DIPLOMARBEIT

Dezentrale Erkennung von gestörtem Netzbetrieb ohne Kommunikation mit der Leitwarte

Ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Diplom-Ingenieurs (Dipl.-Ing.)
unter der Leitung von

Univ.Prof. Dr.-Ing. Wolfgang Gawlik,
Dipl.-Ing. Michael Schrammel,
Dipl.-Ing. Irina Lupandina

Institut für Energiesysteme und Elektrische Antriebe

eingereicht an der

Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von

Tobias Gasteiger BSc
1326449

Wien, im Juni 2021 _____

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Arbeitsgruppe für Elektrische Anlagen
A-1040 Wien, Gusshausstr. 25, Internet: http://www.ea.tuwien.ac.at

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mir während des Studiums beigestanden sind und mich unterstützt haben.

Besonders Bedanken möchte ich mich bei meinen Betreuern DI Michael Schrammel und DI Irina Lupandina, die mich trotz Corona Krise und der dadurch erschwerten Kommunikation großartig unterstützt haben. Die kurzen Antwortzeiten sowie eine gute Kommunikationsbasis ermöglichten ein schnelles Weiterkommen sowie einen reibungslosen Ablauf.

Ein besonderer Dank gilt ebenfalls meinen Eltern, Josef und Maria Gasteiger, für die bedingungslosen finanziellen und sozialen Rückhalt. Ohne diese Unterstützung wäre mein Leben als Student in dieser Form nicht möglich gewesen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Kurzfassung

Transport-, Übertragungs- und Verteilnetze werden mit Netzleitsystemen von der Netzleitstelle aus geführt. Die Netzleitstelle benötigt zur Führung die Informationen aus den im Netz befindlichen Umspannwerken und Schaltanlagen. Bei der Erkennung von Fehlern und Leitungsausfällen ist sie stark von einer funktionierenden Kommunikation mit diesen weit verteilten Komponenten abhängig.

Diese Arbeit beschäftigt sich in diesem Zusammenhang mit der Erkennung von Ausfällen in vermaschten Netzen nur anhand der in einem Umspannwerk zur Verfügung stehenden Messgrößen. Es werden also nicht die Daten vieler Umspannwerke ausgewertet, sondern nur die Messgrößen eines Umspannwerks herangezogen. Die Erkennung soll mit Hilfe von Machine-Learning-Algorithmen ermöglicht werden.

Zunächst wird ein Überblick über die aktuelle Führung des Netzes sowie eine kurze Einführung in das Thema Machine-Learning gegeben.
Um den Kern der Problematik etwas besser zu verstehen, wird anschließend die Fragestellung sowie der Lösungsansatz etwas detaillierter beschrieben. Weiters wird die Art und Weise der Datengenerierung, Datenanalyse sowie die Vorgehensweise der Implementation der Machine-Learning-Algorithmen beschrieben. Die Methodik wird dann an einem spezifischen Netz, dem IEEE-14-Bus System, umgesetzt.

Die Ergebnisse zeigen, dass gewisse Ausfälle in vermaschten Netzen anhand der im Umspannwerk zur Verfügung stehenden Daten erkannt werden können. Weiters wurde nachgewiesen, dass eine Zuordnung des Ereignisses zu einzelnen Betriebsmitteln möglich ist. Die Genauigkeit der Erkennung hängt allerdings stark von der Lage des Ausfalls zum Umspannwerk ab. Da nicht alle Messgrößen in Bezug auf die Ausfälle Informationen enthalten, müssen sie sorgfältig ausgewählt werden, um eine Erkennung zu ermöglichen. Bei der Klassifizierung liefert vor allem der einfache K-Nearest-Neighbor Algorithmus sehr gute Ergebnisse. Es ging bei der Arbeit bewusst darum keine Transienten heranzuziehen und lediglich „langsame“ Abtastung zu nutzen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Abstract

Transport, transmission and distribution grids are managed via grid control systems through the grid control center. To control and observe the grid the control center needs the information from the substations located within the grid. The detection of faults and line failures, it is heavily dependent on a functioning communication with these widely distributed components.

In this context, this work deals with the detection of failures in meshed grids only on the basis of the measured data available locally in a substation. Instead of evaluating the data from many substations, only the measurements from one substation are used for detection. The detection will be achieved with the help of machine learning algorithms.

First, an overview of the current management of the grid and a brief introduction to the topic of machine learning is given.
In order to understand the core of the problem a little better, the problem and the solution approach are described in more detail. Furthermore, the way of data generation, data analysis as well as the procedure of the implementation of the machine learning algorithms are described. The methodology is then implemented on a specific grid, the IEEE-14 bus system.

The results show that certain outages in meshed grids can not only be detected but also classified based on the data available at the substation. Yet, the accuracy of the detection strongly depends on the location of the outage regarding to the substation. Since not all measured variables contain information concering to the outages, they must be carefully selected to enable detection. For the classification, especially the simple K-Nearest-Neighbor algorithm is showing remarkable results.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Inhaltsverzeichnis

Kurzfassung iii

Abstract v

1 Einleitung 1

2 Theoretische Grundlagen und Beschreibung der Rahmenbedingungen 3
2.1 Netzleitttechnik . . . 3
2.1.1 IEC 61850 . . . 4
2.2 Maschinelles Lernen . . . 7
2.2.1 Überwachtes Lernen . . . 7
2.2.2 Teilüberwachtes Lernen . . . 17
2.2.3 Unüberwachtes Lernen . . . 17
2.2.4 Verstärktes Lernen . . . 18

3 Methodik 19
3.1 Allgemeine Beschreibung . . . 19
3.2 Daten . . . 23
3.2.1 Generierung der Daten . . . 23
3.2.2 Analyse der Daten . . . 25
3.3 Erkennung vom gestörten Netzbetrieb . . . 28
3.3.1 Auswahl des Algorithmus . . . 28
3.3.2 Anpassung des Algorithmus . . . 29
3.4 Umsetzung der Methodik an einem spezifischen Netz . . . 34
3.4.1 Auswahl der Ausfälle . . . 36
3.4.2 Erfassung der Betriebszustände . . . 37
3.4.3 Implementierung in MATLAB . . . 38

4 Auswertung und Validierung der Ergebnisse 41
4.1 Analyse der Daten . . . 41
4.1.1 Ausfall eines Systems des Doppelsystems zwischen SS1 und SS2 . . . 41
4.1.2 Ausfall von L23 . . . 44
4.1.3 Ausfall von L34 . . . 46
4.1.4 Gemeinsame Betrachtung der Ausfälle . . . 48
4.2 Ergebnisse der Maschine-Learning-Algorithmen . . . 50
4.2.1 Ausfall eines Systems des Doppelsystems zwischen SS1 und SS2 . . . 50
4.2.2 Ausfall von L23 . . . 53
4.2.3 Ausfall von L34 . . . 55

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.2.4 Gemeinsame Betrachtung der Ausfälle 57

5 Zusammenfassung und Ausblick **65**

Literaturverzeichnis **71**

A Anhang **73**

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

1 Einleitung

Um Transport,- Übertragungs- und Verteilnetze mit alle ihren räumlich verteilten Umspannstationen und Schaltanlagen erfassen, führen und steuern zu können sind Netzleitsysteme notwendig. Ein Netzleitsystem besteht grundsätzlich aus der Netzleitwarte, mehreren Umspannstationen und deren Schaltfeldern. Die Netzleitwarte eines Netzes (z.B. ein 110kV Netz) kommuniziert ihrerseits mit den Netzleitstellen anderer Netze sowie mit Kraftwerken. Die Komponenten solcher Netzleitsysteme sind weit verteilt. Damit diese Komponenten miteinander kommunizieren können ist die sogenannte Fernwirktechnik notwendig. Dabei kommuniziert die Netzleitwarte über Fernwirklinien mit den Umspannwerken und deren Schaltanlagen. Durch die Vielzahl an Komponenten und Prozessleitsystemen, die miteinander in Verbindung stehen, entsteht ein komplexes Kommunikationsnetzwerk. In den Umspannstationen ihrerseits befindet sich ein Stationsleitrechner, welcher mit den Feldrechnern der einzelnen Schaltfeldern kommuniziert.

Die Netzleitstellen haben also einen globalen Blick auf das von ihnen geführte Netz. Wie jedoch leicht zu erkennen ist, sind Prozessleitsysteme wie das Netzleitsystem sehr stark von der funktionierenden Kommunikation zwischen den Prozesskomponenten abhängig. Bei einer Kommunikationsunterbrechung zwischen der Netzleitstelle und einer Umspannstation ist sie in Bezug auf diesen Teil des Netzes sozusagen blind und eine Fernsteuerung ist nicht mehr möglich. Mit der steigenden Komplexität des Netzes steigt auch der Informationsverkehr zwischen Netz und Leitwarte und immer aufwendigere Datenaufbereitungs- und -analysierungsstrategien sind nötig. Problemen dieser Art kann durch dezentrale Regelung und Fehlererkennung entgegengewirkt werden.

In diesem Zusammenhang beschäftigt sich diese Arbeit damit, inwieweit es möglich ist Fehler und Ausfälle von Leitungen auch lokal in einem Umspannwerk und ohne Kommunikation mit der Warte erkennen zu können. Ein spezielles Augenmerk wird dabei auf entfernte Fehler in vermaschten Netze gelegt, die außerhalb der Schutzzonen der herkömmlichen Schutzgeräte, wie z.B. dem Distanzschutz, liegen. In radialen Netzen ist der Informationsfluss ähnlicher zum Lastfluss. Die Energieverteilung erfolgt von einem Punkt aus, somit ist ein Fehler in einem Versorgungsbereich durch die Zustände der Schutzsysteme erkennbar. Durch ihre Struktur ist diese Erkennung in vermaschten Netzen um einiges komplexer. Es stellt sich die Frage, wie weit anhand der im Umspannwerk vorhandenen Messwerten in das vermaschte Netz hineingesehen werden kann. Ist es möglich Fehler zu erkennen, die sich in einiger Entfernung vom Umspannwerk im vermaschten Netz befinden? Falls dies möglich ist erhebt sich weiter die Frage, ob die einzelnen Ausfälle unterschieden und klassifiziert werden können.

Auf Basis dieser Überlegungen wird in dieser Arbeit ein vermaschtes Hochspannungsnetz modelliert, in welchem die Auswirkungen von Ausfällen von Übertragungsleitungen im Umspannwerk gemessen werden. Es wird der Informationsgehalt der einzelnen Mess-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

größen bestimmt und infolgedessen die wichtigen Parameter definiert. Um dies zu erreichen ist eine Vielzahl an Daten notwendig. Maschinelles Lernen soll anschließend dazu dienen, mit Hilfe von den relevanten Messwerten eine Unterscheidung zwischen fehlerhaftem und Normalbetrieb treffen zu können.

Der Aufbau der Arbeit gliedert sich dabei wie folgt: Im Kapitel 2 wird zuerst ein Überblick über die aktuelle Führung des Netzes gegeben. Dabei wird vor allem auf die Art und Weise der momentanen Führung sowie auch auf die Kommunikationsarchitektur IEC61850 eingegangen. Außerdem erfolgt ein grober Überblick über die existierenden Machine-Learning-Algorithmen. Im Kapitel 3 wird die Herangehensweise zur Lösung der Problemstellung, die Methodik, erläutert. Es wird versucht einen möglichst allgemeinen Zugang zur Problematik, dem Lösungsansatz, der Erhebung der Daten sowie der Auswahl der Messgrößen und der Vorgehensweise bei der Implementierung eines Machine-Learning-Algorithmus gegeben. Dieser allgemeine Zugang wird anschließend in einem spezifischen Netz, dem IEEE14-Bus-System, umgesetzt. Im Kapitel 5 erfolgt schließlich die Analyse der erzielten Ergebnisse.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2 Theoretische Grundlagen und Beschreibung der Rahmenbedingungen

Im ersten Teil dieses Kapitels erfolgt eine Beschreibung der Netzleittechnik. Es muss geklärt werden, wie und mit welchem Kommunikationsstandard kommuniziert wird. Dies dient vor allem dazu, zu verstehen auf welchem Fundament die Arbeit aufbaut. Im zweiten Teil wird auf ein Teilgebiet der Künstlichen Intelligenz, dem Maschinellen-Lernen eingegangen. Da die Forschungsfrage mithilfe von Machine-Learning-Algorithmen gelöst wird, ist eine Beschreibung dieser Methoden unverzichtbar. Im Speziellen wird dabei auf die in der Arbeit implementierten Algorithmen eingegangen.

2.1 Netzleitttechnik

Die Transport-, Übertragungs- und Verteilnetze von elektrischen Energiesystemen werden mit Hilfe von Netzleitsystemen geführt. Netzleitsysteme bestehen grundsätzlich aus einer Netzleitstelle, mehreren Umspannstationen und deren Schaltfeldern. Die räumliche Dispersion macht die schon in der Einleitung erwähnte Fernwirktechnik mit ihren Fernwirklinien notwendig.

Im Umspannwerk (UW) selbst befinden sich Schaltanlagen, die von der Warte aus ferngesteuert werden können. Die in den Umspannstationen befindlichen Stationsleitrechner kommunizieren ihrerseits mit den jedem Schaltfeld zugeordneten Feldrechnern (Feldleittechnik). Die Kommunikation des Stationsleitrechners mit der Feldleittechnik erfolgt in der Regel über ein lokales Busnetzwerk oder ebenfalls über Fernwirklinien. Die drei Leittechniken bilden zusammen eine 3-schichtige Pyramidenstruktur [16]:

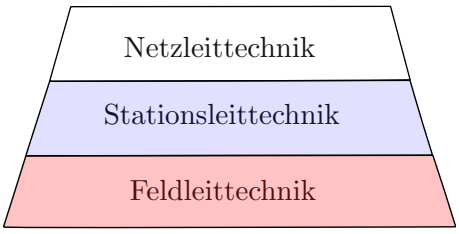


Abbildung 2.1: Pyramidenstruktur eines Netzleitsystems einer Spannungsebene

Um die Netze führen zu können verfügt die Netzleitstelle über Funktionen, die sich in SCADA-Funktionen sowie höherwertige Entscheidungs- und Optimierungsfunktionen unterteilen lassen [16]. SCADA Funktionen dienen zum einem dem Überwachen und Melden der Prozesszustände und zum anderen dem Steuern und Regeln in weiträumig verteilten Prozessen und sind somit essenziell für den Betrieb eines Prozesses bzw. Netzes. Die Netzleitstelle erhält also Informationen bzw. Meldungen vom Netz, wie z.B. „Schutz ausgelöst“,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

„Leistungsschalter ein“ etc., als auch Messwerte wie Ströme, Spannungen und Frequenz. Bei größeren Störungen wie Kurzschlüssen treffen in der Leitwarte nahezu gleichzeitig eine Vielzahl von Meldungen ein, die an das Wartenpersonal hohe Anforderung stellt [16]. Um das Personal zu entlasten, wurden auf dieser Ebene schon intelligente Alarmprozessoren eingesetzt, die mittels Expertensystemen Meldungen filtern und analysieren, um anschließend nur relevante und konzentrierte Information zu präsentieren. Gleichzeitig sendet die Netzleitstelle Steuer- (z.B. Ein-, Ausschalten von Leistungsschaltern) und Regelbefehle (z.B. Erhöhen- Vermindern der Stufe bei Regeltransformatoren) an die Umspannwerke. Die Netzleitstelle hat sozusagen einen „globalen“ Blick und Eingriffsmöglichkeiten auf das von ihr zu führende Netz (beispielsweiße ein 110kV Netz). Im Fall eines Fehlers oder einer Überlast ist es durch diese gesamtheitliche Betrachtung des Systems möglich, die Ursache der Störung ausfindig zu machen und durch entsprechende Schalthandlungen das Netz in den regulären bzw. stabilen Betrieb zurückzuführen. Zusammengefasst hat sie also die Aufgabe der Fernsteuerung und Fernüberwachung der in ihrem Netz befindlichen Umspannstationen.

Aus steuerungstechnischer Sicht ist die zweite Ebene der Pyramide (die Stationsleittechnik) für die Nahsteuerung und Nahüberwachung in den Umspannstationen zuständig. Der Leitrechner in den Stationen speichert alle Messwerte, Systemzustände und Systemfehler mit Zeitstempel. Er hat sozusagen den lokalen Blick auf das Netz. Die unterste Ebene der Pyramide, die Feldleittechnik, dient zur „Vor-Ort“ Steuerung und Überwachung einzelner Schaltfelder.

2.1.1 IEC 61850

Um die Zusammenarbeit von Geräten von verschiedenen Herstellern in Umspannstationen überall auf der Welt zu ermöglichen, ist ein globaler Standard von Nöten. Der Standard IEC61850 beschreibt ein allgemeines Übertragungsprotokoll für die Schutz- und Leittechnik in elektrischen Schaltanlagen. Die Motivation für einen globalen Standard ist neben den anwendungstechnischen Vorteilen auch die Einsparung von Kosten. Durch einen globalen Standard wird der Markt an verfügbaren Geräten größer, wodurch die Investitionskosten optimiert werden können. Außerdem kann ein Standard zur Kostensenkung der Betriebs- und Wartungskosten beitragen. Die grundlegenden Ziele des Standards IEC 61850 [4] sind

- Anwendungsbereich: Es müssen alle Funktionen unterstützt werden, die in einer Schaltanlage vorkommen
- Interoperabilität: Geräte von verschiedenen Herstellern sind in der Lage, Informationen auszutauschen und diese nutzen zu können
- Konfigurationsfreiheit: Der Standard muss unterschiedliche Betriebsphilosophien unterstützen und eine freie Anordnung von Funktionen ermöglichen (z.B. zentral oder dezentral).
- Langzeitstabilität: Der Standard muss zukunftssicher sein

Der Standard selbst ist sehr umfassend und besteht aus vielen verschiedenen Teilen. Im Folgenden wird nur das Grundprinzip des Standards kurz erläutert.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Der Ansatz des IEC 61850

Beim Standard IEC61850 werden alle Funktionen einer Schaltanlage aus Sicht der Kommunikation in kleinste Einheiten (sogenannte logische Knoten) zerlegt, die miteinander kommunizieren und auch alle zu kommunizierenden Daten enthalten. Ein Gerät kann dabei mehrere dieser logischen Knoten aufweisen. Ein logischer Knoten ist z.B. eine Drei-Phasen-Strommessung, ein Leistungsschalter oder ein Distanzschutz.
Mit standardisierten Diensten kann auf die logischen Knoten und ihre Daten zugegriffen werden oder die logischen Knoten melden sich mit standardisierten Diensten. Die Datenstruktur ist objektorientiert. Dieses Funktionsmodell ist in der Software von Geräten realisiert. Logische Knoten werden weiter in logischen Geräten gruppiert. Abbildung 2.2 soll dies veranschaulichen [11].
Die Funktionen, Strukturen und Randbedingungen der Schaltanlage sind nicht standardisiert, geben aber die Anforderungen an die Kommunikation vor.

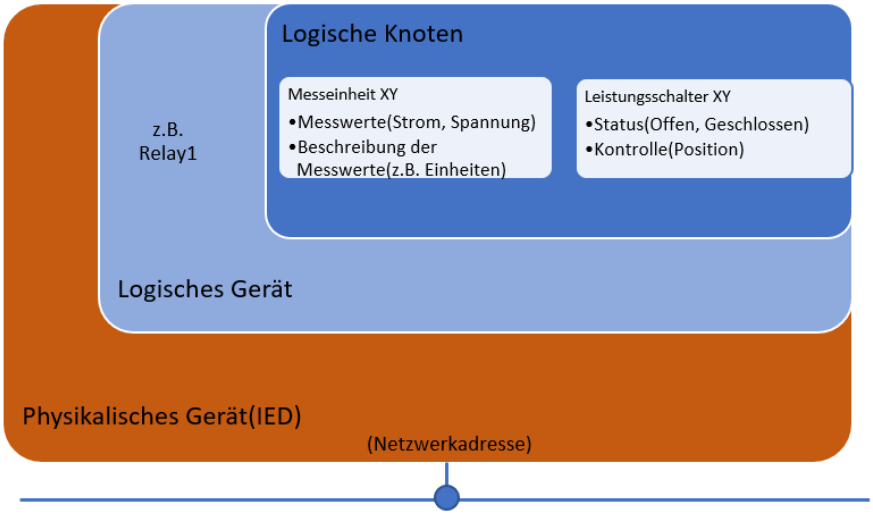


Abbildung 2.2: IEC61850 Objekt Modell

Kommunikationsarchitektur

Der IEC61850 definiert eine Kommunikationsarchitektur, die in die Prozess-, Feld- und Stationsebene unterteilt wird [8](Abb. 2.3). Die Prozessebene ist für die Datenerfassung und die Ansteuerung von Leistungsschaltern und anderem Hochspannungsequipment zuständig. Die, über die Strom- bzw. Spannungswandler erfassten analogen Signale, werden abgetastet, in digitale Form konvertiert, formatiert und schließlich an die Geräte der Feldebene gesendet.

Die Feldebene besteht aus den Intelligent-Electronic-Devices(IEDs), die Messwerte von dem Prozesslevel sammeln. Sie können lokale Kontrollentscheidung treffen und senden/-empfangen Daten anderer IEDs oder des SCADA Systems.

Die Stationsebene ist für das Monitoring, die Datenspeicherung, Steueroperationen uvw.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

zuständig und beinhaltet Informationen für SCADA und die Geräte der Feldebene. Auf dieser Ebene befinden sich Geräte wie der Stationsleitrechner, das Human-Machine-Interface (HMI), das SCADA-System und Remote-Terminal-Units (RTUs).

Für den Informationsaustausch zwischen der Feld- und Prozessebene definiert der Standard IEC61850 das SV (Sampled Value) – Protokoll. Es handelt sich dabei um eine Kommunikation vom Typ Publisher/Subscriber. Publisher senden dabei periodische Nachrichten in exakt definierten Zeitintervallen. Die Zeitintervalle hängen dabei von der Signalfrequenz und den Abtastwerten pro Periode (Samples per Period, SPP) ab. Hat das zu messende Signal eine Frequenz von z.B. 50 Hz und eine Abtastrate von 80 SPP, dann beträgt das Zeitintervall 250µs. Der IEC 61850-9-2 definiert einen Abtastrate von 80 SPP für grundlegende Schutz- und Kontrollfunktionen und eine Abtastrate von 256 SPP für Hochfrequenzanwendungen wie das Überprüfen der Versorgungsqualität oder hoch aufgelöste Oszillographie. Die Abtastrate für das 50 Hz Netz befindet sich also zwischen 4kHz und 12,8kHz [2].

Die Kommunikation zwischen der Stationsleitstelle und den IEDs (Intelligent Electronic Devices) läuft hauptsächlich nach dem bekannten Client/Server Prinzip über den Stationsbus ab, während die IEDs auch untereinander mithilfe von GOOSE (Generic Object Oriented Substation Event) Nachrichten über den Stationsbus kommunizieren. GOOSE

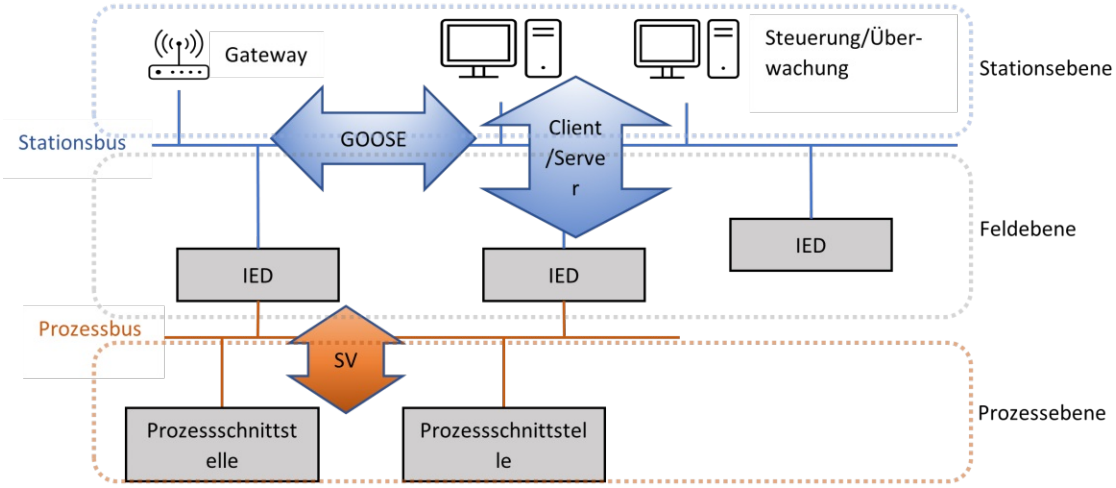


Abbildung 2.3: IEC61850 Architektur

Nachrichten ermöglichen Echtzeitkommunikation zwischen den IEDs. Sie sind Eventgesteuert und werden per Multicast nach dem Publisher/Subscriber Prinzip übertragen, d.h. die Nachrichten werden an alle IEDs am Bus gesendet, aber sie werden nur von jenen Busteilnehmern empfangen, die den Sender „abonniert“ haben.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.2 Maschinelles Lernen

Maschinelles Lernen ist ein Teilgebiet der Künstlichen Intelligenz. Beim maschinellen Lernen geht es darum, dem Computer eine Problemstellung (Daten) zu übergeben, anhand derer er sich eigenständig eine Lösung sucht. Der Computer „lernt“ also anhand der Daten die ihm zur Verfügung stehen. Je nach Bedarf ist es möglich Muster in Daten zu erkennen, Anomalien ausfindig zu machen oder auch Klassifizierungsaufgaben durchzuführen. Beim maschinellen Lernen werden die Algorithmen zuerst hinsichtlich ihrer Lernverfahren bzw. Lernmethoden klassifiziert. Abbildung 2.4 gibt einen Überblick über die existierenden Lernverfahren, welche im Folgenden kurz erklärt werden. Das Hauptaugenmerk liegt dabei auf dem überwachten Lernen, da dieses Lernverfahren im Zuge dieser Arbeit verwendet wird.

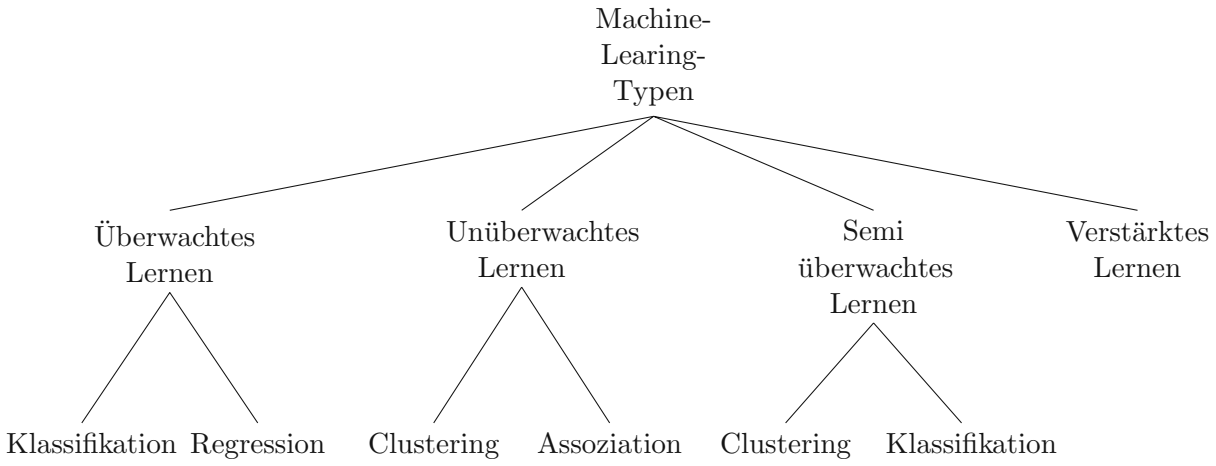


Abbildung 2.4: Übersicht Machine-Learning-Typen

2.2.1 Überwachtes Lernen

Beim überwachten Lernen werden dem Algorithmus nur kategorisierte Daten übergeben, d.h. es werden sowohl die Eingangsdaten als auch das gewünschte Ergebnis übergeben. Mithilfe dieser Daten versucht der Computer eine Verknüpfung zwischen Input und Output zu generieren. Das überwachte Lernen wird weiter unterteilt in Methoden der Regression und der Klassifikation. Die Regression zielt darauf ab, Merkmale in Abhängigkeit zu setzen. Durch das Trainieren von Regressionsmodellen erhält man Schätzfunktionen, die sich dann auf neue Daten anwenden lassen. Als Output ergibt sich dabei ein Wert (z.B.: aus bestimmten Input Daten wird der Wert eines Hauses geschätzt). Bei der Klassifikation werden die Daten in Klassen unterteilt (z.B.: „Objekt von Interesse“ oder „Kein Objekt von Interesse“). Aus mathematischer Sicht müssen sich die beiden Zugänge gar nicht allzu sehr unterscheiden. Viele Verfahren können mit nur wenig Modifikation sowohl für die Klassifikation als auch für die Regression eingesetzt werden. Der Unterschied liegt vor allem in der Art der Anwendung. Die Verfahren des überwachten Lernens sind aufgrund ihrer Strukturiertheit gut nachvollziehbar und vergleichbar. Es ist jedoch ein hoher manueller Aufwand zur Aufbereitung der Daten notwendig. Im Folgenden wird näher auf jene

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Algorithmen des überwachten Lernens eingegangen, welche später in der Arbeit verwendet werden.

