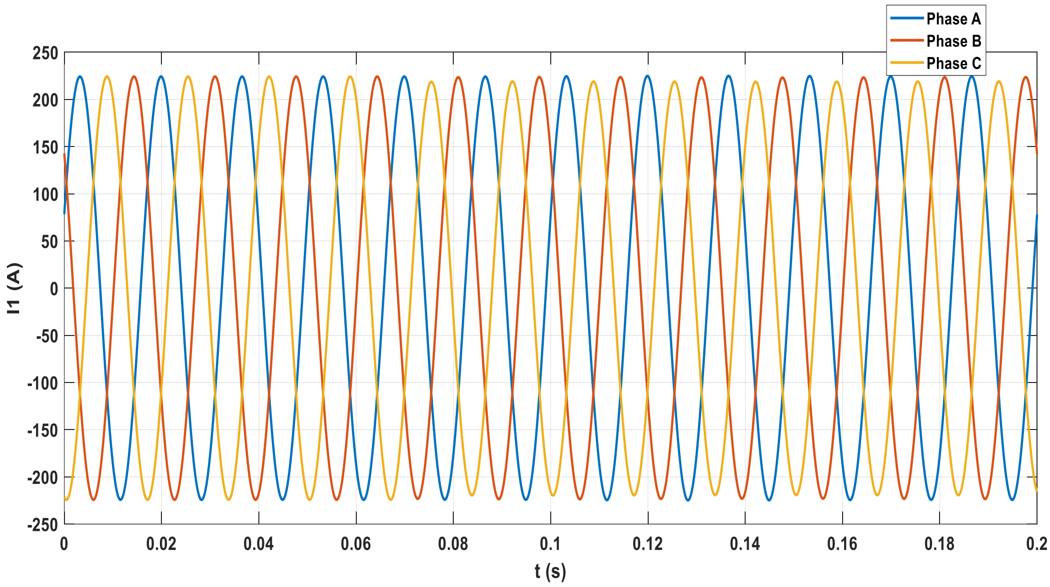


Figure III.3: Ccurrent of healthy system

III.3.2 Phase to phase (A-B) fault

The graph illustrates the characteristics of voltages and currents during a line-to-line (A-B) fault scenario. During this fault, the current reaches its peak, while the C phase remains unaffected. Conversely, the voltage increases due to the absence of grounding, with the C phase maintaining its normal condition.

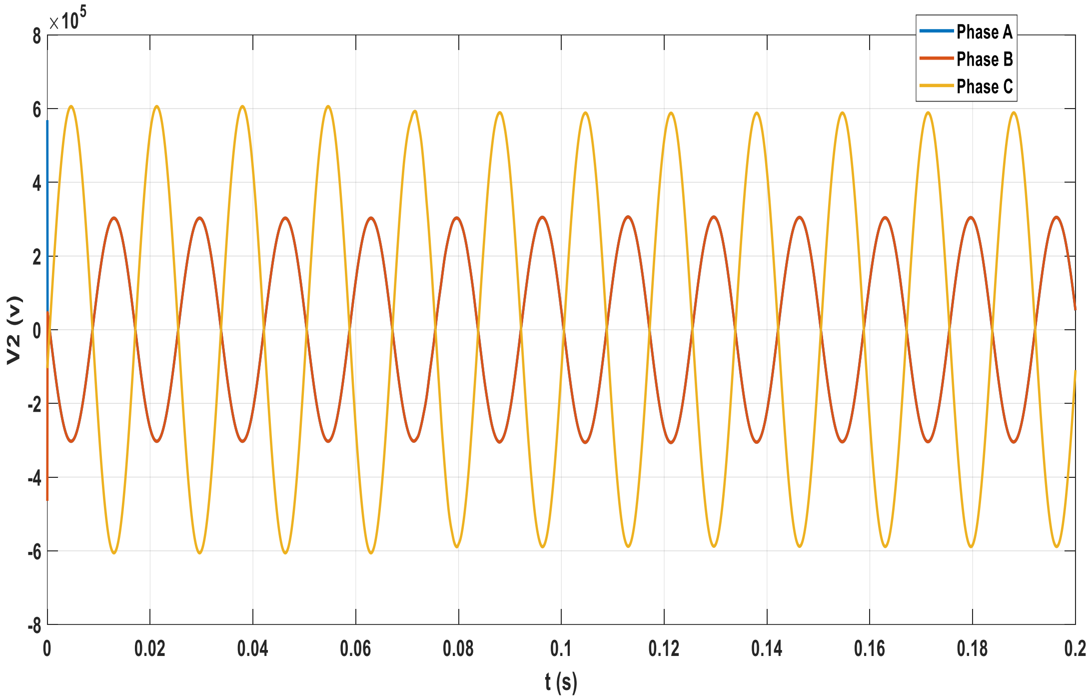


Figure III.4: Voltage of phase to phase fault (A-B)

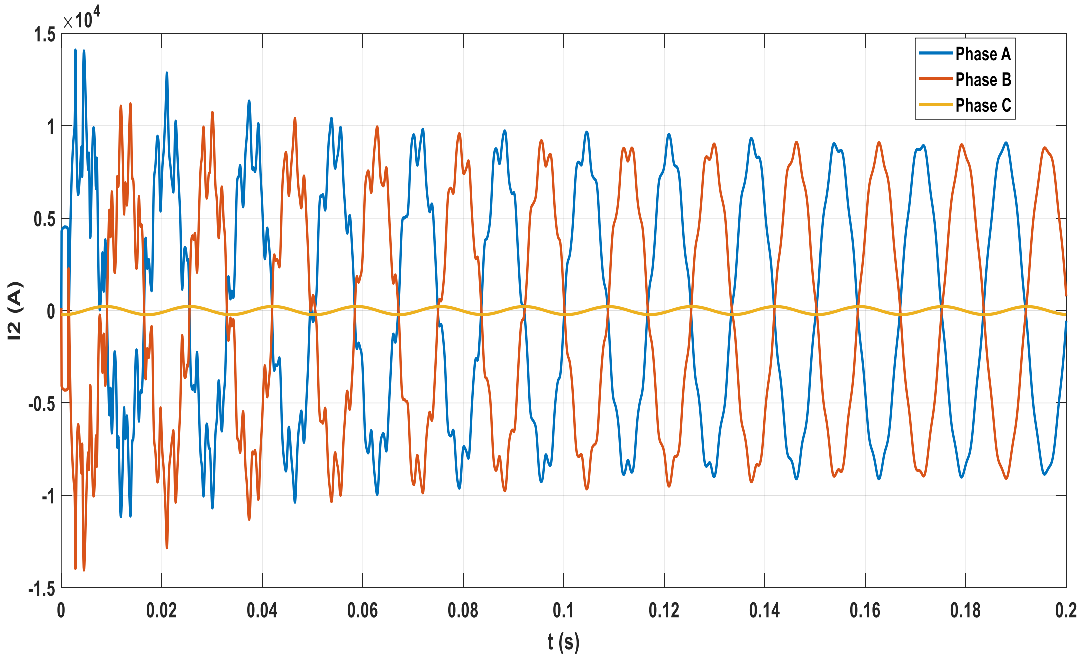


Figure III.5: current of phase to phase fault (A-B)

III.3.3 Single phase to ground (A-G)

The graph illustrates the behavior of voltages and current in a single line-to-ground fault (A-G). During this fault, the current peaks while the other phases remain unaffected. Regarding voltage, phase A drops to zero, while the other phases maintain their normal levels.

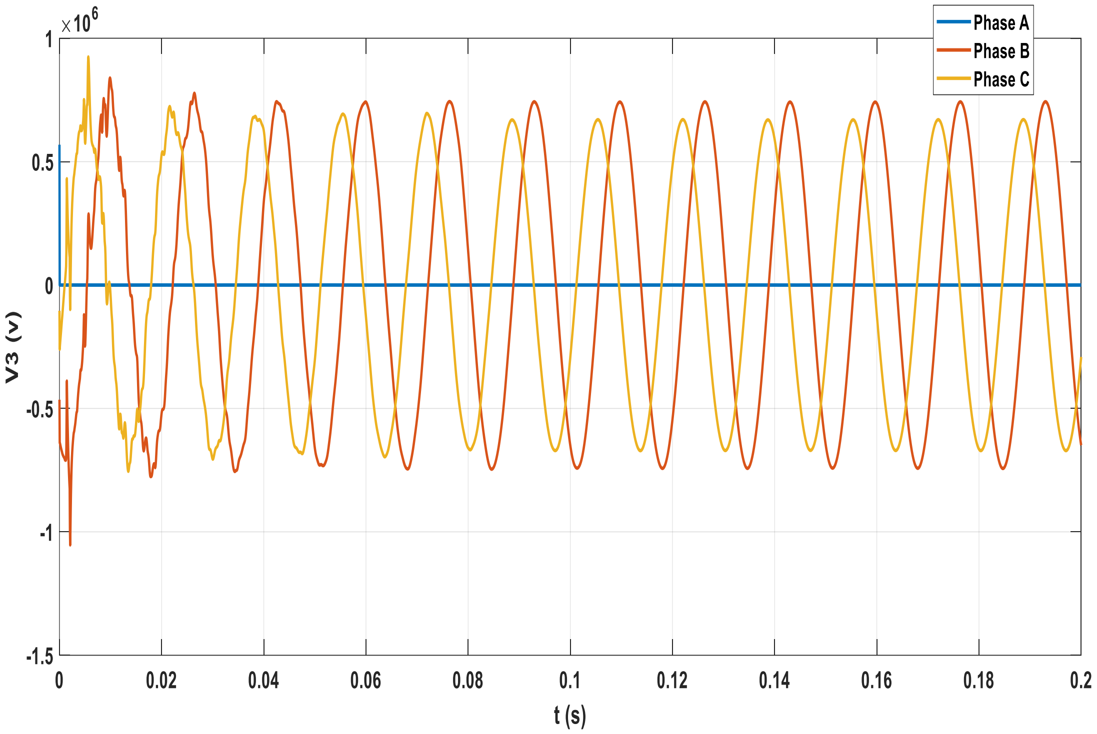


Figure III.6: Voltage of phase to ground fault (A-G)

