ANKARA YILDIRIM BEYAZIT UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES



A COMPREHENSIVE REVIEW OF BIG DATA APPLICATIONS FOR POWER SYSTEMS

M.Sc. Thesis by

Okan ŞEN

Department of Energy Systems Engineering

December, 2019

ANKARA

A COMPREHENSIVE REVIEW OF BIG DATA APPLICATIONS FOR POWER SYSTEMS

A Thesis Submitted to

The Graduate School of Natural and Applied Sciences of

Ankara Yıldırım Beyazıt University

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Energy Systems Engineering, Department of Energy Systems Engineering

by

Okan ŞEN

December, 2019

ANKARA

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"A COMPREHENSIVE REVIEW OF BIG DATA APPLICATIONS FOR POWER SYSTEMS"** completed by **OKAN ŞEN** under the supervision of **PROF. DR. KAMİL ÇAĞATAY BAYINDIR** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Kamil Çağatay BAYINDIR

Supervisor

Prof. Dr. Abdullah YILDIZ

Jury Member

Assoc. Prof. Dr. Ahmet ÖNEN

Jury Member

Prof. Dr. Ergün ERASLAN

Director

Graduate School of Natural and Applied Sciences

ETHICAL DECLARATION

- I hereby declare that, in this thesis which has been prepared in accordance with the Thesis Writing Manual of Graduate School of Natural and Applied Sciences,
- All data, information and documents are obtained in the framework of academic and ethical rules,
- All information, documents and assessments are presented in accordance with scientific ethics and morals,
- All the materials that have been utilized are fully cited and referenced,
- No change has been made on the utilized materials,
- All the works presented are original,

and in any contrary case of above statements, I accept to renounce all my legal rights.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Kamil Çağatay BAYINDIR for his tremendous support and motivation during my study. His immense knowledge and precious recommendations constituted the milestones of this study. His guidance assisted me all the time of my research and while writing this thesis.

2019, December **Okan ŞEN**

A COMPREHENSIVE REVIEW OF BIG DATA APPLICATIONS FOR POWER SYSTEMS

ABSTRACT

By this study, author aims for reader to get basic knowledge about energy systems, relation between data and energy systems, big-data analytics concept. Concept of intelligent energy systems, namely smart grid is described. Concept of big-data is studied. Benefits of data-driven systems are searched.

A detailed information about applications of big-data on energy systems is given. Sources of energy big-data is studied. Many applications studied before are analyzed.

Finally, we point out the challenges of managing big data-driven energy systems. Effect of big-data analytics of energy efficiency. Examples of studies to increase energy efficiency.

Challenging aspects of managing such a huge amount of data. Gives some requirements of s IT infrastructure to achieve applications on data.

A comprehensive review on big data driven smart energy management applications is presented.

Keywords: Big-data, Energy Systems, Information Technology, Security, Energy Management, Data Analytics

GÜÇ SİSTEMLERİNİN BÜYÜK VERİ UYGULAMALARININ KAPSAMLI OLARAK İNCELENMESİ

ÖZ

Bu çalışmada, yazar okuyucunun enerji sistemleri, veri ile enerji sistemleri arasındaki ilişki, büyük veri analitiği hakkında temel bilgilere sahip olmasını amaçlamıştır. Başta smart grid olmak üzere akıllı enerji sistemleri açıklnamıştır. Büyük veri kavramı üzerinde çalışılmıştır. Veri sürümlü sistemler araştırılmıştır.

Enerji sistemlerinde büyük veri uygulamaları hakkında detaylı bilgi verilmiştir. Sistmlerin büyük veri kaynakları üzerinde çalışlımıştır. Daha önce yapılmış çalışmalar ile ilgili bilgi toplanmıştır.

Sonuçta veri sürümlü sistemlerin yönetimindeki zorlukları belirliyoruz. Enerji verimliliği üzerindeki katkılarını çalıştık. Verimlilik artışı yapmak için yapılmış çalışmalar üzerinde durduk.

Çok miktardaki verilerin yönetilmesindeki zorluklar üzerinde duruldu ve bu sistemlerin bilgi teknoljisi altyapısı ile ilgili bilgi verildi.

Güç ve enerji sistemlerinin veri yönelimli uygulamaları ile ilgili kapsamlı bir çalışma yapıldı.

Anahtar Kelimeler: Büyük-Veri, Enerji Sistemleri, Bilgi Teknolojisi, Güvenlik, Enerji Yönetimi, Veri Analitiği

CONTENTS

M.Sc. THESIS EXAMINATION RESULT FORM ii
ETHICAL DECLARATION iii
ACKNOWLEDGEMENTS iv
ABSTRACT v
ÖZ vi
CONTENTS vii
NOMENCLATURE x
LIST OF TABLES xiv
LIST OF FIGURES xv
CHAPTER 1 - INTRODUCTION 1
1.1 Background 1
1.2 Purpose of Study 2
1.3 Research Question 2
1.4 Thesis Outline 2
CHAPTER 2 - BIG DATA 4
2.1 Data Analytics and Big Data 4
2.2 Brief History of Big Data 4
2.3 Big data characteristics 5
2.3.1 Volume 5
2.3.2 Velocity 5
2.3.3 Variety 6
2.3.4 Veracity 6
2.3.5 Value 6
2.4 Sources of big data 7
2.5 Computation Techniques 7
2.6 Application areas of big data 9
2.6.1 Understanding and Targeting Consumers 10
2.6.2 Self-Optimization 10
2.6.3 Improvement of Healthcare 10

2.6.4 Security and Crime Determination 11
2.7 Big data for power systems 11
CHAPTER 3 - SMART SYSTEMS 14
3.1 Smart System Basics and Concepts 16
3.1.1 Internet of Things 16
3.1.2 Energy Internet and Digitalization in Energy 19
3.2 Smart System Components 20
3.2.1 Smart meters 20
3.2.2 Infrastructures 25
3.2.3 Communication 25
3.3 Smart System Characteristics 26
3.3.1 Real-Time 26
3.3.2 Secure & Reliable 26
3.3.3 Manageable & Visualizable 30
3.3.4 Ecologically Sensitive & Clean 33
3.3.5 Technological Assets of a Smart System 35
CHAPTER 4 - BIG DATA APPLICATIONS 39
4.1 Smart Grid 39
4.1.1 Generation 39
4.1.2 Transmission 56
4.1.3 Distribution 56
4.2 Renewable Energy Systems 58
4.2.1 Solar Energy 61
4.2.2 Wind Energy 62
4.3 Demand Side Management and Demand Response 64
4.3.1 Demand Forecasting 64
4.3.2 Price Prediction 71
4.4 Energy Management 72
4.4.1 Residential, Commercial & Industrial Consumption 72
4.4.2 Objectives 75
4.4.3 Dynamic Energy Management 85

4.5 Power Quality 86
4.6 Fault Diagnosis and Asset Management 86
4.7 Smart and Sustainable Cities 87
4.8 Electrical vehicles 89
4.9 Challenges and future perspectives 91
CHAPTER 5 - CONCLUSIONS 94
REFERENCES 95
CIRRICULUM VITAE 109

NOMENCLATURE

Acronyms

IoT	Internet of Things
GrC	Granular Computing
IT	Information Technologies
SCADA	Supervisory Control And Data Acquisition
GPU	Graphics Processing Unit
DNA	Deoxyribonucleic Acid
AMI	Advanced Metering Infrastructure
SMASH	Smart Meter Analytics Scaled by Hadoop
ANFIS	Adaptive Neuro Fuzzy Inference System
CVVM	C-vine copulas framework
ICT	Information and Communication Technologies
NIST	National Institute of Standards and Technology
CCHP	Combined Cooling, Heating and Power Systems
RMT	Random Matrix Theory
MSR	Mean Spectral Radius
CIM	Common Information Model
SG	Smart Grid
PCA	Principal Component Analysis
FRA	Fast Recursive Algorithm
TLBO	Teaching-Learning Based Optimization
AVC	Automatic Voltage Control
EBDG	Event-Based Data Gathering
DSS	Decision Support System
EMS	Energy Management System
SoC	System on Chip
BI	Business Intelligence
HVAC	Heating, Ventilation and Air Conditioning
SDSR	Small Disturbance Stability Region
SVM	Support Vector Machine

DR	Demand Response
EIBD	Energy Internet Big Data
DEM	Dynamic Energy Management
CPS	Cyber-Physical System
IDS	Intrusion Detection System
DGA	Dissolved Gas Analysis
TSA	Transient Stability Assessment
PMU	Phasor Measurement Units
CVM	Core Vector Machine
IEEE	Institute of Electrical and Electronics Engineers
OSI	Open System Interconnection
GIS	Geographic Information System
DFA	Detrended Fluctuation Analysis
PDFA	Parallel Detrended Fluctuation Analysis
SVR	Support Vector Regression
OPF	Optimal Power Flow
RT	Real-time
GMM	Gaussian Mixture Model
GABEM	Genetic Algorithm-Based Expectation Maximization
SDP	Semi Definite Programming
ADMM	Alternating Direction Method of Multipliers
IS	Information Systems
CB	Competitive Benchmarking
TAC	Trading Agent Competition
MRV	Monitoring, Reporting and Verification
IMO	International Maritime Organization
EI	Energy Internet
BD	Big Data
EV	Electric Vehicles
GEI	Global Energy Interconnection
BDEv	Big Data Evaluator
GPS	Global Positioning System

SWOT	Strengths Weaknesses Opportunities Threats
DER	Distributed Energy Resource
MCDM	Multi-Criteria Decision Making
MOGLP	Multi-Objective Grey Linear Programming
HGV	Heavy good vehicles
K-LIWC	Korean-Linguistic Inquiry and Word Count
SNS	Social Networking Services
WSN	Wireless Sensor Networks
PV	PhotoVoltaic
RMSE	Root Mean Square Error
FPTAS	Workflow Based Polynomial-Time Approximation Scheme
TESEMA	Techno-Econo-Socio-Environmental Multivariate Analysis
DNN	Deep Neural Network
DI	Development Index
ELM	Extreme Learning Machine
NRMSE	Normalized Root Mean Square Error
CPS	Cyber Physical System
FFO	Fruit Fly Optimization
BRR	Best Resource Ratio
UAV	Unmanned Aerial Vehicle
Rcnn	Region-based Convolutional Neural Networks
Caffe	Convolutional Architecture for Fast Feature Embedding
CPU	Central Processing Unit
HDD	Hard Disk Drive
SDD	Solid State Disk
RF	Random Forest
GCA	Grey Correlation Analysis
KPCA	Kernel function and Principle Component Analysis
I2S	Intelligent Immune System
ANN	Artificial Neural Network
SM	Smart Manufacturing
HDFS	Hadoop Distributed File System

LASSO	Least Absolute Shrinkage Selection Operation
MRE	Mean Relative Error
MLP	Multi-Layer Perceptron
MLR	Multi-Variable Linear Regression
BEGIN	Big data enabled EnerGy-efficient vehIcular edge computiNg
NN	Neural Network
CC	Cloud Computing
FC	Fog Computing
ML	Machine Learning
LSTM	Long Short-Term Memory
ECE	Economic Commission of Europe
V2G	Vehicle-to-grid technology
SCF	Schedulable Capacity Forecasting
PDTR	Parallel Decision Tree Regression
PRFR	Parallel Random Forest Regression
IEN	Intelligent Energy Network
DSM	Demand Side Management
API	Application Programming Interface

LIST OF TABLES

Table 2.1 Potential application areas of big data in smart grid 13

LIST OF FIGURES

Figure 2.1 Major elements of big data 6
Figure 2.2 Delivering benefits through big data analytics 12
Figure 3.1 Computational capability per kWh in ages.....15
Figure 3.2 IoT architecture layers and components 16
Figure 3.3 System architecture.....18
Figure 3.4 Working principle of data distribution between smart meters.....22
Figure 3.5 Comparison among the clustering techniques 25
Figure 3.6 History of power system protection.....27
Figure 3.7 Intrusion detection model achitecture.....28
Figure 3.8 Energy systems & data flow 29
Figure 3.9 Architecture of big data platform.....32
Figure 3.10 Big data platform of power grid fault 36
Figure 4.1 Three stages of an electrical power system 39
Figure 4.2 Summary of K-LIWC on the collected data 41
Figure 4.3 Solution for obtaining the optimal electricity generation mix. 43
Figure 4.4 A typical smart grid model 45
Figure 4.5 A Smart grid controlled CCHP system.....46
Figure 4.6 Apache spark framework.....48
Figure 4.7 Web based portal 50
Figure 4.8 Data fusion and specific data structure of power equipment.....51
Figure 4.9 Flowchart of EBDG. 54
Figure 4.10 Big data for distribution system parameter estimation 57
Figure 4.11 BigDataOcean wave power value chain 60
Figure 4.12 Output losses in the grid for different PV capacity ranges. 61
Figure 4.13 Scatter plot of wind power versus wind speed loads. 63
Figure 4.14 Ultra-short-term forecasting error distribution after error correction.....63
Figure 4.15 Neural network training.....65
Figure 4.16 Workflow of the proposed dynamic ensemble 67
Figure 4.17 Comparison between actual and forecast loads. 69

Figure 4.18 Comparison on accuracy among HSEC and benchmark frameworks. 72
Figure 4.19 The data monitoring/processing topology 74
Figure 4.20 Design of big data structure 76
Figure 4.21 Decision-making layers for smart buildings in a smart city. 79
Figure 4.22 Workflow of Job Scheduling 81
Figure 4.23 Big-data pipeline 83
Figure 4.24 Building blocks of dynamic energy management 85
Figure 4.25 Main data mining tasks 87
Figure 4.26 Data processing procedure 91
Figure 4.27 Future directions on implementing data analytics 93

CHAPTER 1

INTRODUCTION

1.1 Background

Energy sector has a key role for a sustainable development, since continuous energy supply is mandatory for societies and economies. Previously, data produced by an energy system was only used to see errors occurred while system is not monitored. Works on recent years show that data collection is much more important than it is thought before. Moreover, it is understood that data production is inevitable for energy systems. Now, benefits about work on data analytics is seen clearly.