K-Nearest-Neighbor (KNN)

K-Nearest-Neighbor (zu deutsch K-Nächste-Nachbarn) , oder kurz KNN, bezeichnet einen Algorithmus, der zur Klassifikation die Testdaten jeweils mit den K nächsten Nachbarn vergleicht. Bei KNN handelt es sich um keinen lernenden Algorithmus im herkömmlichen Sinn, da keine Funktion aus den Trainingsdaten gelernt wird. Er „erinnert“ sich an die Testdaten und entscheidet dann anhand dieser.

Abbildung 2.5 zeigt beispielhaft die Funktion des KNN Algorithmus. Das schwarze Kreuz stellt dabei den Punkt dar, welcher zu klassifizieren ist. Er kann entweder der Klasse Stern oder der Klasse Kreis zugeordnet werden. Wird k gleich 3 gesetzt entscheidet sich der Algorithmus für die Klasse Kreis, da seine nächsten drei Nachbarn zwei Kreise und nur ein Stern sind. Wird jedoch die Anzahl der nächsten Nachbarn gleich 5 gesetzt, so entscheidet er sich für die Klasse Stern. Die Kreise und Sterne stellen dabei die gelabelten Daten dar, mit welchen der Algorithmus trainiert worden ist. Schon aus diesem einfachen Beispiel ist ersichtlich, wie wichtig die richtige Auswahl der Anzahl der Nachbarn ist. Die Abbildung

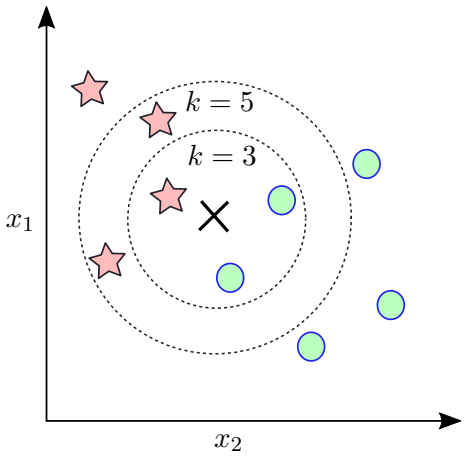


Abbildung 2.5: Idee der KNN - Methode

zeigt nur eine vereinfachte Methode der Klassifizierung durch den KNN. Will man einen KNN-Algorithmus implementieren müssen folgende Punkte berücksichtigt werden. Als Erstes muss die Art der Distanzbestimmung festgelegt werden. Es existieren verschiedene Methoden, wie die Distanz zwischen dem zu klassifizierenden Punkt und seinen Nachbarn bestimmt wird. In einem zweiten Schritt muss entschieden werden, ob jeder Nachbar gleich viel „wert“ ist, oder ob die Punkte die am nächsten zum zu klassifizierenden Punkt sind mehr zählen als jene die weiter davon entfernt sind. Dies wird mithilfe des Parameters Distanzgewicht eingestellt. Weiters kann bestimmt werden, ob die Daten standardisiert oder normal übergeben werden. Standardisierung ist eine Skalierungstechnik, wobei sich der neue Wert der Variable aus dem Mittelwert μ und der Standardabweichung σ wie folgt berechnet:

$$x_{standardisiert} = \frac{x_{normal} - \mu(\mathbf{x})}{\sigma(\mathbf{x})} \tag{2.1}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Im Folgenden erfolgt eine kurze Beschreibung der zwei Distanzbestimmungen, die in der Arbeit verwendet worden sind. Es existieren noch mehr Methoden der Distanzbestimmung, welche aber hier nicht angeführt werden.

Euklidsche Distanz:

Zur Bestimmung der nächsten Nachbarn, und somit der Daten die den Testdaten am ähnlichsten sind, können verschiedene Methoden zum Einsatz kommen. Eine populäre Methode ist die Bestimmung der euklidschen Distanz zwischen den Test und den Trainigsdaten. Die euklidsche Distanz $d_E(\mathbf{p}, \mathbf{q})$ ist dabei definiert als [14]

$$d_E(\mathbf{p}, \mathbf{q}) = \sqrt{(\mathbf{p} - \mathbf{q})^T \cdot (\mathbf{p} - \mathbf{q})} = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} \tag{2.2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \tag{2.3}$$

wobei $\mathbf{p} = [p_1, p_2, \ldots, p_n]^T$ und $\mathbf{q} = [q_1, q_2, \ldots, q_n]^T$ zwei Datenpunktvektoren darstellen. Grafisch interpretiert entspricht die Distanz der Länge der geraden Verbindungslinie zwischen den zwei Punkten p und q wie es in Abbildung 2.6 beispielhaft dargestellt ist. Ein Nachteil der Euklidschen Distanz ist, dass nur der Abstand zwischen zwei Punkten

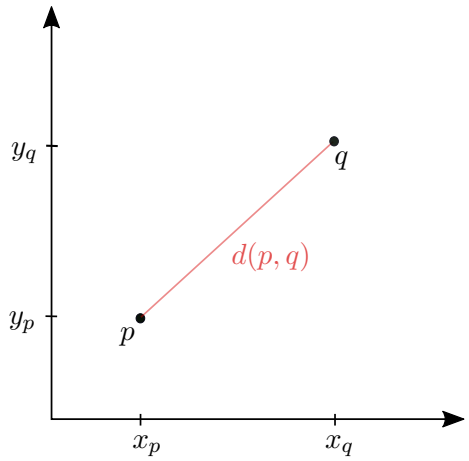


Abbildung 2.6: Euklidsche Distanz

bestimmt wird, während die Verteilung der restlichen Daten im Set keine Rolle spielt.

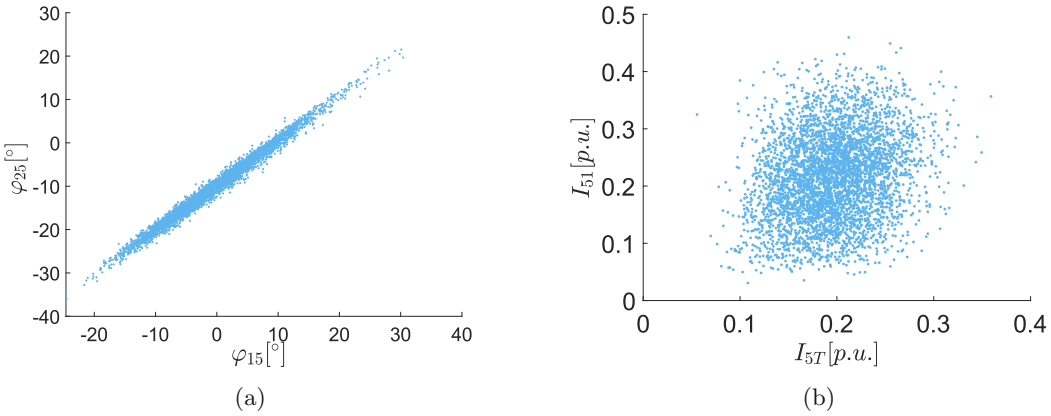
Mahanabolis Abstand:

Die zweite Methode der Abstandsbestimmung ist die Mahanabolis Distanz $d_M(\mathbf{p}, \mathbf{q})$. Im Vergleich zur Euklidschen Distanzbestimmung berücksichtigt sie die Korrelation zwischen den Variablen. Sie ist nicht der Abstand zwischen zwei Punkten, sondern der Abstand zwischen einem Punkt und einer Verteilung. Mathematisch ist sie ist definiert als [18]

$$d_M(\mathbf{p}, \mathbf{q}) = \sqrt{(\mathbf{p} - \mathbf{q})^T \cdot \mathbf{S}^{-1} \cdot (\mathbf{p} - \mathbf{q})} \tag{2.4}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

wobei $\mathbf{p}$ und $\mathbf{q}$ wiederum Datenpunktvektoren darstellen und $\mathbf{S}^{-1}$ der inversen Kovarianzmatrix der Variablen entspricht. Sind die Variablen in dem Datenset stark korreliert wie in Abbildung 2.7(a) , dann ist der Wert der Kovarianzmatrix hoch. Durch die Division wird der Wert für die Distanz reduziert. Sind hingegen die Variablen unkorreliert(Abb. 2.7(b) wird die Distanz hingegen nicht so stark reduziert. Wenn sie komplett unkorreliert



(a) (b)

Abbildung 2.7: Korrelierte(a) und unkorrelierte(b) Attribute

sind entspricht die Mahanabolis Distanz der Euklidschen Distanz.

Die Abbildung 2.8 zeigt schematisch den Unterschied zwischen der Euklidschen und der Mahanabolis Distanzbestimmung. Während bei der Euklidschen Methode die Nachbarn mit dem geringsten Abstand unabhängig von der Richtung zum Testpunkt ausgewählt werden, scheinen bei Mahanabolis Methode jene Nachbarn, die sich rechts und links vom Testpunkt befinden näher als jene die sich ober und unterhalb befinden. Dadurch kann in diesem Beispiel mithilfe der Mahanabolis Methode der Testpunkt eindeutig als grüner Kreis klassifiziert werden, während es bei der euklidschen Methode ein Kopf an Kopf Rennen der beiden Klassen ist.
Sind nun die Daten komplett unkorreliert so nähert sich die Ellipse der Mahanabolis Methode ebenfalls einem Kreis und es werden dieselben Punkte als Nachbarn ausgewählt.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

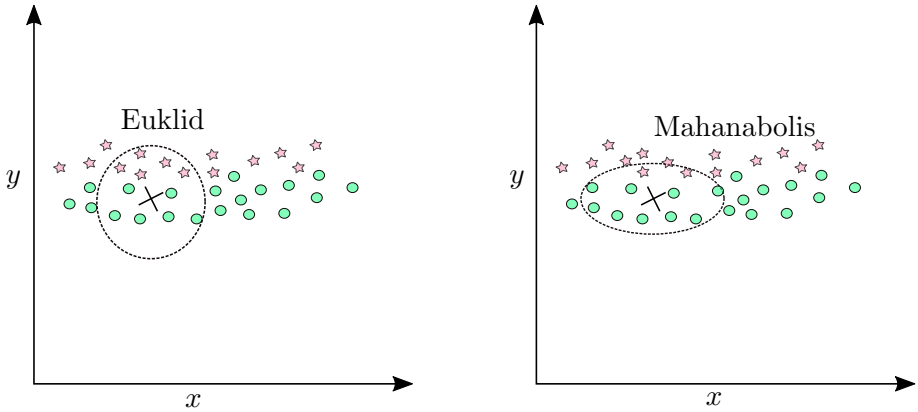


Abbildung 2.8: Vergleich der Abstandsbestimmungen

Die Support-Vector-Machine

Die Support-Vector-Machine(SVM)[3] gehört ebenfalls zu der Klasse der überwachten Lernverfahren. Das Ziel der SVM ist es die optimale Hyperebene zu finden, welche die Klassen am besten teilt. Das Grundprinzip ist in Abbildung 2.9 grafisch dargestellt. Im Zweidimensionalen entspricht diese Hyperebene einfach einer Linie mithilfe derer die Klassen getrennt werden. Um dies zu erreichen bedient man sich sogenannten Stützvektoren.

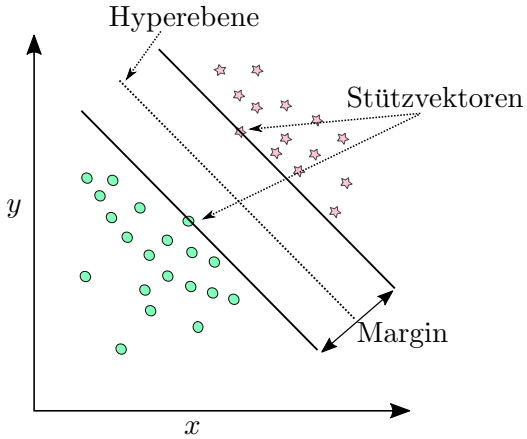


Abbildung 2.9: Idee der Support-Vektor-Machine

Als Stützvektoren (im englischen Support-Vectors) werden jene Punkte bezeichnet, welche sich am nächsten zur Hyperebene befinden. Die Distanz zwischen diesen Punkten wird als Margin bezeichnet. Je weiter diese Support Vektoren von der Hyperebene entfernt sind, also je größer die Margin ist, desto wahrscheinlicher ist es, dass neue Daten richtig klassifiziert werden. Bei der SVM handelt es sich somit um ein Optimierungsproblem mit dem Ziel, die Margin zu maximieren. Die Hyperebene befindet sich immer in der Mitte der beiden Stützvektoren.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die Hyperebene in mathematischer Form ist wie folgt definiert:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{2.5}$$

wobei $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$ den Variablenvektor darstellt. Der Vektor $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T$ ist jener Vektor der senkrecht auf die Hyberebene steht. Sowohl $\mathbf{w}$ als auch b sind unbekannt und müssen erst bestimmt werden. Für alle Punkt die nicht auf der Hyberebene liegen, ist der Wert ungleich 0. Da die Hyperebene zwei Klassen teilt, ist sie für eine Klasse positiv und für die andere negativ. Diese Klassenzugehörigkeit wird durch die Variable $y_i = \pm 1$ ausgedrückt (z.B. Klasse Kreis $y_i = -1$ und Klasse Stern $y_i = +1$). Damit ergeben sich die formelmäßigen Bedingungen:

$$\mathbf{w} \cdot \mathbf{x}_+ + b \geq +1 \text{ für } y_i = +1 \tag{2.6}$$

$$\mathbf{w} \cdot \mathbf{x}_- + b \leq -1 \text{ für } y_i = -1 \tag{2.7}$$

Die beiden obigen Gleichungen können in der Gleichung

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \tag{2.8}$$

zusammengefasst werden. Für die Punkte auf der Linie gilt

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 = 0 \tag{2.9}$$

Der Abstand zwischen den beiden ergibt sich unter dem Einhalten der obigen Bedingungen

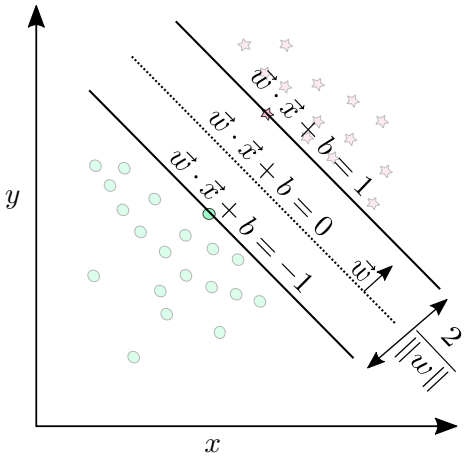


Abbildung 2.10: Support-Vektor-Machine

zu:

$$Margin = (\mathbf{x}_+ + \mathbf{x}_-) \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \tag{2.10}$$

Wie oben schon angemerkt wurde, ist das Ziel der SVM diese Weite zu maximieren(Abb. 2.10). Das Optimierungsproblem kann dann in der Form

$$\max_{w} \quad \frac{2}{\|\mathbf{w}\|} \tag{2.11}$$

$$\text{u.B.v.} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \tag{2.12}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

oder äquivalent dazu als Minimierungsproblem

$$\min_{w} \quad \frac{\|\mathbf{w}\|^2}{2} \tag{2.13}$$

$$\text{u.B.v.} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \tag{2.14}$$

angeschrieben werden. Diese Formulierung wird als primale Form der SVM bezeichnet. Bei der Herleitung wurde davon ausgegangen, dass die Klassen immer streng separierbar sind. Dies ist aber nicht immer der Fall. Tritt dieser Umstand auf, so muss das Optimierungsproblem derart angepasst werden, dass es geringe Verletzungen zulässt, da sonst möglicherweise gar keine oder nur eine schlechte Hyperebene gefunden werden kann.
Jedes Opimierungsproblem kann in zwei Formen, seiner pirmalen und dualen Form, formuliert werden. Die leicht abgeänderte duale Form der SVM lautet

$$\max_{\alpha} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{2.15}$$

$$\text{u.B.v.} \quad 0 \leq \alpha_i \leq C \text{ für alle i = 1,2,..n und } \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{2.16}$$

$$\tag{2.17}$$

Diese Formulierung hat den großen Vorteil, dass nicht auf irgendwelche Bedingungen wie in der primalen Formulierung geachtet werden muss. Sie wird mit Hilfe von Lagrange-Multiplikatoren hergeleitet. Der Vektor $\mathbf{w}$ wird dabei als Linearkombination aus den Trainingsdaten gewonnen. Auf die genaue Herleitung wird hier aber nicht mehr eingegangen. Sind die Daten generell nicht linear separierbar, wird der sogenannte Kernel Trick angewendet. Dies wird in der obigen Formel durch $K(\mathbf{x}_i, \mathbf{x}_j)$ berücksichtigt. Bei diesem Trick werden künstlich neue Dimensionen geschaffen mit dem Ziel anhand der hinzugefügt Dimensionen die Daten trennen zu können. Die populärsten Kernel sind:

- linearer Kernel,
- polynominale Kernel und
- RBF-Kernel.

Auf die genaue Beschreibung der verschiedenen Kernel wird hier nicht mehr eingegangen. Sie kann unter [3] nachgelesen werden.

Entscheidungsbäume

Entscheidungsbäume [13] können sowohl für Regression als auch für Klassifikation verwendet werden. Sie haben grundsätzlich die in Abbildung 2.11 gezeigte Struktur. Zur grafischen Darstellung ist es üblich sie auf den Kopf zu stellen. Bäume bestehen aus einer Wurzel, mehreren internen Knoten, Zweigen und den Blättern. Der Baum wird immer von oben nach unten durchlaufen. Basierend auf den internen Knoten zweigen sich die Bäume auf, wobei jeder Zweig für eine Entscheidung(z.B. ja oder nein) steht. Die Blätter des Baumes stehen für das Ergebnis. Im Beispiel der Abbildung erfolgt eine Zuordnung zu den fünf Klassen A, B, C, D und E anhand der zwei Variablen x und y, die jeweils

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

mit bestimmten Werten X_1, X_2, Y_1 und Y_2 verglichen werden. Die Aufspaltung nach den internen Knoten können wie in der Abbildung binär oder auch mehrwegig erfolgen. Dieser Baum stellt ein Klassifikationsbaum dar. Regressionsbäume werden auf die gleiche Art präsentiert mit dem Unterschied, dass sie kontinuierliche Werte als Ausgangswerte(z.B. einen Preis) haben. Je mehr Splits durchgeführt werden und je „tiefer“ der Baum ist, de-

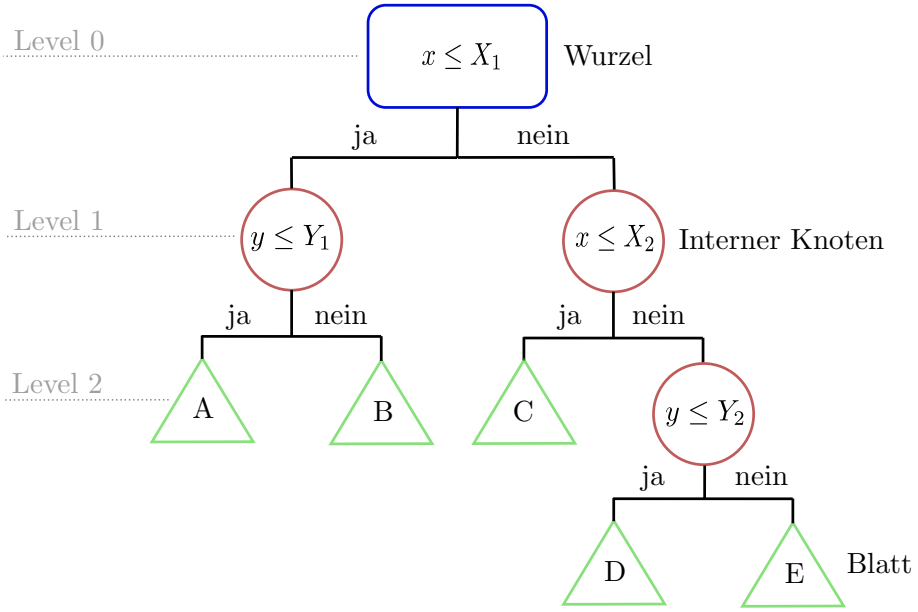


Abbildung 2.11: Entscheidungsbaum

sto höher ist die Komplexität des Baums. Die Tiefe des Baumes wird in Level angegeben.

<u>Entwurf eines Baumes</u>:

Will man einen Baum entwerfen, stellt sich die Frage, welche Attribute wie und wann gesplittet werden sollen und wie viele Level der Baum besitzen soll. Natürlich erfolgt dies nicht manuell sondern durch einen Algorithmus. Die hauptsächlich verwendet Algorithmen sind dabei:

- CHAID
- CART
- ID3
- C4.5

Alle haben die Aufgabe aus den Daten diejenigen Attribute herauszufiltern, die das beste Ergebnis beim Aufspalten der Daten in Untergruppen liefern. Es können aber Obergrenzen wie die Maximale Anzahl der Splits angegeben werden. CART z.B. verwendet für Klassifikationsaufgaben eine Metrik namens Gini-Index. Dieser Index gibt an wie „rein“ oder „unrein“ die Daten nach dem Split sind (Abb. 2.12). Der Index ist Null, wenn der die Daten nach dem Split direkt in die Klassen geteilt sind. Ein Gini-Index von Eins stellt

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

hingegen den schlechtesten Fall dar, da die Klassen nach dem Split immer noch stark durchmischt sind. Es wird also danach gestrebt, nach jedem Split einen möglichst kleine Unreinheit der Klassen zu bekommen. Auf die mathematische Darstellung sowie auf die genaue Beschreibung der einzelnen Algorithmen wird hier nicht eingegangen. Genaueres kann in [17] nachgelesen werden. Entscheidungsbäume werden häufig verwendet da sie

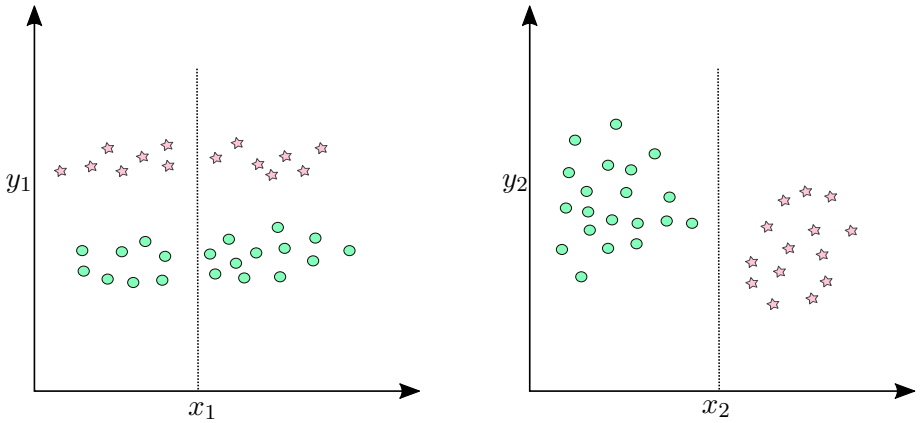


Abbildung 2.12: CART - Algorithmus: hoher Gini Index in der linken Abbildung, perfekter Split in der rechten Abbildung mit Gini Index gleich Null

einfach zu verstehen und zu implementieren sind. Sie haben aber auch einige Nachteile, welche nicht außer Acht gelassen werden dürfen. Eine kleine Änderung in den Daten kann zu ganz anderen Splits führen(hohe Varianz). Zudem sind sie sehr empfindlich in Hinsicht auf Überanpassung uvm. Um den Nachteilen entgegenzuwirken kann ein Ensemble an Bäumen verwendet werden.

Ensemble Methoden:

Diese Methoden kombinieren mehrere Entscheidungsbäume um die Performance zu verbessern und den erwähnten Nachteilen entgegen zu wirken. Die wahrscheinlich gängigsten Techniken dabei sind Bagging und Boosting.

Beim Bagging [5] werden die Daten des Trainingssets in mehrere Subsets (oder auch „Bags“ genannt) zerlegt. Die Daten der Subsets werden dabei zufällig ausgewählt, wobei jeder Datenpunkt des Trainingssets die gleiche Wahrscheinlichkeit aufweist, in einen Subset aufzutauchen. Mit jedem Subset wird ein eigenes Modell trainiert, d.h. am Ende erhält man so viel Modelle wie Subsets. Neue Datenpunkte werden nun allen Modellen als Eingang übergeben, mit welchen jedes Modell ein Vorhersage trifft. Das letztendliche Ergebnis bildet sich aus dem Mittelwert der Vorhersagen aller Modelle.

Boosting teilt die Trainingsdaten ebenfalls in mehrere Subsets. Die Art wie die Daten geteilt werden unterscheidet sich jedoch vom Bagging. Jedes Modell wird dabei unter Berücksichtigung der Performance des vorhergehenden Modells trainiert, d.h. das nachgehende Modell wird vorrangig mit jenen Datenpunkten trainiert, bei den das vorangegangene Modell schlecht abgeschnitten hat. So erhält man eine Kette von Modellen die aufeinander aufbauen und quasi immer versuchen die „Fehler“ des Vorgängers auszubes-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

sern. Die Auswahl der Daten aus dem Trainingsset ist dabei immer noch zufällig mit dem Unterschied, dass jene Daten bei denen der Vorgänger schlecht abgeschnitten hat mit einer höheren Wahrscheinlichkeit gewichtet werden und somit eher zum Training herangezogen zu werden. Das Ergebnis wird analog zum Bagging aus dem Mittelwert der Vorhersagen aller Modelle gebildet. Es existieren verschiedene Methoden, wie diese Gewichtung der Daten usw. erfolgt. Einige Beispiele sind die XGBoost-, Gradient Boost- und AdaBoost-Methode. Auf die genaue Beschreibung dieser Methoden wird hier verzichtet. Sie kann unter [10], [15] und [7] nachgelesen werden.

Logistische Regression

Logistische Regression [9] ist unter der Sparte der Klassifikationsalgorithmen in den überwachten Lernverfahren zu finden. Der Wert der Zielvariable Y wird dabei anhand einer oder mehrerer unabhängigen Variablen $\mathbf{X} = (x_1, x_2, \ldots, x_k)^T$ geschätzt. Mathematisch kann logistische Regression wie folgt ausgedrückt werden:

$$\log\left(\frac{p(\boldsymbol{X})}{1 - p(\boldsymbol{X})}\right) = \beta_0 + \boldsymbol{\beta}_1 \mathbf{X} \tag{2.18}$$

Dabei beschreibt $p(\boldsymbol{X})$ die Wahrscheinlichkeit, dass $\mathbf{X}$ zu einer gewissen Klasse gehört, während $1 - p(\mathbf{X})$ die Wahrscheinlichkeit angibt, dass sie zur anderen Klasse gehört. β ist der Vektor der Regressionskoeffizienten $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, \ldots, \beta_k)^T$. Diese Gleichung stellt die sogenannte Logit-Funktion dar. Wird diese Gleichung umgeformt erhält man folgende Sigmanoid-Funktion (Abb. 2.13)

$$p(\boldsymbol{X}) = \frac{e^{\beta_0 + \boldsymbol{\beta}_1 \mathbf{X}}}{1 + e^{\beta_0 + \boldsymbol{\beta}_1 \mathbf{X}}} \tag{2.19}$$

Mit dieser Formel kann anhand der Eingangsdaten dann eine bestimmte Klasse mit einer gewissen Wahrscheinlichkeit bestimmt werden. Die Sigmanoid-Funktion bildet eine beliebe Anzahl von realen Werten auf einen Wert zwischen 0 und 1 ab. Mithilfe einer Trennlinie (z.B. bei $p = 0.5$) wird der Punkt klar einer Klasse zugeordnet.

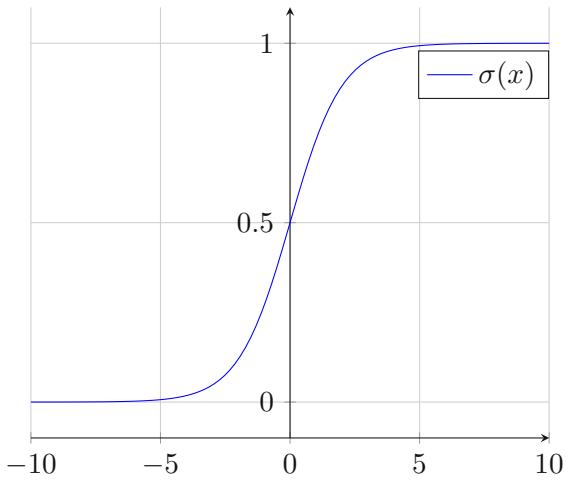


Abbildung 2.13: Sigmanoid Funktion

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Um die Koeffizienten β zu bestimmen wird eine Maximum-Likelihood-Lösung geschätzt. Im Fall von binären Daten sieht die Likelihood-Funktion dabei folgendermaßen aus:

$$\mathcal{L}(\boldsymbol{\beta}|\boldsymbol{y}) = \prod_{i=1}^{N} (\frac{\pi_i}{1-\pi_i})^{y_i}(1-\pi_i) \tag{2.20}$$

π stellt dabei die Wahrscheinlichkeit einer Zugehörigkeit zu einer bestimmte Klasse da. Bei der Bestimmung des Parameter handelt es sich um ein iteratives Verfahren. Es werden Parameter angenommen, die Likelihood bestimmt und die Parameter anschließend angepasst bzw. geändert um einen neuen Lauf durchzuführen und erneut eine Likelihood berechnet. Verwendet wird am Schloss das Set an Parametern bei dem die Likelihood am größten ist.