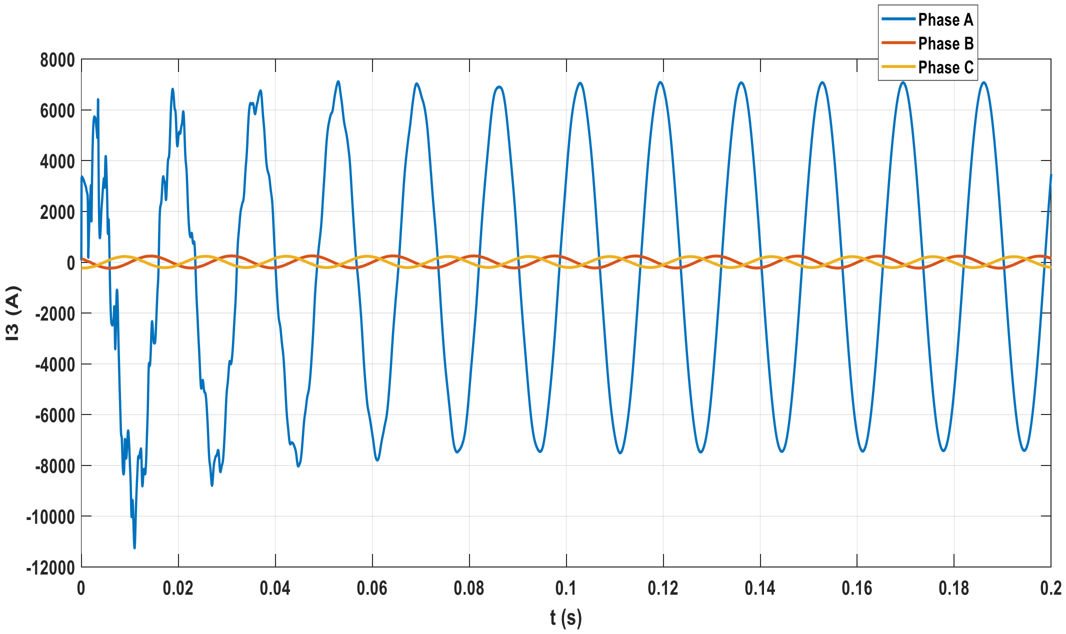


Figure III.7: Current of phase to ground fault (A-G)

III.3.4 Double phase to ground fault (AB-G)

The graph depicts the characteristics of voltages and current during a double line to ground fault (AB-G). In this scenario, the current peaks, while phase C remains unaffected. Voltage drops to zero, but phase C maintains its normal state.

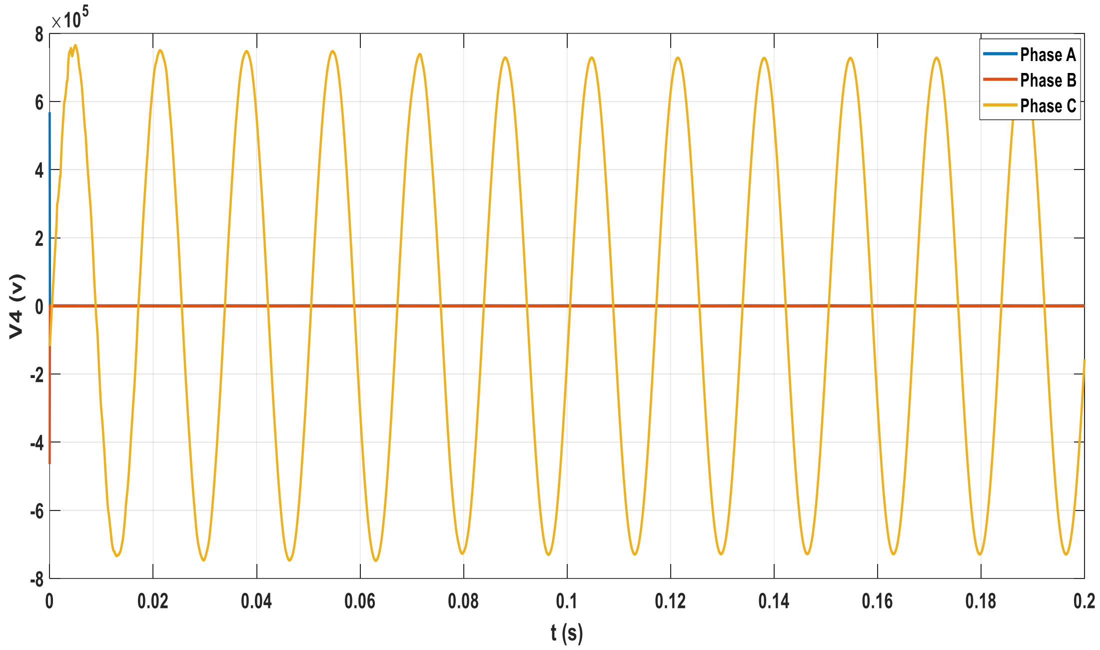


Figure III.8: Voltage of double phase to ground fault (A-B-G)

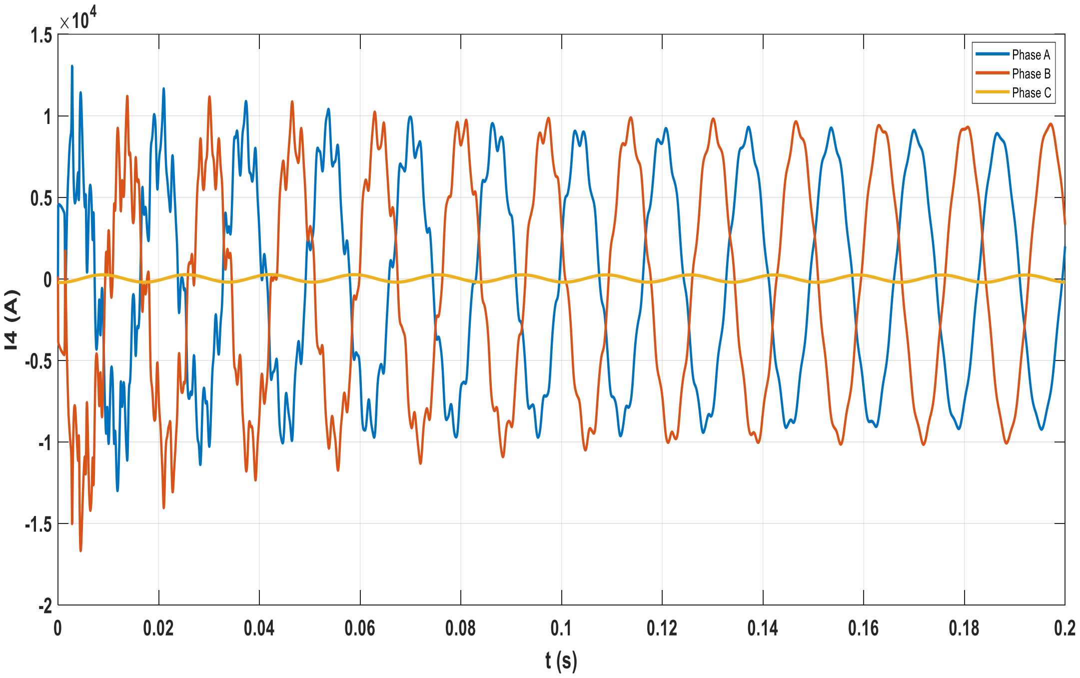


Figure III.9: Current of double phase to ground fault (A-B-G)

III.3.5 Three phase fault

The graph illustrates the behavior of voltages and current in a three-phase fault. During this fault, the current peaks, reaching its highest value, while the voltage drops to zero.