Especially, after development of intelligent energy network systems, data size extracted is getting bigger. To enable system to be “smart” a huge amount of data is exchanged between components and the enterprise systems managing these components. Smart grid is the brain of an intelligent system which requires reliable real-time data for processing. As a result, energy systems are designed such to produce increasing amount of data to be used for advanced statistical machine learning applications.

Data analysis is the key lever for optimization of energy management. It can provide benefits by means of system measurement, operation and control. By processing huge amount of data, especially with usage of neural network techniques demand management can be performed in an effective way.

Dealing with data has some difficulties also. Size of data is one of the main or maybe the first challenging tasks of data consideration. By continuous application of sensors, transmission logs, network communication and error events large amounts of data are increasingly accumulated in energy sector. Collection, integration and sharing data may cause some standardization problems. As data security is a basic consideration for every system, it is a big task for power systems also.

In our age, addition to classical energy sources, renewable energy sources are began to be used. If this variety is extended to transmission and distribution systems the complexity of power network systems can be understood. It is natural that this complexity causes power data also to become a big and complex data.

A smart system cannot be imagined without prediction of demand. This is social demand of power and energy. By help of estimation of weather conditions, decision and aggregation of data calculations, demand can be managed feasibly.

1.2 Purpose of Study

The purpose of this study is to analyze how big datasets become an advantage for power systems. Analysis of such big data yields results to gain insight and obtain prediction in the consumption data and current inefficiencies, consumers consumption behavior. These results will be used also for increasing energy saving abilities and design and construction of cost-efficient new age energy systems.

1.3 Research Question

The research questions can be proposed as follows;

How big data is obtained in recent power systems

What is the advantage of big data for power systems?

By which aspects power systems are developed by big data analysis techniques?

In which areas of power systems big data is used?

What are the future applications of big data on power systems?

1.4 Thesis Outline

Following chapters are organized as follows;

In Chapter 2 big-data concept is presented. In this chapter big data is defined, a chronological development of big data technique is described and some areas other than power systems are listed and explained.

Chapter 3 is the main chapter of thesis. In this chapter, after discussing network elements of that causes formation of big data and energy generation is classification, applications of big data on power systems is explained in details.

Chapter 4 stands for declaration of conclusion and future works on studies of the subject of thesis.

CHAPTER 2

BIG DATA

2.1 Data Analytics and Big Data

The “Big Data” concept, the act of gathering and storing large amounts of information for useful analysis. Big data is a form of data that is obtained from different sources that is continuously being recorded. With the improvement of technology and internet, the concept gained momentum in the early 2000s and the power of information has come to around today. Now it is understood that if all the data that streams into a certain business is captured, analytics can be applied and significant value can be get from data.

Databases and existing data in it are being classified in relational databases. Private and public organizations make decisions due to reports generated through reporting systems running in these databases. But there are many more data sets that we cannot fully classify and relate to. Until the concept of big data is introduced, all of this information has been called dump, because it was very difficult to store this data in the databases and use it in reporting systems.

Executives had to wait technological improvements for their data to start inspiring solutions to their organizations’ problems. Now they data as a treasure. The growing concept “Big Data” has brought much success in the field of data science; it provides data scalability in a variety of ways that empower data science.

2.2 Brief History of Big Data

The applications of data and search to understand available data is present for a long time, but officially the term “Big Data” is first declared at beginning of 2000’s.

At 2005 a database is designed to index entire world wide web.

After this time social networks were rapidly increasing and large amounts of data were being created on a daily basis. Businesses and governments alike began to establish big data projects.

2009 the Indian government stored fingerprint and iris scans of all of its citizens. This was the largest biometric database ever created. At year 2010 it is told that 5 exabytes of data stored since the beginning of time up to the year 2003. by the year 2016, the same amount of data would normally be created every two days.

Over the past number of years, there have been various organizations that have come up in an attempt to deal with big data, These organizations' business is aiding other businesses to understand big data. Everyday more and more businesses are moving towards the acceptance and exploitation of big data.

Although it seems like big data has been around for a long time now and that we are getting closer to the pinnacle, big data may just be at its formidable stages. Big data in the near future may end up making big data now seem like a poultry amount.

2.3 Big data characteristics

The characteristics of big data is defined by five major elements. These are **V**olume, **V**elocity, **V**ariety, **V**eracity and **V**alue. These are 5Vs of big data.

2.3.1 Volume

Big data has large volume of data. This is a consequence of data being built by large number of deployed smart metering devices and a wide-range level of measurement. This presents also new opportunities to detail analysis, but also a challenging task to store and processing of data.

2.3.2 Velocity

This refers to speed for collecting, processing and usage of big data of systems. Time intervals of data collection should be as short as possible to obtain data in details. Processing time should be short to decrease response time to an acceptable level. Power systems are required to work in real-time or almost real-time manner for known reasons.

2.3.3 Variety

This stands for having to consider different types of data. Structured or unstructured data may be in consideration for a system due to different levels of recording and at different time. Since standardization is reduced, processing algorithms have to adapt to this situation.

2.3.4 Veracity

This represents uncertainty of data. Especially in power systems, sometimes generation, sometimes demand and mostly both of them is related to natural events. Overall system of processing of data must assure to regard this veracity.

2.3.5 Value

This refers to be able to extract valuable information from data. Endless amount of data collected is useless unless it is turned into value. The opportunity of big data applications is directly related with the value hidden inside it.

Figure 2.1 Major elements of big datashows these major elements of big data concept.

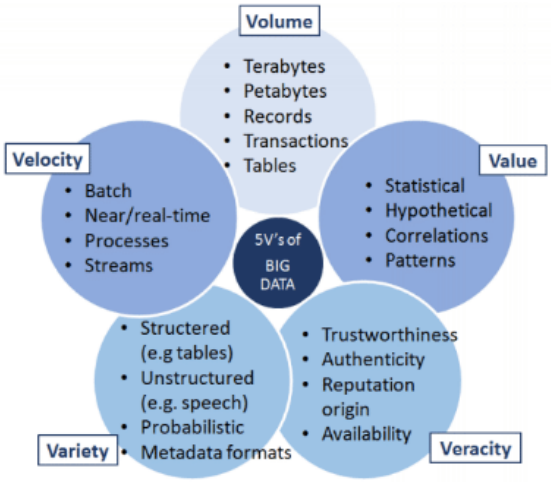


Figure 2.1 Major elements of big data

There are there are three major obstacles to the development and implementation of big data analytics in power distribution systems [1].

(i) Lack of innovative use cases and application proposals that convert big data into valuable operational intelligence.

(ii) Insufficient research on big data analytics system architecture design and advanced mathematics for this size of data.

(iii) Risk of failing to adhere to data privacy and data protection standards.

2.4 Sources of big data

Power systems collects data that is large in volume continuously in different format from different sources [2]. Among many others some data source examples can be given;

- Traditional Utilities data on customers and service connection and assets
- Smart meter data allowing better insight on customer usage and service quality
- Smart grid devices data bringing extended information and control over Utilities grids
- Customer interaction channels data over voice, internet, mobile
- Home automation and intelligent home devices data feeds
- New technologies requiring additional monitoring – Electric Vehicles, Wind Generation, City Lighting monitoring, Photo Voltaic panels, micro-grids integration…
- Localized weather data
- Socio-demographic information and other publicly available information

2.5 Computation Techniques

There are some techniques which are used by modern technologies to analyze big data. Some of them adore to be explained briefly here. These are cloud computing, Internet of things, (IoT), granular computing, data center and quantum computing. [3]

Cloud computing: Cloud computing is delivery of applications and services through Internet. It is closely related to big data. Cloud computing aims to use distributed computing and storage resources at the same time to provide big data applications [4]. The occurrence of big data concept also accelerates the development of cloud computing. By the help of distributed storage technology and the parallel computing big data can be analyzed effectively. Cloud computing increased requirement of transferring and sharing data also.

Via customer energy meters, energy generation units in the customer area and third party players a smart grid collects large amount of data from every substations of the energy utilization network. Not only utilizer to consumer but a two directional flow of information exists here. In this context, appropriate communication, reliable, distributed, parallel and scalable processing of data is needed. However, usually power system is lacking such computing capacity to satisfy requirements. Researchers suggest cloud computing to solve this problem. Analyzations on the application of cloud computing for power system application is done. Apprise on feasibility of the available cloud computing tools is conducted [5].

Internet of Things (IoT): The Internet of Things (IoT) is a network of things by name. "Things" stands for physical objects implemented with electronics, software and sensors. By exchanging data in between “things” (IoT) achieves a greater service than sum of independent network sources can do. (IoT) consists of high number of sensors embedded in devices in network. Deployed sensors in different fields collect various kinds of data. Naturally, IoT enables advanced applications for smart grid, since different types of data can be collected, which has heterogeneity and variety. It is expected that by 2030, the IoT data will be the most important part of big data [3].

Granular computing: Granular computing (GrC), assumes classes, clusters, subsets, groups and intervals as granules and are used to build computational models for complex applications. It offers a solution to uncertainty of big data which often contains a significant amount of unstructured, uncertain and imprecise data. GrC aims to find a suitable level of granularity of a problem level of which can be adjusted according to the degree of complexity of the problem.

Data center: Acquiring, managing & organizing data is performed by datacenters. Data centers are infrastructure and basic for data collection. Data centers should be modernized continuously, due to rapid growth of IT technologies. As grids gets smarter, explosive growth of data is experienced. As a consequence enhancement of infrastructure and related software of data center is a must. The development of data centers should be taken into consideration to improve the capacity of rapidly and effectively processing of big data.

Quantum computing: Quantum computing studies quantum computers that use quantum-mechanical b [6]. Unlike digital computers Quantum computers require data to be encoded into binary digits. These are called quantum bits. Actual quantum computers have newly begun to develop, but experiments have been carried out for developing. It is calculated that large-scale quantum computers will be able to solve problems much more quickly than classical computers.

Some studies considers converged computing method and an associated model that create SCADA systems based on wireless networks for the energy industry [7]. For aggregation of primary data, fog computing is used as well as to warn about abnormal and emergency situations on SCADA systems objects. Large size of sensor data, integral indicators and heterogeneous information from various sources is processed via GRID computing model. GRID computing model has three level architecture, which includes the main server, a cluster of servers and a lot of GPU video card with support for Compute Unified Device Architecture. The implementation of covered computing models for the operation of components of SCADA system is the shortest way and nearest approach to distributed sensor data processing.

2.6 Application areas of big data

Big data analytics helps organizations work on their data and use it to identify new opportunities. These may lead to smarter business moves, more efficient operations, higher profits and other advantages. As advantages of big data is being understood by time, there has been a shift of interest in the subject. big data analysis is started to be used in new areas. Some examples can be given that big data is being applied.

2.6.1 Understanding and Targeting Consumers

This is a common area of big data usage. Consumers' behaviors and interests are focused for better understanding to develop business. Incorporating the use of sensor data and data obtained from social media, browser logs, a clearer picture of consumers need can be obtained. Once this is understood, a business will target to right position of consumer needs and create predictive models.

Analyzing audience's interests may be a subject of data analytics also. A politician or a leading person of an organization even behave as to influence society by the help of big data.

2.6.2 Self-Optimization

Big data does not only applicable to a business, but also can be applied to an individual also. Mobile phones and watches and other smart devices generate data about a person who use them. For example amount of calories that a person intake, activity levels in a day and sleep patterns are recorded daily or hourly. With the analysis of data collected over a certain period of time, one can adjust his personal life in order to be more healthy, productive and to acquire sufficient amounts of sleep.

2.6.3 Improvement of Healthcare

Every member of a society carries a complex and big data on his body. This is genetic codes of that member. Today by the help of computers, analyzing and decoding DNA issues can be achieved in very short time periods. This will allow us to discover diseases faster then ever. In addition, this gives opportunity to healthcare providers to predict the patterns of diseases. Therefore, measures can be set up to prevent further spread of the diseases. Easy collection of data from mobile phones forms a base for various health studies. Applications executing on devices can inform user about his life style, nutrition faults and treatment services.

The increased number of users means increased data variety, increased number of samples for intelligent systems to learn and consequently increased quality of services. Probability of having illness for a person or more generally epidemic risks in a society can be predicted and precautions taken.

2.6.4 Security and Crime Determination

Another area where big data applied is in the security enhancement and law enforcement. Official institutions, use big data to detect potential criminal activities and terrorism. By the help of big data the officers of security departments predict and deter criminal activities. Group of people who carry criminal risk can be generated by a computer through the analysis of big data. The officers can execute a proactive process by offering them information about jobs and training programs.

In technological age, on the other hand, big data is used for prevention of cyber attacks and unauthorized access.

2.7 Big data for power systems

Appropriate big data analytics capabilities will be required to be able to capture, manage, and analyze large set of information in a logical and relational manner. Big data Analytics extract value across utilities of operations examples of which are given below [2];

- Anticipate failures of distribution assets. Transformers are examples. Before they occur, to structure proactive maintenance operations.
- Improvement of balancing generation and demand. Better timely forecasting and demand response management is required.
- Energy planning forecasts to decrease energy costs on markets, allowing conventional and distributed generation plants to be used efficiently.
- Improve Customer Service Quality by identifying power cuts more accurately for faster restoration.
- Enable optimization and control of mobile generation at the time during which the system is in acceptable mode by the connection of massive number of Electrical Vehicles and Renewable Energy Sources.
- Supporting optimized Low Voltage distribution network operation, including applications for voltage optimization, outage management, extended network monitoring and better connection services.

- Enable better way of consuming by enabling Energy Use advisory based on anonymous customer benchmarking and Energy saving.
- Structure effective Energy Saving policies and campaigns to foster rational use of energy resources and enable adoption of energy management services and technologies.
- Improve customer service with error free metering and billing services and anticipating consumptions anomalies as early as possible.
- Allowing Utilities to identify and reduce losses accurately enabling them to address them and reduce the overall overhead cost charged on the user community.