2.2.2 Teilüberwachtes Lernen

Beim teilüberwachten Lernen werden dem sowohl gelabelte als auch ungelabelte Daten übergeben. Dieses Lernverfahren stellt sozusagen eine Mischung aus überwachtem und unüberwachtem Lernen dar. Die Methoden des teilüberwachten Lernen arbeiten meist auf Grundlage einer kleinen Menge an gelabelten Daten und einer großen ungelabelten Datenmenge. Dies hat den Vorteil, dass der hohe Aufwand zum Labeln auch bei großen Datensets klein bleibt. Im Allgemeinen arbeiten teilüberwachte Lernverfahren auf folgende Weise. Als erstes wird anhand der gelabelten Daten ein Modell trainiert. Dieses Modell wird anschließend verwendet um die ungelabelten Daten zu labeln. Dies Daten werden als „pseudo-gelabelte“ Daten bezeichnet. Zum Schluss werden die zwei Datensets kombiniert, um einen einzigartigen Algorithmus zu kreieren.

2.2.3 Unüberwachtes Lernen

Beim unüberwachten Lernen werden dem Algorithmus nur nicht kategorisierte Daten übergeben. Sie lernen quasi automatisch. Das unüberwachte Lernen ist daher mit wesentlich weniger manuellen Aufwand möglich. Das Hauptaugenmerk dieser Techniken liegt vor allem im Finden von Strukturen und Zusammenhängen in rohen Daten. Unüberwachtes Lernen wird weiter in Segmentierung (Clustering), Assoziationsanalyse und Anomalie-Erkennung gruppiert. Clustering versucht Strukturen bzw. Muster in den unbekannten Daten zu finden. Beim Clustering werden Gruppen von ähnlichen Daten gefunden. Dabei ist nicht festgelegt, welche Merkmale die Ähnlichkeiten oder Unterschiede der Gruppen aufweisen. Wird es z.B. auf Bilder mit Hunden und Katzen angewendet, wird es zwar zwei verschiedene Muster erkennen, kann diese jedoch nicht kategorisieren. Die Assoziationsanalyse hingegen versucht Verknüpfungen zwischen Variablen in großen Datensets zu finden (z.B. Personen die Wein kaufen, kaufen wahrscheinlich auch Chips).

Eine weiter Sparte der unüberwachten Lernverfahren ist die Anomalie-Detektion. Sie wird häufig dem unüberwachten zugeschrieben, da oft nicht genügend Daten von Anomalien für das Trainingsset vorhanden sind. Dabei werden Daten gesucht, die nicht zu anderen Mustern in den Daten passen. Man greift vor allem auf solche Verfahren zurück, wenn nicht ausreichend Testdaten für Anomalien vorhanden sind. Ein Beispiel wäre hier die Abnutzung einer rotierenden Maschine, da man die Maschine erst abnützen müsste, um zu wissen welche Auswirkungen es im Bezug auf die Daten hat. Zur Abhilfe verwendet

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

man in solchen Fällen Algorithmen wie z.B. den Autoencoder oder die PCA. Diese entwerfen Wahrscheinlichkeitsverteilungen innerhalb derer sich die Datenpunkte normalerweise befinden. Je weiter ein Datenpunkt von der Mitte der Verteilung entfernt ist, desto wahrscheinlicher handelt es sich bei diesem Punkt um eine Anomalie. Im Fall der rotierenden Maschinen würden sich hier also die Datenpunkte mit steigendem Alter und damit dem Fortschreiten der Abnützung der Maschine immer weiter von der Wahrscheinlichkeitsverteilung des Normalzustandes entfernen und könnte so als anomaler Zustand der Maschine aufgefasst werden. Nachteilig bei unüberwachten Lernmethoden ist vor allem, dass sie kein spezifisches Ergebnis haben und daher der Output oft schwer zu bewerten ist.

2.2.4 Verstärktes Lernen

Bei Verstärkte Lernmethoden lernt ein sogenannter Agent selbstständig eine Strategie. Er wird quasi „belohnt“ falls er etwas richtig macht und „bestraft“ falls er etwas falsch macht. Der Agent steht für ein Programm das trainiert wird, um eine bestimmte Aufgabe zu erfüllen. Die Umgebung bezeichnet den Ort, an welchem der Agent seine Aktionen ausführt. Eine Aktion ist eine Handlung des Agenten, bei dem sich der Zustand der Umgebung ändert. Abbildung 2.14 zeigt die grundsätzliche Funktionsweise des verstärkten Lernens. Der Agent führt ausgehend vom jetzigen Zustand eine Aktion A_t aus einer Menge von Aktionen aus, wodurch sich der Zustand S_t der Umgebung ändert zu S_{t+1}. Für diese Aktion erhält er dann eine Belohnung R_{t+1} oder eine Strafe. Er ist dabei bestrebt den größtmöglichen Gewinn zu erzielen. Somit lernt der Agent während des Trainings einzuschätzen, wie sich seine Aktionen auf die Umgebung auswirken.

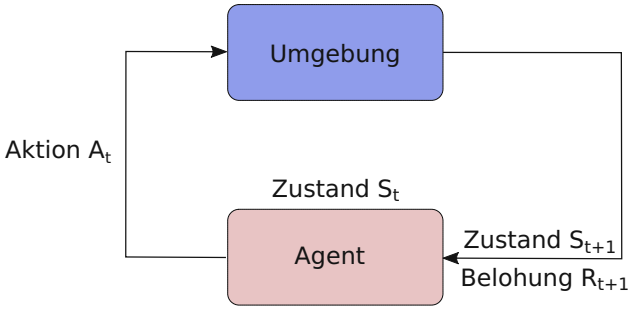


Abbildung 2.14: Verstärktes Lernen Grundidee

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3 Methodik

In diesem Kapitel erfolgt die Beschreibung der Methodik, d.h. die Vorgehensweise zur Lösung der Problemstellung und zur Gewinnung der Erkenntnisse. Das Kapitel setzt sich aus vier Abschnitten zusammen. Zuerst erfolgt eine allgemeine Beschreibung des Problems. Im zweiten Abschnitt wird die Art und Weise der Datenerhebung und Datenanalyse näher erklärt. Der dritte Teil gibt Auskunft über die Vorgehensweise bei der Implementierung der Machine-Learning-Algorithmen. Der letzte Abschnitt zeigt die Umsetzung der Methodik an einem spezifischen Netz.

3.1 Allgemeine Beschreibung

In dieser Arbeit gibt es keinen Abschnitt über die bereits existierenden Arbeiten zu diesem Thema. Das liegt daran, dass es kaum Literatur zur Forschung in diese Richtung gibt. Die existierenden Fehlererkennungen in Verbindung mit KI in elektrischen Energiesystemen zielen hauptsächlich auf die Auswertung von vielen Daten aus mehreren UWs ab [6]. Der Blick auf das Netz ist dabei also eher „global“ und stellt einen grundsätzlich anderen Zugang zur Fehlererkennung dar.

In dieser Arbeit wird danach gestrebt eine dezentrale Erkennung von gestörtem Netzbetrieb ohne Kommunikation mit der Leitwarte zu realisieren. Ohne Kommunikation mit der Leitwarte bedeutet, dass zur Fehlererkennung nur Messwerte von einem bestimmten geographischen Punkt wie z.B. einem Umspannwerk (UW) verwendet werden können, da die Kommunikation mit der Leitwarte nicht möglich ist. Dieser Satz impliziert bereits die Dezentralität der Erkennung, da die Netzleitwarte den zentralen Punkt bei der Führung des Netzes darstellt und daher ohne sie gar keine zentrale Erkennung möglich ist. Im Vergleich zu den übrigen Methoden wird also hier versucht, nur anhand der lokal zu Verfügung stehenden Messwerten einen Blick in das Netz zu werfen.

Aus prozessleittechnischer Sicht würde sich die 3-schichtige Pyramide aus Abbildung 2.1 hier durch das Wegfallen der Netzleitebene in eine zweischichtige Pyramide wie in Abbildung 3.1 dargestellt umwandeln. Die Abhängigkeit von der funktionierenden Kommu-

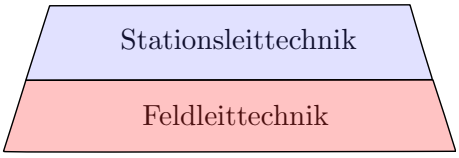


Abbildung 3.1: Leittechnik im UW

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

nikation über weite Wege im Fall der Netzleitebene beschränkt sich hierbei nur auf die Abhängigkeit der Kommunikation der im Vergleich kurzen Distanzen im UW. Außerdem müssen zur Fehlererkennung nicht eine Vielzahl an Messwerten, Systemgrößen usw. aus vielen verschiedenen UWs verarbeitet werden, sondern nur die Mess- und Systemgrößen von einem UW. Diese werden bereits in der Regel im Stationsleitrechner abgespeichert.

Gestörter Netzbetrieb

Weiters ist die Frage zu beantworten, was mit einem gestörten Netzbetrieb in dieser Arbeit gemeint ist. Es existiert eine Vielzahl gestörter Betriebszustände im Netz, von Kurzschlüssen und Ausfällen von Generatoren, bis hin zum Frequenzabfall bei zu hoher Last, welche nicht alle betrachtet werden können. Grundsätzlich beschränkt sich diese Arbeit in diesem Zusammenhang auf den Ausfall von Leitungen. Wird also im Folgenden von einem gestörten Netzbetrieb gesprochen, so ist irgendwo im Netz eine Leitung ausgefallen. Wodurch dieser Ausfall der Leitung verursacht wird, ist zunächst nicht von Interesse. Das Gegenteil, also wenn alle Leitungen in Betrieb sind, wird im Folgenden als Normalzustand oder auch Normalbetrieb bezeichnet.

Der Grund für diese Auffassung des gestörten Betriebes ist, dass versucht wird mit möglichst einfachen Methoden Ausfälle erkennen und klassifizieren zu können. Einfach bedeutet in diesem Zusammenhang den Versuch Ausfälle anhand der Messwerte im eingeschwungenen Netzzustand zu erkennen. Das Prinzip ist in Abbildung 3.2 grafisch dargestellt. Findet also

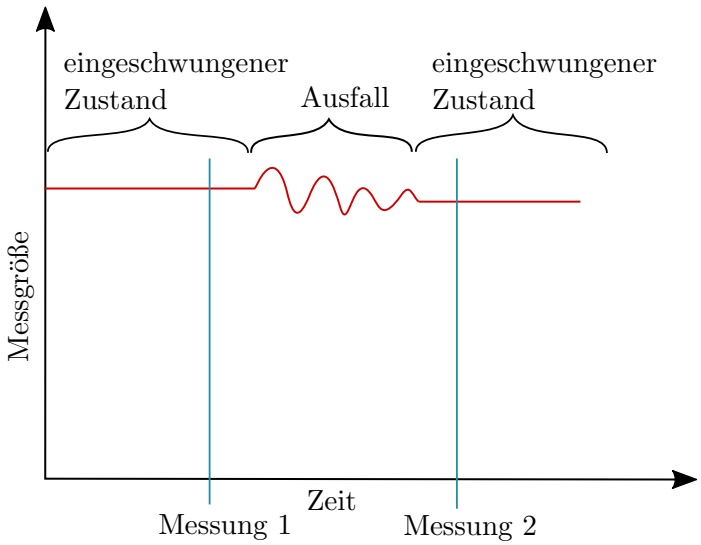


Abbildung 3.2: Messungen werden im eingeschwungenen Netzzustand aufgenommen

Beispielsweise ein Ausfall einer Leitung aufgrund eines Kurzschlusses statt, so interessiert uns hier nicht der Verlauf des Kurzschlussstromes, sondern ob die Zustände vor dem Ausfall der Leitung und nach dem Abklingen der Ausgleichsvorgänge zu unterscheiden sind.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Sind mithilfe dieser Methoden keine Erkennungen möglich, kann eine schrittweise Steigerung der Komplexität in Richtung dynamischer Vorgänge genommen werden. Das ist deshalb wichtig anzumerken, da die spätere Modellierung und Auswahl eines spezifischen Netzes stark davon abhängt, ob transiente Vorgänge oder eingeschwungenen Zustände betrachtet werden. Dynamische Vorgänge möglichst realistisch zu modellieren und simulieren stellt hohe Anforderungen an den Entwurf des Netzes.

Im nächsten Schritt muss geklärt werden, welche Ausfälle denn nun überhaupt erkannt werden sollen. Nicht alle Ausfälle sind für diese Untersuchung interessant. Von Interesse ist, Ausfälle von Leitungen zu erkennen welche nicht direkt mit dem UW verbunden sind. Die Ausfälle der direkt angeschlossenen Leitungen werden durch die Information der bereits installierten Einrichtungen erkannt. Dabei müssen z.B. nur die Meldungen der Schutzgeräte der Abzweige betrachtet werden, um zu überprüfen, ob eine Leitung aktiv ist oder nicht. Uns interessieren in dieser Arbeit viel mehr Ausfälle die entfernt in der Masche oder in Nachbarmaschen passieren, mit denen das UW nicht direkt verbunden ist. Damit soll herausgefunden werden, wie weit es möglich ist auf Basis der Messdaten im UW Ausfälle von entfernten Leitungen abzuleiten. Es wird davon ausgegangen, dass mit zunehmender Entfernung der Ausfälle vom UW auch die Erkennung immer schwieriger wird. Ausfälle die sich nahe am UW befinden sollten anhand der Messwerte im UW leichter zu erkennen sein.

Die Reichweite der Information in einem elektrischen Netz soll im Folgenden an einem kleinen, stark vereinfachten Beispiel etwas näher gebracht werden. Nehmen wir an wir haben die Situation wie sie in Abbildung 3.3 dargestellt ist. Die Impedanz $\underline{Z}_1$ kann z.B. eine Leitung darstellen. Der Strom am Eingang $\underline{I}_E$ berechnet sich zu

$$\underline{I}_E = \frac{\underline{U}_E}{\underline{Z}_1 + \underline{Z}_2} \tag{3.1}$$

Er stellt in unserem Beispiel die Messgröße dar. Nun wird die Impedanz $\underline{Z}_2$ variiert und der Strom am Eingang beobachtet. Ist die Impedanz $\underline{Z}_1$ klein, so ändert sich der Strom $\underline{I}_E$ wesentlich, wenn $\underline{Z}_2$ variiert wird . Die Auswirkungen der Änderung von $\underline{Z}_2$ sind also anhand des Stromes deutlich „sichtbar“. Wird jetzt für $\underline{Z}_1$ eine größere Impedanz angenommen, so ändert sich der Strom bei der gleichen Variation von $\underline{Z}_2$ weniger als im vorigen Fall. Vom Blickwinkel der Spannungsquelle ist die Variation von $\underline{Z}_2$ anhand der Messung Stromes also nicht mehr so gut zu sehen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

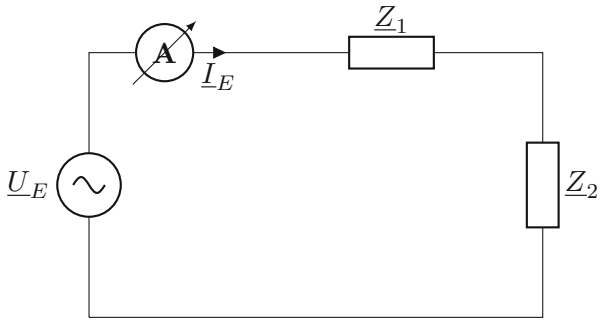


Abbildung 3.3: Beispiel zur Erläuterung des Begriffes der Sicht in einem elektrischen System

Erkennung

„Erkennung“ bedeutet in diesem Zusammenhang, dass ein gestörter Netzbetrieb, also der Ausfall einer Leitung, auch als solcher klassifiziert werden kann. Diese Klassifizierung erfolgt allerdings nicht durch den Menschen sondern wird von einem Machine-Learning-Algorithmus durchgeführt. Interessant ist dabei nicht nur die Erkennung der zuvor beschriebenen Ausfälle einzeln, sondern auch ob die Ausfälle in gemeinsamer Betrachtung voneinander unterscheidbar sind.

Um einen solchen Algorithmus zu implementieren, müssen viele Daten generiert und analysiert werden. Die Beschreibung, wie dies von statten geht wird im nächsten *Abschnitt 3.2* beschrieben. Die Daten sind dabei nichts anderes als Messwerte in einem UW. Der dritte Abschnitt in diesem Kapitel zeigt anschließend wie diese Erkennung mit einem Machine-Learning-Algorithmus realisiert wurde.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.2 Daten

Um eine Fehlererkennung implementieren zu können, müssen sowohl eine Vielzahl an Daten generiert sowie auch analysiert werden. Dieser Abschnitt gibt Aufschluss über die Art und Weise wie die Daten erzeugt und verarbeitet wurden, um schließlich einem Machine-Learning-Algorithmus übergeben zu werden.

3.2.1 Generierung der Daten

Die Generierung der Daten unterteilt sich in zwei weitere Teile. Zum einen muss geklärt werden, wie und was gemessen wird und zum anderen muss gezeigt werden wie denn überhaupt die Betriebszustände erzeugt werden, die gemessen werden.

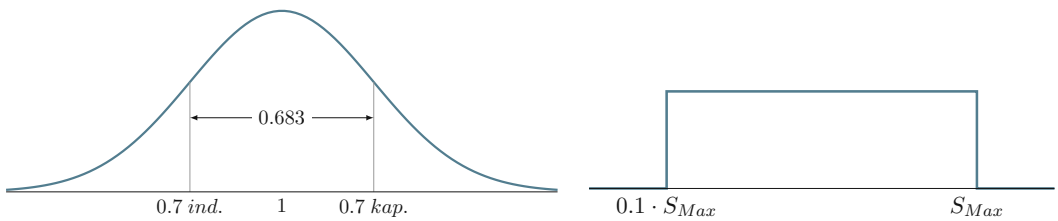
Erzeugung der Betriebszustände

Um aussagekräftige Ergebnisse zu erhalten müssen möglichst alle möglichen Szenarien, d.h. Netzzustände durchgespielt werden. Diese verschiedenen Betriebszustände des Netzes werden durch die Variation der Verbraucher im Netz erreicht. Grundsätzlich werden die Lasten dabei durch die Vorgabe der Wirk- und Blindleistung bestimmt. Diese werden allerdings nicht direkt vorgegeben sondern über den Betrag der Scheinleistung S und den Lastwinkel $\cos(\varphi_L)$ berechnet, da so eine einfachere Erzeugung der Daten nach den definierten Kriterien ermöglicht wird. Bei der Variation der Lasten müssen eine Reihe von Annahmen getroffen bzw. Begrenzungen eingehalten werden.

Als Erstes muss darauf geachtet werden, dass keine Spannungsbandverletzungen auftreten. Das vorgeschriebene Spannungsband muss von den Netzbetreibern eingehalten werden und darf nicht verletzt werden, um z.B. Betriebsmittelbeschädigungen bei zu hohen Spannungen oder die Gefahr eines Spannungskollapses bei zu niedriger Spannung zu verhindern. Die Lasten können also nur so lange variiert werden, bis eine Spannungsbandgrenze erreicht wird. Zudem werden Annahmen für die Ausprägung des $cos(\varphi_L)$ der Lasten getroffen. Es wird angenommen, dass nur in den äußert unwahrscheinlichsten Fällen Belastungen mit einem $cos(\varphi_L)$ kleiner 0,5 auftreten. Der Großteil der Belastungen tritt mit einem $cos(\varphi_L)$ zwischen 0,7 induktiv und 0,7 kapazitiv auf.

Da es viel zu umständlich ist, alle möglichen Betriebszustände manuell zu erstellen werden die Lasten nach einem Zufallsprinzip generiert. Allerdings unterscheidet sich dabei die Art wie die Scheinleistung und der Lastwinkel generiert wird. Während der Bestimmung der Scheinleistung eine Gleichverteilung zugrunde liegt, folgt der $cos(\varphi_L)$ einer Normalverteilung(Abb. 3.4). Für die Scheinleistung S_L der Lasten bedeutet dies, dass die Beträge der Scheinleistung zwischen $0{,}1 \cdot S_{Max}$ bis S_{Max} alle mit der gleichen Wahrscheinlichkeit auftreten. S_{Max} bezeichnet dabei den Maximalwert der Scheinleistung der jeweiligen Last. Da es sich bei diesen Lasten um größere Verbraucher handelt (z.B. untergeordnete Netze oder Industrieanlagen) und nicht um einen Haushaltskunden, wird zu keinem Zeitpunkt angenommen dass der Verbrauch bei Null liegt. Für den $cos(\varphi_L)$ wird die vorhin getätigte Annahme durch die Verwendung einer Gaussverteilung erreicht. Damit stellt sich im Großteil der Fälle ein $cos(\varphi_L)$ um 1 mit einer Standardabweichung von 0,3 ein. Mit dieser Verteilung sind auch $cos(\varphi_L)$ bis 0,5 induktiv sowie 0,5 kapazitiv möglich, allerdings nicht in dem Ausmaß wie zwischen 0,7 induktiv bis 0,7 kapazitiv.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) $\cos(\varphi_L)$ (b) Scheinleistung S_L

Abbildung 3.4: Wahrscheinlichkeitsdichtefunktionen für die Bestimmung der Lasten

Nachdem die Scheinleistungen S_L der Lasten sowie deren $cos(\varphi_L)$ bestimmt wurden kann die Wirkleistung P_L über

$$P_L = S_L \cdot cos(\varphi_L) \tag{3.2}$$

und die Blindleistung Q_L über

$$Q_L = S_L \cdot sin(\varphi_L) \tag{3.3}$$

berechnet werden. Mit dieser Vorgehensweise können tausende von Betriebszuständen simuliert werden.

Erfassung der Betriebszustände

Um den Zustand des elektrischen Energiesystems bestimmen zu können, sind Messgeräte notwendig. Natürlich existiert eine Vielzahl verschiedener Messgeräte, von der Messung des Stroms und der Spannung, von Wirk- und Blindleistungsflüssen, bis hin zu Messung von Temperaturen von Betriebsmitteln, welche nicht alle in Betracht gezogen werden können. Diese Arbeit beschränkt sich in diesem Zusammenhang auf die Messung von Spannungen und Strömen sowie den Winkel zwischen den Beiden. Wie schon vorher erwähnt, wird versucht, anhand grundlegender und einfacher Methoden, das Problem zu lösen. Sollte anhand dieser Messungen die Problemstellung nicht gelöst werden, können auch hier schrittweise neue Messungen hinzugefügt werden.

UWs bestehen grob beschrieben aus ober- und unterspannungsseitigen Sammelschienen von denen jeweils Leitungen zu anderen UWs, Schaltanlagen oder Lasten abzweigen. Zwischen den Sammelschienen befindet sich ein Transformator (auch Umspanner genannt). Um nun den elektrischen Zustand dieses UWs erfassen zu können, werden an allen Abgängen die Ströme sowie die Spannungen an den SS auf beiden Seiten des Umspanners gemessen. Zudem wird der Strom über den Transformator bestimmt. Der Impedanzwinkel φ ergibt sich anschließend aus der Winkeldifferenz zwischen der Spannung an den SS und den Strom des jeweiligen Abgangs.
Wie schon im vorigen Abschnitt und in Abbildung 3.2 gezeigt wurde, werden immer Momentaufnahmen vor und nach dem Ausfall durchgeführt. Dadurch ergeben sich für jeden Belastungsfall und Ausfall zwei „Snapshots“ der Zustände des UWs.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.2.2 Analyse der Daten

Nachdem tausende von Simulationen durchgeführt und Unmengen an Daten generiert wurden geht es darum herauszufinden, welche Messungen für welchen Ausfall herangezogen werden bzw. Information enthalten. Die Rohdaten, ohne vorherige Verarbeitung, einfach einem Maschine-Learning Algorithmus zu übergeben, liefert in der Regel keine guten Ergebnisse. Das liegt daran, dass viele Eingangsvariablen(z.B. Strom und Spannung),oder auch Attribute genannt, wenig bis gar keine Information enthalten und dadurch eher zu einer Verwirrung als zu einer Erkenntnis für den Algorithmus führen. Die im Folgenden beschriebenen einzelnen Schritte bei der Analyse der Daten sind auf sich aufbauend.

In einem ersten Schritt werden die Messgrößen einzeln betrachtet. Dabei werden zuerst einzelne Simulationen durchgeführt und die Wirkung des Ausfalls der Leitung auf die einzelnen Messgrößen analysiert, noch bevor die Masse an Daten bzw.Messwerten verwendet wird. Es werden alle Verbraucher gemeinsam sowie einzelne Verbraucher einzeln variiert, mit dem Ziel herauszufinden, welche Lasten einen hohen Einfluss auf die Messgrößen bei dem Ausfall der Leitung haben. Zudem wird der Lastwinkel variiert und die Auswirkung auf die Messung untersucht. Bei diesen Untersuchungen geht es vor allem darum, ein Bild für Ursache (z.B. Ausfall dieser und jener Leitung) und Wirkung(z.B. Strom erhöht sich) und ein grundsätzliches Verständnis für das Verhalten des Systems zu bekommen. Würden gleich alle Daten herangezogen, ist es nur mehr schwer nachzuvollziehen, was genau bei einem Ausfall passiert und in welchem Sinne sich die Messgrößen ändern.

Erst nach diesem Schritt wird die Gesamtheit der generierten Daten herangezogen. Dabei wird jedes Attribut zuerst einzeln beleuchtet. Es wird die generelle Verteilung dieses Attributs sowie der Unterschied zwischen dem Normalbetrieb und dem gestörten Betriebsfall analysiert. Abbildung 3.5 zeigt eine solche Analyse, wie sie im Laufe der Arbeit durchgeführt wurde. Diese stellt die Verteilung der Beträge der Ströme im Normalbetrieb dar, welche mit den Werten bei gestörtem Betrieb verglichen werden. Verglichen werden dabei Mittel-, Maximal- und Minimalwerte sowie die Standardabweichung.

Um das Verhalten vom Netz zu untersuchen und die Auswirkungen bestimmter Fehler auf die Messgrößen im UW zu bestimmen ist es hilfreich, Stark- und Schwachlastzeiten getrennt voneinander zu untersuchen. Durch die geringeren Belastungen zu Schwachlastzeiten ist auch die Auswirkung eines Ausfalls auf die Messwerte im UW kleiner. Zeigen die Messgrößen für diese Betriebszustände keine Unterschiede, so ist möglicherweise keine Erkennung im Schwachlastfall implementierbar. Sollte dies der Fall sein kann der anschließend implementierte Algorithmus dementsprechend angepasst werden um trotzdem zumindest eine Erkennung im Starklastfall zu ermöglichen.