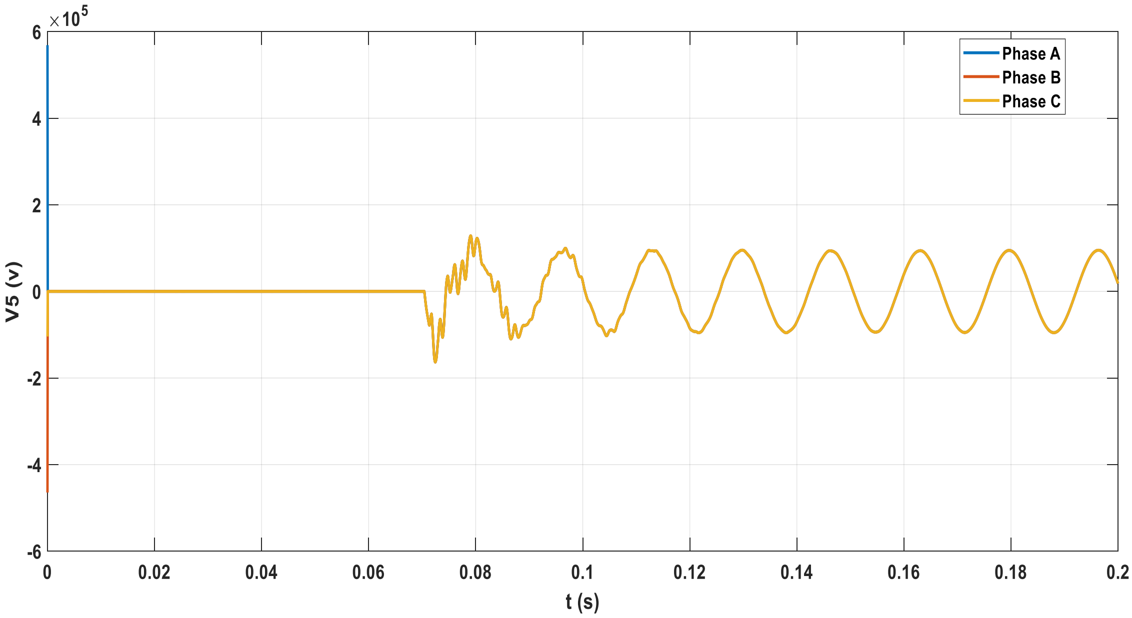


Figure III.10: Voltage of three phase fault (A-B-C)

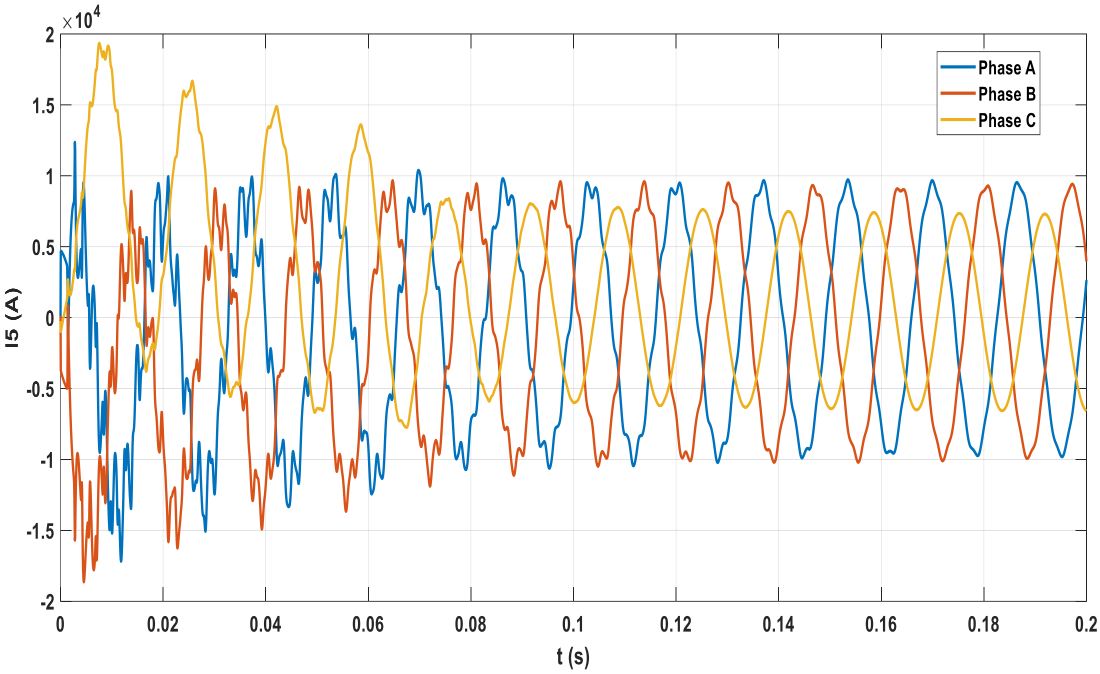


Figure III.11: Current of three phase fault (A-B-C)

III.4 Dataset

In this section, we have collected a dataset of voltages and currents in various states of the electrical system, such as the healthy state, single-phase to ground fault, phase-to-phase fault, double-phase to ground fault, and three-phase fault. Each of the aforementioned states is distinguished by a number in the output, which allows us to identify the type of fault from the output.

Table III.1: Healthy system dataset

Va	Vb	Vc	Ia	Ib	Ic	sumVabc	K1
547457.5	-465830	-81627.3	80.98813	142.8301	-223.818	-1.06E-07	0
566484.3	-471598	-94886.1	84.88781	139.5427	-224.43	-1.12E-07	0
558770.5	-478893	-79877.7	88.79993	136.2084	-225.008	-1.08E-07	0
555394.9	-485728	-69667.3	92.67739	132.8255	-225.503	-1.06E-07	0

Table III.2: Phase to phase fault dataset (A-B)

Va	Vb	Vc	Ia	Ib	Ic	sumVabc	K2
547457.5	-465830	-81627.3	80.98813	142.8301	-223.818	-1.06E-07	1
43914.29	43905.9	-87832.4	4274.572	-4050.06	-224.466	-1.22E+01	1
40506.72	40498.17	-81015.8	4358.528	-4133.44	-225.047	-1.09E+01	1
34600.96	34592.33	-69202.4	4398.163	-4172.59	-225.537	-9.14E+00	1

Table III.3: Phase to ground fault dataset (A-G)

Va	Vb	Vc	Ia	Ib	Ib	sumVabc	K3
547457.5	-465830	-81627.3	80.98813	142.8301	-223.818	-4.07E-10	2
35.22253	-637728	-253546	3285.02	139.2622	-224.691	-8.91E+05	2
35.97208	-641424	-243571	3354.872	135.8532	-225.364	-8.85E+05	2
36.05324	-645407	-228861	3363.244	132.3581	-225.972	-8.74E+05	2

Table III.4: Two phase to ground fault dataset (AB -G)

Va	Vb	Vc	Ia	Ib	Ic	sumVabc	K4
547457.5	-465830	-81627.3	80.98813	142.8301	-223.818	-4.07E-10	3
8.266243	-0.11947	-107885	4468.365	-3856.27	-224.455	-1.08E+05	3
8.146715	-0.40456	-99188.3	4542.552	-3949.42	-225.049	-9.92E+04	3
7.67438	-0.95568	-84595.4	4557.666	-4013.09	-225.553	-8.46E+04	3

Table III.5: Three phase fault dataset (A-B-C)

Va	Vb	Vc	Ia	Ib	Ic	sumVabc	K5
547457.5	-465830	-81627.3	80.98813	142.8301	-223.818	-1.06E-07	4
4.549937	-3.83578	-0.71416	4631.327	-3693.31	-938.02	1.32E-07	4
4.61014	-3.94114	-0.669	4692.479	-3799.49	-892.989	1.29E-07	4
4.603337	-4.02672	-0.57661	4685.67	-3885.08	-800.586	1.31E-07	4

III.4 Data analysis

Data processing and analysis is crucial because it generates a feature space that distinguishes between various operating conditions, specifically differentiating between healthy and faulty modes. After classifying the data collected from the system into the five cases mentioned in the previous section.

The next step is to plot the feature space to examine the data statistics and understand its characteristics. This allows us to see how all the cases are distributed, identify overlapping cases or cases with similar features, and distinguish between important and unimportant features.

Figure III.12: The figure represents the location and shape of the voltage in various states of the previously studied system.