Figure 2.2 shows benefits of big data driven systems.

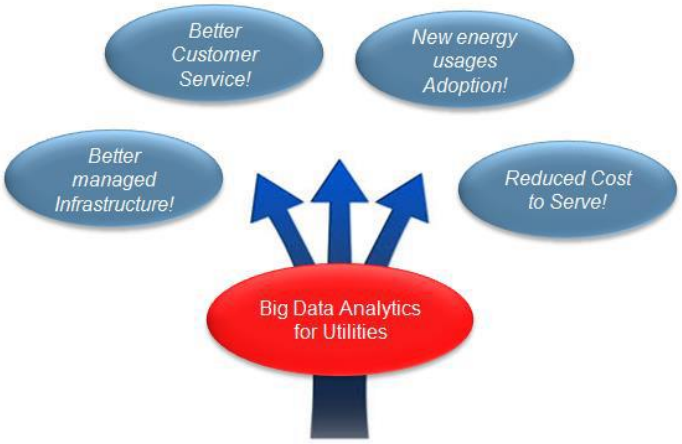


Figure 2.2 Delivering benefits through big data analytics

All these cases refer to a management of customer data. Privacy is a key and real concern. Banking industry, Telecom industry and many other industries have already come up with similar ‘data explosions’ and are successfully operating based on the European Privacy framework. These success stories can be supported to ensure customer privacy protection at high levels. Table 2.1 summarizes some potential application areas of big data in smart grid.

Table 2.1 Potential application areas of big data in smart grid

1	Prediction and analysis of economic situation and social impact
2	Development of scientific reasoning for decision making
3	Performance analysis for generation and storage systems
4	Load management with demand response and energy utilization & efficiency analytics
5	Consumer behavior analysis
6	Using AMI and intelligent electronic devices for state estimation
7	Pricing analytics and incentive implementation analysis
8	Grid infrastructures optimization
9	Demand and generation forecast under high uncertainties
10	Asset management
11	Service quality analytics

CHAPTER 3

SMART SYSTEMS

Energy Efficiency is a phrase that we come across. Efficiency can be defined as amount of work done with a unit energy. Some searches on energy is on Energy-Efficiency Trends [8]. Digital computers of 1940s performed approximately one operation consuming per Watt-hour (Wh). Those machines were bulky, unreliable, were not user-friendly and had many problems. Moreover they were energy inefficient. Koomey's law [9] is known as the exponential decline of the per-computation energy requirement over decades. That is to say also there is an exponential rise in the number of operations performed per unit of energy in the same period of time. The exponential rise of computational capability per kWh of energy is observed in Figure 3.1

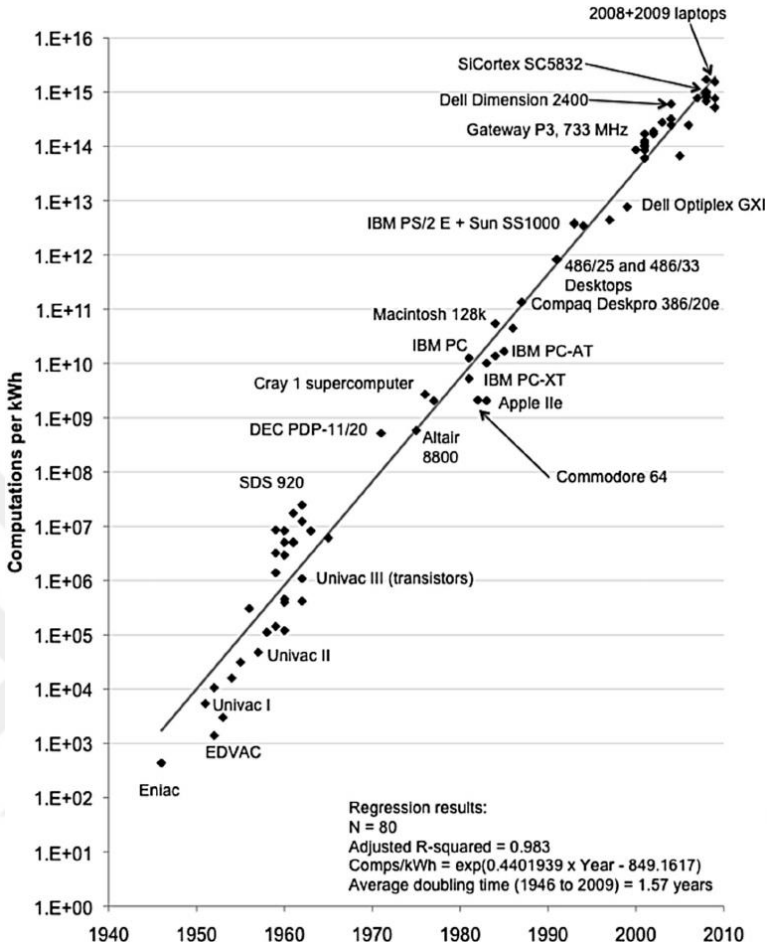


Figure 3.1 Computational capability per kWh in ages

Compared to classical energy systems, in the modern industrial systems data analytics is playing a more important role. With development of information and communication technology, Information layer is a part of a modern electrical transmission system. Such a system has wide installation of smart meters and sensors [10]. Installation of Smart Meter was a significant development, which establishes two-way communication between the service and the consumers. The Smart meters collect data at high speed leading to large volume of data and have been classified as big data.

3.1 Smart System Basics and Concepts

3.1.1 Internet of Things

Concept of Internet of Things (IoT) refers to an intelligent network based on technologies of infrastructure, database and communication. First of all, the level of the power system will be promoted, when such a modern concept is considered to be used. Moreover, it is applied to systems to improve stability and security of the power system. Protecting the environment and meet the renewable development requirements may be other purposes of usage. Figure 3.2 shows layers and components of IoT architecture.

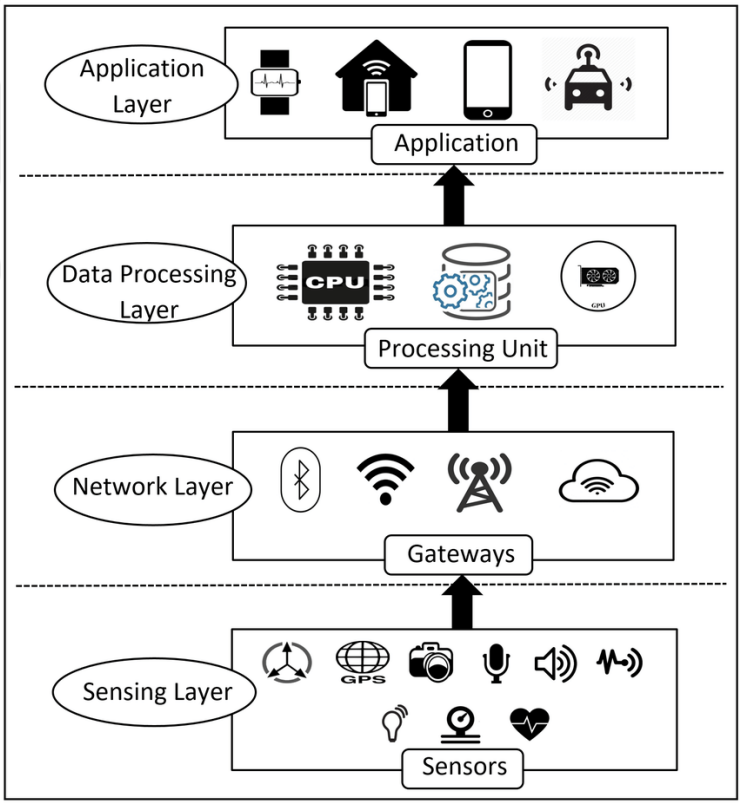


Figure 3.2 IoT architecture layers and components

Big data is one of ascendant technological concepts and includes smart and intelligent energy services, management, energy consumption prediction and exploitation of (IoT) solutions. Because of this, big data technologies will have significant effects in the energy sector. A big

data platform architecture is proposed to support the development, maintenance and exploitation of smart energy services [11]. The platform also enables the simplification of the procedure gathering by multiple sources. A Decision Support System (DSS) working on web has been developed according to the proposed architecture, Exploitation of multi-sourced data within a smart city context towards the creation of energy management action plans included. Application DSS is tested in three European cities during work and results discussed.

Improvement in energy efficiency is important for energy-intensive industries, an example of which is China, consumption of which is said to be %51. No doubt, cleaner production is as an effective way to improve energy efficiency and reduce energy consumption. However, there is a lack of manufacturing data due to the difficult implementation of sensors. Sensing high temperature, high pressure, high acid, high alkali, and smoky environment is important, since these data by big data techniques lead to implementation of the cleaner production strategy. Internet of Things brings many data can be sensed and collected in the manufacturing processes. A big data driven analytical framework is proposed to reduce the energy consumption in manufacturing industries [12]. A big data analytics framework is proposed consisting two key technologies, name of which is energy big data acquisition and energy big data mining. An application scenario is presented to demonstrate the proposed framework. The results of experiments conducted show that the energy consumption and energy costs are reduced by 3% and 4% respectively. These improvements can promote the implementation of cleaner production strategy which will definitely contribute to the sustainable development of energy intensive manufacturing industries.

Internet of Things (IoT) technology brings an assessment to power data recorded anywhere. Any power data can be analyzed through internet and be analyzed for real-time monitoring. To approach a real-time power monitoring platform an efficient implementation techniques has been developed [13]. The collected data stored by the Hadoop subsystem and Spark unit is used for data processes. This research adopts multilayer software modules to real-time power monitoring with high efficiency, high feasibility and low cost. At last step experiments are conducted to verify the query-response efficiency, and performance evaluations for the proposed real-time power monitoring platform. Experiment results reveals the high feasibility for the target research goals.

A smart Energy Management System (EMS) can reduce the costs of energy demand. The recent technologies like Internet of Things (IoT) and big data can be utilized to manage energy in a better way in the axis of consumption in residential, commercial, and industrial sectors. A paper presents an Energy Management System (EMS) for smart homes [14]. This system proposes each home device to be interfaced with a data acquisition module. Module is an IoT object and has a unique IP address which identifies it in a large mesh wireless network of devices. System on Chip (SoC) is the data acquisition module and collects energy consumption data from each device. Data is sent to a centralized server for further analysis and processing. The proposed EMS also utilizes Business Intelligence (BI) and big data analytics software packages. Manage energy for better consumption and to meet consumer demand requirements are satisfied by these utilizations. Validation of the proposed system HVAC units of Arab Gulf countries is taken as a case study. A prototype was built and tested in the lab. System architecture is shown in Figure 3.3.

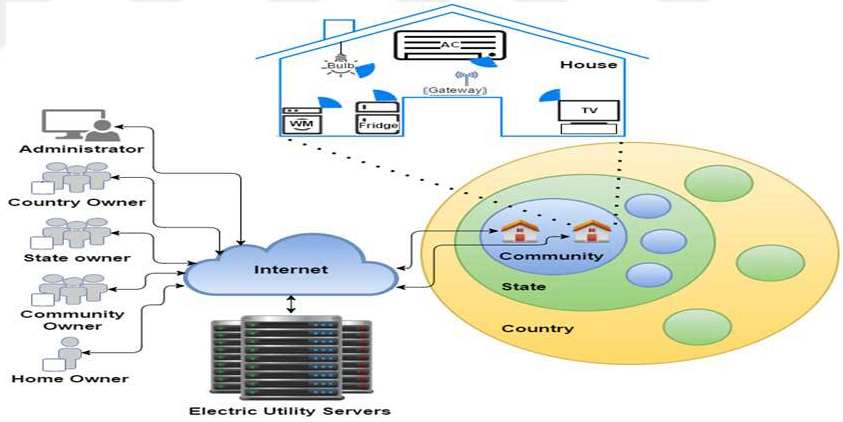


Figure 3.3 System architecture

IoT smart energy management is to maximize energy efficiency in energy management systems by its smart energy platform technology. A paper shows IoT energy management platform with energy big data system [15].

3.1.2 Energy Internet and Digitalization in Energy

Energy internet concept provides efficient utilization of wide area renewable clean energy. This contribution of energy internet is combined with real-time big data platform to enhance the architecture which is ready to solve energy system management problems. Three key technologies are analyzed in a paper [16]. These technologies are in use for the implementation of real-time big data platform of energy internet. These are

- unified communications and data standards of heterogeneous devices and networks,
- large-scale stream data real-time analysis,
- high-performance data processing based in-memory computing technologies.

Analysis results enable us to propose a feasible platform architecture.

Paper concerns Energy Internet with the system-level stability issues of it [17]. Study conducted to maintain a stable and healthy energy network environment. A system-level stability evaluation model of Energy Internet is proposed. Model explores small disturbance stability region (SDSR), SDSR can be acquired via estimating the operational data threshold of distributed generations. The threshold is estimated by energy consumption which reduces the computation complexity. Simulations executed as single and multiple dominant oscillation modes, which showed the advantages and superiority of method.

The data generated from different smart devices of smart cities is heterogeneous in nature and exhibits the properties of all essential V's for big data. To come against processing difficulties this a tensor-based big data management technique is presented to reduce the dimensionality of data gathered from the Internet-of-Energy (IoE) environment in a smart city [18]. After reducing the dimensionality of data, it is used to provide some services and Support vector machine (SVM)-based classifier is used to classify the end-users into categories named normal, overloaded and underloaded. Overall load is optimized. Results obtained indicate the supremacy of the proposed tensor-based scheme over the traditional scheme for DR management.

Energy and Internet is deeply integrated to form Energy Internet concept. In addition, Information Technology, Energy Industry and Energy Internet Big Data are core technologies that bind energy, information, economic interconnection to each other. This structure of various concepts improve the development and evolution of Energy Internet. A paper published to describe the concept and characteristics of Energy Internet Big Data (EIBD) [19]. Feasibility of applying EIBD to integrated energy market is discussed. Based on this, typical application and technical system based on EIBD in integrated energy market is proposed for multi-subjects of Energy Internet. These subjects provides a reference for the analysis and decision of integrated energy market in Energy Internet.