Im nächsten Schritt werden die Attribute kombiniert. Durch diese Kombination können die Zusammenhänge der Messgrößen herausgearbeitet werden, wodurch es zu einem zusätzlichen Informationsgewinn kommt. Konkret bedeutet dies, dass durch die Kombination von Attributen eine Erhöhung der Dimension erreicht wird und somit ausgehend vom eindimensionalen Fall(Abb.3.5), eine zwei- (Abb. 3.6) oder mehrdimensionale Darstellung erreicht wird. Als erstes erfolgt die Hinzunahme einer Dimension. Im günstigsten Fall sind anhand dieser zweidimensionalen Darstellung bereits Unterschiede erkennbar. Die Abbildung

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

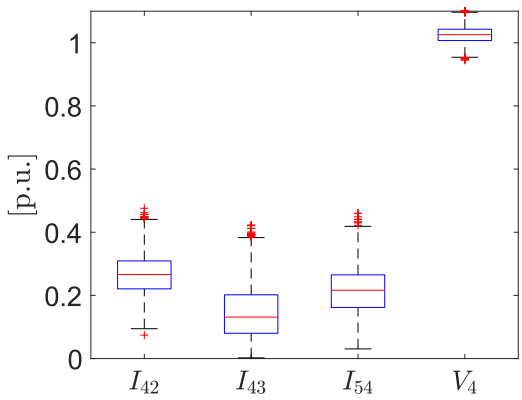


Abbildung 3.5: Boxplot der Ströme und der Spannung zur Analyse der Daten

3.6(a) zeigt beispielhaft einen Fall, wo die Unterscheidung leicht fällt, während in (b) nur schwer eine Trennung möglich ist. Diese Darstellungen sind für die Analyse durch den Menschen sehr praktisch, da auf einem Blick eine Aussage über Unterschiede getätigt werden kann. Wie in der linken Abbildung ersichtlich ist, variiert der Strom sowohl im Fehler-

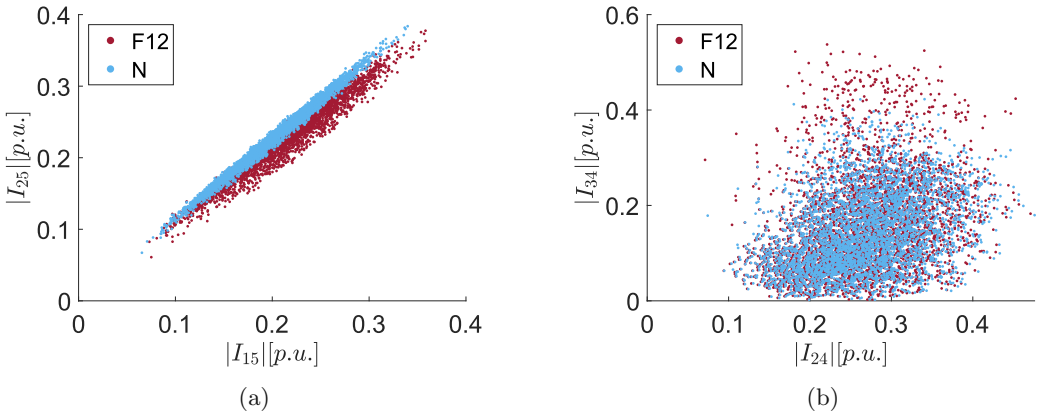


Abbildung 3.6: 2 dimensionale Analyse der Daten

fall(blau) als auch im Normalbetrieb(violett) zwischen 0.05 p.u. und 0.4 p.u. Das heißt, es kommt durch den Ausfall einer Leitung nicht zu einer übermäßigen Erhöhung oder zu einem Abfall des Stromes, sodass nur anhand eines einzelnen Attributs eine Unterscheidung getroffen werden könnte. Den Unterschied macht in diesem Fall die Kombination der Ströme I_{52} und I_{51}. In diesem Fall kann klar eine Trennlinie zwischen den beiden Zuständen gezogen werden. Der Informationsgewinn durch die Erhöhung der Dimension ist offensichtlich. Anders ist es bei der Abbildung 3.6(b), in welcher auch durch Kombination der Ströme keine Unterscheidung getroffen werden kann. Die Messwerte sind diffuser verteilt und es ist kein wesentlicher Zusammenhang zwischen den Parametern erkennbar. Es existieren zwar kleine Bereiche in denen eine Unterscheidung möglich wäre, jedoch überlappt sich der Großteil der Punkte. In solchen Fällen ist es hilfreich, Dimensionen hinzuzunehmen oder erstmal Messungen anderer Zweige zu begutachten. Dimensionen

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

hinzuzufügen bedeutet wieder nichts weiter als weitere Messgrößen dazuzugeben. Wird mehr als eine weitere Dimension hinzugefügt, ist keine visuelle Darstellung wie im zwei oder drei dimensionalen Fall mehr möglich. Hier müssen andere Methoden angewendet werden.

Diese Schritte sind essenziell, um eine erfolgreiche Erkennung zu implementieren. Messgrößen, die keine Information enthalten, können dadurch identifiziert und entfernt werden. Anzumerken ist jedoch, dass dies auch mit einigem Aufwand verbunden ist. Sowohl für jeden Ausfall, der betrachtet werden soll, sowie auch für jede Messgröße und deren Kombination muss eine Überprüfung durchgeführt werden. Außerdem müssen die Daten der verschiedenen Ausfälle auch gemeinsam analysiert werden. Nachdem die informationsbehafteten Attribute bestimmt sind, können sie einem Machine-Learning-Algorithmus übergeben werden.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.3 Erkennung vom gestörten Netzbetrieb

Wie in der Einführung erwähnt wurde, bedeutet Erkennung im Sinne dieser Arbeit die erfolgreiche Klassifikation der Betriebszustände des Netzes. Abbildung 3.7 zeigt grafisch das angestrebte Ziel. Auf Grund der vorher ausgewählten Attribute und deren Daten lernt der Algorithmus zwischen normalem und gestörtem Betrieb zu unterscheiden. Werden ihm neue Messwerte übergeben, soll der Maschine-Learning-Algorithmus eine Klassifikation treffen.
Um dieses Ziel zu erreichen, müssen mehrere Schritte durchgeführt werden. Als erstes müssen in Frage kommende Algorithmen ausgewählt werden. Anschließend müssen die Algorithmen angepasst, trainiert und getestet werden. Die einzelnen Schritte werden im Folgenden etwas detaillierter beleuchtet.

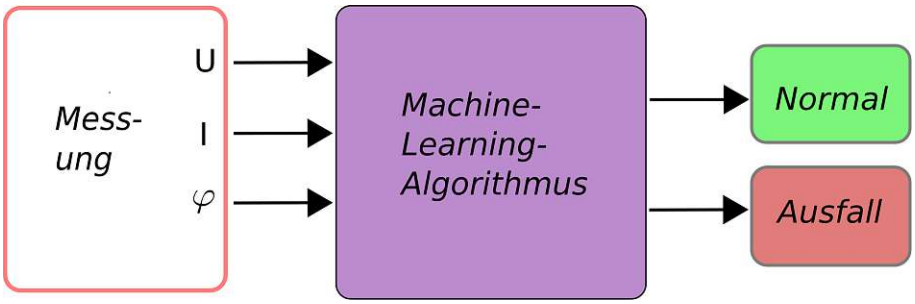


Abbildung 3.7: Klassifizierung des Netzbetriebs auf Grundlage der Messwerte durch einen Machine-Learning-Algorithmus

3.3.1 Auswahl des Algorithmus

Wie schon im Abschnitt 2.2 *Machinelles Lernen* gezeigt wurde, existiert eine Vielzahl von Machine-Lerning-Algorithmen. Um nachvollziehen zu können, warum und wieso im Laufe dieser Arbeit bestimmte Algorithmen bevorzugt wurden, wird im Folgenden der Gedankengang zur Auswahl dieser beschrieben.

Die erste Unterscheidung erfolgt anhand des Lernverfahrens. Dabei wird in Abbildung 2.4 der Zweig in Richtung der überwachten Lernverfahren genommen. Beim Training werden dem Algorithmus also vollständig gelabelte Daten übergeben. Methoden des überwachten Lernverfahrens weisen im Allgemeinen höhere Genauigkeiten der Erkennung auf im Vergleich zu den Methoden anderer Lernverfahren. Bei selbstlernenden Algorithmen ist zudem nicht immer klar, wonach die Daten aufgeteilt werden. Die Voraussetzung für die Implementation ist, wie schon angesprochen, das Vorhandsein von gelabelten Daten. Da es in dieser Arbeit ohne größeren Aufwand möglich ist die Daten zu labeln, spricht nichts gegen ein überwachtes Lernverfahren.

Als nächstes muss zwischen den Methoden der Klassifikation und der Regression entschieden werden. Da wir einen Algorithmus implementieren wollen, welcher eindeutig die Aussage liefert: „Normalbetrieb“ oder „Ausfall einer Leitung“ kommen hier nur die Methoden der Klassifikation in Frage.
Auch hier wird versucht, mit möglichst einfachen Algorithmen zu beginnen, und, sollte

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

es nicht funktionieren, schrittweise komplexere Algorithmen zu implementieren. Bei den verwendeten Algorithmen handelt es sich um:

- KNN
- Logistische Regression
- Lineare SVM
- Entscheidungsbäume

KNN ist zum einen ein einfach aufgebauter Algorithmus und zum anderen ist er in der Lage Erkennungen zu erzielen, wo andere Algorithmen mit ihren entworfenen Funktionen keine Klassifizierung treffen können. Die Logistische Regression ist interessant, da sie als Linearkombination der Attribute dargestellt werden kann. Das bedeutet, sollte eine Erkennung möglich sein, können die Betriebszustände anhand einer einfachen Linearkombination der Messgrößen unterschieden werden. Entscheidungsbäume sind aufgrund ihres einfachen Aufbaus und weiten Anwendungsgebieten vielversprechend. Mit der SVM wird noch ein etwas aufwendigerer Algorithmus hinzugefügt.
Im folgenden wird darauf eingegangen, wie diese Algorithmen angepasst und trainiert wurden. Die Beschreibung erfolgt dabei allgemein und nicht für jeden Algorithmus separat.

3.3.2 Anpassung des Algorithmus

Bei der Auswahl und Implementation eines Machine-Learning-Algorithmus handelt es sich um ein iteratives Verfahren. Dieses Verfahren ist in Abbildung 3.8 grafisch dargestellt. Der gewünschte Algorithmus wird ausgewählt, seine Parameter eingestellt und anschließend trainiert. Nach dem Training werden dem Algorithmus neue Daten zur Validierung übergeben und die Klassifikationsgenauigkeiten ausgewertet. Es gibt mehrere Methoden wie die Daten in ein Trainings- und ein Validierungsset aufgespalten werden. Dies wird im folgenden Unterpunkt *Training* erläutert. Sind die Genauigkeiten der Erkennung gering, werden die Parameter des Algorithmus geändert und ein neuer Lauf durchgeführt. Liefert ein bestimmter Algorithmus auch nach N Iterationen der Parameter keine zufriedenstellenden Ergebnisse, wird er ausgetauscht.

Genauigkeit:

Die Genauigkeiten berechnen sich aus Klassifikationsergebnissen des Algorithmus. Die Ergebnisse können mit der sogenannten Kontingenztafel abgebildet werden. In der Tabelle 3.1 ist eine solche Kontingenztafel für eine binäre Klassifizierung dargestellt. True-Positiv

Vorhersage \Klasse	positiv	negativ
positiv	T.P.	F.P.
negativ	F.N.	T.N.

Tabelle 3.1: Kontingenztafel

(T.P.) und True-Negativ(T.N.) bezeichnen dabei immer den Fall, dass eine Klasse richtig

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

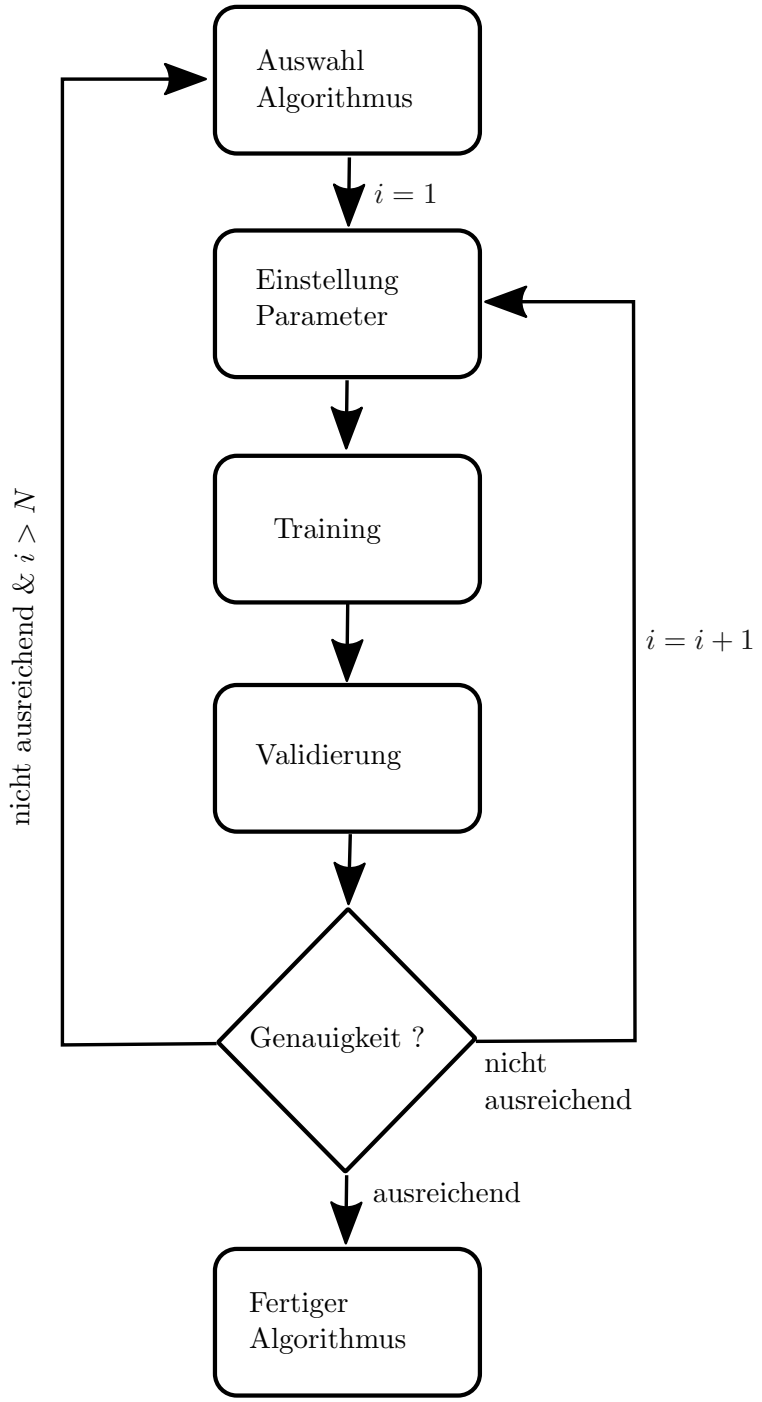


Abbildung 3.8: Iteratives Verfahren bei der Anpassung der Machine-Learning-Algorithmen

klassifiziert wird, d.h. die wirkliche Klasse mit der vorhergesagten übereinstimmt. False-Positiv(F.P.) und False-Negativ(F.N.) stehen für die falsch Klassifizierten der jeweiligen

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Klasse. Mit diesen Werten ergibt sich die Genauigkeit zu

$$Genauigkeit = \frac{T.P. + T.N.}{T.P. + T.N. + F.P. + F.N.} \tag{3.4}$$

Training

Für das Trainieren und Anpassen des Algorithmus ist ein Trainings- und ein Validierungsset, wie in Abbildung 3.9 gezeigt, notwendig. Das Testset dient zum Testen des Algorithmus, nachdem dieser fertig trainiert und angepasst wurde.(Abb. 3.9). Das Trai-

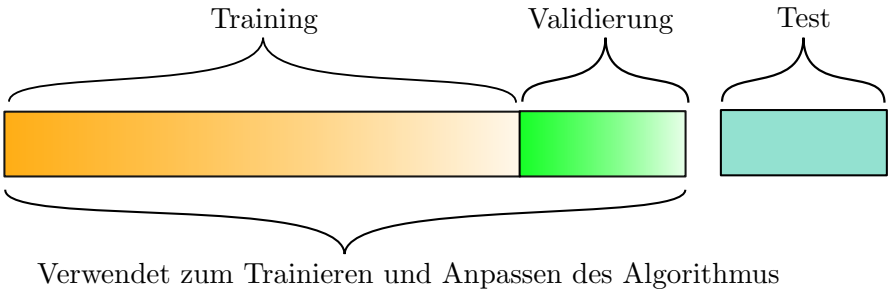


Abbildung 3.9: Aufspaltung der Daten in ein Traings-, Validierungs- und Testset (Hold-Out-Validierung)

ningsset wird, wie schon der Name verrät, zum Trainieren der Algorithmen verwendet. Der Algorithmus erlernt anhand dieser Daten z.B. eine Funktion, mit welcher er die Klassen unterscheiden kann. Übergeben werden ihm nur Attribute, welche sich in der vorigen Datenanalyse herauskristallisiert haben.

Das Validierungsset dient anschließend zur Überprüfung, ob die Parameter der Algorithmen richtig eingestellt wurden und ob der Algorithmus überhaupt eine Erkennung ermöglicht. Diese Art der Aufspaltung der Daten wird Hold-Out-Validierung genannt. Ein anderes Verfahren stellt die, in Abbildung 3.11 gezeigte, Kreuzvalidierung dar.

Beim Training der Algorithmen muss darauf geachtet werden, dass es zu keiner Unter- oder Überanpassung kommt. Die beiden Fälle sind beispielhaft in Abbildung 3.10 dargestellt. Unteranpassung tritt z.B. auf, wenn das Modell die vorliegenden Daten nicht genau genug beschreiben kann. Die Ursachen können entweder das Vorhandensein von zu wenig Trainingsdaten sein oder das Beschreiben oder die Verwendung eines Linearen Modells für einen nicht linearen Datensatz. Eine Überanpassung tritt auf, wenn ein Modell die Details und das Rauschen in den Trainingsdaten in einem Maße erlernt, dass es die Leistung des Modells auf neuen Daten negativ beeinflusst, d.h. das Modell lernt quasi „zu genau“. In der Regel kann der Überanpassung durch die Verwendung von mehr Daten, Reduktion der Dimension oder durch das bereits erwähnte Kreuzvalidierungsverfahren entgegengewirkt werden. Es existieren noch mehrere Verfahren die jedoch hier nicht angeführt sind.

Beim erwähnten Kreuzvalidierungsverfahren werden die Trainingsdaten (orange und grün in Abb. 3.9) in mehrere gleich große Sets zerlegt(siehe Abb. 3.11) . Von diesen kleineren Sets wird eines als Validierungsset (orange) und die übrigen als Trainingsset(leer) spezifiziert. Mit den Trainingssets wird anschließend der Algorithmus trainiert und mit dem

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

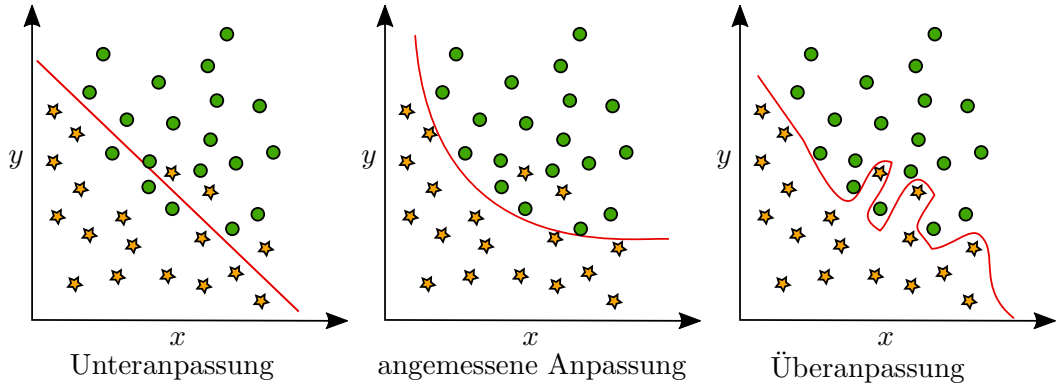


Abbildung 3.10: Anpassung an die Daten

Validierungsset getestet. Anschließend beginnt der Zyklus von neuem, bis jedes Set einmal als Validierungsset verwendet worden ist (im Bsp. werden also vier Läufe durchgeführt). Das Ergebnis ergibt sich aus dem Mittelwert aller Durchgänge. Mit dieser Methode können Fähigkeiten eines Machine-Learning-Modells besser eingeschätzt werden als durch ein einfaches Aufspalten in ein fixes Trainingsset und ein Validierungsset wie bei der in Abbildung 3.9 gezeigten Hold-Out Validierung.

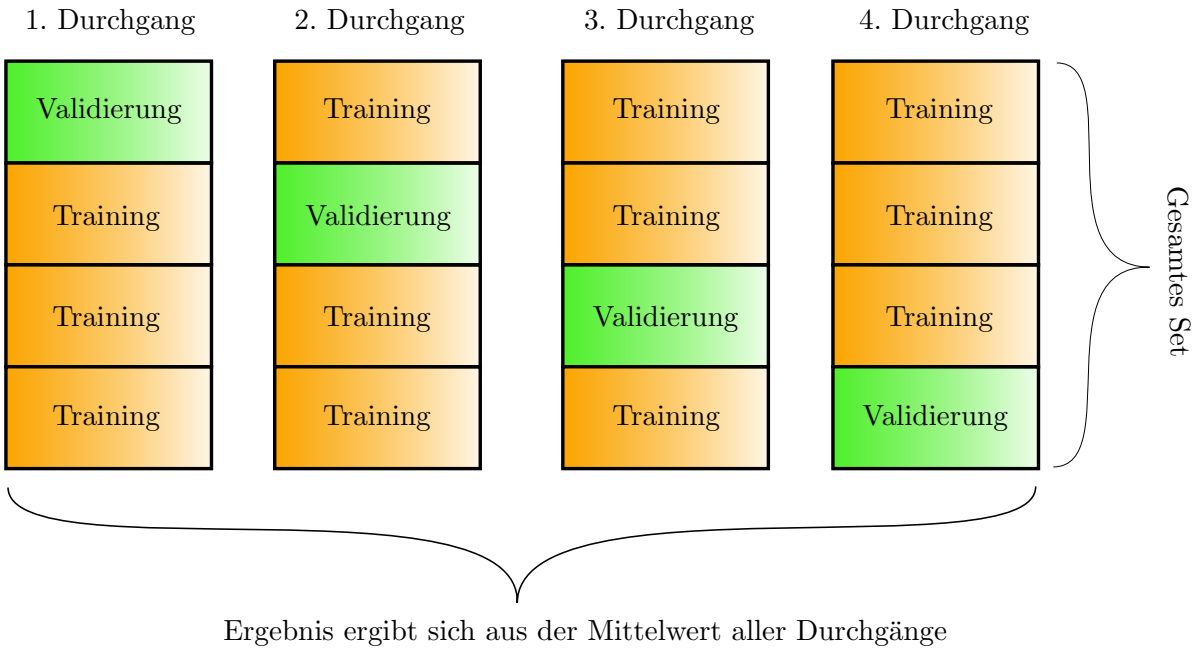


Abbildung 3.11: Kreuzvalidierungsverfahren

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Variation der Parameter

Liefert ein Algorithmus keine akzeptablen Ergebnisse, werden seine Parameter geändert. Diese Parameter unterscheiden sich natürlich je nach verwendetem Algorithmus. Die richtige Konfiguration ist essenziell, um eine effiziente Erkennung zu erzielen. Sie werden so lange verändert bis ein Optimum erreicht ist, d.h. die Erkennung die bestmögliche Genauigkeit erreicht hat. Um später in den Ergebnissen zu verstehen, warum die Parameter so eingestellt wurden, wird es hier kurz am Beispiel des KNN gezeigt.

Im Abschnitt 2.2 *Maschinelles Lernen* wurde angekündigt, dass bei den KNN-Algorithmen vor allem auch die Anzahl der Nachbarn, die Distanzbestimmung und das Distanzgewicht einmal außer Acht gelassen, von entscheidender Bedeutung ist. Um die optimale Anzahl der Nachbarn zu bestimmen, werden mehrere Läufe durchgeführt und dabei die Anzahl der Nachbarn variiert. Dabei wird zunächst eine grobe Analyse durchgeführt, um eine ungefähre Ahnung von der besten Anzahl zu erhalten. Abbildung 3.12 zeigt eine solcherart Erstanalyse. Es ist deutlich zu erkennen, dass das Optimum irgendwo zwischen 0 und 50 Nachbarn liegt. In diesem Bereich wird anschließend noch eine feine Analyse durchgeführt. Die Anzahl, welche das beste Ergebnis liefert, wird schlussendlich verwendet. Das Optimum ist nicht in allen Fällen so eindeutig erkennbar wie in dem gezeigten Beispiel. Die Optima hängen zum Teil auch vom verwendeten Validierungsset ab. Auch hier ist wieder ein Kreuzvalidierungsverfahren von Vorteil, um nicht von einem einzigen Validierungsset abhängig zu sein.

Analog wird bei der Bestimmung weiterer Modellparameter vorgegangen. Ein Optimum für den Machine-Learning-Algorithmus zu finden, in dem mehrere Parameter variierbar sind, erweist sich allerdings nicht immer als einfach. Die Parameter beeinflussen sich gegenseitig im Bezug auf die erzielbare Genauigkeit. Wird ein Parameter verändert, ändert sich auch das Optimum für den anderen Parameter und umgekehrt.

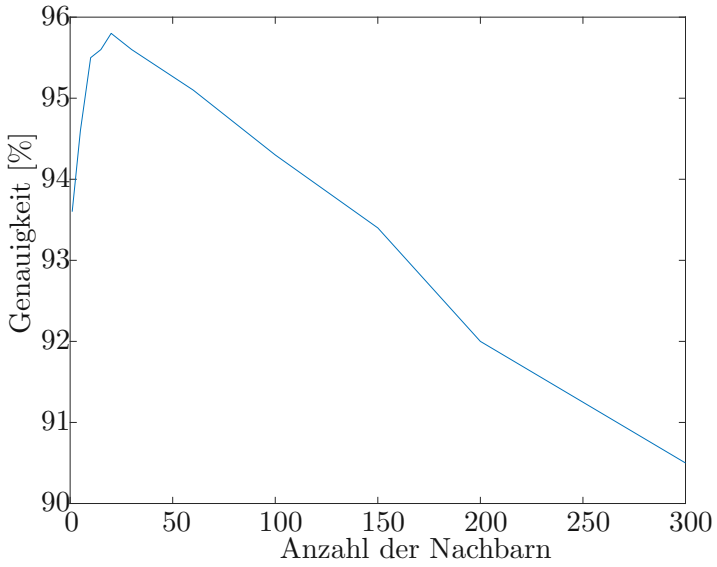


Abbildung 3.12: KNN Erkennungsgenauigkeiten bei Variation der Nachbarn

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.4 Umsetzung der Methodik an einem spezifischen Netz

Als Modell für die Untersuchung wurde das IEEE-14-Bus-System ausgewählt(Abb. 3.13). Das Netz repräsentiert einen Teil des amerikanischen Stromnetzes im Jahr 1962. Für diese Auswahl gibt es mehrere Gründe.

Der Entwurf eines realistischen Netzes ist mit hohem Aufwand verbunden. Die Bestimmung der Topologie, die Verteilung der Generatoren und Lasten sowie die Länge und Art der Leitungen und vieles mehr müsste bestimmt werden. Durch die Verwendung eines bereits existierenden Netzes wird der Aufwand reduziert und damit Zeit eingespart, wobei die Aussagekraft der Ergebnisse mindestens genauso hoch ist wie bei einem selbst entworfenen Netz. Weiters hat ein publiziertes Netz wie dieses eine besondere Stellung in der Forschung, da es durch die Verbreitung in den Publikationen äußerst gut analysiert ist und damit eine Vergleichbarkeit mit anderen Publikationen leichter ist bzw. sich die Ergebnisse so leichter in ein Gesamtbild einfügen lassen.

Zweitens erfüllt es alle oben genannten Bedingungen. Das Netz ist sowohl auf der höheren sowie auf der niedrigeren Spannungsebene vermascht ausgeführt, wobei sich hier die Untersuchungen auf die Hochspannungsseite begrenzen. Außerdem enthält es zwei Umspannwerke, sowie auch Doppelleitungen die für die Untersuchung von Interesse sind.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

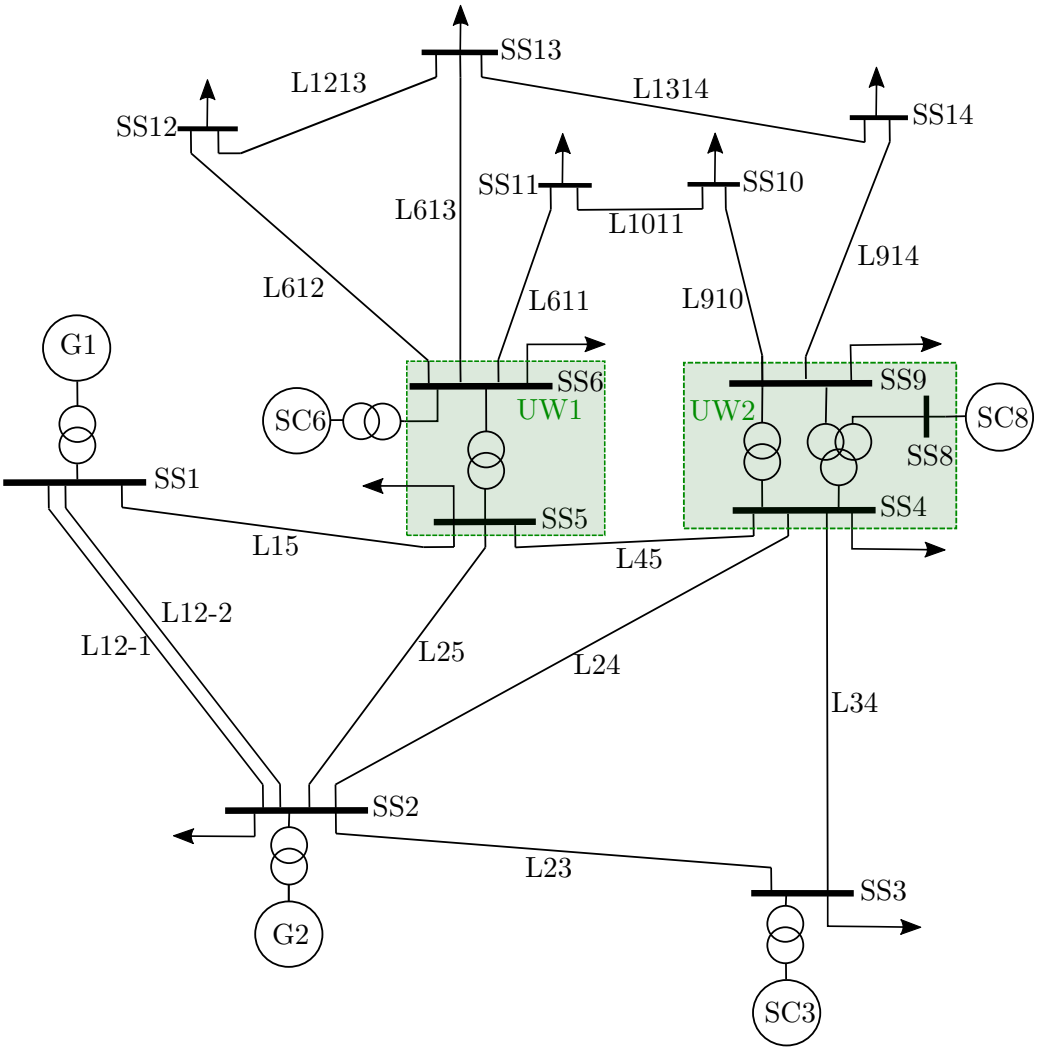


Abbildung 3.13: IEEE 14-Bus-System

Insgesamt sind an das Netz zwei Einspeiser $G1$ und $G2$, drei Synchronkompensatoren $SC3, SC6$ und $SC8$ sowie elf Verbraucher (dargestellt durch die abgehenden Pfeile an den SS) angeschlossen. Die Daten der Elemente befinden sich im Anhang. Die Generatoren sind jeweils über einen Transformator ans Netz gekoppelt. Die Ausnahme bildet SC8, der über einen Dreiwickeltransformator an das Netz gekoppelt ist.