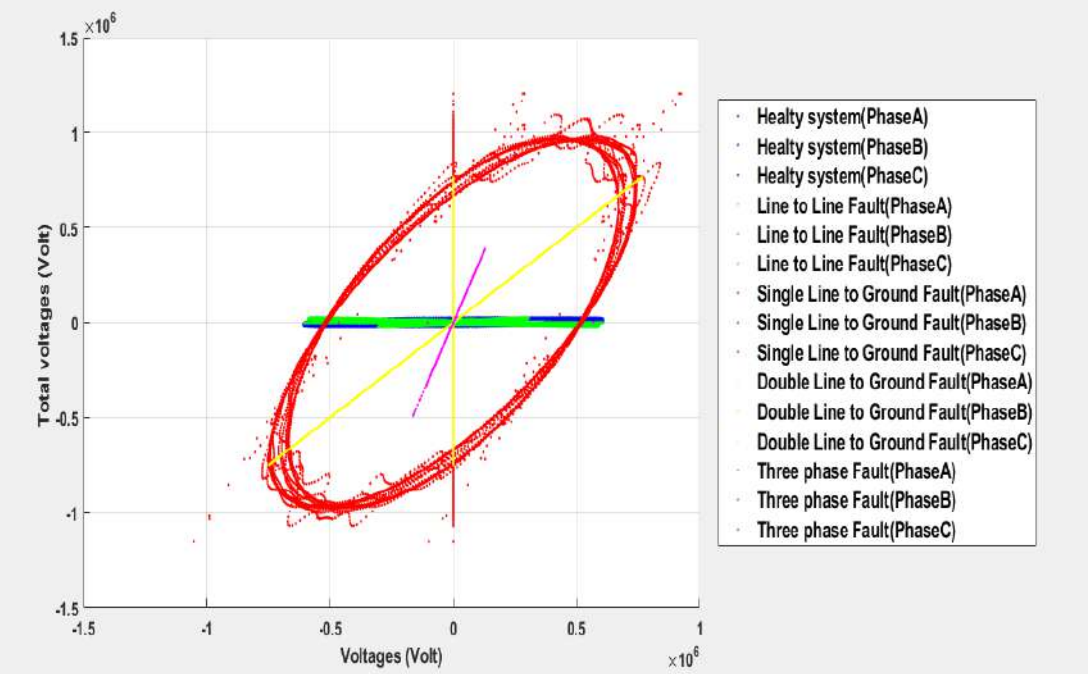


Figure III.12: Form of the voltage in different system states

Figure III.13: The feature spaces representing all healthy or faulty conditions are visualized, revealing that the feature spaces for healthy conditions appear as small systems nestled within the broader feature space. The spatial characteristics of distinct system states are clearly distinguished, although there are some areas where they overlap or intersect. To address this issue, it is recommended to introduce system status as a third axis (see figure III.14). This final figure demonstrates that all the modes are clearly separated without any ambiguity. These results indicate that these feature spaces enable the detection or localization of various faults.

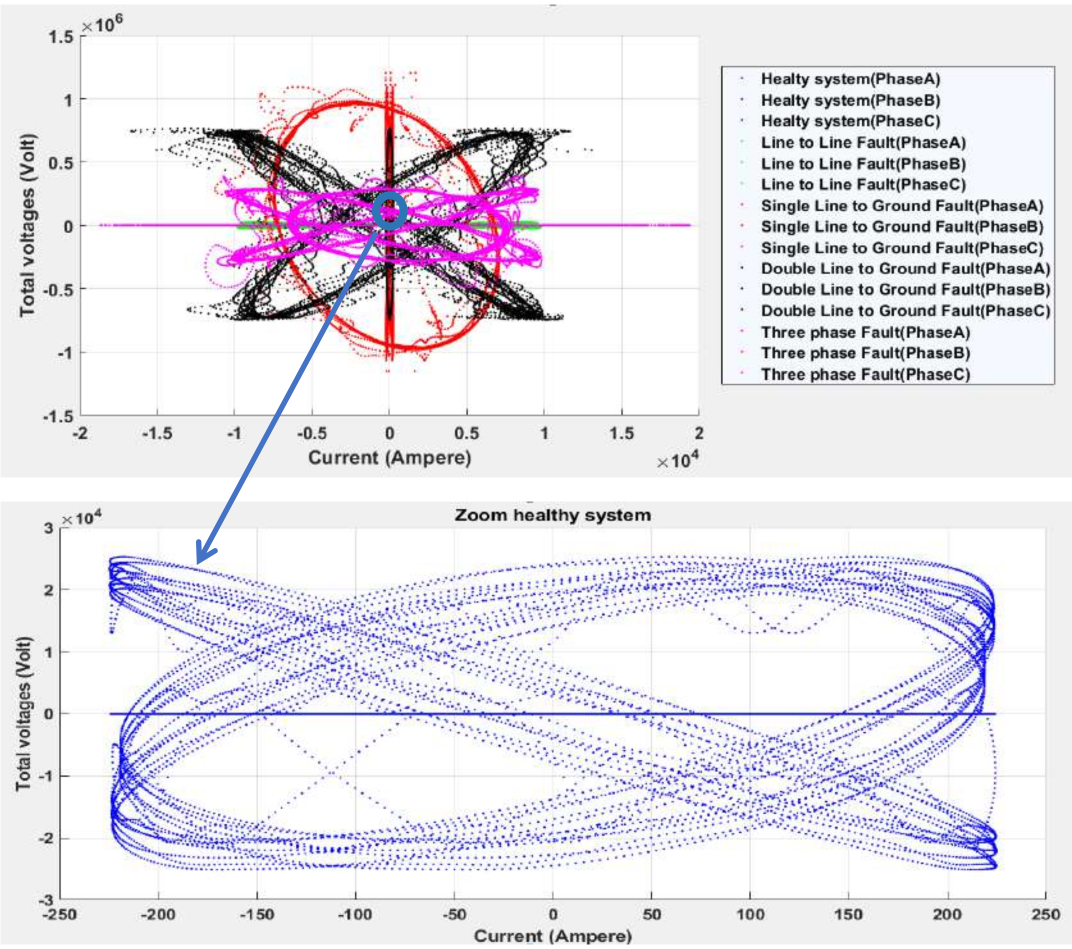


Figure III.13: Form of the current in different system states

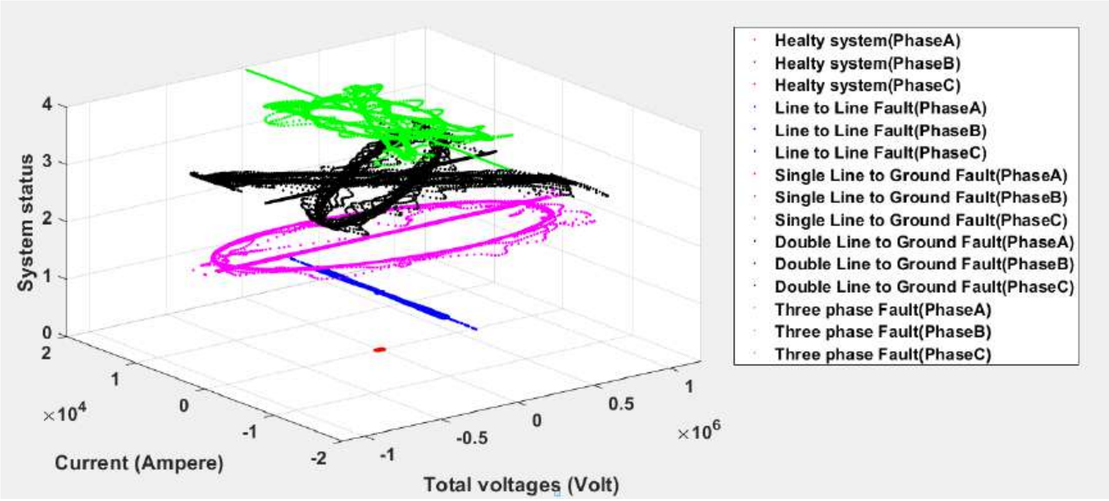


Figure III.14: 3D form of the current and total voltages in different system states

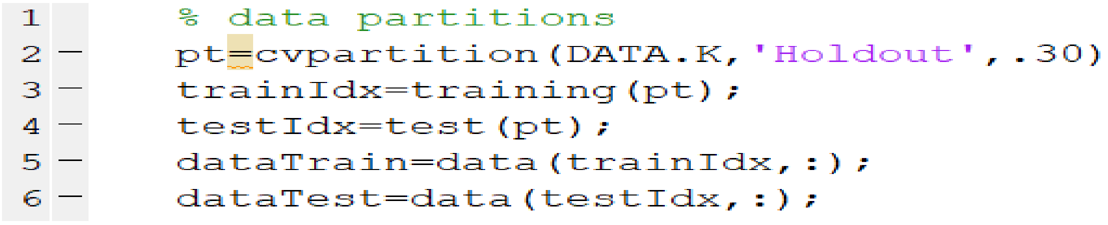
III.5 Application of KNN algorithm

The previous steps focused on preparing the data, followed by integrating the data to apply machine learning using the K-Nearest Neighbors (KNN) algorithm.

- The data was categorized into 5 classes: healthy mode and 4 faults modes.
- Labeling the data as well we use the supervised learning, we chose:
  - "0": Healthy system.
  - "1": Phase to phase fault (A-B).
  - "2": Phase to ground fault (A-G).
  - "3": Two phases to ground fault (AB -G).
  - "4": Three phases fault (A-B-C).