3.2 Smart System Components

3.2.1 Smart meters

Advanced sensor systems such as advanced metering infrastructure (AMI), high-frequency overhead and underground current and voltage sensors have been increasing significantly in power distribution systems over the past few years.

The smart meter is a key element for the smart grid and is expected to provide benefits for power systems. A key factor that will determine the success of smart meters is data analytics. An article is published after a comprehensive survey of smart electricity meters and their utilization [20]. Key aspects of the metering process is focused. Different stakeholder interests, and the technologies used to satisfy stakeholder interests are covered. Also, challenges as well as opportunities due to developing big data technologies are highlighted.

Smart meters are widely used in energy management systems. Smart meters are advanced meters having capability of measuring customer energy consumption at a period of time. For example an interval of 15 minutes [21].

A new core-broker-client system architecture is proposed for big data analytics. A platform named Smart Meter Analytics Scaled by Hadoop (SMASH) is implemented [22]. SMASH is tested on 20 TB scale datasets and it is seen to be is able to perform data storage, query, analysis and visualization tasks on such a large dataset. The performance of SMASH is

compared with the published results provided by Google Cloud, IBM, MongoDB, and AMPLab. Its performance is found to be competitive and easily operable in managing big energy data and supporting data-intensive decision making.

Feasible construction of the prediction model of Smart Meter life, the operational data of Smart Meter in running state has to be analyzed. A research on subject includes collection of all the data from running smart meters [23]. Data is recorded due to operational faults, maintenance management and their application data from the Power Information Collection System. Then the operational data in running state is analyzed. There prediction methods are considered. These are element stress method, reliability prediction method based on reliability test, and reliability prediction method based on reliability verification. These methods are examined to open front of design of a new modal.

Smart grids having smart sensor technology have some advantages. Some of them are;

- power monitoring
- demand-side energy management
- coordination of distributed storage
- integration of renewable energy generators

Smart sensor networks are used because of their low cost and their ease of deployment. A large size of data is a result. Processing and analyzing these data requires special techniques to utilize a power grid having better performance. Technology needed for collection of data is available, whereas managing this data and getting useful information from data is still a challenging task. A paper has this field of discussion [24]. Also recommendations provided on future works. In addition, different applications of smart sensor networks are searched. Eventually, a discussion on big data generated by sensors and meters and the techniques used to manage them is done.

Energy resource utilization can be efficiently obtained by smart grids by recent technological developments. Unlike traditional power systems large number of sensors are used in network to improve its efficiency and reliability which enable communication between generators and consumers. By this controlling, visualization and management of system is enhanced.

However in this case a question arose as to where in the network to process and analyze the data and how to perform this analysis. The data both can be processed by one processor or by distributed more than one processors. A thesis about subject proposes a distributed architecture, which contains home sensors in communication with a gateway in home. After data is locally processed by this gateway, it is transmitted to a central processor for detailed analysis [25]. The same study involves designation of a test bed which validates advantages of distributed smart grid architecture. Result is obtained by comparing central and local processing of data. Figure 3.4 shows detailed working of smart meter implementation of proposed approach.

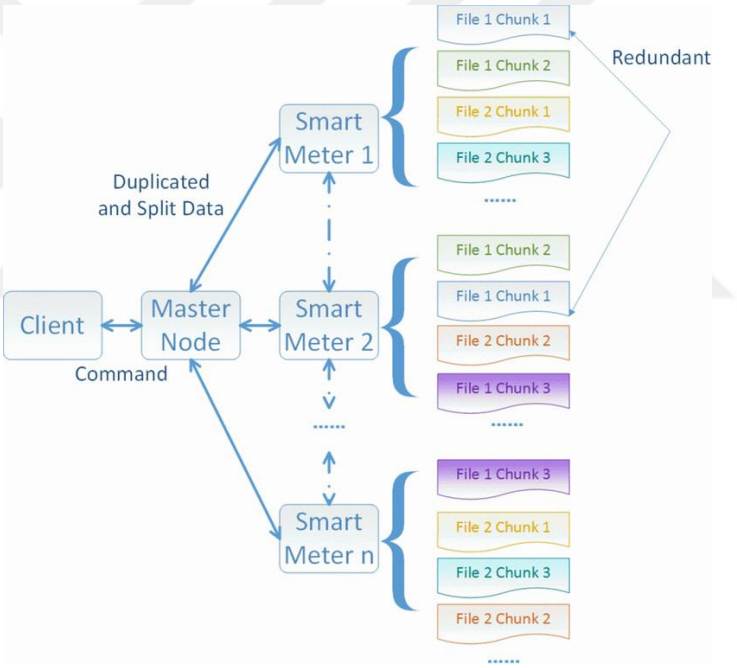


Figure 3.4 Working principle of data distribution between smart meters

Afterwards, thesis focuses on the distributed architecture. Some operations that can be performed by architecture is discussed. It is proven that the local processing of data positively effects efficiency when available network bandwidth is utilized. Furthermore, local processing is advantageous for time critical smart grid operations, since it has less delay for data communication. Finally performing energy usage prediction locally while central processor is responsible for coordination between local processors is discussed.

Uncovering useful information data collected from Smart Meters is a big data challenge. A paper deals with subject of big data to forecast the average electricity load for every hour on daily basis [26]. Adaptive Neuro Fuzzy Inference System (ANFIS) method is used to predict the load in the next 24 hours. The experimental results are calculated to be successful with the overall prediction accuracy of 84.02%.

An intelligent power system has a sizable data from various sources. Data of type social economic metrics such as geographic information, user properties, consumer behaviors are collected. This complexity makes data management a challenging issue. From another point of view, data mining and cloud computing technologies make the analysis of this complex data possible. That is the benefit energy management is improved. Some papers focuses on building energy [21]. Such a paper propose a software solution, which aims to streamline smart meter data. By this way it deals with the complexity of data processing and data analytics. At rest of document an information integration pipeline is proposed to process data. This is a scalable platform for preprocessing and collecting data sets. A web portal is implemented for visualization of results. Hybrid technologies is used in system. Big data technologies Spark and Hive are integrated to RDBMS PostgreSQL, which is a database having add-ons of machine learning toolkit. MADlib is one of used technologies, which is able to satisfy requirements.

With developing of smart metering technology huge datasets are collected by advanced sensor systems. This contributes an opportunity to enhance management of power systems. Effectivity in operation, planning and utilization of various energy sources is improved. Increasing operational uncertainty due to non-continuous sources, unlocalized load elements and the growing interconnection is a big problem to solve. Statistical model construction is fundamental for building city analysis frameworks which explores this uncertainty and real-time decision. A dissertation is delivered about data analytics in power systems [27]. It proposes composite models using dimensional reduction, clustering and parametric modelling techniques.

The proposed methods are compared to alternative high dimensional modelling techniques and shown to be feasible. Benefits of this model are shown on security assessment based on

decision trees. A finite mixture modelling framework based on consumer categorization is presented. The great flexibility of proposed framework is the key parameter to be superior to alternatives. Proposed C-vine copulas (CVMM) framework is compared to traditional methods by using real demand records across 2,613 households in a London smart-metering trial.

Four validity indicators are analyzed for comparison. A classification module is developed for partitioning new consumers which improved predictive performance of framework. Different loading conditions and different sets of large numbers of households are considered for case studies. Demand measurements across 2,613 households in London is used for dataset. Demand diversity and other metrics are collected for different customer classes to find out correlation between wealth and peak demand behavior. Furthermore another TNEP framework that accommodates multiple sources of operational stochasticity is presented. A multivariate Gaussian copula is applied to record inter-spatial dependencies between loads in various locations and output non-continuous generation components. This model is the basis of a Monte Carlo analysis framework for studying the statistical uncertainty. Splitting the investment and operation problems is conducted by Benders decomposition. By means of a case study on a IEEE 118-bus system the advantages of the model is shown.

At last in thesis transmission expansion problem scenarios to obtain an accurate solution is selected by a proposed framework. This selection process is done by operating costs and investment decisions and reduced number of states. Several clustering and classification techniques are compared to find out efficient combination for considered problem. Comparison among the clustering techniques is shown in Figure 3.5.

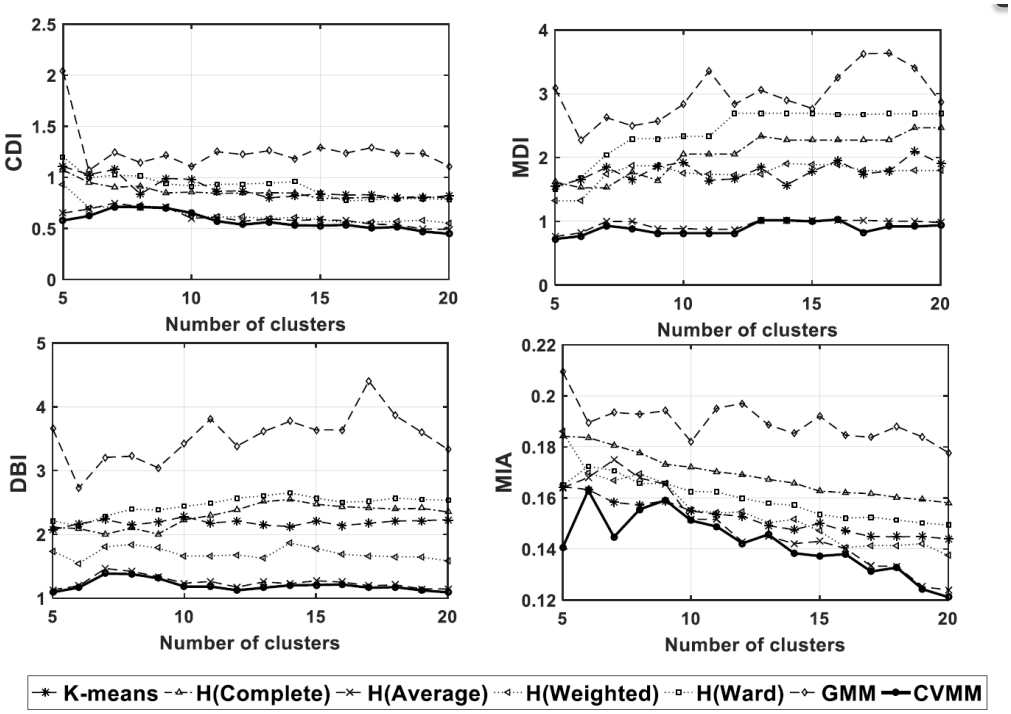


Figure 3.5 Comparison among the clustering techniques

3.2.2 Infrastructures

A cloud computing system provides different types of services. One of these services is infrastructure. This service should bw designed suitable for big data engineering and is called infrastructure as a service (IaaS) [28]. This service includes monitoring, storage, computing and networking.

3.2.3 Communication

Data sources have the potential to provide incomplete or noisy data, which increase veracity complexity. Smart meters is an example of situation which measure the load at a very local level for individual households, also they communicate it to other parts of the system. So it is said that communication might affect the veracity of the recorded data [29] .

Consequently new generation of power systems are anticipated to be supported by an extensive information and communication technology. [30] This will be one of the key factors, that gives rise to smart grid paradigm.

3.3 Smart System Characteristics

3.3.1 Real-Time

A distributed and highly scalable machine learning platform called Tensorflow has been studied for electricity price forecasting. Tensorflow is an open source platform for artificial intelligence systems [31]. The advantage of this package ability to run parallel computations on distributed systems on a scale of thousands cores. So, nowadays big data analytics for real time applications are done with this package. A case study is held considering a day-ahead price forecasting for power market of two European countries have been studied via Tensorflow. Dataset is obtained from renewable energies and conventional generators, as well as weather forecasts. These factors affects day-ahead price of the market. In the results, improvement in the convergence, speed and prediction accuracy is seen.

3.3.2 Secure & Reliable

Not only control and execution of the power systems, protection of them is a significant task as well [32]. Authors pointed to the deficiency of the present protection and control system to handle a real-time data of interaction between the substations and the wide area protection and control system. A new concept of integrated wide area protection and control is introduced. According to the concept a hierarchical protection and control system provides the protection and control for wide area or regional power substations/plants and their associated power networks. The system is mainly divided into three levels: the local, the substation/plant, and the wide area/regional. Furthermore, the development of information technology has resulted in the interests in utilizing cloud computing and big data techniques, as shown in Figure 3.6, to improve the performance of power system protection and control.

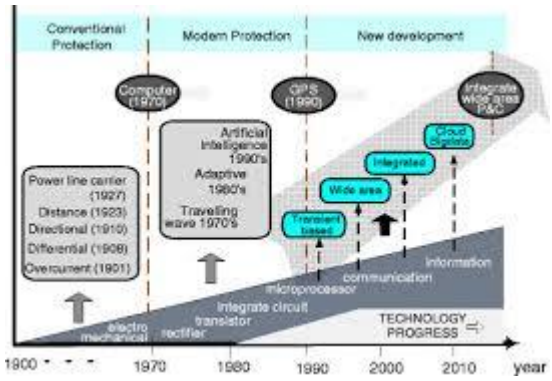


Figure 3.6 History of power system protection

The main challenges of smart grids is managing different types of front-end intelligent devices efficiently and processing a huge amount of data received from these devices. A paper is to propose a cloud computing based framework for big data processing. At the same time a security solution is designed which has identity-based encryption, signature and proxy re-encryption to address critical security issues of framework [33].

Power system data has some features. Some of these are large quantity, complicated data item, complex processing logic, long storage cycle and high frequency calculation. Big data is generated by cyber devices and physical components. This is a cyber-physical system (CPS). This type of power system is facing intrusions which can be more damaging, complicated and wide spreading. Most power systems use manual network intrusion detection systems. In this case, the detection knowledge is provided by a security expert and applied into the network intrusion detection system (IDS). However, detection systems needs to be continuous and knowledge concerning the intrusion detection should be up to date. A paper proposed a hierarchic IDS that integrates misuse detection to abnormal detection for Power System [34]. Using data mining algorithms, rules are defined. Rules come out by studying and analyzing historical monitor date. A prototype implemented to prove the model. It is seen that model can detect cyber-attacks accurately with low false positive and false negative rate. Figure 3.7 shows the modal.