Die 14 Sammelschienen (SS) sind über 15 Leitungen miteinander verbunden. Die Abbildung 3.13 zeigt das Einlinienschaltbild ohne Repräsentation der Schalter. SS1 und SS2 sind über zwei Systeme (L12-1 und L12-2) miteinander verbunden. Alle anderen SS sind über einzelne Leitungssysteme verbunden. Die Synchronkompensatoren (SC) dienen zur Blindleistungskompensation am jeweiligen Anschlusspunkt. Die SC-en sind im Prinzip nichts anderes als Synchronmotoren dessen Welle keine mechanische Last antreibt. Im SC-Betrieb muss Wirkleistung aufgenommen werden, da die Lagerreibung und der strömungdynamische Widerstand des im gasgefüllten Raum drehenden Rotors überwunden

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

werden muss. Das Blindleistungsverhalten der Synchrongeneratoren lässt sich über die Erregung beeinflussen. Die Aufgabe des SC3 an der SS3 ist die Blindleistungskompensation und damit die Spannungsstützung der großen Last an dieser Stelle, welche den größten Verbraucher im Netz darstellt..

3.4.1 Auswahl der Ausfälle

Im *Abschnitt 3.1* wurde erläutert, dass Ausfälle untersucht werden, welche nicht direkt mit den UWs verbunden sind. In der Arbeit ist der Zustand von Bereichen des Netzes interessiert, der aufgrund fehlender direkter Verbindung zum UW nicht aus Messwerten die im UW vorhanden sind, direkt abgeleitet werden kann. Für das UW1 werden die Leitungen L12-1, L23 und L34 herangezogen. Für das UW2 die Leitungen L12-1 und L23. L34 ist mit dem UW 2 direkt verbunden und daher für dieses UW nicht von Interesse.

Interessant ist vor allem auch der Ausfall einer der Leitungen des Doppelsystems zwischen SS1 und SS2, da er keinen kompletten Ausfall der direkten Verbindung zwischen zwei SS bedeutet, sondern eine direkte Verbindung zwischen den SS verbleibt.

Direkte und entfernte Masche:

Als direkte Maschen werden dabei die in Abbildung 3.14 gezeigten Maschen bezeichnet. Die blauen Leitungen inklusive der Violetten stehen für die direkten Maschen des UW1. Die rot markierten Leitungen inklusive der violetten Leitungen zeigen die direkten Maschen des UW2. Umgekehrt bedeutet z.B. der Ausfall der Leitung L23 für das UW1 einen Fehler in einer entfernten Masche, während er für das UW2 einen Ausfall in der direkten Masche repräsentiert. Diese zwei Begriffe werden eingeführt, um eine Unterscheidung zwischen Ausfällen die nahe am UW liegen und Ausfällen die etwas weiter entfernt sind zu ermöglichen. Es wird nämlich angenommen, dass die Ausfälle von Leitungen welche sich in der direkten Masche des UWs befinden für das UW leichter zu erkennen sind als Leitungsausfälle in entfernten Maschen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

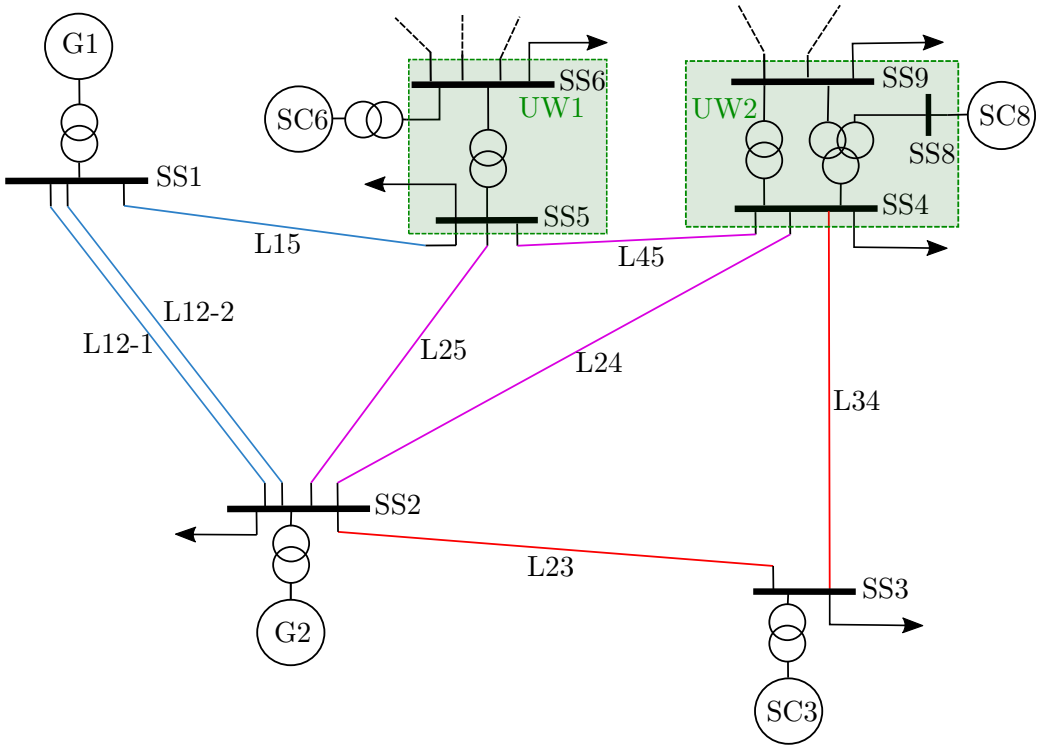


Abbildung 3.14: Direkte Maschen der UWs: blau(UW1), rot(UW2), violett(UW1 & UW2)

3.4.2 Erfassung der Betriebszustände

Für jede durchgeführte Netzberechnung werden die Ströme aller Abgänge sowie die Spannungen an den SS aufgenommen. Berücksichtigt werden auch die Ströme über die zwei Transfomatoren zwischen der SS4 und SS9. Zur Beschränkung der Arbeit sowie der Daten wird nur eine Spannungsebene betrachtet und die Untersuchungen auf diese Spannungsebene konzentriert. Die betrachteten Ausfälle auf der höheren Spannungsebene wirken sich generell wesentlich geringer auf die Messwerte der niederspannungsseitigen Sammelschiene aus als auf die der eigenen Abgänge.

Abbildung 3.15 zeigt schmatisch welche Größen aufgenommen wurden. Die beiden Ziffern stehen immer für die SS, mit welchen die Abzweigen verbunden sind (z.B. bezeichnet I_{51} den Strom der im gegeben Bezugssinn von SS5 nach SS1 fließt). Steht an zweiter Stelle ein L, so befindet sich an diesem Abgang eine Last. In der Abbildung sind die Bezugssinne der Ströme sowie deren Bezeichnung angeführt.

Der Winkel φ ergibt sich aus der Phasendifferenz zwischen Spannung und den jeweiligen Strömen. Es wird, wie im Abschnitt 3.1 *AllgemeineBeschreibung* beschrieben, jeweils eine Messung im Normalzustand und eine nach dem Ausfall der Leitung aufgenommen. Wichtig ist hier nochmal anzumerken, dass keine Zeitreihen gemessen werden und somit keine transienten Vorgänge für die Erkennung der Netzzustände herangezogen werden.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

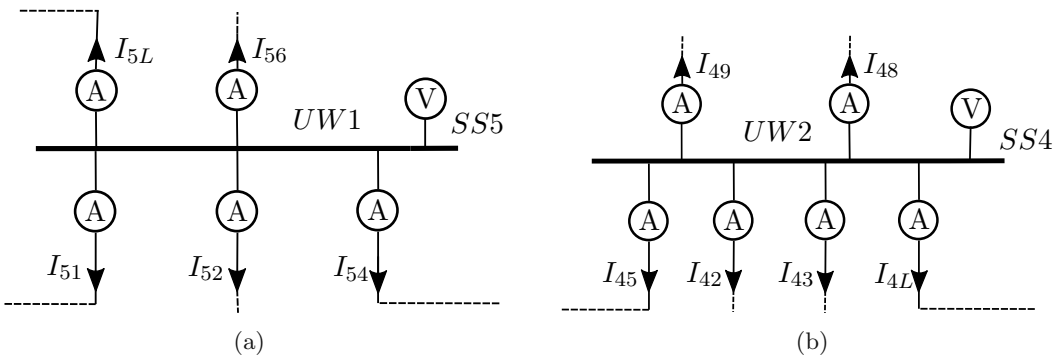


Abbildung 3.15: Messgrößenaufnahme und Darstellung der Bezugssinne im UW1 und UW2

3.4.3 Implementierung in MATLAB

Die Simulationen wurden von Grund auf in $MATLAB^{®}$ / $SIMULINK^{®}$ unter zu Hilfenahme von $SimscapeElectrical^{TM}$ durchgeführt. $SimscapeElectrical^{TM}$ bietet Komponentenbibliotheken für die Modellierung und Simulation elektrischer Stromsysteme.
Bei jedem Simulationsstart wird dabei ein spezieller Initialisierungsprozess gestartet. Dieser Prozess bildet das elektrische System auf ein Gleichungssystem in Zustandsraumdarstellung ab, welches von Simulink simuliert werden kann. Dieser Prozess läuft grundsätzlich in drei Schritten ab [12]. Zuerst werden die Parameter der Blöcke eingelesen sowie die Netztopologie evaluiert. Die Blöcke werden in lineare und nichtlineare Blöcke unterteilt. In einem zweiten Schritt wird das Zustandsraummodell (A,B,C, D Matrizen)

$$\dot{\mathbf{x}} = \mathbf{A} \cdot \mathbf{x} + \mathbf{B} \cdot \mathbf{u} \tag{3.5}$$

$$\mathbf{y} = \mathbf{C} \cdot \mathbf{x} + \mathbf{D} \cdot \mathbf{u} \tag{3.6}$$

für den linearen Teil des Systems bestimmt. Im letzten Schritt wird das Simulink Modell entworfen. Abbildung 3.16 zeigt schematisch das komplette Simulink Modell samt der Verbindungslinien der einzelnen Blöcke. Das nichtlineare Modell ist als Rückkopplung zwi-

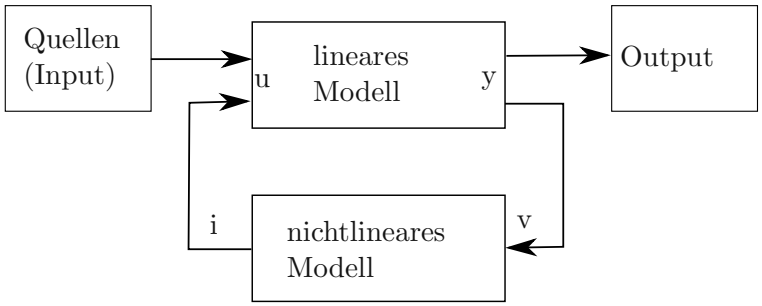


Abbildung 3.16: Verbindung zwischen linearem und nichtlinearem Modell

schen Spannungsausgang und Stromeingang des linearen Modells geschaltet. Nichtlineare Modelle werden als Stromquellen simuliert. Die Lasten sind als äquivalente Impedanzen (d.h. linear) modelliert. Ist der Initialisierungsprozess abgeschlossen startet die Simulation.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Eine genauere Beschreibung dieses Prozesses ist in [12] zu finden. Als Lösungsverfahren wurde das Tustin/Backward-Euler (TBE) Verfahren verwendet.

Das IEEE14-Bus Simulationsmodell wurde von [19] entnommen, wobei die SCs wurden als Synchronmotoren ohne mechanische Last modelliert wurden. Als Regler der SCs wurden IEEE-Type-DC1A-Excitation Systeme [1] verwendet.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

4 Auswertung und Validierung der Ergebnisse

In diesem Kapitel werden die Ergebnisse für das in Kapitel 3 definierte Modell präsentiert. Im ersten Teil erfolgt die Präsentation und die Analyse der durch die Simulation des IEEE14-Bus-Systems gewonnenen Daten. Der zweite Teil zeigt die Ergebnisse der Machine-Learning-Algorithmen. In beiden Abschnitten erfolgt zuerst eine Unterteilung in die ausgefallenen Systeme und eine weitere Unterteilung in die beiden Umspannwerke. Die einzelnen Ausfälle werden in einem ersten Schritt getrennt voneinander betrachtet um Unterschiede vom gestörten Betriebsfall zum Normalbetrieb zu zeigen. In einem zweiten Schritt erfolgt die gemeinsame Betrachtung der Ausfälle.

4.1 Analyse der Daten

Die Analyse der Daten erfolgt nach der Beschreibung im Kapitel 3 *Methodik*. Grafisch werden die Ergebnisse allerdings nur zweidimensional dargestellt, weil höherdimensionale Darstellungen nicht mehr grafisch aussagekräftig oder gar möglich sind. Insgesamt wurden 6900 Betriebszustände des Normalzustandes sowie 6700 Betriebszustände für jeden der Ausfälle aufgenommen.

4.1.1 Ausfall eines Systems des Doppelsystems zwischen SS1 und SS2

Bei einem Ausfall einer Leitung zwischen Sammelschiene 1 und Sammelschiene 2 ändert sich der Lastfluss über die Leitung 1 und 2 durch die Erhöhung der Impedanz 12.

Umspannwerk 1

An den Messergebnissen ist erkennbar, dass mit der erfolgten Methode der Lastgenerierung alle angestrebten Betriebszustände erreicht werden. Die Beträge der Winkel φ sowie der Ströme sind vor allem durch die Lastsituation bestimmt. Der Ausfall der Leitung hat Auswirkungen auf die Beziehung der einzelnen Messgrößen zueinander. Bei Ausfall einer der Doppelleitung zwischen Bus 1 und 2 ist auffallend, dass sich besonders die Impedanzwinkel φ_{15} und φ_{25} vom Normalbetrieb unterscheiden (Abb. 4.2(a)). Auch die Kombination der Beträge der Ströme I_{15} und I_{25} unterscheidet sich in der Abbildung 4.2(b) sichtbar. Sie haben im Bezug auf diesen Fehler den höchsten Informationsgehalt. Dass der Ausfall genau an diesen Messgrößen sichtbar ist ergibt dahingehend Sinn, da die Leitungen L15 und L25 am anderen Ende über die Sammelschienen direkt mit dem Doppelsystem verbunden sind. Sie sind daher am stärksten von dem Ausfall eines dieser Systeme betroffen. Der Betrag der Spannung an der SS5 hingegen hat nur wenig Aussagekraft. Die Änderung der Werte an den anderen Strömen fällt weniger eindeutig aus. Sie sind im Vergleich nicht so stark von dem Ausfall betroffen wie die Leitungen L15 und L25. Wie in den Abbildungen zu erkennen ist besteht eine positive Kovarianz zwischen den Variablen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

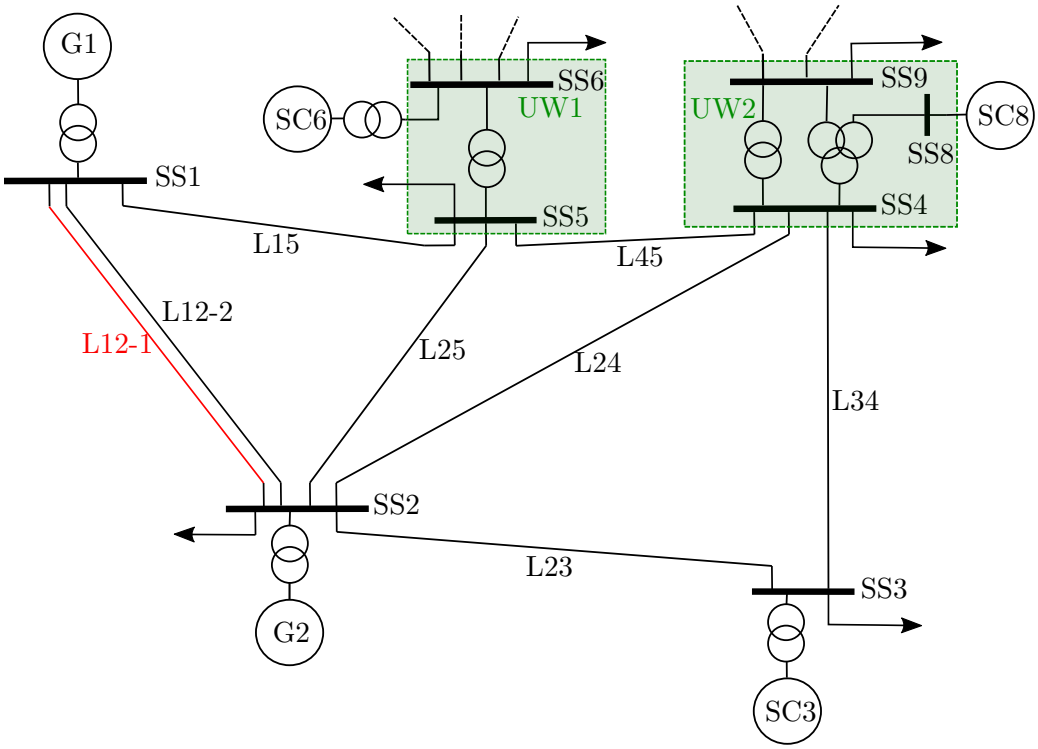


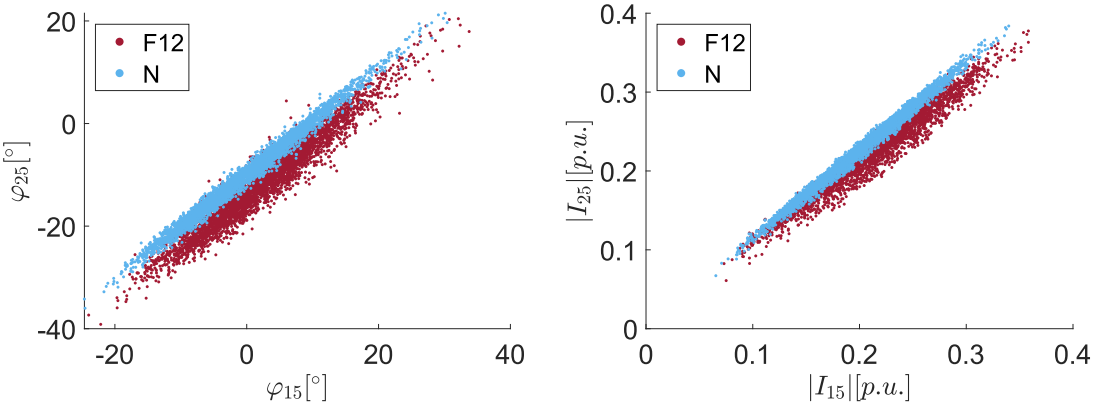
Abbildung 4.1: Ausfall der Leitung L12-1

Dies ist deshalb wichtig anzumerken, da es bei der Implementation der Machine-Learning-Algorithmen berücksichtigt werden muss. Einige Algorithmen berüchsichtigen diese Kovarianz und können in solchen Fällen eine höhere Genauigkeit erzielen. Die Messgrößen weisen aber auch Überschneidungen der beiden Betriebszuständen auf (Abb. 4.3). Das bedeutet, anhand zweier Messgrößen wird zwar ein grobe Unterscheidung möglich sein, eine komplette Teilung in zwei Klassen erfordert aber die Kombination von mehr Attributen. Durch diese Kombination kommt es zu einem weiteren Informationsgewinn, wodurch eine bessere Unterscheidung der Klassen ermöglicht wird. Ausfälle von einem System eines Doppelsystems in der direkten Masche eines UWs zeigen zum Teil eindeutige Unterschiede zu den Daten des normalen Betriebszustandes.

Umspannwerk 2

Im UW2 ist der Ausfall dieses Systems nicht so eindeutig erkennbar. Wie in den Abbildungen 4.3 ersichtlich, ist zumindest im zweidimensionalen euklidschen Vektorraum keine klare Trennung der Daten in den Normal- und Fehlerzustand mehr möglich. Die anderen Attribute zeigen ein ähnliches Bild. Bis auf die gemeinsame Betrachtung des Stromes I_{24} und des Winkel φ_{24} liefert keine Kombination eine aussagekräftige Information. Dass genau diese zwei Attribute etwas Information enthalten ergibt sich aus der Tatsache, dass die Leitung zwischen Sammelschiene 2 und 4 direkt von der Änderung des Lastflusses durch das Ausfallen der Leitung 12 betroffen sind. Allerdings ist dieser Zustand nur bedingt von den normalen Betriebszuständen unterscheidbar.

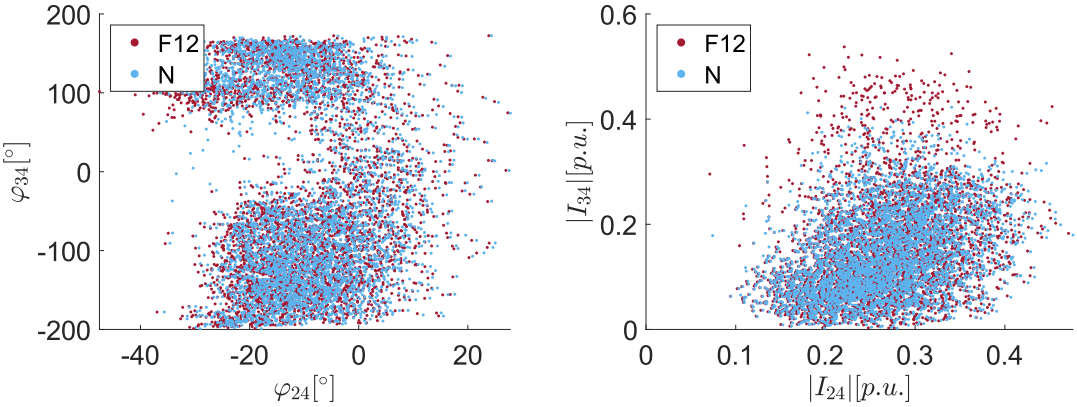
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Abbildung 4.2: Simulationsergebnisse UW1: Impedanzwinkel und Ströme der Leitungen 15 und 25

Es können keine Bereiche gefunden werden die eindeutig einem Betriebszustand zuordenbar wären. Auch im Bezug auf Standardabweichungen und Mittelwert zeigen die Attribute kaum Unterschiede. Durch die Hinzunahme von Attributen hin zu höheren Dimensionen kann es jedoch zu einem Informationsgewinn kommen. Dies wird sich dann in den Erkennungsgenauigkeiten der Algorithmen zeigen. Allgemein lässt sich aber anhand der gemes-



(a) (b)

Abbildung 4.3: Simulationsergebnisse UW2: Impedanzwinkel und Ströme der Leitungen 24 und 34

senen Werte sagen, dass der Ausfall eines Systems eines Doppelsystems in einer entfernten Masche (vom UW2 aus) keine bis nur sehr kleine Änderungen mit sich zieht, wodurch eine Erkennung bzw. Klassifizierung dieses Ausfalls schwierig wird.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.1.2 Ausfall von L23

Ein Ausfall der Leitung zwischen 2 und 3 zieht eine starke Änderung des Lastflusses mit sich (Abb. 4.4). Im Vergleich zu vorher handelt es sich nicht lediglich um eine Impedanzerhöhung einer Leitung zwischen zwei SS, sondern um eine grundsätzliche Änderung der Topologie. Um Last 3 versorgen zu können, muss sich der Lastfluss statt über Leitung 23 einen anderen Weg suchen. Hinzu kommt, dass Last 3 den größten Verbraucher im gesamten Netz darstellt. Wird diese Last z.B. auf das 10-fache reduziert, so verkleinen sich die Auswirkungen des Leitungsausfalls auf die Messwerte im Umspannwerk um die Hälfte im Vergleich zum Ausfall bei einem voll belasteten Netzes. D.h. das Ausmaß der Messwertänderung bei Ausfall der Leitung hängt stark von der Größe dieser Last ab.

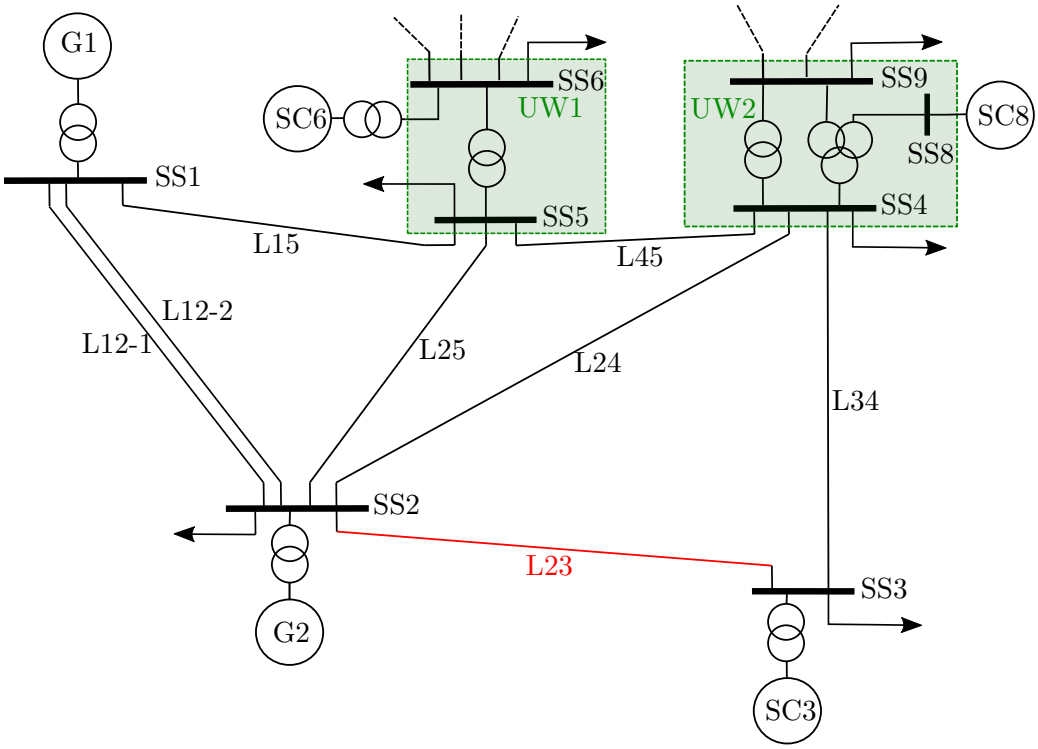


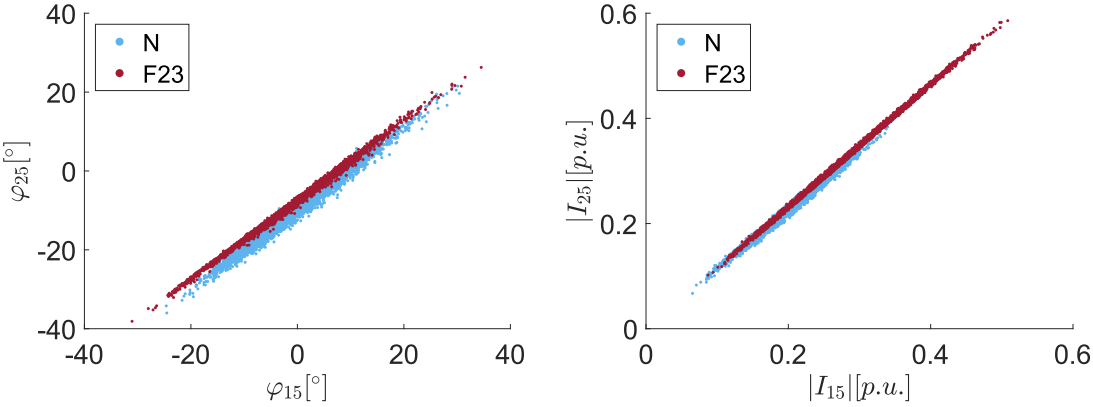
Abbildung 4.4: Ausfall der Leitung L23

Umspannwerk 1

Der unterbrochene Stromfluss zwischen SS3 und SS4 zieht eine Erhöhung der Ströme der anderen Leitung mit sich. Diese Erhöhung der Ströme bewirkt einen erhöhten Spannungsabfall über die Leitungen und somit zu einem Abfall der Spannung an der SS5, welchem die Synchronkompensatoren mit der aktuellen Reglereinstellung nur bedingt entgegenwirken können.