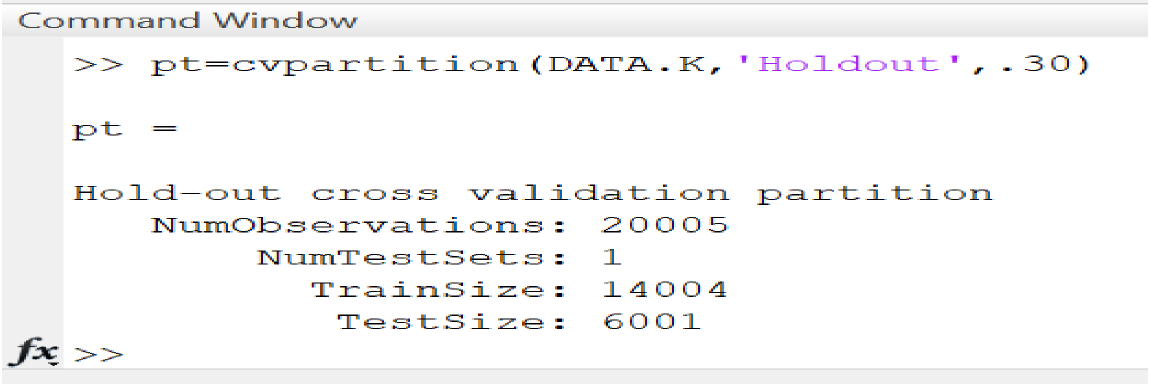
The next step which is split the data into two part the training set and the testing set:

- Training data represent 70% of all the data.
- Testing data represent 30% of all the data.



```
% data partitions
pt=cvpartition(DATA.K,'Holdout',.30)
trainIdx=training(pt);
testIdx=test(pt);
dataTrain=data(trainIdx,:);
dataTest=data(testIdx,:);
```

Figure III.15: Prepare data



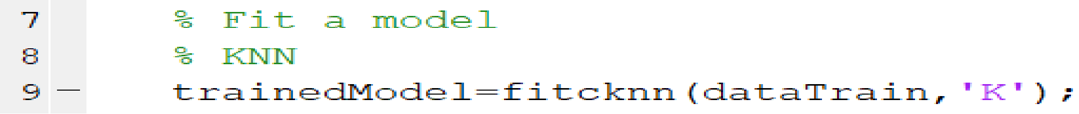
```
>> pt=cvpartition(DATA.K,'Holdout',.30)

pt =

Hold-out cross validation partition
   NumObservations: 20005
       NumTestSets: 1
         TrainSize: 14004
          TestSize: 6001
>>
```

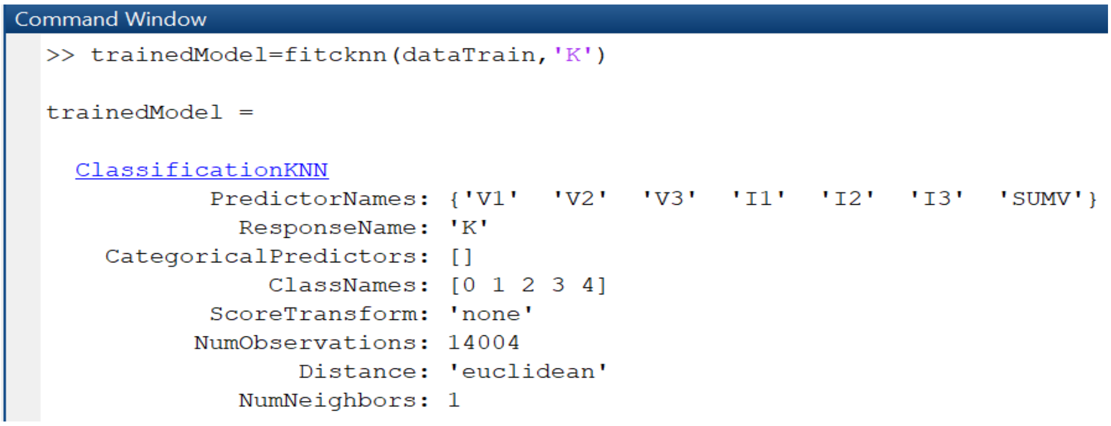
Figure III.16: Partition results

After splitting the data into the two conditions, a model was created using the K-Nearest Neighbors (KNN) method.



```
% Fit a model
% KNN
trainedModel=fitcknn(dataTrain,'K');
```

Figure III.17: Choose classifier KNN



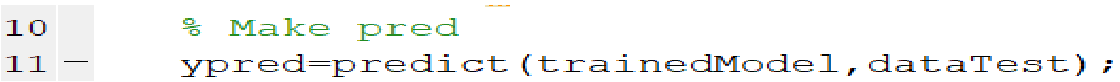
```
>> trainedModel=fitcknn(dataTrain,'K')

trainedModel =

  ClassificationKNN
           PredictorNames: {'V1'  'V2'  'V3'  'I1'  'I2'  'I3'  'SUMV'}
             ResponseName: 'K'
    CategoricalPredictors: []
               ClassNames: [0 1 2 3 4]
           ScoreTransform: 'none'
          NumObservations: 14004
                 Distance: 'euclidean'
             NumNeighbors: 1
```

Figure III.18: Model data

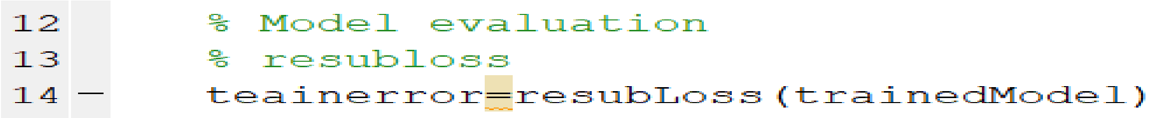
Make predictions



```
% Make pred
ypred=predict(trainedModel,dataTest);
```

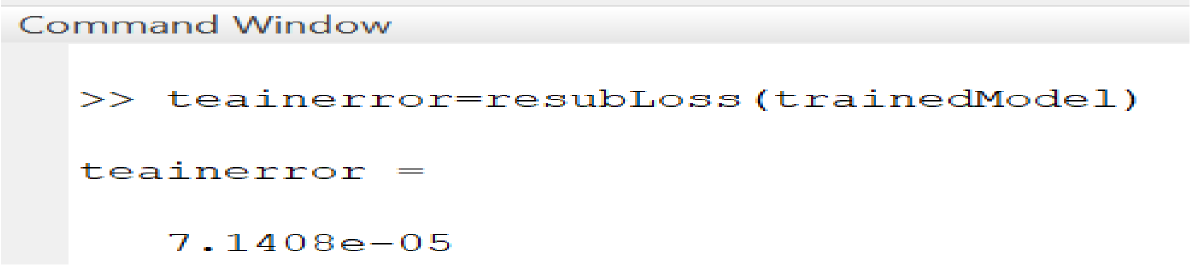
Figure III.19: Model testing

By knowing the error value of the algorithm, we can better estimate its performance and understand the extent of its accuracy in predicting data.



```
% Model evaluation
% resubloss
teainerror=resubLoss(trainedModel)
```

Figure III.20: Model evaluation



```
>> teainerror=resubLoss(trainedModel)

teainerror =

   7.1408e-05
```