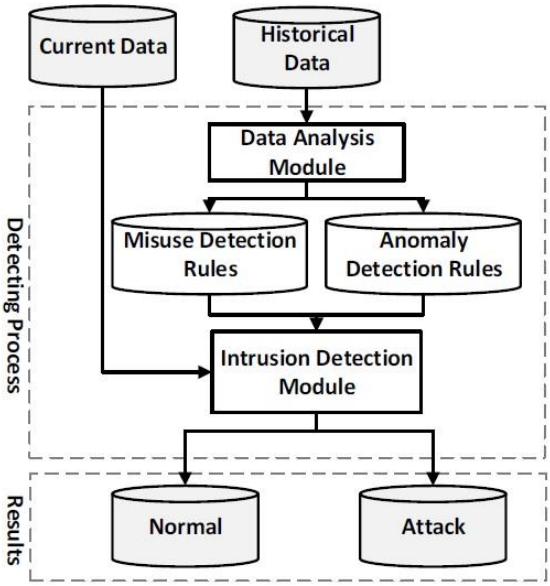


Figure 3.7 Intrusion detection model achitecture

Because of both increase in extreme weather conditions and deterioration of old infrastructure, electricity network cut become a more usual event. 75% of power outages caused by these. Whereas renewable energy generation increased, their intermittent nature forms another problem also. A chapter about subject wants to find an answer to the question of how such big data can be integrated, and analyzed in real-time, in order to improve reliability of modern electricity network in dealing with weather caused emergencies [35].

To detect the potential faults of the transformer fault diagnosis methods are used. Dissolved Gas Analysis (DGA) is one of but most significant of them. In view of DGA big data's characteristics, a new combined fault diagnosis method for transformer is proposed in a paper [36]. Support vector machine is used to make an extra diagnosis. The method overcomes the deficiency of single diagnosis model in transformer fault diagnosis, and improves the diagnostic accuracy and the scope of application of the model. The training and deployment strategy of the combined diagnosis model is designed based on Storm and Spark platform.

Critical infrastructures are highly interdependent due to both consumed services and provided services. Unlike previous studies a paper is focusing on cyber interdependency. Depending on the information transmitted through, infrastructures have cyber interdependency. The

communication network is the backbone of smart energy systems. Data flow in energy systems is shown in Figure 3.8. This network is responsible for transmission of data of all substations in both directions to utility side and to client side. In a smart system, data is generated at different levels because of having been consisting of complex sub systems. The data at each layer depends on the information transmitted through the information infrastructure. The data from different layer of system has its own cyber characteristics. These characteristics and interdependencies at each layer gives a glue of designing an appropriate ICT architecture that fits that segment (layer) rather than having an generic ICT architecture. A paper outlines a new approach by focusing only on the cyber interdependencies in smart energy systems is published [37].

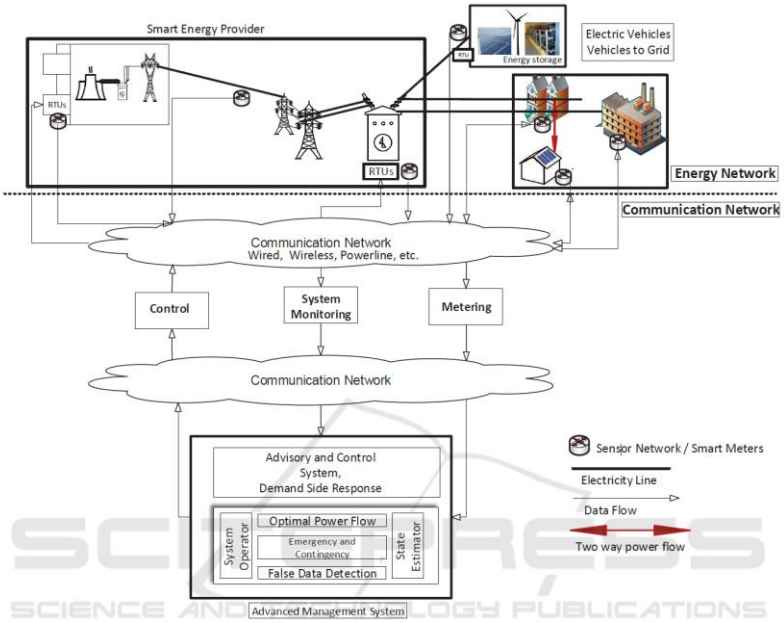


Figure 3.8 Energy systems & data flow

A publication mapped online power system transient stability assessment (TSA) problem as a two-class classification problem and a data mining algorithm [38]. To solve the problem based on phasor measurement units (PMUs) big data the core vector machine (CVM) is proposed. First offline training of system is held, an online application framework is proposed for which 24 features are selected to present the system status. After the PMU big data is generated by time domain simulation a CVM model is trained and tested. The evaluation

indices are calculated and compared with other support vector machines. Proposed CVM based assessment algorithm has the higher precision. The speed and effectiveness of the proposed algorithm is proved by case studies on the IEEE New England 39-bus system, and real systems in China and the U.S.

3.3.3 Manageable & Visualizable

A big data visualization platform is designed to discover the hidden useful knowledge for smart grid (SG) operation, which is still a challenging issue [39]. The Apache Spark, an open source cluster computing framework, is used to process the big-data is a subject of some studies. In such a study a high-speed communication architecture is used. The model is designed with Open System Interconnection (OSI) to transmit the data to a visualization platform which uses Google Earth and a geographic information system (GIS). A link is setup between SG data and geological information of visualization. As test bench The University of Denver's campus grid is used to present platform.

A paper investigates how to set up a correlation between big data from the lightning surge and the traveling wave measurements setup [40]. Wave measurement factors are time and space. This correlation will improve fault location accuracy. The integration of big data to time, space pair is achieved by Global Positioning System and Geographic Information System. A system is proposed which has a pair of main benefits. These are;

- The decision process can be accelerated through automation
- Better accuracy of fault location result can be provided due to the data correlation.

Phasor Measurement Units (PMUs) have high sampling rates. Hence their deployment in power grids becomes common. They offer accurate visibility of the power grids than classical SCADA systems. Beneficial high sampling rates however causes two major challenges that should be handled for effectivity. These challenges are;

- Any transient events captured in the PMU measurements can negatively impact the performance of steady state analysis.

- Processing the high volumes of PMU data in a timely manner poses another challenge in computation.

A parallel detrended fluctuation analysis (PDFA) approach is proposed to be a solution to these problems [41]. Evaluation of proposed approach performance is done from the aspects of speedup, scalability and accuracy. It is also compared to standalone DFA approach.

Building heat demand is the leading reason for need of sharing total global final energy consumption. Building stock models using high spatio-temporal resolution will be a powerful solution tool to investigate the effects of new building policies. Usage of these models aimed to increase energy efficiency, to develop of new heating technologies, to integrate buildings with renewable energy sources. A paper proposes a method integrated with a geographic information system (GIS). This method stands to model building heat demand of large regions with a high temporal resolution [42]. Considered approach allows to derive the envelope of all buildings from digital elevation models. It also models location dependent influences such as shadowing and climate conditions. Spatio-temporal climate data for temperature and solar radiation is integrated to model the effects of climate changes. The model is validated with data from 1845 buildings from a city St. Gallen, Switzerland and 120 buildings from Zernez, Switzerland. The proposed model is able to satisfy targets. The validation resulted in an average goodness of fit (R2) of 0.6.

A dissertation about renewable energy integration in distribution systems, their synchrophasor sensor based big data analysis and their visualization is established [43]. The hybrid smart grid (SG) operation is mainly featured by distributed renewable generation, data visualization, data prediction in high accuracy and operation cost minimization. Both by the customer and the grid side has large volume of heterogeneous data provided. Hence a big data visualization platform should be built to discover the hidden useful knowledge for smart grid (SG) operation, control and situation awareness. An Apache Spark , which is an open source cluster computing framework is considered to discover the hidden knowledge of the big data. Between database and data visualization platform data is transmitted with a Open System Interconnection (OSI) model in a high-speed communication architecture. Google Earth and global geographic information system (GIS) are used to design the visualization platform. A

short-term load forecasting approach is used to provide a high accuracy load forecasting for the network reconfiguration. Design is based on Support Vector Regression (SVR) technique. The three-phase balanced optimal power flow is converted to an Optimal Power Flow (OPF) problem first. Network reconfiguration is done in parallel manner with the limited switches and the strong computing capability. The system cost with renewable generation is reduced at the substation level. The given error distribution model of renewable generation is simulated with a chance constraint, and the derived deterministic form is modeled with Gaussian mixture model (GMM) with genetic algorithm-based expectation maximization (GAEM). The system cost is further reduced with the OPF in real-time (RT) scheduling. The semi definite programming (SDP) is used to relax the nonconvexity of the three-phase unbalanced distribution system into a convex problem, which make sure to achieve the global optimal result. In the parallel manner, the ADMM is realizing getting the results in a short time. Architecture of platform is shown in Figure 3.9.

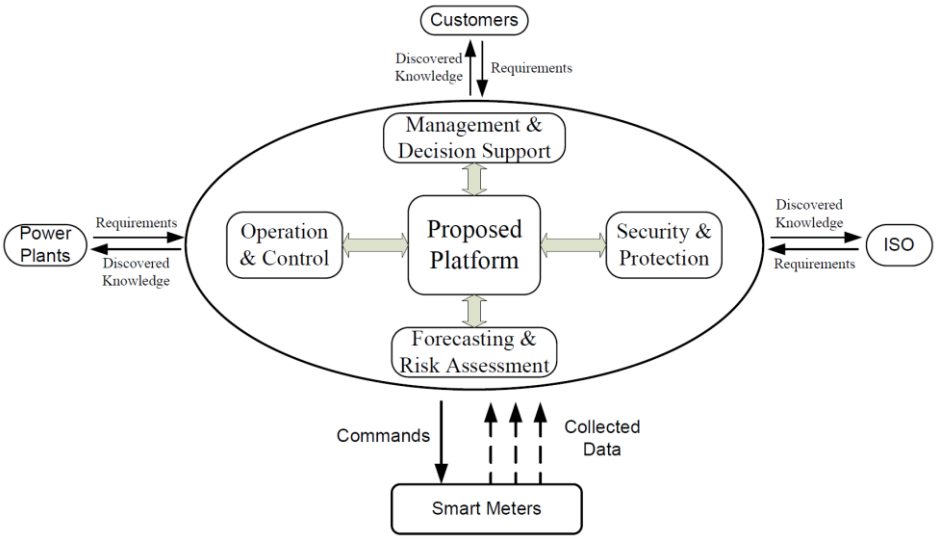


Figure 3.9 Architecture of big data platform

Sustainable energy and financial market stability wicked problems are societal challenges that arise from complex socio-technical systems. Understanding and reducing them requires research methods by inquiry of Information Systems (IS) “individuals, organizations, and markets” and that deliver solutions in addition to insights. An approach is presented to address

these challenges through Competitive Benchmarking (CB) [44]. The CB platform presented generates data that can be used to improve operational strategies and judge the effectiveness of regulatory regimes and policies. Experiments conducted applying CB to the sustainable energy challenge in the Power Trading Agent Competition (Power TAC). More than a dozen research included.

A paper about a work on power equipment monitoring is in literature [45]. Work done finds that random matrix theory is suitable for modeling the massive data sets in this situation. The statistical correlation is extracted between key states and parameters. Entries of large dimensional random matrices are modeled as time series. By eigenvalue spectrum distribution and the single ring law are derived from these entries. The single ring law used as statistical analytics to characterize the measured data. The evaluation of key state and anomaly detection are accomplished through the comparison of those statistical analytics.

Distributed energy resources of monitoring systems with high frequency synchronized wide area awareness systems are generating immense volume of energy data. However still there is a question whether the structures and the tools being developed for power system data are indeed suitable for big data science. In the domain of power systems an assessment of the distinguished aspects in big data analytics developments is provided [46]. Identification and classifications of the existing and the missing elements in the structures and methods associated with big data analytics in power systems are given.

3.3.4 Ecologically Sensitive & Clean

Nowadays, concept of green energy is widely discussed. Hence solar power generation is a must-be trend and approach. For the countries with nature resource shortage the importance of the subject is higher. However, it is still a crucial issue that how solar power plants to build with the best power generation efficiency in limited spaces. A paper proposing the approach of finding the optimum models of generating solar power to build solar power plants for different environments in Taiwan is established [47]. Existing solar power farms gives some glues about subject including (i) design methods of power generation, (ii) actual power generation, and (iii) surrounding environments. They collect data on these subjects. After a series of preprocessing steps and system analysis, the optimal models of generating solar

power is identified. Results show that, we observe that using the majority voting strategy improves the system accuracy and helps engineers build solar power plants with the maximum power generation.

Around 3% of global carbon emissions are caused by shipping which is heavily regulated industry [48]. Now the industry is expected to turn off due to different situations like upcoming regulations, climate change, energy shortages and technological revolutions. Ships generate a large volume of data from different sources and in different formats. So big data concept become in talk of the industry nowadays. Big data analysis is used to discover correlations between parameters to extract hidden patterns and trends. By this way, big data gives decision-making support to the ship operator. It will also increase the capability of performance monitoring in maritime industry. However, there are some challenges that industry will have to face such as data processing, reliability, and data security. Ship data should be regulated for Monitoring, Reporting and Verification (MRV) to quantify the CO2 emissions for ships above 5000 gross tonnage. The MRV is a data-oriented regulation requiring ship operators to capture and monitor the ship emissions for the International Maritime Organization (IMO).