Durch Variation der Belastungssituation im Netz zeigen sich Differenzen in den Impedanzwinkeln (zwischen Normalbetrieb und gestörtem Betrieb). Jedoch gibt es auch Betriebszustände, die nicht vom Normalzustand zu unterscheiden sind (Abb. 4.5). Die Beziehung der Winkel φ_{15} und φ_{25} zueinander ändert sich beim Ausfall von L23 nur geringfügig. Vor

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

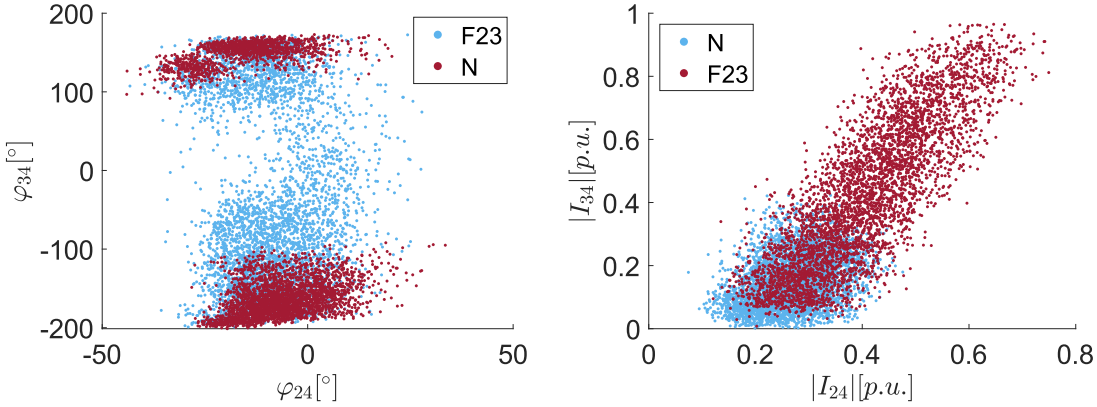
Abbildung 4.5: Simulationsergebnisse UW1: Impedanzwinkel und Ströme der Leitungen L15 und L25

allem bei hohen Belastungen ist eine eindeutige Unterscheidung anhand der Beträge der Ströme möglich.
Ein Ausfall einer solchen Leitung in einer entfernten Masche zeigt also durchaus Unterschiede zum Normalbetrieb. Das bedeutet, dass eindeutige Messwertänderungen und damit Informationen über die Änderung des Betriebszustandes vorhanden sind.

Umspannwerk 2

Die Simulationsergebnisse sind in Abbildung 4.6 dargestellt. Die Verteilung der Daten ist etwas diffuser als im UW 1. Es zeigen sich aber durchaus Unterschiede. Wie auch im vorhergehend Punkt sind Unterscheidungen bei größerer Last allein anhand der Ströme I_{24} und I_{34}. möglich. Bei Ausfall der Leitung lässt sich ein linearer Zusammenhang der Ströme feststellen. Im Schwachlastbereich hingegen erweist sich die Unterscheidung als schwieriger. Die einzelnen Messgrößen des UWs zeigen jedoch in Hinsicht auf Mittelwert und Standardabweichung durchaus Unterschiede. Auch in den Winkeln existieren relativ große Bereiche die sich auch Unterscheiden lassen. Aus den Winkeln ist die Tatsache ersichtlich, dass der Lastfluss der Leitung L34 bei diesem Ausfall nur in Richtung SS3 zeigt, während er sich im Normalbetrieb in manchen Fällen umdreht.
Um in allen Belastungsszenarien eine Erkennung des Ausfalls von L23 im UW2 zu ermöglichen, müssen also Ströme und Winkel zusammen herangezogen werden.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN



(a) (b)

Abbildung 4.6: Simulationsergebnisse UW2: Impedanzwinkel und Ströme der Leitungen 24 und 34

4.1.3 Ausfall von L34

Wie auch bei dem Ausfall von L23 ändert sich der Lastfluss bei einem Ausfall von L34 massiv. Die Last am SS3 kann nur mehr über die Leitung 23 versorgt werden. Dieser Ausfall wird nur im UW 1 untersucht. Eine Untersuchung im UW 2 wäre wie erwähnt überflüssig, da die Leitung direkt mit diesem Umspannwerk verbunden ist (siehe Abb. 4.7). Vor allem der Vergleich zum Ausfall von L23 ist hier interessant.

Umspannwerk 1

Durch den Ausfall von L34 ergeben sich geringfügige Änderungen an allen Messwerten des Umspannwerks. Diese, vor allem zu Schwachlastzeiten, nicht sehr ausgeprägte Änderung macht eine Unterscheidung der Betriebszustände schwierig, da die Daten kaum vom Normalbetrieb bei verschiedenen Belastungen zu unterscheiden sind (siehe Abbildung 4.8). In der Abbildung ist unschwer die Überlappung der Daten zu erkennen. Nur im äußersten Starklastfall ($I_{15} \geq 0.3$ p.u.) ist ein Unterschied erkennbar. Die anderen Messwerte zeigen ähnliches Verhalten. Im Vergleich zu den vorhergehenden Ausfällen erweist es sich auch durch Kombination der Messgrößen als schwierig, optisch aus Mensch due Klassengrenzen zu ziehen. Die Beziehung der Ströme I_{15} und I_{25} zueinander ändert sich beim Ausfall von L23 stärker als bei diesem Ausfall. Ähnliches gilt für die Impedanzwinkel φ_{15} und φ_{25}. Im UW1 ist nicht ausreichend Information in den Messwertänderungen vorhanden um einen Fehler zu detektieren.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

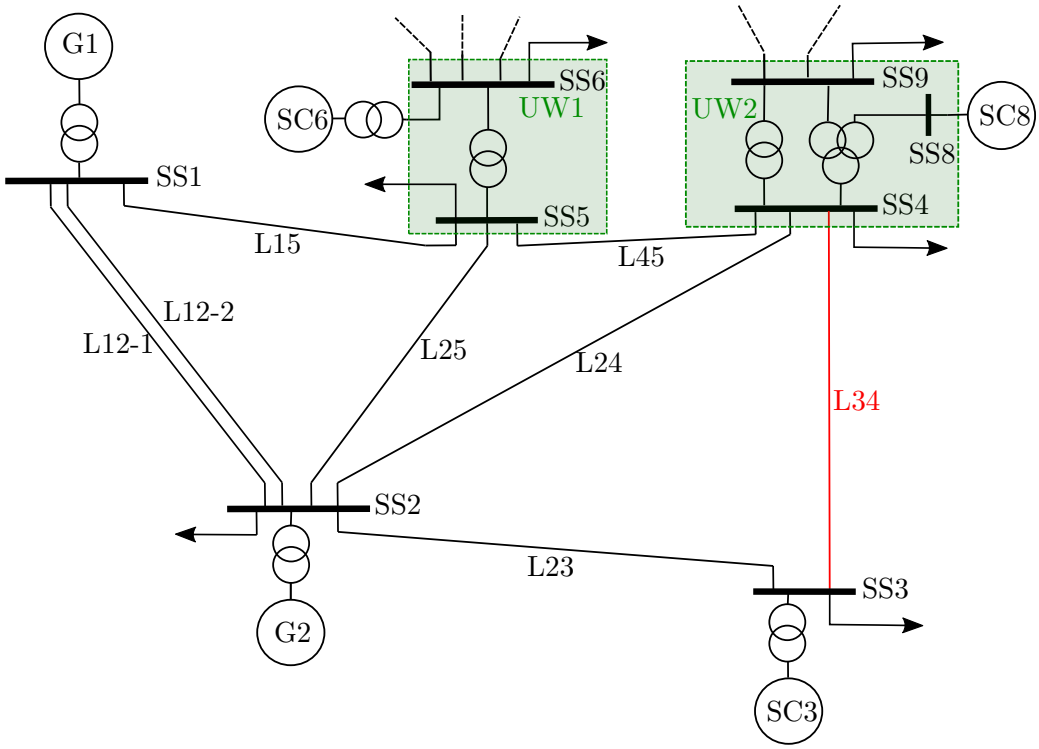
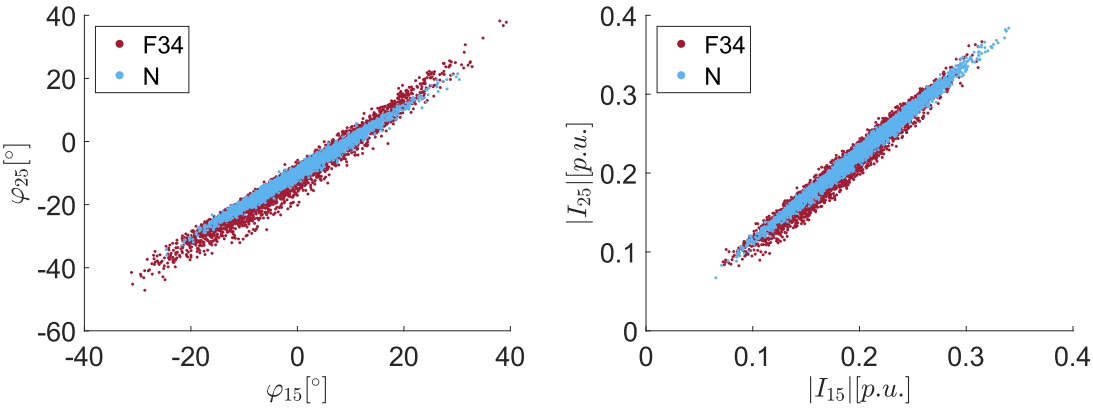


Abbildung 4.7: Ausfall der Leitung L34



(a) (b)

Abbildung 4.8: Simulationsergebnisse UW1: Impedanzwinkel und Ströme der Leitungen 24 und 34

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

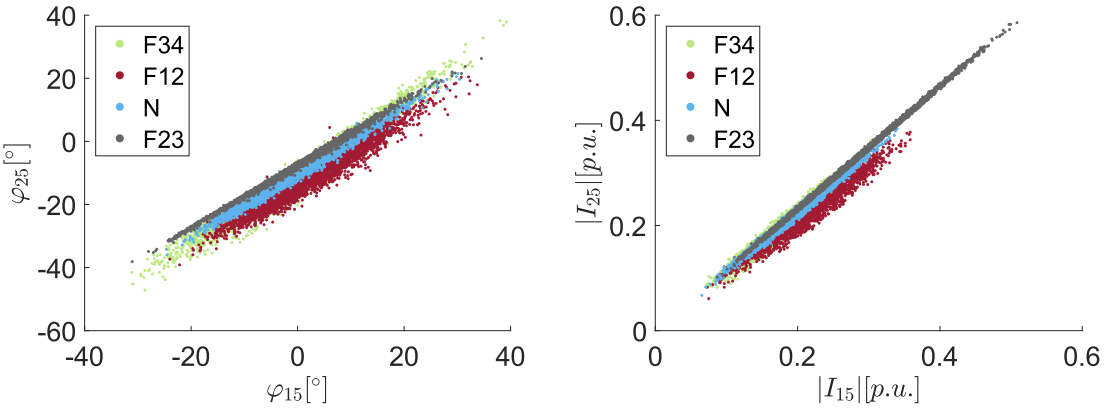
4.1.4 Gemeinsame Betrachtung der Ausfälle

Nach der getrennten Analyse der Ausfälle werden hier alle Ausfälle gemeinsam betrachtet. Dabei wird untersucht, ob sich die Messwerte der verschiedenen Ausfälle ähneln oder überlappen und daher in weiterer Folge ein Absenken der Genauigkeit in den Erkennungen resultiert.

Umspannwerk 1

Die Simulationsergebnisse sind in Abbildung 4.9 ersichtlich. Unterschiede in den Daten sind mit freien Auge und anhand einer dieser zweidimensionalen Grafik erkennbar. Dargestellt sind wiederum die Impedanzwinkel φ_{15} und φ_{25} sowie die Ströme I_{15} und I_{25}. Der Fehler 34 wird hier nur hinzugefügt, um seine Auswirkungen auf die Erkennung der anderen Zustände zu analysieren. Im vorigen Abschnitt habe wir gesehen dass er nur mit sehr viel Unsicherheit vom Normalzustand zu unterscheiden ist.

Die Ausfälle von L23 und L12-1 zeigen deutliche Unterschiede. Dies ergibt auch durchaus Sinn, da sie sich insgesamt um zwei Netzelemente unterscheiden und sich der Lastfluss daher grundsätzlich anders aufteilt. Vorteilhaft ist auch, dass sich der Ausfall von L23



(a) Winkel (b) Ströme

Abbildung 4.9: Gemeinsame Betrachtung der Ausfälle

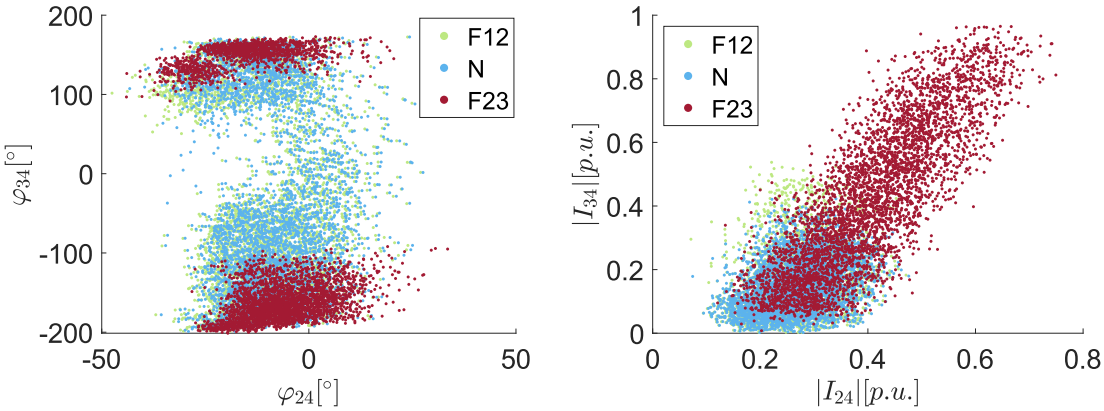
auf z.B. den Strom I_{54} stark auswirkt, während sich der Ausfall L12-1 vor allem an den Messgrößen der Abgänge 51 und 52 zu sehen ist.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Umspannwerk 2

Wie schon gezeigt wurde, ist die Unterscheidung des Ausfalls von L12-1 im UW2 vom Normalzustand nur sehr schwierig möglich. Der Fehler wird in diesem Abschnitt aber miteinbezogen, um zu untersuchen, ob er Überlappungen mit den Messwerten vom Ausfall zwischen SS2 und SS3 aufweist. In Abbildung 4.10 sind die Ergebnisse dargestellt. Abgebildet sind die Ströme der Leitungen 24 und 34 sowie die Impedanzwinkel φ_{24} und φ_{34}.

Es sind durchaus Unterschiede zwischen den beiden Ausfällen erkennbar. Laut den Daten sollte es also zu keinen allzu starken Beeinflussungen des Ausfalls von L12-1 auf die Erkennung des Fehlers 23 und umgekehrt kommen. Die Auswirkung der Daten des Ausfalls von L12 sollten kaum Auswirkungen auf die Erkennung des Ausfalls von L23 haben.



(a) Winkel (b) Ströme

Abbildung 4.10: Gemeinsame Betrachtung der Ausfälle im UW2

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.2 Ergebnisse der Maschine-Learning-Algorithmen

In diesem Abschnitt werden die Klassifizierungsergebnisse der Machine-Learning-Algorithmen präsentiert. Dabei werden sowohl die Anzahl der Eingangsparameter sowie auch die Erkennungsalgorithmen variiert. Die verwendeten Attribute werden in den Tabellen jeweils angegeben. Als Ergebnis dient die Genauigkeit der Erkennungsalgorithmen. Sie wird in Prozent angegeben und unterteilt sich jeweils in True-Positiv(T.P.), True-Negativ(T.N.), False-Positiv(F.N.) und False-Negativ(F.N.). True-Positiv bezeichnet dabei den Fall, dass ein Ausfall einer Leitung auch als solcher erkannt wurde. True-Negativ bezeichnet dasselbe für den Normalzustand. False-Positiv und False-Negativ bezeichnen jeweils die Fälle in denen die Betriebszustände falsch klassifiziert wurden. Die folgenden Abschnitte sind gleich unterteilt wie bei der vorigen Analyse der Daten.

4.2.1 Ausfall eines Systems des Doppelsystems zwischen SS1 und SS2

Wie im vorigen Abschnitt gezeigt wurde ist der Ausfall der L12-1 an den Messwerten im UW1 deutlich erkennbar während im UW2 nur schwer eine Unterscheidung zu treffen ist.

Umspannwerk 1

Die Genauigkeiten der einzelnen Algorithmen sind in Tabelle 4.1 ersichtlich. Wie sich zeigt

Tabelle 4.1: Ergebnisse Machine-Learning-Algorithmen im UW1 bei Ausfall der Leitung zwischen SS1 und SS2

Algorithmus	**Input**	**T. P.**	**T. N.**	**F. P.**	**F. N.**	**Genauigkeit**
KNN-Euklid	$I_{51}, I_{52}, \varphi_{51}, \varphi_{52}$	$99,6\%$	$97,6\%$	$2,4\%$	$0,4\%$	$98,6\%$
	$I_{51}, I_{52}, \varphi_{51}, \varphi_{52}, V_5$	$98,9\%$	$99,2\%$	$0,8\%$	$1,1\%$	$99,0\%$
	alle Messwerte	$89,0\%$	$87,8\%$	$12,4\%$	$11,0\%$	$88,4\%$
KNN-Mahalanobis	$I_{51}, I_{52}, \varphi_{51}, \varphi_{52}$	$99,2\%$	$99,9\%$	$0,1\%$	$0,8\%$	$99,5\%$
	$I_{51}, I_{52}, \varphi_{51}, \varphi_{52}, V_5$	$99,8\%$	$100,0\%$	$0,0\%$	$0,2\%$	$99,9\%$
Logistische Regression	$I_{51}, I_{52}, \varphi_{51}, \varphi_{52}$	$98,9\%$	$99,1\%$	$0,9\%$	$1,1\%$	$99,0\%$
	$I_{51}, I_{52}, \varphi_{51}, \varphi_{52}, V_5$	$99,5\%$	$99,2\%$	$0,8\%$	$0,5\%$	$99,3\%$
Linear SVM	$I_{51}, I_{52}, \varphi_{51}, \varphi_{52}$	$98,9\%$	$98,7\%$	$1,3\%$	$1,1\%$	$98,8\%$
	$I_{51}, I_{52}, \varphi_{51}, \varphi_{52}, V_5$	$99,3\%$	$99,6\%$	$0,4\%$	$0,7\%$	$99,4\%$

liefern mehrere Machine-Learning-Algorithmen hohe Genauigkeiten. Werden alle gemessenen Größen als Eingangsparameter verwendet, verschlechtert sich die Erkennung erheblich. Dies zeigt die Wichtigkeit der zuvor getätigten Analyse der Daten. Vor allem die Winkel φ_{51} und φ_{52} weisen für diesen Fehlerfall und dieses Umspannwerk einen hohen Informationsgehalt auf. Alleine mit ihnen sind Genauigkeiten von ca. 90% erreichbar. Es folgen die Ströme, welche etwas weniger Information enthalten. Ihre Hinzunahme führt insgesamt zu einer Verbesserung der Erkennung um einige Prozent. Die Spannung an der Sammelschiene 5 dagegen bringt nur eine Steigerung von maximal einem Prozentpunkt. Auf jeden Fall ist unter Verwendung dieser Parameter eine effiziente Erkennung mit allen verwendeten Algorithmen möglich.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

KNN: Die KNN-Methode liefert vielversprechende Ergebnisse bei der Erkennung dieses Ausfalls. Die Parameter der Erkennung wurden dabei wie in Tabelle 4.2 eingestellt. Un-

Tabelle 4.2: Einstellungen der Parameter des KNN-Algorithmus

Distanzmetrik	**Nachbarn**	**Distanzgewicht**	**Daten**
Euklid	10	inverses Abstandsquadrat	Standardisiert
Mahalanobis	120	inverses Abstandsquadrat	Standardisiert

ter Verwendung der Mahalanobis Distanz werden alle Normalzustände richtig klassifiziert und nur $0,2\%$ der Fehlerzustände falsch klassifiziert. D.h. ein einfacher Algorithmus wie der KNN kann eindeutig eine Unterscheidung zwischen Normalbetrieb und dem Betrieb bei dem Ausfall der Leitung treffen. Die Euklidsche Methode ignoriert die Kovarianz der Variablen und schneidet daher etwas schlechter ab.

Logistische Regression: iefert sehr gute Ergebnisse unter der Verwendung von 5 Parametern. Dem Algorithmus ist es also möglich eine Sigmanoid-Funktion zu generieren, mithilfe derer sich die Zustände unterscheiden lassen. Die Hinzunahme der Spannung bringt nur eine minimale Verbesserung der Erkennungsgenauigkeiten.

Lineare SVM: Die Lineare SVM zeigt ebenfalls gute Ergebnisse. Sowohl der Normalbetrieb sowie auch der Fehlerbetrieb können mit über 99%iger Wahrscheinlichkeit festgestellt werden. D.h. die Klassen können mittels einer Hyperebene im vier- oder fünfdimensionalen Raum getrennt werden. Es muss eine Soft-Margin gewählt und damit gewisse Fehler zugelassen werden, damit es dem Algorithmus möglich ist diese Hyberebene zu entwerfen.

Umspannwerk 2

Die Ergebnisse der Machine-Learning-Algorithmen sind in Tabelle 4.3 dargestellt. Der Ausfall einer Leitung zwischen SS1 und SS2 ist hier nur mehr mit sehr viel Unsicherheit feststellbar. Dies spiegelt sich in den Genauigkeiten der Erkennung verschiedener Algorithmen wider. Es ist ihnen anscheinend nicht möglich, eine eindeutige Differenzierung zwischen Normal- und Fehlerzustand zu treffen. Zum Trainieren werden nur der Strom I_{42} und der Winkel φ_{42} herangezogen. Die weitere Hinzunahme des Winkels φ_{45} zieht kaum eine Steigerung der Trefferquote mit sich. Auch durch das Heranziehen aller Attribute zeigt sich keine Verbesserung.

KNN: Mittels der KNN-Methode können ca. dreiviertel der Zustände richtig klassifiziert werden. Die Methode mit der höchsten Genauigkeit liefert gerade einmal $75,3\%$ Prozent. Im Vergleich zu den anderen Algorithmen ist allerdings anzumerken, dass mittels KNN ca. 20% mehr Zustände richtig klassifiziert werden. Werden im Vergleich alle Messwerte miteinbezogen, so verschlechtert sich die Genauigkeit auf $59,4\%$. Da in den Daten kein signifikanter Zusammenhang zwischen den Parametern ablesbar ist, scheidet auch die beiden Distanzbestimmungen gleich gut ab.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Lotistische Regression und Lineare SVM: Die Machine-Leraning-Algorithmen können keine Sigmanoidfunktion oder Hyberebene entwerfen, welche eine sichere Unterscheidung ermöglicht. Das bedeutet, die Daten können nicht linear auseinandergehalten werden. Beide Algorithmen klassifizieren fast die Hälfte der Betriebszustände falsch. Mit dieser Trefferquote ist keine sichere Erkennung möglich und damit auch keine Implementierung im Umspannwerk 2 sinnvoll.

Insgesamt ergibt sich aber dennoch ein erster Eindruck inwieweit es möglich ist, vom Umspannwerk 2 aus weiter entfernte Fehler in vermaschten Netzen erkennen zu können. Ergeben sich nur kleine Änderungen wie der Ausfall eines Systems eines Parallelsystems in einer nicht direkt mit dem Umspannwerk verbundenen Masche, ist unter Verwendung der aufgelisteten Algorithmen keine Erkennung bzw. nur eine Erkennung mit sehr viel Unsicherheit möglich. Mit KNN können aber immerhin dreiviertel der Ausfälle erkannt werden.

Tabelle 4.3: Ergebnisse Machine-Learning-Algorithmen im UW2 bei Ausfall der Leitung zwischen SS1 und SS2

Algorithmus	**Input**	**T. P.**	**T. N.**	**F. P.**	**F. N.**	**Genauigkeit**
KNN-Euklid	I_{42}, φ_{42}	$74,8\%$	$74,6\%$	$25,4\%$	$25,2\%$	$74,6\%$
	$I_{42}, \varphi_{42}, \varphi_{45}$	$74,9\%$	$75,5\%$	$24,5\%$	$25,1\%$	$75,2\%$
	alle Messwerte	$59,4\%$	$59,4\%$	$40,6\%$	$40,6\%$	$59,4\%$
KNN-Mahanabolis	I_{42}, φ_{42}	$75,0\%$	$74,6\%$	$25,4\%$	$25,0\%$	$74,6\%$
	$I_{42}, \varphi_{42}, \varphi_{45}$	$75,9\%$	$74,7\%$	$25,4\%$	$24,1\%$	$75,3\%$
Logistische Regression	I_{42}, φ_{42}	$53,0\%$	$53,1\%$	$46,9\%$	$47,0\%$	$53,1\%$
	$I_{42}, \varphi_{42}, \varphi_{45}$	$54,3\%$	$54,5\%$	$45,5\%$	$45,7\%$	$54,4\%$
Linear SVM	I_{42}, φ_{42}	$56,3\%$	$51,5\%$	$48,5\%$	$43,7\%$	$52,5\%$
	$I_{42}, \varphi_{42}, \varphi_{45}$	$54,5\%$	$54,6\%$	$45,4\%$	$45,5\%$	$54,5\%$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.2.2 Ausfall von L23

Der Ausfall von L23 hat starke Auswirkungen auf die Messwerte beider Umspannwerke.

Umspannwerk 1

Die Genauigkeiten der Erkennung beim Heranziehen der Daten aus dem Umspannwerk 1 sind in Tabelle 4.4 ersichtlich. Auch bei diesem Ausfall sind gute Klassifizierungsergebnisse erreichbar. Die Grundlage für die Erkennung bilden wiederum die Winkel $\varphi_{51}, \varphi_{52}, \varphi_{54}$ sowie die Ströme I_{51}, I_{52} und I_{54}. Auch unter Verwendung aller Messwerte ist eine hohe Genauigkeit erreichbar. Dies lässt darauf schließen, dass fast alle Attribute zumindest geringfügig Information enthalten. Dies wurde auch schon in der Datenanalyse gezeigt. Durch den starken Eingriff in die Topologie ist das auch durchaus nachvollziehbar.