Figure III.21: Mistake percentage

III.5.1 Check performance per class in the confusion matrix

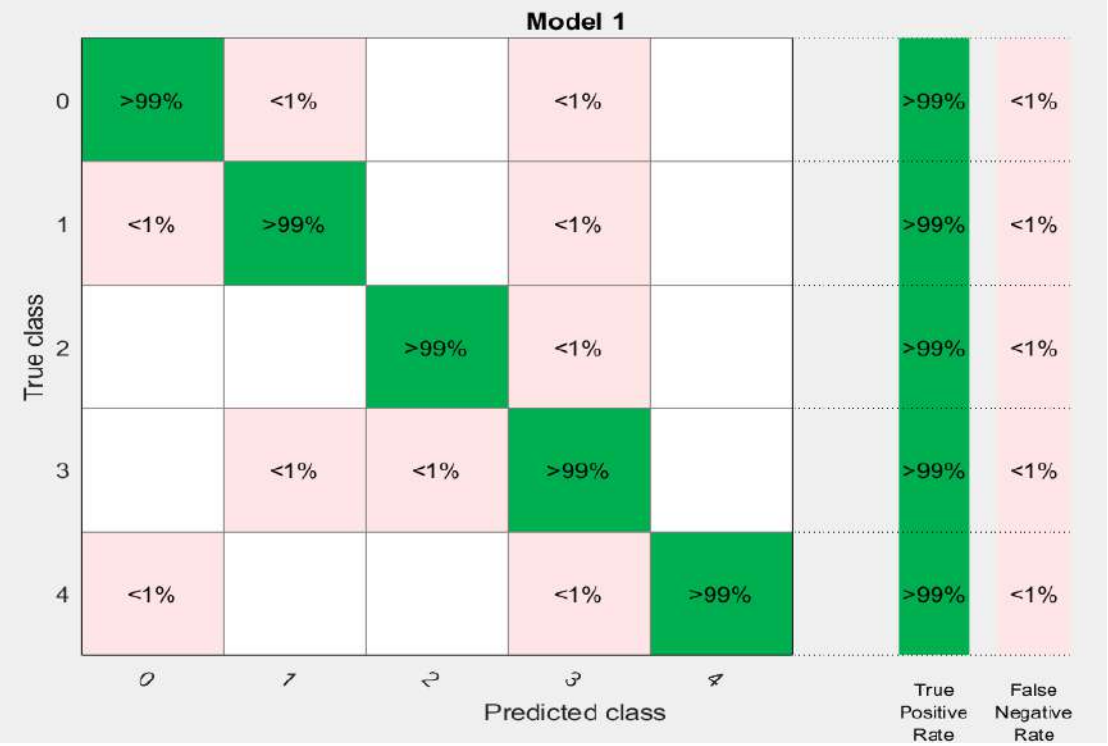
The confusion matrix represents the set of correct and incorrect classifications, where the diagonal represents the number of correct classifications. We observe that the majority of classifications are correct, while the other elements of the matrix are few in number, which demonstrates the effectiveness of the K-Nearest Neighbors algorithm.

	0	1	2	3	4
0	3999	1		1	
1	1	3998		2	
2			4000	1	
3		2	1	3998	
4	1			2	3998

Figure III.22: Confusion matrix

Figure III.23: The top row shows all the data with the correct class "0". The columns display the predicted classes. In the top row, 99% of the healthy system is correctly classified, so 99% is the true positive rate for the points correctly classified in this category, as indicated in the green cell in the true positive rate column.

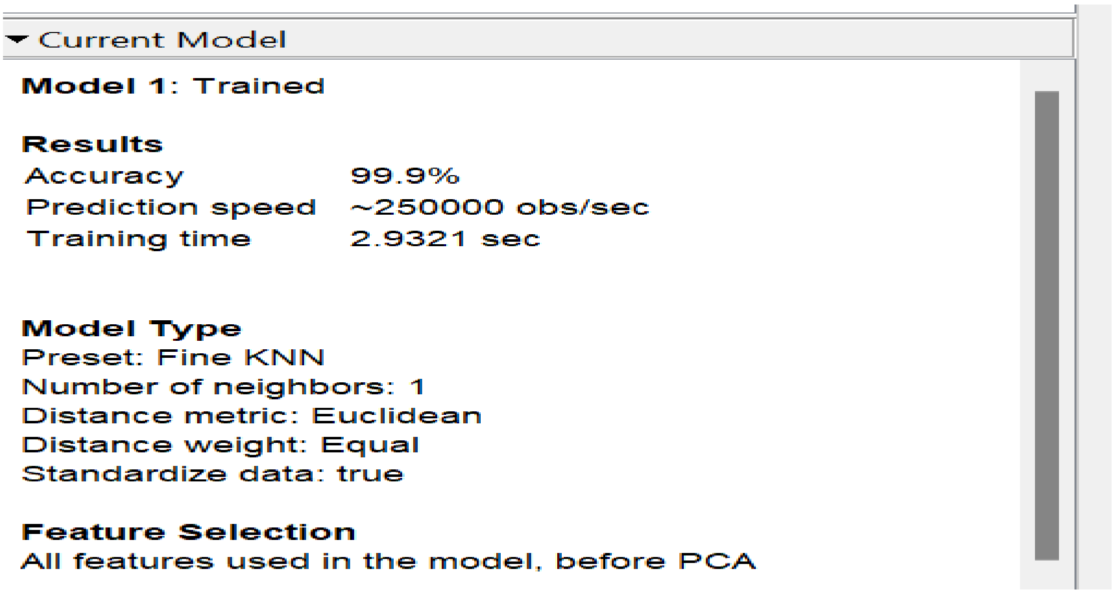
The other data in the healthy system "0" row were misclassified: less than 1% of the data were incorrectly classified as phase-to-phase fault (A-B), and less than 1% were classified as Two Phase to Ground fault (AB-G). And the total of 1% is the false negative rate for the points incorrectly classified in this category, as indicated in the pink cell in the false negative rate column.



True class	0	1	2	3	4	True Positive Rate	False Negative Rate
0	>99%	<1%		<1%		>99%	<1%
1	<1%	>99%		<1%		>99%	<1%
2			>99%	<1%		>99%	<1%
3		<1%	<1%	>99%		>99%	<1%
4	<1%			<1%	>99%	>99%	<1%

Figure III.23: The confusion matrix in percentage terms

Figure III.24: We observe that the accuracy reaches 99.9%, the prediction speed reached 250,000 observations per second, and the training time took 2.9321 seconds.



Model 1: Trained

Results

Accuracy	99.9%
Prediction speed	~250000 obs/sec
Training time	2.9321 sec

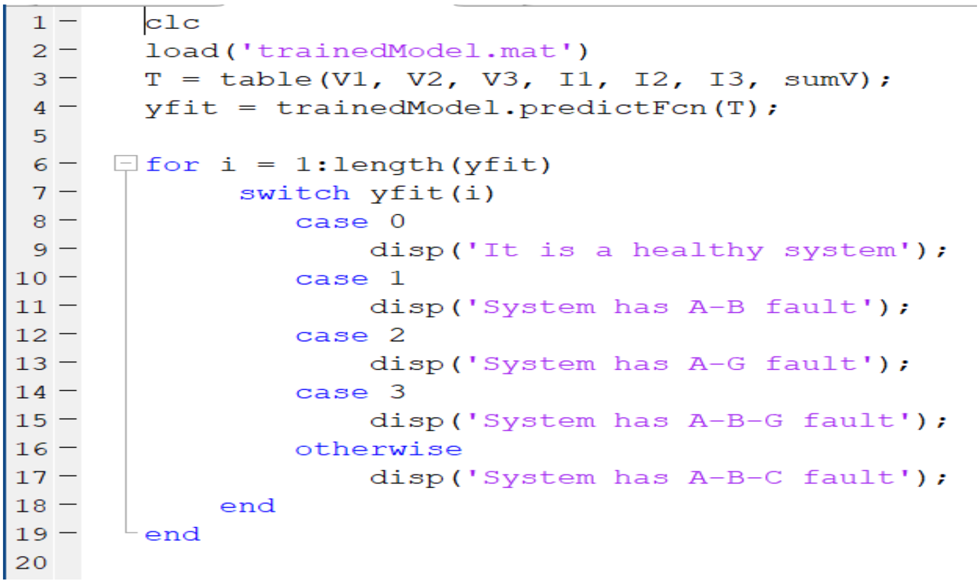
Model Type
Preset: Fine KNN
Number of neighbors: 1
Distance metric: Euclidean
Distance weight: Equal
Standardize data: true

Feature Selection
All features used in the model, before PCA

Figure III.24: Characteristics and accuracy of the KNN algorithm

III.5.2 The fault display algorithm

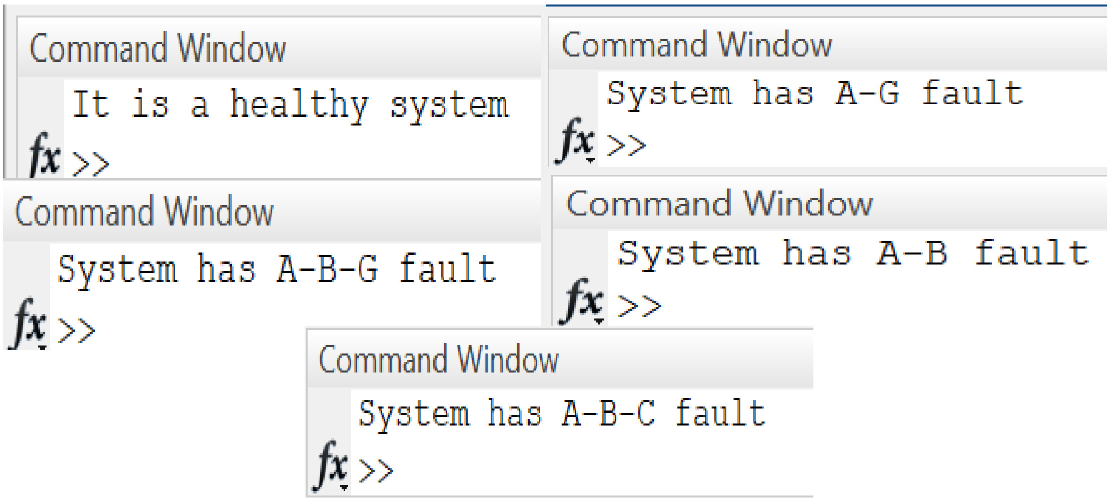
This algorithm identifies the type of fault from the provided data and displays it, making it easier for us to determine the fault category.



```matlab
clc
load('trainedModel.mat')
T = table(V1, V2, V3, I1, I2, I3, sumV);
yfit = trainedModel.predictFcn(T);

for i = 1:length(yfit)
    switch yfit(i)
        case 0
            disp('It is a healthy system');
        case 1
            disp('System has A-B fault');
        case 2
            disp('System has A-G fault');
        case 3
            disp('System has A-B-G fault');
        otherwise
            disp('System has A-B-C fault');
    end
end
```

Figure III.25: The algorithm displays the type of fault



```
It is a healthy system
fx >>
```

```
System has A-G fault
fx >>
```

```
System has A-B-G fault
fx >>
```

```
System has A-B fault
fx >>
```

```
System has A-B-C fault
fx >>
```

Figure III.26: The results display

III.6 Conclusion

Depends on factors such as the choice of distance metric, the value of k, and the distribution of the data. While KNN can achieve high accuracy in many cases, it may struggle with high-dimensional data or when there is overlap between classes. Overall, KNN remains a valuable tool in the machine learning toolkit.