Production and consumption of energy negatively impacts the environment Searches on cleaner production should be done. A study investigates sustainable energy consumption in the industrial and mining sectors [49]. Although these sectors are experiencing increasing data growth, adoption of systems to big data technologies is not enough. Enhance modularity and extensibility of using micro services and containers, big data methods were applied. Polyglot persistence, Data lakes and Lambda architecture is integrated to form a hybrid architecture. MongoDB is the primary data to store records of an engineering services company. Seven case studies are obtained, which illustrate the impact of big data on decision-making. These improved decisions and initiated suitable actions to achieve consumption reductions. The overall impact of the case studies was an energy saving of 75 622 MWh, a reduction of 74 111Mg of carbon-dioxide-emission, and a saving of 108 141m3 of water.

China revised its energy management by increase of energy utilization efficiency, the improvement of energy efficiency level, and the reduction of pollutant emission. The concept

of Energy Internet (EI) is supported by series of energy technologies. By help of these technologies system is developing more diversified, cleaner, and highly efficient direction. With parallel network of distributed energy and implementation application of various smart watches and energy. As other big data based systems workload of EI on data information processing is increasingly growing. This paper focuses on EI and big data (BD) based on data mining techniques and model of electric vehicles (EV's) charging and discharging data analysis [50].

Specialists on subject say that Construction of Global Energy Interconnection (GEI) is an effective way to utilize clean energy. As one of the enabling technologies for GEI, big data is accompanied with the comprehensive application of energy related data all over the world. The technology innovation direction of GEI has become subject of some papers [51]. Advantages of big data technologies in supporting GEI development is discussed. Some typical application scenarios given to illustrate the application value of big data. Finally, the architecture for applying random matrix theory in GEI is presented.

3.3.5 Technological Assets of a Smart System

A new power grid fault trace method based on big data platform is developed [52]. Spark technology is used to deal with fault date after extension of data source to transformer substation. This method is used to solve the problem of mass data processing. Tracing the malfunction of protection or circuit breaker and their reason is identified by decision tree. Compared with the traditional fault diagnosis systems, this method can use every monitoring data in substation and give the reason of fault. Big data platform of power grid fault is shown in Figure 3.10.

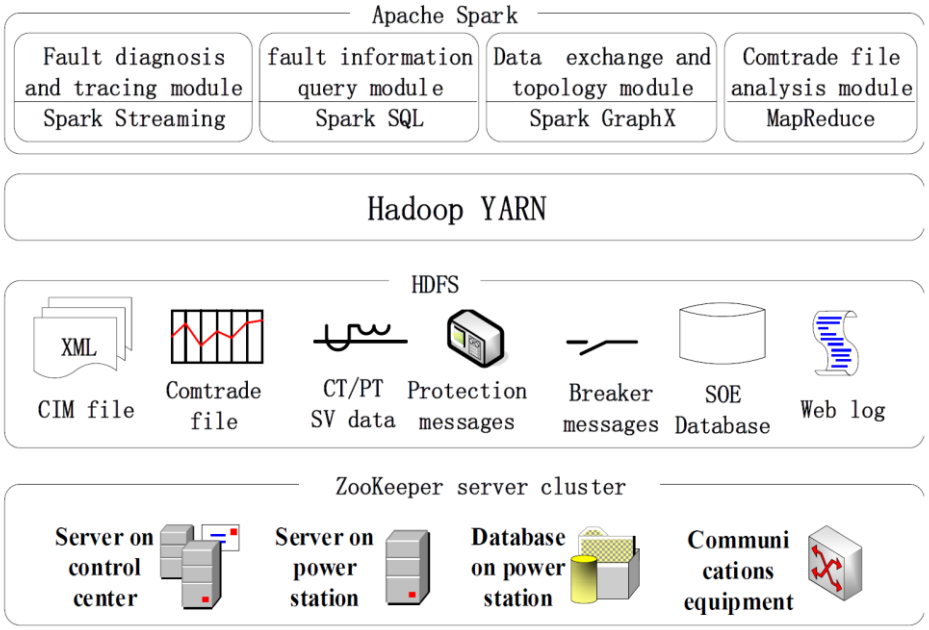


Figure 3.10 Big data platform of power grid fault

The design idea of wireless big data platform is discussed [53]. Data acquisition, data transmission, data network construction and business coordination operation mechanisms of big data system are the main aspects of this discussion. After some examples of the monitoring business of power grid is given, simulations held. The simulation experiment shows that the choose of the relevant feature variables can support the need of the whole wireless big data network to monitor the power system with high precision.

In a smart grid large scale data are collected by sensors and measurement devices. Smart meters can record fine-grained information about electricity consumption in near real-time. This forms a smart meter big data which has provided new opportunities for electric load forecasting, anomaly detection, and demand side management. However, the high-dimensional and massive smart meter big data creates great pressure on data transmission lines and cause enormous storage costs. Therefore, to reduce the transmission pressure and storage overhead, efficiency should be improved to optimize potential of smart meter big data. A study presents on the compression techniques for smart meter big data [54]. The development of smart grids and the characteristics and application challenges of electric power big data are first introduced, followed by analysis of the characteristics and benefits of

smart meter big data. Finally, the potential data compression methods for smart meter big data is discussed, and the evaluation methods for smart meter big data compression are considered.

Works on identification of potential performance difficulties of frameworks about processing of large datasets continues academically. However, most of the existing works generally focus only on execution time and resource utilization. In fact, analyzing some other important metrics is key to fully understanding the behavior of these frameworks. Giving an example, microarchitecture-level events can bring meaningful insights to characterize the interaction between frameworks and hardware. Also increasing attention paid to energy consumption as systems scale to thousands of cores. A discussion held on the current state of the art in evaluating distributed processing frameworks [55]. This work based on Big Data Evaluator Tool (BDEv) to extract energy efficiency and microarchitecture-level metrics from the execution of representative big data workloads. Final experimental evaluations on popular frameworks such as Hadoop, Spark and Flink using BDEv demonstrates its usefulness to bring meaningful information.

Challenges in power transmission grid occurs as data sizes have increased rapidly. This data is large in volume, generated fast and recorded in different format, and come from different sources such as electrical substations. Traditional relational databases are not sufficient in response time and have negative impact on performance. To overcome this inadequatancy, the big data implementations are developed new technologies such as NoSQL data stores. A research paper has a goal of improving this execution by modeling and processing those data using Neo4j database [56]. An architecture with two power transformers is modeled and implemented. According to the requirement of business, a new power transformer is added to simulate features got from Neo4j database.

Rapid growth of big data analytics in scientific applications and online services, increased popularity of MapReduce and its open-source implementation Hadoop. Hadoop is a highly feasible solution for building big data analytics platforms. Some lacks of it are recorded also such as data management, resource management, scheduling policies. These usually cause high energy consumption in MapReduce jobs in Hadoop clusters. A paper is published to

review the studies on improving energy efficiency of Hadoop clusters [57]. They are summarized in five categories;

- energy-aware cluster node management
- energy-aware data management
- energy-aware resource allocation
- energy-aware task scheduling
- other energy-saving schemes

Work is aimed to inspire the relevant research on improving the energy efficiency of Hadoop in big data analytics.

In the scope of power system Phasor Measurement Units (PMUs) are valuable devices. They are implemented on buses of the power systems and responsible for collecting situational information. On the other hand, term “synchrophasor data” stands for data produced by PMUs which are synchronized by GPS timing and involving plenty of information about the power system. Generally synchrophasor data is used offline in various areas such as model validation and post-event analysis. A thesis published for contribution by proposing a different approach in synchrophasor data analytics [58]. Unlike previous studies in which synchrophasor data is considered as extra sensor data, this thesis accepts data as a type of big data and utilizes some methods to find solutions.

CHAPTER 4

BIG DATA APPLICATIONS

4.1 Smart Grid

A typical power system has three stages. The generation stage, the transmission stage and the distribution stage. These stages for typical electrical power system are illustrated in Figure 4.1.

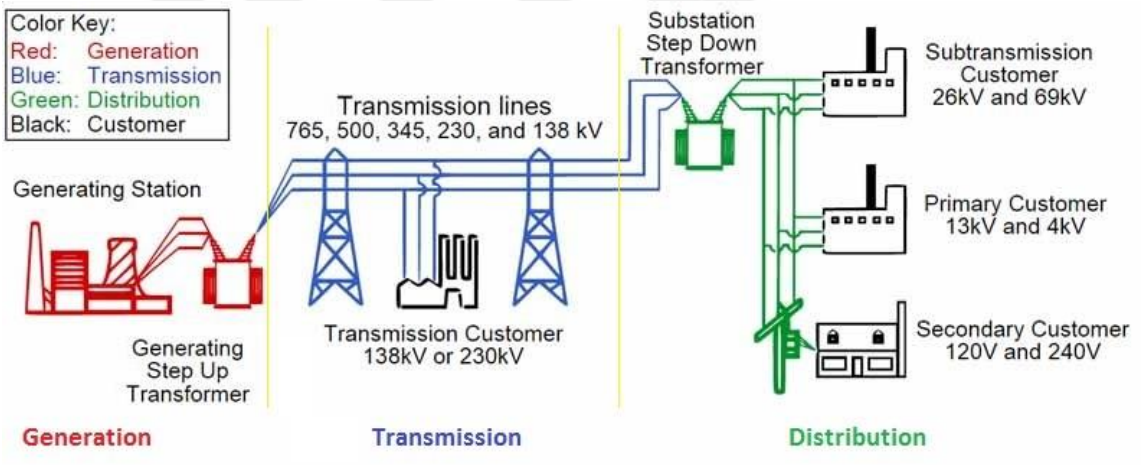


Figure 4.1 Three stages of an electrical power system

4.1.1 Generation

The generation is either done by a power station or from renewable energy farms. The voltage is generated at 25kV and then goes through a step up transformer to increase the voltage of the power generated.

The oil and gas industry collects massive datasets for many years, but “Big Data” is a new concept that has the potential to significantly re-shape the industry. Despite the huge amount

of data, data collected within the oil and gas sector is not appropriate for big data analytics. A paper examines previous data management practices in oil and gas industry and compares them to practices in area of philosophies [59]. The comparison shows that, unlike in companies that are widely considered to be leaders in big data analytics, data is not regarded as a value in companies of the oil and gas industry. In these companies data is regarded as descriptive information about a physical asset. Then it is discussed how industry could potentially extract more value from data. Finally document ends with a series of policy related questions to be answered in future works.

An essential but uncertain factor in the oil supply chain is the oil consumption. Estimating uncertain factors is a result that big data studies want to arrive at. Google trends are used for an online big data-driven oil consumption forecasting model [60]. This proposed model is executed in two steps namely, relationship investigation and prediction improvement. Predictive power of Google trends are tested by cointegration tests and a Granger causality analysis is executed. Tests showed that proposed online big-data-driven forecasting system works with Google trends improved prediction capacity.

Does method of big data statistics bring opportunities and prospects for the power-to-gas (PtG) device in the energy system? Also answer of this question is analyzed [61]. Regional difference, seasonal difference and peak-valley difference are the three scenarios for PtG application. There is a relation between the comfort degree and temperature range. To analyze correlation load demand is used and statistical results are analyzed.

A study is performed to identify the accuracy of the HDM-4 fuel consumption model calibrated for the United Kingdom [62]. Study also considers whether a further calibration is needed or not heavy good vehicles (HGV) are focused during work and estimates compared with estimates made by HDM-4 to measurements from a large fleet of vehicles driving on motorways in England. The data was obtained from the telematic database of truck fleet managers. 19,991 records from 1645 trucks are included. These records represent trucks driving at constant speed which is 85 km/h along two certain motorways. These conditions have been simulated in HDM-4. Estimates are compared to real measurements under two

separate sets of assumptions. As a result it is seen that the current calibration of HDM-4 for the United Kingdom already requires a recalibration.

Nuclear accidents changed acceptance of people toward nuclear energy technologies negatively. For this reason, according to some countries that employ nuclear energy as their key energy resource public attitudes toward nuclear energy facilities is an important factor in national energy and electricity policymaking. However, it is a big challenge to explore such an attitude, since most of previous studies analyzed public attitude with limited number of participants in polls. A study suggests a big data analyzing method with K-LIWC (Korean-Linguistic Inquiry and Word Count), investigates public attitudes, positive and negative opinion about nuclear energy [63]. Investigation performed by means of the collected data sets, social media and network services in Korea. Results show that accidental events related to nuclear energy have at least temporary may be consistent effects on the attitude. Event type is also effective on subject. After results, inspirations, difficulties and future research areas are presented. Figure 4.2 presents summary of the results.

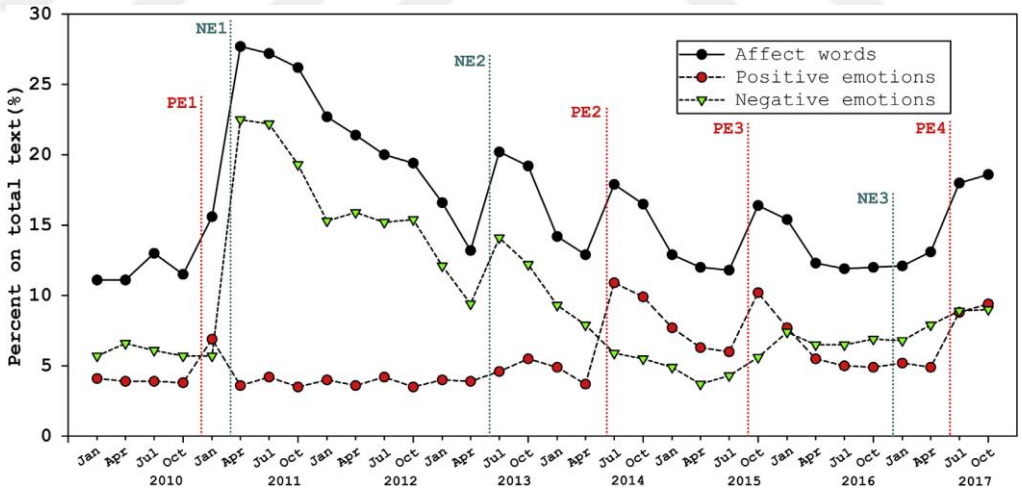


Figure 4.2 Summary of K-LIWC on the collected data

Since consensus is required to drive a nuclear strategy, public acceptance of nuclear power is important for the government. Unlike past researches big data analysis is proposed as an attractive alternative and attempted to identify the attitudes of the public on nuclear power [64]. SNS (Social Networking Services) is used to analyze consumer opinions. The keyword

analysis identified the attitudes of the public toward nuclear power. Results are stated at the end of document. The common public opinion for nuclear power was positive when Korea exported nuclear reactors to the United Arab Emirates. Events like Fukushima accident in 2011 and certain supplier scandals in 2012, degraded public acceptance.