KNN: Wird ein KNN-Algorithmus verwendet, so ist die Einstellung der Distanzmetrik

Tabelle 4.4: Ergebnisse Machine-Learning-Algorithmen im UW1 bei Ausfall der Leitung zwischen SS2 und SS3

Algorithmus	Input	T. P.	T. N.	F. P.	F. N.	Genauigkeit
KNN-Euklid	$\varphi_{51}, \varphi_{52}, \varphi_{54}$	$97,9\%$	$98,4\%$	$1,6\%$	$2,1\%$	$98,2\%$
	$I_{51}, I_{52}, I_{54}, \varphi_{51}, \varphi_{52}, \varphi_{54}$	$94,4\%$	$97,3\%$	$2,7\%$	$5,6\%$	$95,9\%$
KNN-Mahanabolis	$\varphi_{51}, \varphi_{52}, \varphi_{54}$	$97,6\%$	98.1%	$1,9\%$	$2,4\%$	$97,9\%$
	$I_{51}, I_{52}, I_{54}, \varphi_{51}, \varphi_{52}, \varphi_{54}$	$99,0\%$	$99,3\%$	$1,0\%$	$0,7\%$	$99,2\%$
	alle Messwerte	$98,9\%$	$98,8\%$	$1,2\%$	$1,1\%$	$98,9\%$
Logistische Regression	$\varphi_{51}, \varphi_{52}, \varphi_{54}$	$93,7\%$	$94,4\%$	$5,6\%$	$6,3\%$	$94,1\%$
	$I_{51}, I_{52}, I_{54}, \varphi_{51}, \varphi_{52}, \varphi_{54}$	$97,5\%$	$98,4\%$	$1,6\%$	$2,5\%$	$98,0\%$
Linear SVM	$\varphi_{51}, \varphi_{52}, \varphi_{54}$	$93,4\%$	$95,2\%$	$4,8\%$	$6,6\%$	$94,3\%$
	$I_{51}, I_{52}, I_{54}, \varphi_{51}, \varphi_{52}, \varphi_{54}$	$97,1\%$	$99,0\%$	$1,0\%$	$2,9\%$	$98,1\%$

sowie des Gewichts maßgebend. Die Parameter wurden wie in Tabelle 4.5 eingestellt. Wie

Tabelle 4.5: Einstellungen der Parameter des KNN-Algorithmus

Distanzmetrik	Nachbarn	Distanzgewicht	Daten
Euklid	20	inverses Abstandsquadrat	Standardisiert
Mahalanobis	126	inverses Abstandsquadrat	Standardisiert

schon in den vorderen Untersuchungen werden unter Verwendung der euklidschen Distanz weniger Nachbarn benötigt, um die besten Klassifizierungsergebnisse zu erzielen. Durch Einsatz des Mahalanobis Abstandes in Verbindung mit der Squared-Inverse-Gewichtung sind hier Genauigkeiten bis zu $99,2\%$ erreichbar. Interessant ist vor allem, dass bei der Mahalanobis - Methode das Hinzunehmen von Parametern zu einer Verbesserung führt, während es bei der Euklidschen Methode zu einer Verschlechterung kommt. Daraus ergibt sich die Erkenntnis, dass je nach Distanzbestimmungsmethode die Ströme entweder zu einem Informationsgewinn führen oder die Informationen verrauschen, wodurch die Information schwieriger zu extrahieren ist. Es ist je nach verwendetem Algorithmus also sinnvoll oder auch nicht sinnvoll sie zu verwenden. Wie sich zeigt ist, unter der Verwendung nur der Winkel, ein einfacher Algorithmus wie KNN nicht nur ausreichend, sondern

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

liefert im Vergleich zur Support-Vector-Machine und der Logistischen Regression sogar bessere Klassifizierungsergebnisse. Insgesamt lässt sich der Ausfall mittels KNN also mit sehr hoher Wahrscheinlichkeit feststellen.

Logistische Regression: Die Logistische Regression benötigt alle sechs Attribute, um gute Ergebnisse zu erzielen. Der Algorithmus ist in der Lage eine Sigmanoid-Funktion zu entwerfen, welche die beiden Betriebsfälle zu 98% unterscheiden kann. Die zwei Zustände können also mit einer Linearkombination der Attribute auseinandergehalten werden.

LSVM: Es wurde der Versuch unternommen, den Linearen SVM - Algorithmus zu optimieren, um eine höhere Genauigkeit zu erreichen. Trotz all dem war es nicht möglich, gleich hohe Genauigkeiten wie beim KNN zu erzielen. Insgesamt wird dennoch eine Genauigkeit von $98,1\%$ erreicht.

Umspannwerk 2

Die Klassifizierungsgenauigkeiten sind in Tabelle 4.6 dargestellt.

KNN: Die Parameter der KNN wurden dabei wie in der folgenden Tabelle ersichtlich einge-

Tabelle 4.6: Ergebnisse Machine-Learning-Algorithmen im UW2 bei Ausfall der Leitung zwischen SS2 und SS3

Algorithmus	**Input**	**T. P.**	**T. N.**	**F. P.**	**F. N.**	**Genauigkeit**
KNN-Euklid	$\varphi_{42}, \varphi_{43}, \varphi_{45}$	$93,1\%$	$87,9\%$	$12,1\%$	$6,9\%$	$90,5\%$
	$I_{42}, I_{43}, V_4, \varphi_{42}, \varphi_{43}, \varphi_{45}$	$96,5\%$	$99,3\%$	$0,7\%$	$3,5\%$	$97,9\%$
	alle Messwerte	$94,1\%$	$97,6\%$	$2,4\%$	$5,9\%$	$95,9\%$
KNN-Mahanabolis	$\varphi_{42}, \varphi_{43}, \varphi_{45}$	$93,4\%$	$87,4\%$	$12,6\%$	$6,6\%$	$90,4\%$
	$I_{42}, I_{43}, V_4, \varphi_{42}, \varphi_{43}, \varphi_{45}$	$96,7\%$	$98,3\%$	$1,7\%$	$3,3\%$	$97,5\%$
Logistische Regression	$\varphi_{42}, \varphi_{43}, \varphi_{45}$	$56,9\%$	$55,0\%$	$45,0\%$	$43,1\%$	$55,9\%$
	$I_{42}, I_{43}, I_{45}, V_4, \varphi_{42}, \varphi_{43}, \varphi_{45}$	$88,8\%$	$94,4\%$	$5,6\%$	$11,2\%$	$91,6\%$
Linear SVM	$\varphi_{42}, \varphi_{43}, \varphi_{45}$	$66,5\%$	$39,9\%$	$60,1\%$	$33,5\%$	$53,2\%$
	$I_{42}, I_{43}, I_{45}, V_4, \varphi_{42}, \varphi_{43}, \varphi_{45}$	$88,2\%$	$94,7\%$	$5,3\%$	$11,8\%$	$91,5\%$
Booosted Tree	$I_{42}, I_{43}, I_{45}, V_4, \varphi_{42}, \varphi_{43}, \varphi_{45}$	$97,0\%$	$99,0\%$	$1,0\%$	$3,0\%$	$98,0\%$

stellt. Wiederum erreichen die KNN-Algorithmen gute Klassifizierungsgenauigkeiten. Mit

Tabelle 4.7: Einstellungen der Parameter des KNN-Algorithmus

Distanzmetrik	**Nachbarn**	**Distanzgewicht**	**Daten**
Euklid	25	inverses Abstandsquadrat	Standardisiert
Mahalanobis	25	inverses Abstandsquadrat	Standardisiert

den Impedanzwinkeln allein werden immerhin Genauigkeiten bis 90% erreicht. Unter der Berücksichtigung der Ströme I_{42} und I_{43} sowie auch der Spannung an der Sammelschiene 4 steigt die Genauigkeit der Erkennung auf $97,9\%$. Werden nur die Impedanzwinkel betrachtet, werden hier mehr Normalzustände als Fehlerzustände klassifiziert als umgekehrt. Mit der Hinzunahme der Ströme und der Spannung dreht sich diese Aussage aber um,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

d.h. es werden mehr Fehlerzustände als Normalzustände klassifiziert als umgekehrt. Die Unterscheidung erweist sich vor allem bei kleinen Belastungen schwierig, da die Gesamtheit der Messdaten des gestörten Betriebszustandes sehr ähnlich zu Betriebszuständen im Normalzustand sind.

Boosted Tree: Dieser Algorithmus erzielt das beste Ergebnis. Er benötigt jedoch zusätzlich die Messgröße I_{45}, um diese Genauigkeit zu erreichen. Diese Messgröße wurde bei den KNN vernachlässigt, da sie zu einer Verschlechterung führt. Beim Boosted-Tree hingegen führt sie zu einer Verbesserung.

LSVM und Logistische Regression: Sie schneiden bei diesem Ausfall nicht gut ab. Unter Verwendung von nur drei Messgrößen ist mit diesen Algorithmen gar keine Erkennung möglich. Zieht man auch die Ströme und Spannung heran, erreichen sie knapp über 91%. In Bezug auf SVM muss eine relativ hohe Soft-Margin gewählt werden, da sich sonst Probleme beim Training ergeben. Wird sie zu klein gewählt schafft es der Algorithmus nicht die Daten zu trennen.

4.2.3 Ausfall von L34

Dieser Ausfall wird nur für das UW1 untersucht.

Umspannwerk 1

Wie in Tabelle 4.8 ersichtlich, werden hier keine hohen Genauigkeiten erzielt. Der beste KNN - Algorithmus erreicht eine Klassifizierungsgenauigkeit von $87,5\%$.

KNN: Es werden vermehrt fehlerhafte Betriebszustände als Normalzustände klassifiziert.

Tabelle 4.8: Ergebnisse Machine-Learning-Algorithmen im UW1 bei Ausfall der Leitung zwischen SS3 und SS4

Algorithmus	**Input**	**T. P.**	**T. N.**	**F. P.**	**F. N.**	**Genauigkeit**
KNN-Euklid	$I_{51}, I_{52}, I_{54}, V_5, \varphi_{51}, \varphi_{52}$	$78,4\%$	$90,7\%$	$21,6\%$	$9,3\%$	$84,6\%$
	alle Messwerte	$77,1\%$	$89,0\%$	$22,8,9\%$	$11,0\%$	$83,1\%$
KNN-Mahanabolis	$I_{51}, I_{52}, I_{54}, V_5, \varphi_{52}, \varphi_{43}$	$77,3\%$	$97,1\%$	$22,7\%$	$2,9\%$	$87,5\%$
Logistische Regression	$I_{51}, I_{52}, I_{54}, V_5, \varphi_{51}, \varphi_{52}$	$60,5\%$	$64,2\%$	$39,5\%$	$35,8\%$	$62,4\%$
Linear SVM	$I_{51}, I_{52}, I_{54}, V_5, \varphi_{51}, \varphi_{52}$	$53,5\%$	$76,0\%$	$46,5\%$	$24,0\%$	$64,7\%$
Ensemble	$I_{51}, I_{52}, I_{54}, V_5, \varphi_{51}, \varphi_{52}$	$81,7\%$	$88,3\%$	$18,3\%$	$11,7\%$	$85,0\%$

Durch die Verwendung der Mahanabolis Distanz wird im Vergleich zur Euklidschen Distanz eine Verbesserung um 2,9 Prozentpunkte erreicht. Diese Erhöhung liegt an der starken Steigerung der True-Negativ-Rate. Die True-Positiv-Rate nimmt hier sogar um 1,1 Prozentpunkte ab. Die Parameter wurden dabei, wie in Tabelle 4.9 ersichtlich, eingestellt.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Tabelle 4.9: Einstellungen der Parameter des KNN-Algorithmus

Distanzmetrik	**Nachbarn**	**Distanzgewicht**	**Daten**
Euklid	30	inverses Abstandsquadrat	Standardisiert
Mahalanobis	120	inverses Abstandsquadrat	Standardisiert

Logistische Regression: Logistische Regression produziert bei diesem Ausfall keine guten Ergebnisse. Er schneidet im Vergleich zu den anderen Algorithmen am schwächsten ab. Es lässt sich also anhand der vorhanden Daten keine Funktion entwerfen, die ein erfolgreiches Zuordnen zu den jeweiligen Klassen ermöglicht.

LSVM:Die Lineare Support-Vector-Machine ordnet fast die Hälfte der Ausfälle dem Normalbetrieb zu. Sie stellt auch keine effiziente Methode zur Erkennung eines solchen Fehlers dar.

Boosted-Tree: Ein Ensemble von Entscheidungsbäumen erzielt 85% Prozent. Vor allem die True-Positiv-Rate ist im Vergleich zu den anderen Methoden am höchsten.

Es wurden noch andere Algorithmen wie Diskriminierung versucht, jedoch ohne brauchbares Ergebnis. Insgesamt lässt sich aber sagen, dass das UW1 mit einer Genauigkeit von $87,5\%$ einen solchen Ausfall in einer entfernten Masche erkennen kann.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.2.4 Gemeinsame Betrachtung der Ausfälle

Wie gezeigt wurde, sind also Machine-Learning-Algorithmen durchaus in der Lage, die Ausfälle getrennt voneinander zu erkennen. Die Datenanalyse hat darauf hingedeutet, dass sich die Daten der einzelnen Ausfälle nicht sonderlich überlappen. Dies sollte sich auch an den Genauigkeiten der Erkennung der gemeinsamen Betrachtung widerspiegeln. Als Testset dient ein Set aus insgesamt 5525 Betriebszuständen. Dieses setzt sich aus jeweils 1675 Betriebszuständen des Normalbetrieb (N), Ausfall von L12-1 (F12) und 23 (F23) und 600 des Ausfalls von L34 (F34) zusammen.

Umspannwerk 1

In der Datenanalyse wurde gezeigt, dass der Ausfall 34 nur mit sehr viel Unsicherheit vom Normalzustand zu unterscheiden ist. Durch die Hinzunahme der Daten der anderen Ausfälle ist hier keine Verbesserung zu erwarten. Interessant ist aber vor allem, ob es durch Hinzunahme der Daten dieses Ausfalls zu einer Verschlechterung der Klassifizierungsergebnisse der anderen Ausfällen kommt.

KNN:

Als Distanzmetrik wurde der Mahanabolis Abstand gewählt. Die Parameter wurden dabei laut Tabelle 4.10 eingestellt. Die Genauigkeiten der Erkennung mittels KNN sind in Abbildung 4.11 ersichtlich. Als Attribute für den Algorithmus dienen die Ströme I_{51}, I_{52}, I_{54}, die Spannung an der SS5 V_5 sowie die Impedanzwinkel $\varphi_{51}, \varphi_{52}$ und φ_{54}. An den Ergebnissen bestätigt sich die vorhin getätigte Aussage, wonach kein Ausfall von L23 als Ausfall von L12-1 klassifiziert wird und umgekehrt. Auch die Einbindung des Ausfalls von L34 hat nur relativ geringe Auswirkungen auf die Erkennung der übrigen Ausfälle. 3% werden als Ausfall von L23 und 1% wird als Ausfall von L12-1 interpretiert. Wie auch im vorigen Abschnitt werden mit $26,2\%$ viele Ausfälle von L23 mit dem Normalzustand verwechselt. Zu beobachten ist insgesamt ein kleiner Rückgang der True-Positiven. Der Grund hierfür liegt an dem Algorithmus, der implementiert wurde. Während im vorigen Abschnitt die Algorithmen auf den jeweiligen Fehler angepasst wurden, wird hier ein Algorithmus verwendet, welcher im Bezug auf beide Fehler die besten Ergebnisse liefert. Gut zu beobachten ist dieses Phänomen an den Erkennungsgenauigkeiten des Ausfalls von L34. Durch die Implementation eines einfachen KNN-Algorithmus ist also im Umspannwerk 1 eine Erkennung der Fehler 23 und 12 mit durchaus hohen Genauigkeiten möglich. Ebenfalls können im UW1 fast 70% der Ausfälle von L34 richtig erkannt werden.

Tabelle 4.10: Einstellungen der Parameter des KNN-Algorithmus

Distanzmetrik	**Nachbarn**	**Distanzgewicht**	**Daten**
Mahalanobis	120	inverses Abstandsquadrat	Standardisiert

Werden nur die Impedanzwinkel $\varphi_{51}, \varphi_{52}, \varphi_{54}$ verwendet, sinkt die Erkennungsgenauigkeit. Die Klassifizierungsergebnisse für diesen Fall sind in Abbildung 4.12 dargestellt. Unter diesen Umständen ist es aber immer noch ein relativ hoher Wert. Die T.P. für die Fehler 12 und 23 betragen hier immer noch $95,4\%$ und 97%. Auch unter Verwendung nur der Win-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

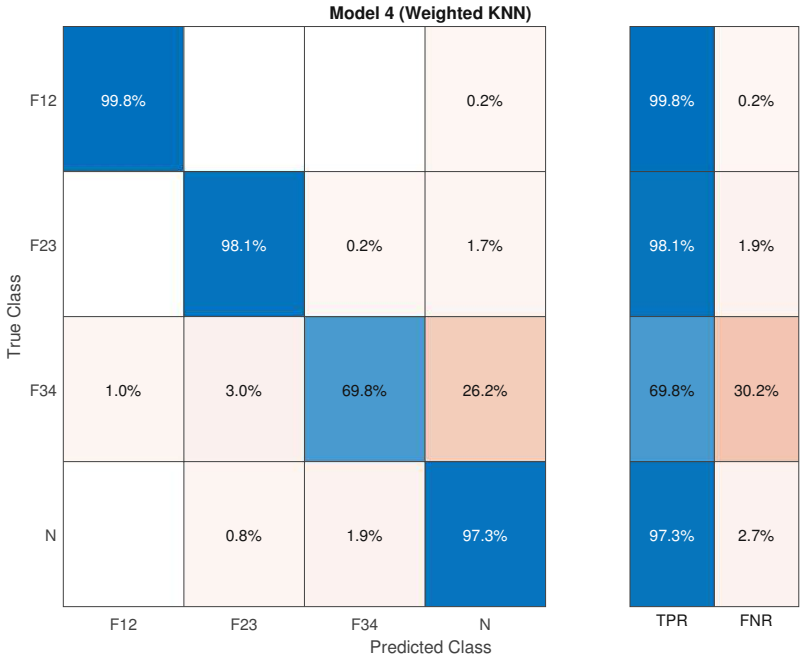


Abbildung 4.11: Klassifizierungsergebnisse mittels KNN, Anzahl Attribute: 7

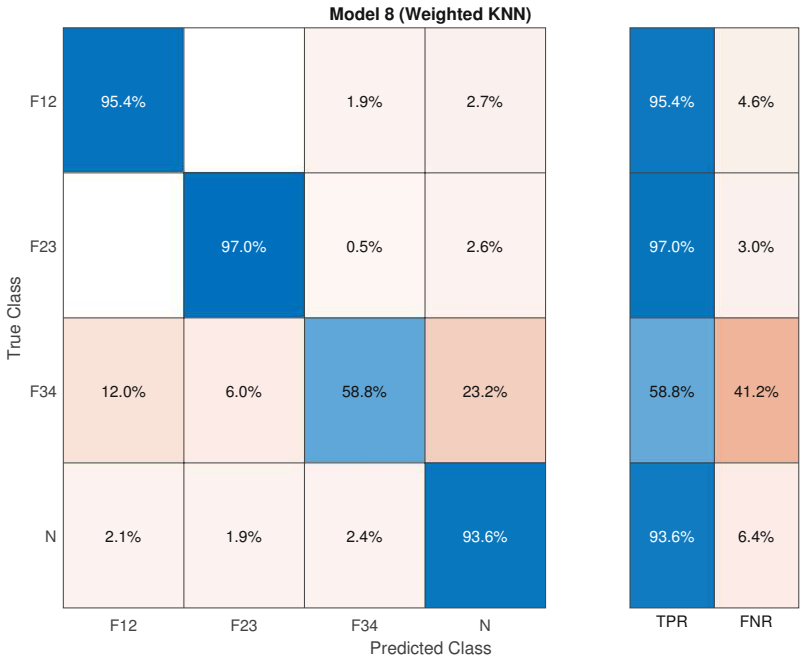


Abbildung 4.12: Klassifizierungsergebnisse mittels KNN, Anzahl Attribute: 3

kel wird kein Ausfall von L23 mit einem Ausfall von L12-1 verwechselt und vice versa. Die Reduktion der Parameter bringt einen Rückgang der Erkennungsraten von Fehler 12 und vom Normalbetrieb. 97% der Ausfälle von L23 können immer noch richtig klassifiziert werden. Für den Ausfall 34 bedeutet die Verwendung einzig dieser Attribute, dass

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

nur mehr 58,8% der Ausfälle richtig erkannt werden. Der Rest wird anderen Ausfällen oder dem Normalzustand zugeschrieben. Für ihn ist also eine solche Implementation nicht sinnvoll.

Zusammengefasst ist es also ohne hohen Aufwand zu betreiben möglich, nicht nur das Auftreten von Ausfällen festzustellen, sondern ausgefallene Leitung auch zu klassifizieren. Die Sicht ist dabei allerdings begrenzt. Nur Fehler die entsprechende Auswirkungen auf die Messwerte haben, können mit Sicherheit erkannt werden. Ausfälle wie der Ausfall von L34 in Nachbarschleifen können nur noch unscharf gesehen werden.

Linear-Support-Vector-Machine:

Abbildung 4.13 zeigt die Ergebnisse unter Verwendung der LSVM. Als Attribute für den Algorithmus dienen wiederum die Ströme I_{51}, I_{52}, I_{54}, die Spannung V_5 sowie die Impedanzwinkel $\varphi_{51}, \varphi_{52}$ und φ_{54}. Ausfälle von L12-1 und Ausfälle von L23 sind auch für diesen Algorithmus eindeutig unterscheidbar. Der Ausfall von L34 wird im Vergleich zum KNN allerdings häufiger als Ausfall von L12-1 und L34 klassifiziert. Fällt aber eine der Leitungen aus, so wird der Ausfall mit hoher Wahrscheinlichkeit erkannt. Der Ausfall 34 kann so gut wie gar nicht erkannt werden. Es würde hier Sinn machen, keine Erkennung des Fehlers 34 zu implementieren und ihn den Daten des Normalzustandes zuzuordnen, da er ohnehin nicht davon unterscheidbar ist. Problematisch ist allerdings, dass der Fehler relativ häufig als Ausfall von L23 interpretiert wird. Dies führt zu einem erhöhten Fehlalarm für diesen Ausfall. Im Vergleich zum KNN erzielt die Support-Vector-Machine also

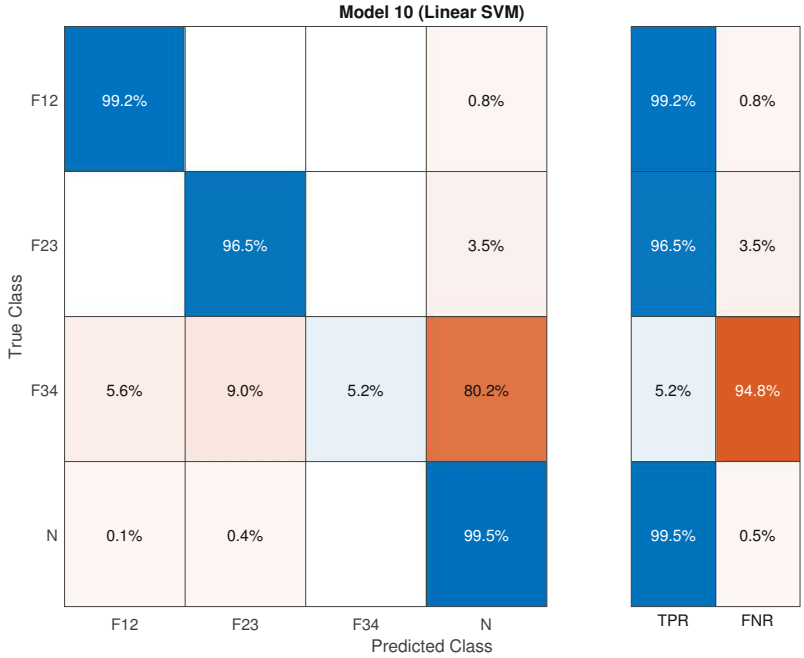


Abbildung 4.13: Erkennung mittels LSVM

etwas schlechtere aber immer noch brauchbare Ergebnisse. Ausfälle von L12-1 werden mit einer Wahrscheinlichkeit von über 99% erkannt. Auch der Ausfall von L23 ist mit einer

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Wahrscheinlichkeit von 96,5% feststellbar.
Unter Verwendung nur der Winkel als Attribute befindet sich der LSVM allerdings auch bei diesen Ausfällen nicht auf Augenhöhe mit dem KNN-Algorithmus.

Umspannwerk 2

Wie bei der einzelnen Erkennung der Ausfälle gezeigt wurde, ist die Erkennung des Ausfalls von L12-1 im UW2 nur mit sehr geringer Wahrscheinlichkeit möglich. Der Ausfall von L23 kann mit hoher Wahrscheinlichkeit erkannt werden.

KNN:

Die Ergebnisse der Erkennung mittels K-Nearest-Neighbour sind in den Abbildungen Abb. 4.15 und Abb. 4.14 ersichtlich. Als Distanzmetrik wurde wiederum Mahalanobis verwendet. Die Parameter des KNN wurden dabei wie folgt eingestellt: Die Anzahl der

Tabelle 4.11: Einstellungen der Parameter des KNN-Algorithmus

Distanzmetrik	**Nachbarn**	**Distanzgewicht**	**Daten**
Mahalanobis	130	inverses Abstandsquadrat	Standardisiert

nächsten Nachbarn wurde um einiges größer gewählt, da mit dieser erhöhten Anzahl bessere Ergebnisse erzielt werden. Die beiden Algorithmen, mit denen diese Ergebnisse erzielt wurden, unterscheiden sich nur in ihrer Kostenmatrix. Die Kostenmatrix sieht unmodifiziert folgendermaßen aus: In dieser Form sind alle Falsch-Positiven gleichwertig gewichtet.

Tabelle 4.12: Kostenmatrix KNN

Tr.C / Pr.C	F12	F23	N
F12	0	1	1
F23	1	0	1
N	1	1	0

Dabei erreicht die True-Positive-Rate des Ausfalls von L23 95,2%, d.h. 95,2% der Ausfälle können richtig klassifiziert werden (Abb. 4.15). Die restlichen 3,6% und 1,1% werden vom Algorithmus mit dem Normalzustand bzw. Ausfall von L12 verwechselt. Umgekehrt ist auffällig, dass nur 0,5% der Ausfälle von L12-1 und 0,8% der normalen Betriebszustände als Ausfall von L23 gedeutet werden. Dieses Ergebnis bedeutet eine kleine Fehlalarmrate. 71% der Ausfälle eines Systems der Doppelsysteme in einer entfernten Masche, können ebenfalls richtig interpretiert werden. Prinzipiell ist das für diesen Ausfall ein gutes Ergebnis. Problematisch ist aber, dass 27,1% der Normalzustände als Ausfall von L12-1 verstanden werden. Da sich das Netz grundsätzlich die meiste Zeit im Normalbetrieb befindet, bedeutet das eine Vielzahl von Fehlalarmen. Klassifiziert der Algorithmus also einen Zustand als Ausfall von L12-1, könnte man dem Ergebnis nicht „trauen“. Wird nun die Kostenmatrix wie folgt modifiziert:

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Tabelle 4.13: Kostenmatrix KNN

Tr.C / Pr.C	F12	F23	N
F12	0	1	1
F23	2	0	2
N	1	1	0

werden die in Abbildung 4.14 gezeigten Ergebnisse erreicht. Der Algorithmus wird also besonders „bestraft“, wenn er einen Betriebszustand fälschlicherweise als Ausfall von L23 klassifiziert. Mit dieser angepassten Kostenmatrix werden $98,1\%$ der Ausfälle von L23 richtig klassifiziert. Im Vergleich zu vorher bedeutet dies eine Steigerung um 2,9 Prozentpunkte. Dass Normalzustände als Fehler der Leitung 12 und umgekehrt erkannt werden, ist hier erstmal uninteressant. Wie sich zeigt, werden $2,3\%$ der Ausfälle von L12-1 fälschlich als Ausfall von L23 interpretiert. Für den Ausfall von L12-1 bleiben sind die Aussagen identisch zum anderen Fall.

Insgesamt muss entschieden werden, ob man einen Ausfall von L23 sicher erkennen will, dafür jedoch auch eine erhöhte Wahrscheinlichkeit eines Fehlalarms in Kauf nimmt oder ob Ausfälle von L23 mit weniger Wahrscheinlichkeit feststellbar sind, dafür aber eine kleine Anzahl an Fehlalarmen auftritt. Eine Implementation der Erkennung des Ausfall von L12-1 macht relativ wenig Sinn.