The simulation results demonstrate the capability of the K-Nearest Neighbors (KNN) algorithm to identify and classify faults. The KNN algorithm technique will be beneficial for data analysis used in the electrical grid system. The KNN method has successfully detected faults in the electrical grid. It can be observed that the KNN method is capable of distinguishing, identifying, and classifying the type of fault effectively.

GENERAL CONCLUSION

General conclusion

Electric transmission lines are an essential and non-negotiable element in the electrical network, facilitating the transfer of electrical energy to customers. Additionally, their cables are uninsulated and directly exposed to harsh weather conditions, making them more susceptible to faults. These faults can cause catastrophic damage if not promptly addressed.

This work presents an opportunity to address key issues related to electrical energy, particularly within the power grid. The study was conducted following a professional project approach, ensuring the achievement of objectives related to fault detection in the electrical grid using KNN machine learning algorithms. KNN relies on a simple principle of calculating the distance between points and determining the class based on its nearest neighbors. It can also work with data that do not follow specific distributions or have complex distributions. KNN results are easy to understand and interpret, as the class is determined based on the most common classes among the nearest neighbors.

This project provided an excellent opportunity to study faults and their impact on voltage and current levels. It can have significant benefits for the electrical energy sector. By identifying the type of fault, it becomes easier to repair it quickly, aiming to protect both human elements and electrical components connected to the grid.

As a future prospect, it is important to expand the application of the K-Nearest Neighbors (KNN) algorithm for identifying fault locations in electrical power grid lines. These expansions can contribute to enhancing the efficiency and accuracy of fault detection systems, thereby reducing downtime and improving maintenance response. KNN applications will play a significant role in enhancing the safety and sustainability of electrical power grids in the future.

References

[1] Short,T.A.(2004).ElectricPowerDistributionHandbook.CRCPressLLC.ISBN0-8493-1791-6.Retrieved15November2014.

[2]ChanF.C."ElectricalPowerDistributionSystems"(PDF).www.eolss.net.EncyclopediaofLifeSupportSystems(EOLSS).Retrieved14November2014.

[3] Kwong, Y. (2015, October 9). File: Electricity grid simple- North America.svg - Wikipedia. Pinterest. https://www.pinterest.com/pin/20407004538857587/

[4] Labadi Khaled Transmission line (TL) Fault Detection using wavelet in MATLAB 2022/2023

[5]"A Primer on Electric Utilities, Deregulation, and Restructuring of U.S. Electricity Markets" (PDF). United States Department of Energy Federal Energy Management Program (FEMP). May 2002. Archived (PDF) from the original on October 9, 2022. Retrieved October 30, 2018.

[6]File:StringofElectricalPylonsinWebster,Texas.jpg.WikimediaCommons.(2018,January22). https://commons.wikimedia.org/wiki/File:String_of_Electrical_Pylons_in_Webster,_Texas.jpg

[7]Short,T.A.(2014).ElectricPowerDistributionHandbook.BocaRaton,Florida,USA:CRCPress. pp.1–33.ISBN978-1-4665-9865-2.

[8] "How Power Grids Work" . HowStuffWorks. April 2000. Retrieved 2016-03-18

[9] "The Bumpy Road to Energy Deregulation". EnPowered. March 28, 2016. Archived from the original on April 7, 2017. Retrieved April 6, 2017

[10]N.AOUZELLAGLAHAÇANI,«RÉSEAUXÉLECTRIQUES,»UniversitéA.MIRA-BEJAIA,BEJAIA

[11]Electrical,B.J.(2020,May25).ElectricalEnergyकाSupplysystemTransmission,Distribution.https://babajielectrical.blogspot.com/2019/01/transmition-distribution-system.html

[12]Boilers,T.(2023,December26).ThermalPowerPlant:HowdoesitWork,Principle&Diagram.ThermodyneEngineeringSystems.https://www.thermodyneboilers.com/components-working-thermal-power-plant/

[13]Technology,E.(2022,September29).ThermalPowerPlant–Components,WorkingandSiteSelection.ELECTRICALTECHNOLOGY.https://www.electricaltechnology.org/2021/07/thermal-power-plant.html

[14] Release, Press. "New modification of Russian VVER-440 fuel loaded at Paks NPP in Hungary".

[15] "PRIS – Home". Iaea.org. Retrieved 17 August 2023.

[16] "World Nuclear Power Reactors 2007–08 and Uranium Requirements". World Nuclear Association. June 9, 2008. Archived from the original on March 3, 2008. Retrieved June 21, 2008.

[17] "Table A.III.1 – Cost and performance parameters of selected electricity supply technologies" (PDF). The Intergovernmental Panel on Climate Change. Retrieved 20 December 2021.

[18]^ReductionofCapitalCostsofNuclearPowerPlants.OECD/NEA.8February2000.doi:10.1787/9789264180574-en.ISBN9789264171442.Retrieved20December2021.

[19] Rueter, Gero (27 December 2021). "How sustainable is wind power?".

[20] "Table A.III.2 – Emissions of selected electricity supply technologies (Gco2eq / kWh)" (PDF). The Intergovernmental Panel on Climate Change. Retrieved 20 December 2021.

[21]Markandya,Anil;Wilkinson,Paul(13September2007)."Electricitygenerationandhealth".TheLancet.370(9591):979–990.doi:10.1016/S0140-6736(07)61253-7.PMID17876910.S2CID25504602.Retrieved20December2021.

[22] "Death rates from energy production per TWh". Our World in Data. Retrieved 22 February 2022.

[23] Martin, W. (2024, March 30). Nuclear power | Definition, Issues, & Facts. Encyclopedia Britannica. https://www.britannica.com/technology/nuclear-power.

[24] "Hydropower Special Market Report – Analysis". IEA. Retrieved 2022-01-30.

[25] Renewables 2011 Global Status Report, page 25, Hydropower, REN21, published 2011, accessed 2016-02-19.

[26] De Faria, Felipe A M; Jaramillo, Paulina; Sawakuchi, Henrique O; Richey, Jeffrey E; Barros, Nathan (2015-12-01). "Estimating greenhouse gas emissions from future Amazonian hydroelectric reservoirs". Environmental Research Letters. 10 (12):124019. doi:10.1088/1748-9326/10/12/124019. ISSN 1748-9326.

[27]Jumpupto:^{ab}IEA(2022),Renewables2022,IEA,Parishttps://www.iea.org/reports/renewables-2022,License:CCBY4.0

[28] ^ "BP Statistical Review 2019" (PDF). Retrieved 28 March 2020.

[29] "Large hydropower dams not sustainable in the developing world". BBC News. 5 November 2018. Retrieved 27 March 2020.

[30] Technology, E. (2022, September 29). Hydropower Plant – Types, Components, Turbines and Working. ELECTRICAL TECHNOLOGY. https://www.electricaltechnology.org/2021/07/hydropower-plant.html

[31] Wolfe, Philip (17 March 2020). "Utility-scale solar sets new record" (PDF). Wiki-Solar. Retrieved 11 May 2010.

[32] "Concentrated solar power had a global total installed capacity of 6,451 MW in 2019". Helio CSP. 2 February 2020. Retrieved 11 May 2020.

[33] Technology, E. (2022, September 29). Solar Power Plant – Types, Components, Layout and Operation. ELECTRICAL TECHNOLOGY. https://www.electricaltechnology.org/2021/07/solar-power-plant.html

[34] Robert Gasch, Jochen Twele (editors). Wind Power Plants: Fundamentals, Design, Construction and Operation. Springer, 2011. p.11

[35] Watts, Jonathan & Huang, Cecily. Winds Of Change Blow Through China As Spending On Renewable Energy Soars, The Guardian, 19 March 2012, revised on 20 March 2012. Retrieved 4 January 2012.