Renewable energies are also known to be clean energy. Especially photovoltaic, have recently become popular and important, since it has benefits for environment, economy and energy security. Effective management and usage efficiency improvement is a subject of a study [65]. Wireless sensor networks (WSNs) are utilized, which efficiently deliver the monitoring data photovoltaic (PV) modules to visualization center. Major goal was detecting the problems of PV modules. This is performed by monitoring big data. By means of a data fusion method the monitoring data at sensor nodes of WSNs are integrated. An innovative semi-supervised Support vector machine (SVM) classifier is designed. Trained by samples solar big data. At this step, trained classifier becomes the prediction model. A detection algorithm is proposed to classify problems of PV modules. This algorithm is to calculate the average value of the data. Success of methods calculate by experiments conducted on a test platform. Results show that the predicted values supports theoretical value of power generation.

Renewable energy generation and management options requires fine-tuning and accurate control to be successful. This brings the need of big-data. In a study UK and its Multi-Criteria Decision Making (MCDM) focused [66]. Data incompleteness and redundancy are handled with using a comprehensive grey TOPSIS model. Evaluation scores used to develop a multi-objective optimization model. Demand uncertainty was handled with an interval range and problem is developed with multi-objective grey linear programming (MOGLP). Developed model is shown in Figure 4.3. This model give successful results for nuclear and renewable energy options in the power generation mix.

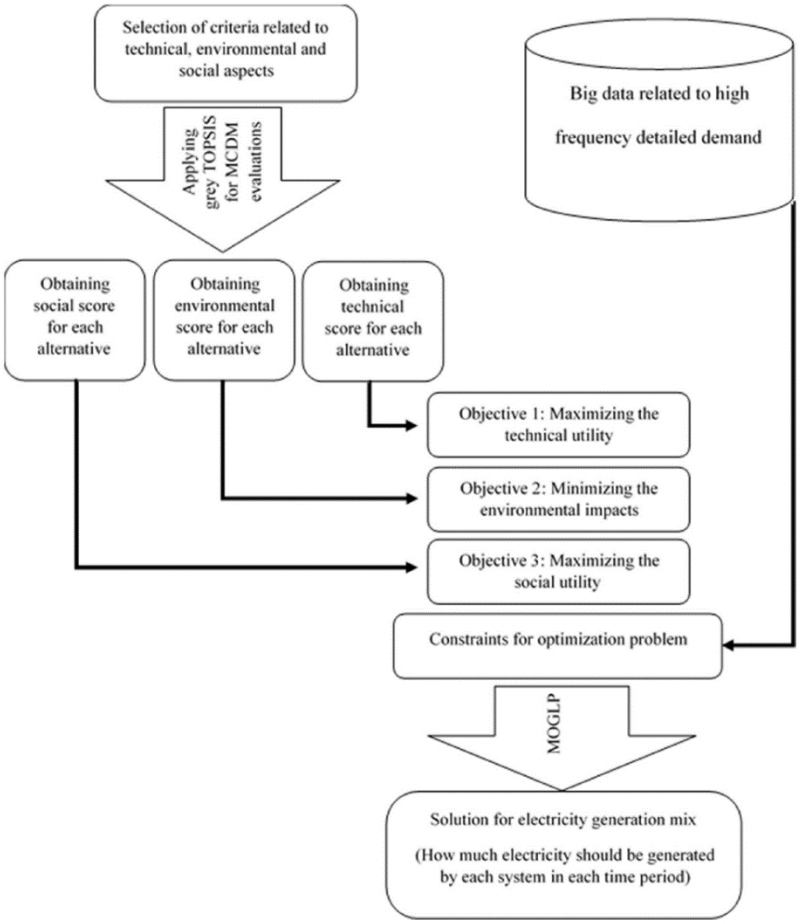


Figure 4.3 Solution for obtaining the optimal electricity generation mix.

Ever increasing human population brings, increasing demand for electricity, consequently it brings the need for modernized equipment and operations. This revolution made grid to incorporate Information and Communication Technologies (ICT) inevitable.

"The grid" alone stands for a network of transmission lines, substations, transformers and more. A grid is used to deliver electricity from power plant to clients. This delivery is performed by transmission lines. Smart grid is a grid which senses values along these transmission lines by means of smart meters and manages two way communication between utility and consumers by means of digital technology.

A smart grid is an electricity network. Unlike classical networks by using digital and other advanced technologies it enhances monitoring and managing electricity transport from

various sources to meet different electricity demands of end-users. Demand response capacity is a concept that smart grid should have, which is critical to provide balanced electrical consumption with supply [33]. There are economic and environmental necessities for the renovation of the conventional power systems, for its replacement with a Smart Electrical Power Grid which is a Smart Grid.

Smart grid enhances reliability, availability, and efficiency in distribution systems. During transmission of energy, carry out testing, technology improvements, consumer education, development of standards and regulations, and information sharing is critical.

Uninterrupted power supply, central management, integrated renewable energy resource as power supply, bidirectional electrical grid, consumption data driven pricing, giving change to grid operator to manage network, tracking power consumption data and analyzing history are some advantages of smart grids.

Benefits associated with the Smart Grid listed as following [67]

- More efficient transmission of electricity
- Quicker restoration of electricity after power disturbances
- Reduced operations and management costs for utilities, and ultimately lower power costs for consumers
- Reduced peak demand, which will also help lower electricity rates
- Increased integration of large-scale renewable energy systems
- Better integration of customer-owner power generation systems, including renewable energy systems
- Improved security

A smart grid not only collects data both from customer and utility side, and also analyze historical and immediate data to perform multi-dimensional and multi-level management to adopt changing demands of consumption.

Smart grid also claimed to be automatic for some of its basic jobs like coordination, awareness, healing and reconfiguration. Additionally, integrating renewable energy resources, enhancing power generation efficiency, transmission and usage can be counted.

Figure 4.4 shows smart grid model updated by NIST.

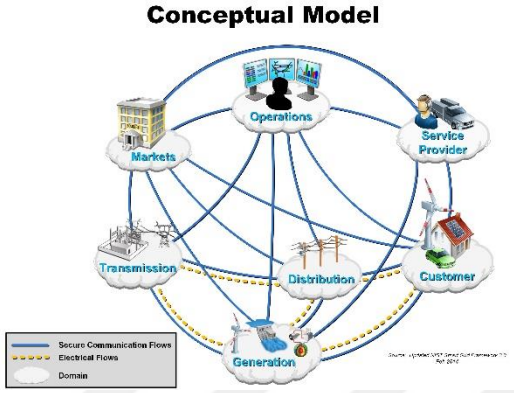


Figure 4.4 A typical smart grid model

Security is another concept that a smart grid should satisfy. Physical and cyber-attacks should be foreseen or response times to these attacks should be shorten as much as possible.

The major goal of big data techniques in smart grid is to support complementary operation of multiform energy system in order to achieve energy cascading utilization and boosts the energy efficiency [68]. Combined Cooling, Heating and Power systems (CCHP) is the main component of a multiform energy system including gas turbines, steam turbines, micro turbines and reciprocating internal combustion engines integrated with thermally activated cooling technologies.

Recently designed CCHP systems are examples of data driven models. They were developed as a supply way which is economical and reliable [69]. Since optimization and economic evaluation is critical for deployment of a CCHP system, new researches are being executed on its design, control, operation and planning.

A data-driven method is better for operating such a multiform energy system, when uncertainty on price and demand is considered. At this point power grids plays a basic role for collecting large amount of data from power systems. Smart grids are designed to contain huge number of intelligent devices which are connected to each other without human intervention.

Then only analyzing of data is left to optimize a system. Figure 4.5 shows a CCHP system controlled by a smart grid.

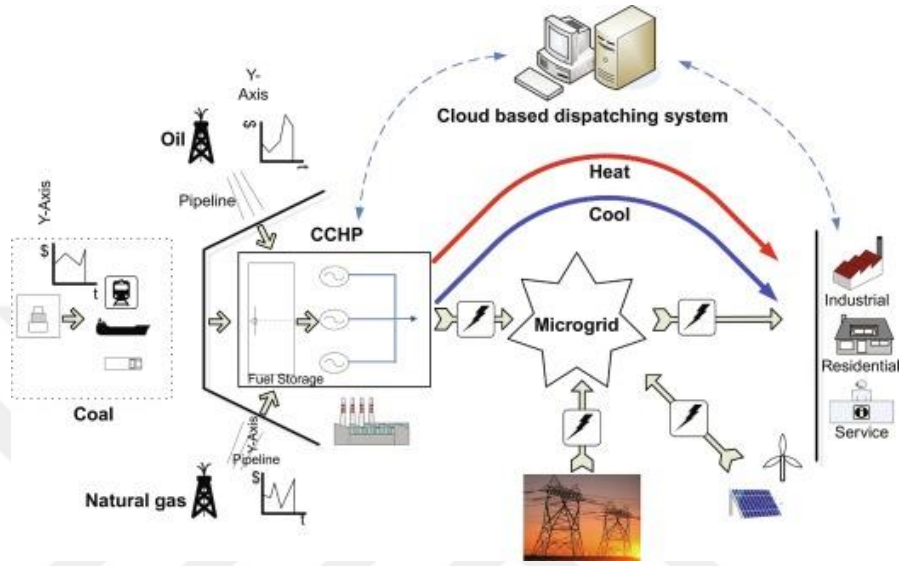


Figure 4.5 A Smart grid controlled CCHP system

As seen the multiform energy system employs more than one thermo-dynamic cycle. The outstanding electricity or energy demand is compensated by power grid.

From Utility, industry, and research points of view presents a review of big data analytics and its applications in smart grids, and also identifies challenges and opportunities. For future research gaps and insights are considered on directions to integrate big data analytics into power systems [70].

An architecture using Random matrix theory (RMT) is proposed to motivate data-driven tools to perceive the complex grids in high-dimension [71]. The architecture performed a high-dimensional analysis and compares the findings with RMT predictions to conduct anomaly detections. To reflect the correlations of system data in different dimensions Mean spectral radius (MSR), as a statistical indicator, is defined. A management mode perspective is discussed to make advanced big data analyses possible. It is proved that proposed architecture, as a data-driven solution, is sensitive to system situation awareness, and practical for real large

scale interconnected systems. Validation of the designed architecture in various fields of power systems accomplished by case studies.

Smart grids bring deeper changes in the information systems that drive them. A multi-player, multi-dimensional system is in consideration. Information flows from the electricity grid, new age producers of renewable energies. Usage of electric vehicles and connected houses. New communicating equipment like smart meters, sensors and remote control points. This will cause a flow of excess data that the energy companies will have to face. Big data technologies offers suitable solutions for utilities. Some studies [72] provide an overview of data management for smart grids, summarize the added value of big data technologies for this kind of data, and discuss the technical requirements, the tools and the main steps to implement big data solutions in the smart grid context.

Big data collected by power grids are supposed to be enormously big and heterogeneous. Because measurement and communication technologies evolved in power grid technologies. In particular, computational complexity, data security, and operational integration is still key challenges of big data integration into power system planning. Business is that to convert this heterogeneous large dataset into actionable outcomes. In one of the reviews done upon this subject is a comprehensive state-of-the-art review of big data analytics and its applications in power grids [73]. Challenges and opportunities are identified from utility, industry, and research points of view. The answer of the question “how utilities can enhance revenue streams and bring disruptive innovation” is tried to be answered.

There are examples of how big data used to monitor complex power system situations in smart grid also [74]. After reviewing the key technologies of big data management and intelligent machine learning methods for complex power systems, in this study several challenges are summarized to unlock the potential of big data technology in the application of smart grid. There is a proposal about a modified and optimized structure of the big data processing platform. Analytic engine of structure consists of numerous open-sourced big data analytical tools and software. Self-developed advanced algorithms are also included. The framework suggested comprises a data interface, a big data management, analytic engine as well as the

applications, and display module. Three major applications are introduced for investigation of structure:

(i) Development of power grid topology and parallel computing using CIM files
(ii) High-efficiency load-shedding calculation
(iii) Power system transmission line tripping analysis using 3D visualization.

Large number of sensors is embedded in a power grid to form a smart grid. Sensors play key role in a smart grid for controlling and monitoring purposes. The data recorded by these sensors have all the characteristics of big data. By means of computation analytics Smart-grid can become more intelligent by processing and extracting useful knowledge from big data in real time. Apache spark is presented in literature as a platform which uses unified cluster computing [75]. This platform is suitable for implementing big data computation techniques which enables platform for automatic demand response and real time pricing. Apache Spark framework is shown in Figure 4.6.

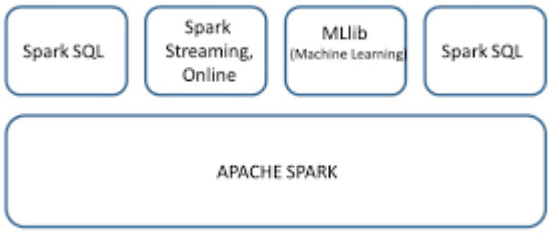


Figure 4.6 Apache spark framework

Because of its advanced sensing infrastructure, Smart Grid has a huge dataset. Such data addressed characteristics of big data, and have a time-series notion as measurements should be made over a time interval. For Demand Response programs it is needed to identify the right customers. For this operation time-series data is valuable for data mining and analytics. Because of their high dimensionality, time series are challenging to mine. To be a solution to problem a paper considers a real application from the smart grid domain [76]. Representations

of time-series data for big data analytics is explored and a proposal is done with clustering method. The aim was to determine natural segmentation of customers and identification of temporal consumption patterns. Proposed method is claimed to be extensible to large-scale, real scenarios. Evaluation is done using real datasets from smart meters. Approximate number of data points was 18,200,000 which is sufficient to show the success of technique.