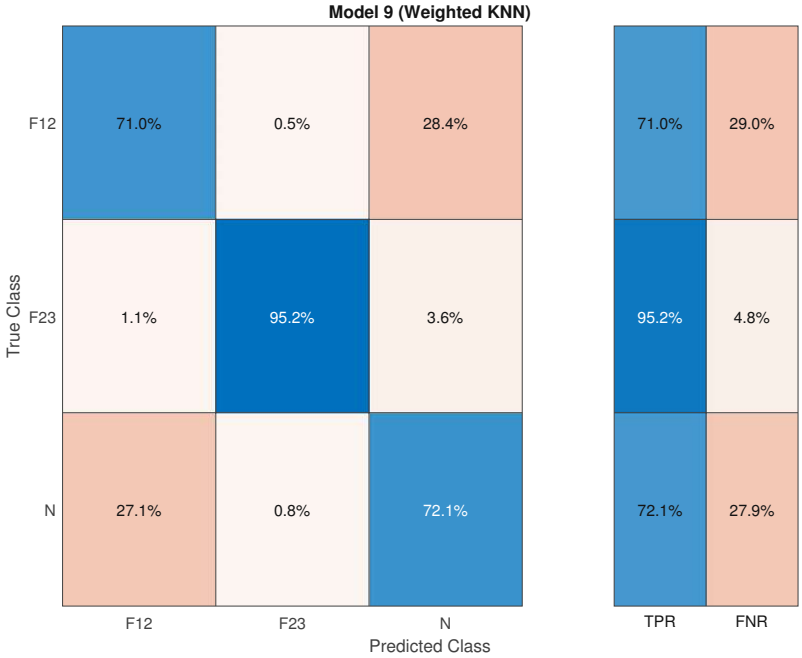


Abbildung 4.14: Genauigkeiten der Erkennung mittels KNN mit Kostenmatrix aus Tab.4.2; Anzahl der Attribute: 5

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

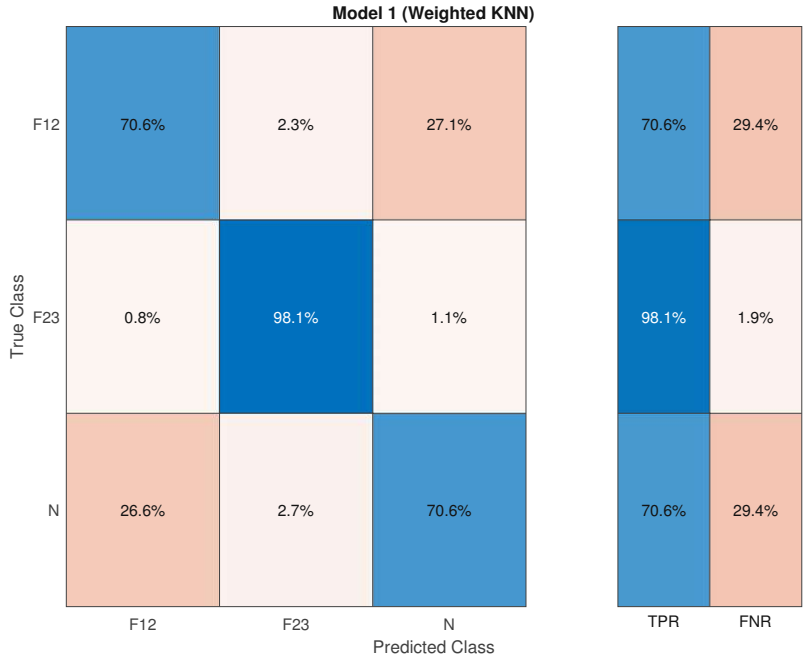


Abbildung 4.15: Genauigkeiten der Erkennung mittels KNN mit Kostenmatrix aus Tab.4.3; Anzahl der Attribute: 5

Linear-Support-Vectormachine:

Die LSVM weist bei dieser Aufgabenstellung Probleme auf. Die Klassifizierungsergebnisse sind in 4.16 dargestellt. Wie erwartet sinken die Erkennungsgenauigkeiten im Vergleich zu den Klassifizierungsergebnissen der einzeln trainierten Algorithmen im vorigen Abschnitt. Konkret bedeutet das für den Ausfall von L23 , dass zusätzlich zu den falsch als Normal klassifizierten Zuständen noch $3,6\%$ als Ausfall von L12-1 erkannt werden. $85,9\%$ der Ausfälle von L12-1 können richtig klassifiziert werden. Im Vergleich zum KNN-Algorithmus schneidet die LSVM also deutlich schlechter ab, sowohl in Hinsicht auf den Ausfall von L23 sowie auch auf den Ausfall von L12-1. Dabei wurden dieselben Attribute wie beim KNN verwendet. Der Algorithmus hat anscheinend Probleme eine Hyperebene zu finden, die die drei Klassen voneinander trennt.
Die Implementation eines solchen LSVM-Algorithmus im UW2 ist nur mäßig sinnvoll. Hier ist eindeutig der KNN-Algorithmus vorzuziehen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

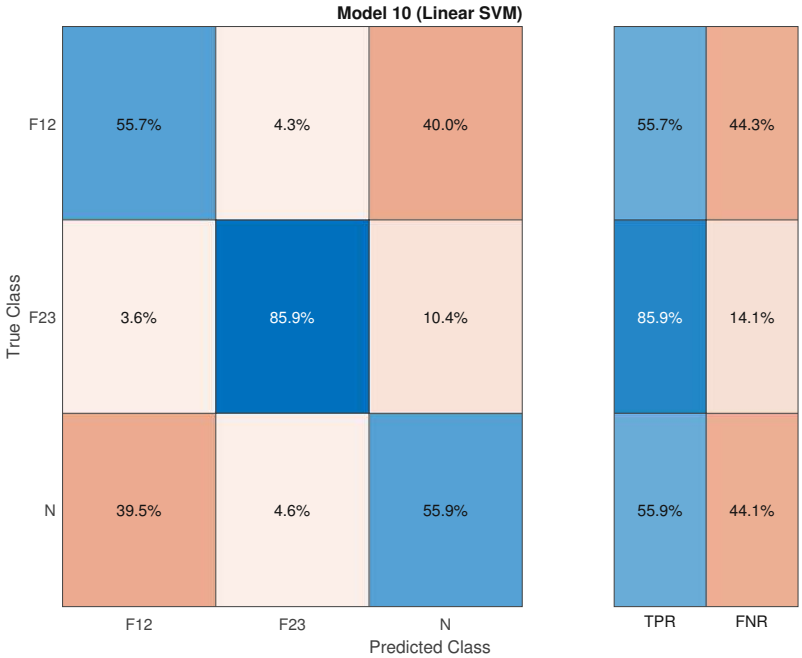


Abbildung 4.16: Erkennung mittels LSVM

Boosted Trees:

Da dieser Algorithmus in der Voruntersuchung vielversprechende Ergebnisse geliefert hat, wird er auch hier miteinbezogen. Die Ergebnisse sind in Abbildung 4.17 und 4.18. Um dieses Ergebnis in Abbildung 4.17 zu erreichen, wurden insgesamt sieben Attribute benötigt. Diese sind die Ströme I_{45}, I_{43}, I_{42}, die Spannung V_4 und die Impedanzwinkel $\varphi_{45}, \varphi_{43}$ und φ_{42}. Mittels der Boosted-Tree Methode können $96,5\%$ der Ausfälle von L23 richtig klassifiziert werden. Diese ist im Vergleich zum KNN-Algorithmus um $1,2$ Prozentpunkte höher. Die Hinzunahme des Ausfalls von L12-1 hat nur geringe Auswirkungen auf die Erkennung vom Ausfall von L23 im UW2. Nur $0,7\%$ der normalen Betriebszustände sowie $0,4\%$ werden fälschlicherweise als Ausfall von L12-1 interpretiert. Der Ausfall von L12-1 wird auch mit diesem Algorithmus oft mit dem Normalzustand verwechselt. Insgesamt wird der Ausfall von L12-1 um $2,6$ Prozentpunkte weniger oft richtig klassifiziert als beim KNN.

Werden die Attribute um zwei verringert, ergibt sich das Ergebnis in Abbildung 4.18. Mit den Attributen I_{42}, I_{43}, V_4 sowie deren Impedanzwinkel φ_{42} und φ_{43} werden also die gleichen fünf wie beim KNN-Algorithmus verwendet. Die Attribute φ_{45} und I_{45} wurden beim KNN nicht herangezogen, da sie für ihn keine Information enthalten. Wie sich in den Ergebnissen zeigt, hat das Vernachlässigen dieser Attribute auf die Erkennung des Ausfalls von L23 so gut wie keinen Einfluss. Allerdings werden die zwei anderen Betriebszustände vermehrt falsch klassifiziert.
Insgesamt kann also der Entscheidungsbaum-Algorithmus für die Ausfalldetektion im UW2 durchaus mit dem KNN-Algorithmus konkurrieren.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

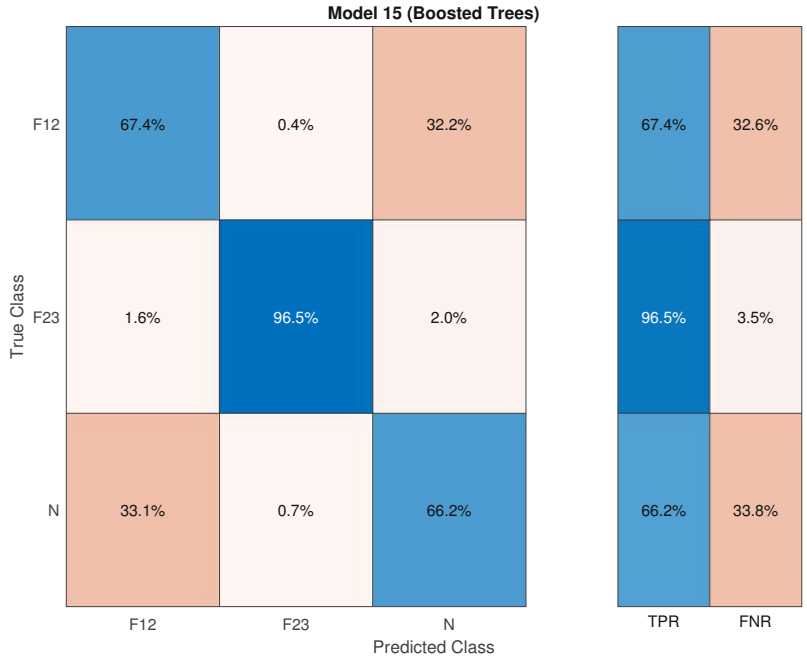


Abbildung 4.17: Genauigkeiten der Erkennung mittels Boosted Trees: Ensemble methode: Bag, Learner Type: Decision Trees, Number of Learners: 15, Anzahl Attribute: 7

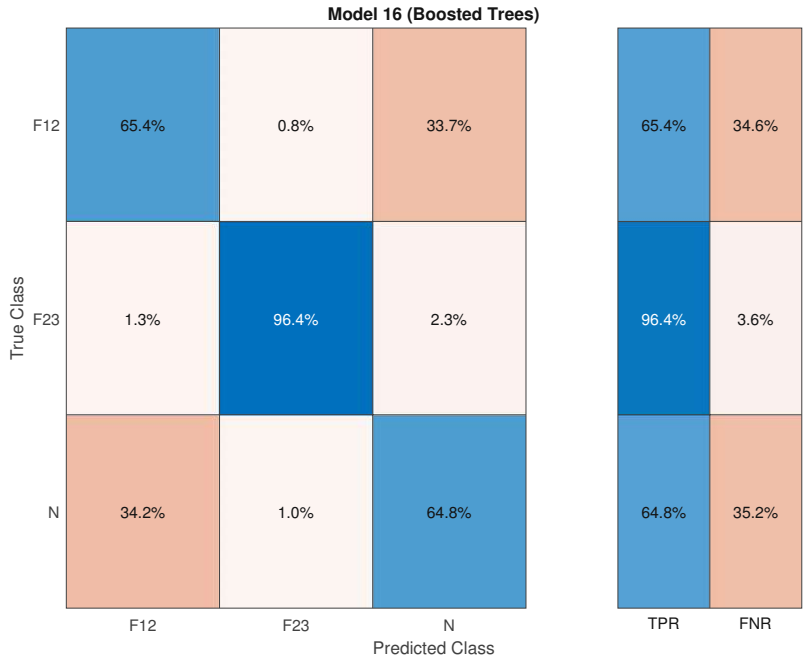


Abbildung 4.18: Genauigkeiten der Erkennung mittels Boosted Trees: Ensemble methode: Bag, Learner Type: Decision Trees, Number of Learners: 15, Anzahl Attribute: 5

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5 Zusammenfassung und Ausblick

Zusammenfassung

Die Erkennung von gestörten Netzbetrieben in kleinen vermaschten Netzen ist grundsätzlich möglich. Die Reichweite der Erkennung ist allerdings begrenzt. Eine wesentliche Rolle spielt dabei die Lage des ausgefallenen Systems im vermaschten Netz zum UW sowie seine Lage selbst im Netz.

Ausfälle von Leitungen in der direkten Masche können nur unter Verwendung der Messgrößen im eingeschwungenen Zustand erkannt werden. Neben dem Ausfall eines ganzen Systems zwischen zwei Sammelschienen ist es auch möglich, den Ausfall nur eines Systems einer doppelsystemigen Verbindung zu erkennen. Werden nur die Impedanzwinkel der Abgänge beobachtet, so können mit Hilfe eines KNN-Algorithmus bereits Genauigkeiten von knapp über 90% erreicht werden. Unter der Hinzunahme von ausgewählten Strömen und der Spannung an der SS werden für beide Ausfälle Genauigkeiten über 98% erreicht.

Die Erkennung der Ausfälle, in den nicht direkten Maschen des UWs, erweist sich als etwas schwieriger. Die Anordnung von den Verbrauchern und der Generatoren hat hier einen starken Einfluss auf die Ergebnisse. Der Ausfall von z.B. direkten Verbindungsleitungen zwischen großen Verbrauchern und Generatoren zieht starke Änderungen auf die Messgrößen in den UWs mit sich und kann daher auch in einiger Entfernung noch mit einer Wahrscheinlichkeit von bis zu 99% erkannt werden.
Ausfälle von Systemen in entfernten Maschen, die nicht dementsprechende Änderungen in den Messgrößen aufweisen, sind schwerer zu erkennen. Immerhin können aber auch hier noch 85% erreicht werden. Zu dieser Gattung von Ausfällen zählt auch der Ausfall eines Systems einer doppelsystemigen Verbindung in einer Nachbarmasche. Maximal 75% der Ausfälle können in diesem Fall richtig klassifiziert werden. Eine Steigerung, zumindest wie es hier implementiert wurde, ist nicht möglich, da die Auswirkungen solcher Ausfälle in entfernten UWs sehr schwer sichtbar sind. Zudem werden Normalzustände zunehmend als Leitungsausfälle klassifiziert.

Das Trainieren der Algorithmen auf die Erkennung aller Ausfälle verschlechtert die Erkennung der einzelnen Aufälle nur sehr gering. Die Ausfälle beeinflussen sich in der Regel nicht sehr stark, da sie sich um insgesamt zwei Schaltzustände voneinander unterscheiden und daher die Auswirkungen auf die Messgrößen in den UWs deutlich unterscheidbar sind. Nur Ausfälle, bei denen es bei den getrennten Untersuchungen schon zu Problemen kam, verschlechtern sich hier etwas deutlicher. Zu Verschlechterung kommt es unter anderem auch, weil die Algorithmen nicht auf die einzelnen Ausfälle optimiert sind, d.h. ihre Parameter nicht auf den jeweiligen Ausfall angepasst sind.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Wie sich gezeigt hat, erzielt vor allem der KNN-Algorithmus gute Ergebnisse. Der Algorithmus ist einfach zu implementieren und die Ergebnisse können relativ leicht nachvollzogen werden. In Hinsicht auf die Erkennung im gestörten Netzbetrieb, wie sie hier untersucht wurde, stellt er ein unverzichtbares Werkzeug zur Meisterung dieser Aufgabe dar. Auch andere Algorithmen wie die LSVM oder Entscheidungsbäume erreichen gute Ergebnisse, jedoch nicht in so einem breiten Spektrum wie der KNN-Algorithmus. Wichtig erweist sich in Hinsicht auf die Implementierung der Machine-Learning-Algorithmen vor allem die richtige Auswahl der Attribute. Durch das Verrauschen der Information durch die Verwendung nicht aussagekräftiger Messgrößen sinken die Genauigkeiten der Algorithmen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Ausblick

Diese Arbeit liefert grundsätzlich den Beweis, dass es für das untersuchte Netz möglich ist gestörte Netzbetriebe, wie die Ausfälle von Leitungen in vermaschten Netzen, nur anhand lokal zur Verfügung stehender Messwerte, erhoben an einem Zeitpunkt ohne mehrere Zeitpunkte in Betracht zu ziehen, zu bestimmen. Die Problemstellung wurde mit möglichst einfachen Mitteln angegangen.

Um noch höhere Genauigkeiten, vor allem bei Ausfällen die hier nur mit sehr viel Unsicherheit erkannt werden können zu erzielen, können in einem nächsten Schritt z.B. der Zustand vor und nach dem Ausfall in Verbindung gesetzt werden. Dies ist in Abbildung 5.1 grafisch anhand aufgenommener Daten dargestellt. Durch diese Verbindung wird mehr Information geschaffen, die einem Algorithmus bei der Klassifizierung durchaus nützlich sein könnten. Zudem kann ein Schritt in Richtung komplexerer Machine-Learning-Algorithmen

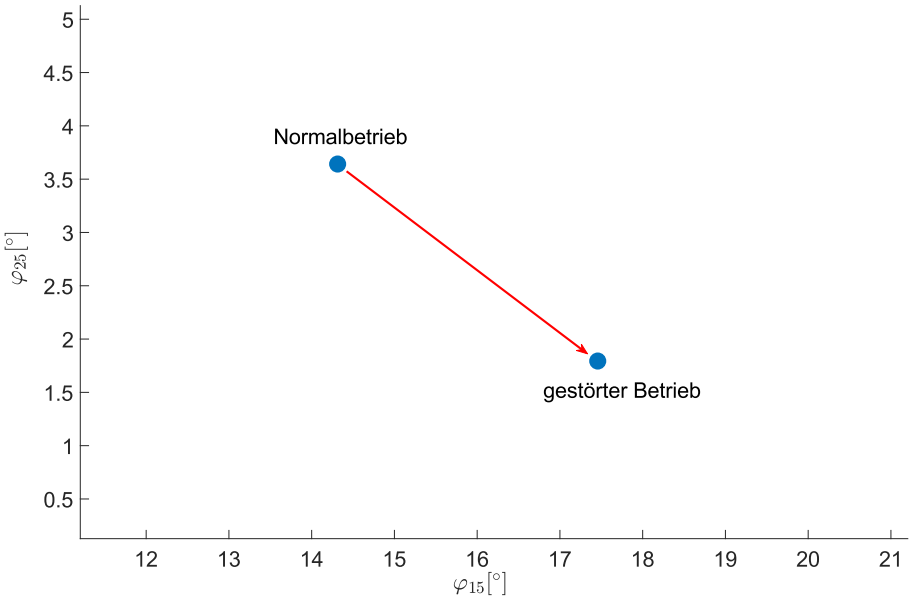


Abbildung 5.1: Zustände vor und nach dem Ausfall miteinander in Verbindung bringen

gesetzt werden. Die Verwendung der immer populärer werdenden Neuronale Netze sei hier als Beispiel genannt. Auch andere Lernverfahren können für weitere Untersuchungen in Betracht gezogen werden. In Fällen, in denen nur die Daten vom Normalzustand vorhanden sind, können die Methoden der Anomaliedetektion in Betracht gezogen werden. Eine Zuordnung von Ausreißern auf bestimmte Leitungen wird sich im Vergleich zur implementierten Methode allerdings als schwierig erweisen.

Weiters sind Untersuchungen für größere Systeme interessant, wie z.B. dem IEEE-30-Bus-System, mit dem Ziel herauszufinden, wie weit die Sicht eines Umspannwerks in einem größeren System reicht.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Abbildungsverzeichnis

2.1 Pyramidenstruktur eines Netzleitsystems einer Spannungsebene 3
2.2 IEC61850 Objekt Modell 5
2.3 IEC61850 Architektur 6
2.4 Übersicht Machine-Learning-Typen 7
2.5 Idee der KNN - Methode 8
2.6 Euklidsche Distanz 9
2.7 Korrelierte(a) und unkorrelierte(b) Attribute 10
2.8 Vergleich der Abstandsbestimmungen 11
2.9 Idee der Support-Vektor-Machine 11
2.10 Support-Vektor-Machine 12
2.11 Entscheidungsbaum 14
2.12 CART - Algorithmus: hoher Gini Index in der linken Abbildung, perfekter Split in der rechten Abbildung mit Gini Index gleich Null 15
2.13 Sigmanoid Funktion 16
2.14 Verstärktes Lernen Grundidee 18

3.1 Leittechnik im UW 19
3.2 Messungen werden im eingeschwungenen Netzzustand aufgenommen 20
3.3 Beispiel zur Erläuterung des Begriffes der Sicht in einem elektrischen System 22
3.4 Wahrscheinlichkeitsdichtefunktionen für die Bestimmung der Lasten 24
3.5 Boxplot der Ströme und der Spannung zur Analyse der Daten 26
3.6 2 dimensionale Analyse der Daten 26
3.7 Klassifizierung des Netzbetriebs auf Grundlage der Messwerte durch einen Machine-Learning-Algorithmus 28
3.8 Iteratives Verfahren bei der Anpassung der Machine-Learning-Algorithmen 30
3.9 Aufspaltung der Daten in ein Traings-, Validierungs- und Testset (Hold-Out-Validierung) 31
3.10 Anpassung an die Daten 32
3.11 Kreuzvalidierungsverfahren 32
3.12 KNN Erkennungsgenauigkeiten bei Variation der Nachbarn 33
3.13 IEEE 14-Bus-System 35
3.14 Direkte Maschen der UWs: blau(UW1), rot(UW2), violett(UW1 & UW2) . 37
3.15 Messgrößenaufnahme und Darstellung der Bezugssinne im UW1 und UW2 38
3.16 Verbindung zwischen linearem und nichtlinearem Modell 38

4.1 Ausfall der Leitung L12-1 42
4.2 Simulationsergebnisse UW1: Impedanzwinkel und Ströme der Leitungen 15 und 25 43

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.3 Simulationsergebnisse UW2: Impedanzwinkel und Ströme der Leitungen 24 und 34 . . . 43
4.4 Ausfall der Leitung L23 . . . 44
4.5 Simulationsergebnisse UW1: Impedanzwinkel und Ströme der Leitungen L15 und L25 . . . 45
4.6 Simulationsergebnisse UW2: Impedanzwinkel und Ströme der Leitungen 24 und 34 . . . 46
4.7 Ausfall der Leitung L34 . . . 47
4.8 Simulationsergebnisse UW1: Impedanzwinkel und Ströme der Leitungen 24 und 34 . . . 47
4.9 Gemeinsame Betrachtung der Ausfälle . . . 48
4.10 Gemeinsame Betrachtung der Ausfälle im UW2 . . . 49
4.11 Klassifizierungsergebnisse mittels KNN, Anzahl Attribute: 7 . . . 58
4.12 Klassifizierungsergebnisse mittels KNN, Anzahl Attribute: 3 . . . 58
4.13 Erkennung mittels LSVM . . . 59
4.14 Genauigkeiten der Erkennung mittels KNN mit Kostenmatrix aus Tab.4.2; Anzahl der Attribute: 5 . . . 61
4.15 Genauigkeiten der Erkennung mittels KNN mit Kostenmatrix aus Tab.4.3; Anzahl der Attribute: 5 . . . 62
4.16 Erkennung mittels LSVM . . . 63
4.17 Genauigkeiten der Erkennung mittels Boosted Trees: Ensemble methode: Bag, Learner Type: Decision Trees, Number of Learners: 15, Anzahl Attribute: 7 . . . 64
4.18 Genauigkeiten der Erkennung mittels Boosted Trees: Ensemble methode: Bag, Learner Type: Decision Trees, Number of Learners: 15, Anzahl Attribute: 5 . . . 64

5.1 Zustände vor und nach dem Ausfall miteinander in Verbindung bringen . . 67

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Tabellenverzeichnis

3.1 Kontingenztafel 29

4.1 Ergebnisse Machine-Learning-Algorithmen im UW1 bei Ausfall der Leitung zwischen SS1 und SS2 50
4.2 Einstellungen der Parameter des KNN-Algorithmus 51
4.3 Ergebnisse Machine-Learning-Algorithmen im UW2 bei Ausfall der Leitung zwischen SS1 und SS2 52
4.4 Ergebnisse Machine-Learning-Algorithmen im UW1 bei Ausfall der Leitung zwischen SS2 und SS3 53
4.5 Einstellungen der Parameter des KNN-Algorithmus 53
4.6 Ergebnisse Machine-Learning-Algorithmen im UW2 bei Ausfall der Leitung zwischen SS2 und SS3 54
4.7 Einstellungen der Parameter des KNN-Algorithmus 54
4.8 Ergebnisse Machine-Learning-Algorithmen im UW1 bei Ausfall der Leitung zwischen SS3 und SS4 55
4.9 Einstellungen der Parameter des KNN-Algorithmus 56
4.10 Einstellungen der Parameter des KNN-Algorithmus 57
4.11 Einstellungen der Parameter des KNN-Algorithmus 60
4.12 Kostenmatrix KNN 60
4.13 Kostenmatrix KNN 61

A.1 Leitungsdaten 73
A.2 Daten Synchronkompensatoren 74

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Literaturverzeichnis

[1] "Ieee recommended practice for excitation system models for power system stability studies.", *IEEE Standard*. 39

[2] A. Adewole und R. Tzoneva, "Impact of iec 61850-9-2 standard-based process bus on the operating performance of protection ieds: Comparative study", *IFAC Proceedings Volumes*, Vol. 47, S. 2245–2252, 2014. 6

[3] K. Bennett und C. Campbell, "Support vector machines: Hype or hallelujah?", *ACM SIGKDD Explorations Newsletter*, 2020. 11, 13

[4] W. Brand, *Der Standard 61850 - Offene Kommunikation in Schaltanlagen im dererguliertem Strommarkt*, 2002. 4

[5] L. Breimann, "Bagging predictors", *Machine Learning*, Vol. 24, S. 123–140, 1996. 15

[6] E. Chai1, P. Zeng, S. Ma, Xing, H. und B. Zhao, "Artificial intelligence approaches to fault diagnosis in power grids: A review", 2019. 19

[7] T. Chen und C. Guestrin, "Xgboost: A scalable tree boosting system", *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, S. 785 – 894, 2016. 16

[8] A. K. Claveria und A. Kalman, "Goose protocol: Ied's smart solution for victoria university zone substation (vuzs) simulator based on iec61850 standard", 2018. 5

[9] S. Czepiel, "Maximum likelihood estimation of logistic regression models: theory and implementation",. [Online]. Available: http://czep.net/stat/mlelr.pdf 16

[10] J. H. Friedman, "Greedy function approximation: a gradient boosting machine.", 2001. 16

[11] M. R. . F. H., "Insight in iec 61850 – what is it?", Vol. NASPI IEC 61850 tutorial, 2012. 5

[12] Hydro-Québec und I. The MathWorks, "Simscape electrical user's guide (specialized power systems)". 38, 39

[13] S. B. Kotsiantis, "Decision trees: A recent overview", *Artificial Intelligence Review*, Vol. 39, 2013. [Online]. Available: https://doi.org/10.1007/s10462-011-9272-4 13

[14] L. Liang, Z. Cai, D. Wang und S. Jiang, "Survey of improving k-nearest-neighbor for classification", *FSKD (1)*, 2007. 9

[15] R. E. Schapire und Y. Freund, "Boosting: Foundations and algorithms". 16

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[16] A. J. Schwab, *Elektroenergiesysteme: Erzeugung, Transport, Übertragung und Verteilung elektrischer Energie*, Reihe Studium und Praxis. Springer, 2012. 3, 4

[17] S. Singh und P. Gupta, "Comparative study id3, cart and c4.5 decision tree algorithm: A survey", *International Journal of Advanced Information Science and Technology (IJAIST)*. 15

[18] S. Xiiang, F. Nie und Z. C, "Learning a mahalanobis distance metric for data clustering and classification", *Pattern Recognition*, Vol. 41, S. 3600–3612, 2008. 9

[19] B. Yk, "Ieee 14 bus system simulink model", *MATLAB Central File Exchange*. 39

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

A Anhang

Tabelle A.1: Leitungsdaten

Leitung	Von SS	Zu SS	R [p.u.]	X [p.u.]	1/2 B [p.u.]
L12-1	1	2	0.01938	0.05917	0.0264
L15	1	5	0.05403	0.22304	0.0246
L23	2	3	0.04699	0.19797	0.0219
L24	2	4	0.05811	0.17632	0.017
L25	2	5	0.05695	0.17388	0.0173
L34	3	4	0.06701	0.17103	0.0064
L45	4	5	0.01335	0.04211	0
L47	4	7	0	2.0912	0
L49	4	9	0	0.55618	0
L56	5	6	0	0.25202	0
L611	6	11	0.09498	0.1989	0
L612	6	12	0.12291	0.25581	0
L613	6	13	0.06615	0.13027	0
L78	7	8	0	0.17615	0
L79	7	9	0	0.1101	0
L910	9	10	0.03181	0.0845	0
L914	9	14	0.12711	0.27038	0
L1011	10	11	0.08205	0.19207	0
L1213	12	13	0.22092	0.19988	0
L1314	13	14	0.17093	0.34802	0

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Tabelle A.2: Daten Synchronkompensatoren

SC	$SC3$	$SC6$	$SC8$
Leistung [p.u.]	0.40	0.25	0.25
Base [MVA]	100	100	100
V [p.u.]	1.01	1.07	1.09
H [s]	1.52	1.2	1.2
r_a [p.u.]	0	0.0025	0.0025
x_d [p.u.]	2.373	1.769	1.769
x_q [p.u.]	1.172	0.855	0.855
x'_d [p.u.]	0.343	0.304	0.304
x'_q [p.u.]	1.172	0.5795	0.5795
x''_d [p.u.]	0.231	0.2035	0.2035
x''_q [p.u.]	0.231	0.2035	0.2035
T'_{d0} [s]	11.6	8.0	8.0
T'_{q0} [s]	0.159	0.008	0.008
T''_{d0} [s]	0.058	0.058	0.058
T''_{q0} [s]	0.201	0.0151	0.0.151

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Eidesstattliche Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit gemäß dem Code of Conduct, insbesondere ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel, angefertigt wurde. Die aus anderen Quellen direkt oder indirektübernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In– noch im Ausland in gleicher oder in ähnlicher Form in anderen Prüfungsverfahren vorgelegt.

Wien, 8. Juni 2021

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.