[36]^Fahey,Jonathan.InPictures:TheWorld'sBiggestGreenEnergyProjects,Forbes,9January2010.Retrieved19June2019.

[37] Kanter, Doug. Gansu Wind Farm, Forbes. Retrieved 19 June 2019.

[38] "World's largest offshore wind farm fully up and running". offshorewind.biz. 30 January 2020. Retrieved 27 December 2020.

[39]Technology,E.(2022,September30).WindPowerPlant–WindTurbines,Generators,SiteSelection&SchemeofGeneration.ELECTRICALTECHNOLOGY.https://www.electricaltechnology.org/2021/08/wind-power-plant.html

[40] T, A. (2017, March 11). Types of Faults in Power System. Circuit Globe. https://circuitglobe.com/types-of-faults-in-power-system.html

[41] Teja, R. (2024, March 25). Types of Faults in Electrical Power Systems. ElectronicsHub USA. https://www.electronicshub.org/types-of-faults-in-electrical-power-systems

[42] Agarwal, T. (2021, January 17). Types of Faults in Electrical Power Systems and Their Effects. ElProCus - Electronic Projects for Engineering Students. https://www.elprocus.com/what-are-the-different-types-of-faults-in-electrical-power-systems/

[43] Sarle, Warren S. (1994). "Neural Networks and statistical models". SUGI 19: proceedings of the Nineteenth Annual SAS Users Group International Conference. SAS Institute. pp. 1538–50.

[44] Russell, Stuart; Norvig, Peter (2003) [1995]. Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall. ISBN 978-0137903955.

[45] Langley, Pat (2011). "The changing science of machine learning". Machine Learning. 82 (3): 275–9. doi:10.1007/s10994-011-5242-y.

[46] Sindhu V, Nivedha S, Prakash M (February 2020). "An Empirical Science Research on Bioinformatics in Machine Learning". Journal of Mechanics of Continua and Mathematical Sciences (7)

[47] Springer,Dordrecht. pp. 151–170. doi:10.1007/978-94-009-0279-4_9. ISBN 978-94-010-6610-5.

[48] IBM. Retrieved 2023-06-27."What is Machine Learning?".

[49] Zhou, Victor (2019-12-20). "Machine Learning for Beginners: An Introduction to Neural Networks". Medium. Archived from the original on 2022-03-09. Retrieved 2021-08-15

[50] Hu, Junyan; Niu, Hanlin; Carrasco, Joaquin; Lennox, Barry; Arvin, Farshad (2020). "Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning". IEEE Transactions on Vehicular Technology. 69 (12): 14413–14423. doi:10.1109/tvt.2020.3034800. ISSN 0018-9545. S2CID 228989788.

[51] Yoosefzadeh-Najafabadi, Mohsen; Hugh, Earl; Tulpan, Dan; Sulik, John; Eskandari, Milad (2021). "Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean?". Front. Plant Sci. 11: 624273. doi:10.3389/fpls.2020.624273. PMC 7835636. PMID 33510761.

[52] Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2

[53] Friedman, Jerome H. (1998). "Data Mining and Statistics: What's the connection?". Computing Science and Statistics. 29 (1): 3–9.

[54] Quantilus . Why is Machine Learning Important and How Will It Impact Business . 1 august 2020. URL : https://quantilus.com/why-is-machine

[55]categoryofmachinelearning.(n.d.).https://www.erswf.buzz/products.aspx?cname=category+of+machine+learning&cid=95

[56] Mr. Monteleone. An Introduction on Artificial Intelligence and Machine Learning. URL: https://www.modernanalyst.com/Resources/Articles/ tabid / 115 / ID / 5468 / An - Introduction - on - Artificial - Intelligence-and-Machine-Learning.aspx.

[57]Banoula,M.(2024,March18).SupervisedMachineLearning:AllYouNeedtoKnow.Simplilearn.com.https://www.simplilearn.com/tutorials/machine-learning-tutorial/supervised-machine-learning

[58]SotirisBKotsiantis,IoannisZaharakis,PPintelas,etal.“Supervisedmachinelearning:Areviewofclassificationtechniques”.In:Emergingartificialintelligenceapplicationsincomputerengineering160.1(2007),pp.3–24.

[59]GavinEdwards.MachineLearning|AnIntroduction.18november2018.URL:https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0#d3ea.

[60] Sun-Chong Wang. “Artificial neural network”. In: Interdisciplinary computing in java programming. Springer, 2003, pp. 81–100.

[61] Geoffrey I Webb, Eamonn Keogh, and Risto Miikkulainen. “Naıve Bayes.” In: Encyclopedia of machine learning 15 (2010), pp. 713–714

[62] Buvaneshwaran, K. (n.d.). Naive Bayes in Machine Learning – CopyAssignment. https://copyassignment.com/naive-bayes-in-machine-learning/

[63] Xiaogang Su, Xin Yan, and Chih-Ling Tsai. “Linear regression”. In: Wiley Interdisciplinary Reviews: Computational Statistics 4.3 (2012), pp. 275–294

[64] Sai Chandra Nerella. Linear Regression. 4July 2021. URL: https://saichandra1199.medium.com/linear-regression-1e279814e2bb.

[65] Vladimir Nasteski. “An overview of the supervised machine learning methods”. In: Horizons. b 4 (2017), pp. 51–62.

[66] Armin Shmilovici. “Support vector machines”. In: Data mining and knowledge discovery handbook. Springer, 2009, pp. 231–247.

[67] M. (2021, December 7). Math behind SVM (Support Vector Machine) - MLMath.io - Medium. Medium. https://ankitnitjsr13.medium.com/math-behind-support-vector-machine-svm-5e7376d0ee4d

[68] Oliver Kramer. “K-nearest neighbors”. In: Dimensionality reduction with unsupervised nearest neighbors. Springer, 2013, pp. 13–23

[69]SebastianRaschka.“STAT479:MachineLearningLectureNotes(2018)”.In:URLhttps://sebastianraschka.com/pdf/lecturenotes/stat479fs18/07_ensembles_notes.pdf.Citadonapág.viii38.

[70] Rupika Nimbalkar. K-Nearest Neighbors Algorithms — Machine Learning. 12 Jul 2021. URL: https://medium.com/appengine- ai/k- nearestneighbors-algorithms-machine-learning-756a7522dccb

[71] Yanli Liu, Yourong Wang, and Jian Zhang. “New Machine Learning Algorithm: Random Forest”. In: Information Computing and Applications. Ed. by Baoxiang Liu, Maode Ma, and Jincai Chang. Springer Berlin Heidelberg, 2012.

[72]Javatpoint.RandomForestAlgorithm.URL:https://www.javatpoint.com/machine-learning-random-forest-algorithm..

[73] Unsupervised Machine learning - Javatpoint. (n.d.). www.javatpoint.com. https://www.javatpoint.com/unsupervised-machine-learning

[74] Lior Rokach and Oded Maimon. “Clustering Methods”. In: Data Mining and Knowledge Discovery Handbook. Ed. by Oded Maimon and Lior Rokach. Springer US, 2005.

[75] Tutorials Point. Clustering Algorithms - Overview. URL: https : / / www . tutorialspoint.com/machine_learning_with_python/clustering_algorithms_overview.htm

[76]Banoula,M.(2023,November7).SupervisedandUnsupervisedLearninginMachineLearning. Simplilearn.com.https://www.simplilearn.com/tutorials/machine-learning-tutorial/supervised-and-unsupervised-learning

[77]WhatIsSemi-SupervisedLearning?|IBM.(n.d.).https://www.ibm.com/topics/semi-supervised-learning

[78]N,Yeshwanth.(2024,February21).SelfTraining.https://www.linkedin.com/pulse/self-training-yeshwanth-n/

[79]June7,2023,PartB.LowDensitySeparation.(n.d.).위키독스.https://wikidocs.net/190180

[80] What is Reinforcement Learning? - Reinforcement Learning Explained - AWS. (n.d.). Amazon Web Services, Inc. https://aws.amazon.com/what-is/reinforcement-learning/

[81]Zaidi,T.(2023,September25).*Howisdynamicprogrammingusedinmachinelearning*.Medium.https://medium.com/@tabish_14839/how-is-dynamic-programming-used-in-machine-learning-b56c35643a53

[82] Pease, C. (2018, September 11). An Overview of Monte Carlo Methods - Towards Data Science. Medium. https://towardsdatascience.com/an-overview-of-monte-carlo-methods-675384eb1694

[83] A. (2023, December 22). Machine Learning Applications and Examples. IABAC®. https://iabac.org/blog/machine-learning-applications-and-examples

[84]Biswal,A.(2023,November7).Top25DeepLearningApplicationsUsedAcrossIndustries.Simplilearn.com.https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-applications

[85] Electronics, K. (2022, December 22). Top 15 Machine Learning Application. Kotai Electronics Pvt. Ltd. https://kotaielectronics.com/machine-learning-application/#applications-of-successful-machine-learning