Smart Grids (SGs) are used to enhance energy efficiency of power system. Classical grids distribute relevant information in real-time mode to different SG components. The Advanced Metering Infrastructure (AMI) of SG has a central component consists of a Control Panel which smart meters communicate data towards. These communicated data forms the big data. Next need is a high computing performance power for processing and mining. A paper deals with interface between SGs and the Cloud. Use of Cloud computing is proposed to provide HPCaaS for SG big data processing [77]. A suitable architecture is designed and the blue print for deploying a real world private cloud testbed is presented. OpenStack is used in testbed. Hadoop, and the MapReduce are included in programming model. Many experiments conducted using big data sets.

Main cause of global warming caused heavily by human energy consumption. Because of urbanization, among all consumption areas building sector is a major human energy consumer. A framework is presented for smart grid big data analytics [78]. Framework is to meet requirements of an energy-saving decision-support system. Layered architecture is preferred. Smart grid, a data collection layer, an analytics bench, and a web-based portal are all layers of system. Web based portal is shown in Figure 4.7. An infrastructure was installed in a building to achieve smart metering for a residential area. Experiments conducted for evaluating the effectiveness of framework. A hybrid system is designed by combining heuristic forecast system and a dynamic optimization algorithm for achieving accurate prediction and optimization of energy consumption. For contribution energy-saving decision process is presented. The framework is thought to be a basis for future full-scale smart decision support systems.

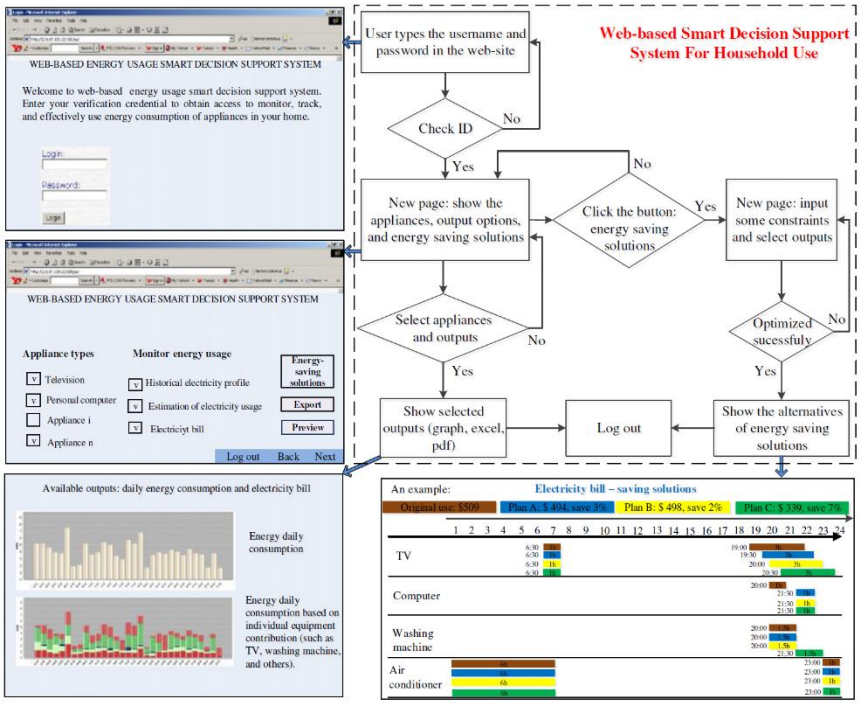


Figure 4.7 Web based portal

As an analysis document of a study a paper published. A platform for big data management system is proposed [79]. The framework consists of four subsystems;

- Acquisition subsystem
- Analysis subsystem
- Decision-making assistance subsystem
- Information integration subsystem

For the big data management system, a novel structure is designed according to three different data resources, including database, data files and data stream. For big data analyzing open-source computation algorithms and self-developed intelligent methods are integrated. These methods, namely Principal Component Analysis (PCA), Fast Recursive Algorithm (FRA), Teaching-Learning Based Optimization (TLBO) are integrated into a self-developed analysis. Hence, with big data acquisition structure and data processing engine, the proposed platform become a powerful tool for big data analytics, which is beneficial for Smart Grid monitoring.

A secure, reliable and economic smart grids operation can only exists after solution of optimization problems. Analysis of data acquired by grid sensors is required to decide actions that minimize system perturbations and enabling grid to be in continuous adaptation to new situations caused by various load and generation patterns. Extracting hidden information from historical datasets is accepted to be a good way to solve this grid management problem. As measurement continues by time, it is required to analyze data with an increasing size. In this context efficient big data computational techniques is used. A paper presented a solution using MapReduce paradigm to search k most similar cases [80]. MapReduce allows distribution of data and tasks over a cluster of computers. Experimental conducted to describe the solution behavior under different situations. The dataset, size, the number of tasks, block size are some conditional cases.

For high reliability, power supply level of urban power grid it is necessary to adopt appropriate methods. Usage of wide and multi-dimensional power system data, brings usage of large data mining which yields exploring the potential law and value of power system equipment. A study is done to integrate the data of power grid equipment environment, based on the monitoring data of main transformer and the records of defects and faults [81]. An association identification algorithm is used to extract the frequent correlation factors of the main transformer whereas the potential dependence of the big data is analyzed by the support and confidence. Then, the integrated data is analyzed by PCA algorithm to construct quantitative scoring model. Study end with providing a new idea for data fusion of smart grid. Figure 4.8 shows the structure of equipment.

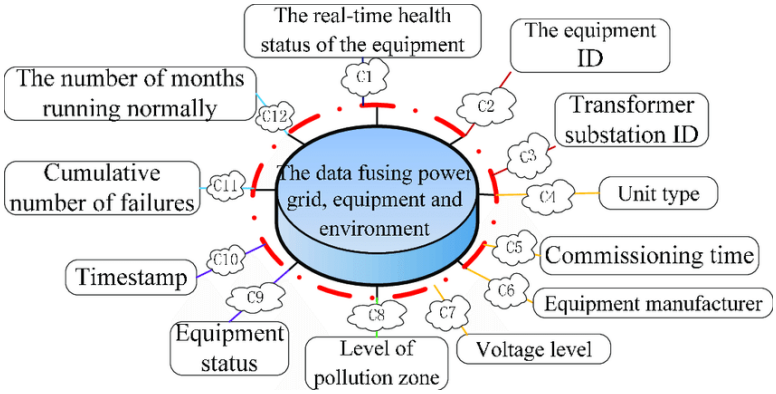


Figure 4.8 Data fusion and specific data structure of power equipment

In a research China's power industry is considered in relation to big data also [82]. This implies a high requirement on management and data servers. The advantage is at using big data technology and cloud computing systems in the electric power industry. Research analyze various applications of big data and cloud computing in power industry and the challenges they faced.

By rapid expansion of urban power grids, the Automatic Voltage Control (AVC) system develops some new characteristics. Based on the technology of big data analysis and mining, a study discusses an AVC system for urban power network with high load [83]. With introduction of several different applications, operation characteristics of a multi-point distribution mode is aimed to show. Hidden troubles and defects of AVC systems are declared. Security check of remote control, self-adaption analysis method and other key technologies are presented.

A tensor-based big data management scheme is proposed for dimensionality reduction problem of big data generated from various smart devices [84]. After application of Frobenius norm to minimize the reconstruction error of the reduced tensors, an empirical probability-based control algorithms designed. Major goal was to minimize the transmission delay incurred. Extensive simulations carried out on the data traces using 'R' programming and Matlab used for evaluation. The big data traces considered for evaluation consist of more than two million entries (2,075,259) collected at one minute sampling.

A review on smart grids integrated with big data is the topic of a document. Available articles in the literature were reviewed in perception with the features of big data and smart grids. Potential issues are discussed. The recent solutions about the subject is presented. The technologies evaluated. The results have been shown that "data" is now a new feature to be added as a main component of power systems [85].

Smart grid is a complicated interconnected power grid that involves sensors, deployment strategies, smart meters, and real-time data processing. It continuously generates large-volume data with high velocity. A paper, gives a brief introduction on big data, smart grid, and big data application in the smart grid scenario [86]. Recent studies and developments are

summarized. Meanwhile, security issues are discussed in detail. Finally, several typical big data applications are introduced.

One of the studies in past reviews the advantages of the use of big data for electric utilities [87]. By big data technology opportunity to better monitoring, correcting, and integrating smart grid technologies with renewable energy is provided. To push the limits of potentials, data management and utilization must be integrated into organizational operations. Electric utilities are heavily regulated and hence conservative. System reliability and overall profitability are two concepts that electrical utilities concern. Thus, technological, economic, institutional, and policy requirements must all be met. As a result of review it is seen that studies about data in the context of electric utilities are rare whereas studies about electric utilities are plentiful.

Nowadays there is an increasing share of renewables and by smart meters more available data is being collected. In this context, management of grids mostly relies on advanced analytics of data which is demand supply forecasting issues. As a consequence effective management of electric grid is a major challenge. A paper presented an energy data management system for the electricity grid [88]. The system was designed by searching the real world. The forecasting of short term energy demand in Vermont is an example of this real world. Data extracted from SCADA system is analyzed. A general data model introduced aiming at supporting advanced data analytics. The data model views an abstract concept that representing measurements or arbitrary operations. The benefits of proposed system is shown.

The representation of energy-based data in low voltage segments of smart grids by using a process-oriented approach is discussed [89]. An interpretation is used to construct an overall framework which contains multiple options for representing energy based data. An established framework defines a specific representation of the information. An enhanced representation of energy-based data is identified by leading to the formulation of the final event-based data gathering (EBDG) scheme, flowchart of which is seen on Figure 4.9. Tests are carried out on data sets for residential customers, including a new challenging benchmark. The results validate the use of the EBDG scheme to mine the energy-varying processes occurring at the nodes of the distribution system.

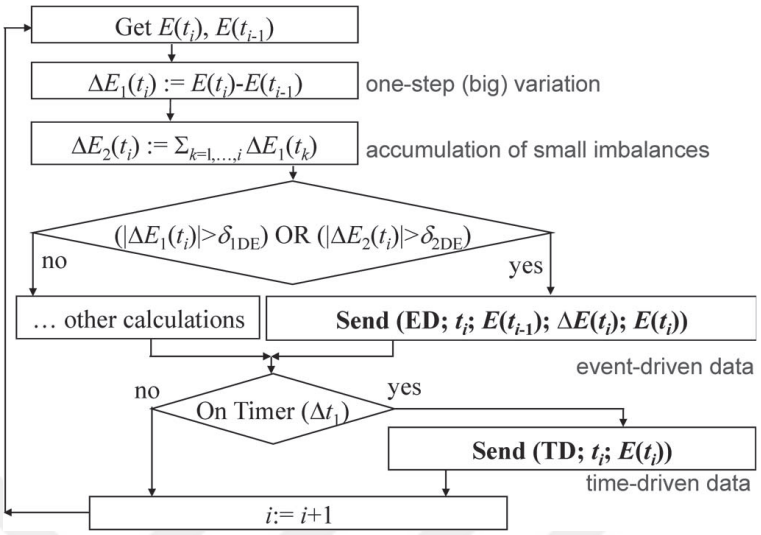


Figure 4.9 Flowchart of EBDG.

In a smart grid large number of sensors are becoming part of the power network to improve its efficiency. These sensors enable communication between home appliances and suppliers to enhance automation. Also monitoring and remote control capabilities arouse by sensors. As smart power grid incorporates a large number of sensors; problem is where in the network to process and analyze the data, and how to perform the analysis. A study is documented by an article [90]. Test bed is discussed to highlight advantages of distributed smart grid architecture. Comparison of central and local processing of Data is made. The advantages of local processing of sensors' generated data is discussed. Focused big data and introduced machine learning algorithms for data processing. Results of test bed, which prototypes distributed smart grid architecture is presented.

It is world widely common that smart meters are being deployed to enable automated collection of energy consumption data. In this case massive amounts of data evolved and it should be managed. Efficiency, reliability and sustainability of the smart grid can only be increased in that way. A framework implemented on a secure cloud-based platform is presented [91]. Two scenarios considered; for a single-house and a smart grid that contains over 6000 smart meters. Framework has been applied on these two scenarios to visualize the energy. Experiment conducted show that the framework is feasible in performing further smart grid data analytics.

It is needed for smart grids to bring changes in the information systems that drive them. With no exception new elements come into consideration; new information flows coming from the grid, new players example of which is producers of renewable energies, new uses such as electric vehicles and connected houses and new technological equipment such as smart meters, sensors and remote control points. All this will cause complexity of data. big data technologies offers suitable solutions for utilities, when feasible decision about big data technology used. In a paper an overview of data management for smart grids is provided [92]. Author summarizes the added value of big data technologies and solutions in the smart grid context.

Smart grid is usually considered as a set of advanced features powered by promising technological solutions. Smart grid is described as a socio-technical transition and illustrate the evolutionary path on which a smart grid can be realized [93]. Role of big data and its support to growth of smart grid is discussed.

Normalized modeled electric power knowledge has some problems for management and analysis of electric power big data. To solve these problems, a new electric power knowledge theory model is proposed. Current modeling techniques are inadequate because of the electric power system data being complex and various. A new modeling approach can be developed with ontology theory and semantic web technologies. A paper in literature structure, elements, basic calculations and multidimensional reasoning method of a new knowledge model is introduced [94]. Regulations are defined in electric power system operation. Different forms of the model are also introduced. These are;

- electric power system standard modeling,
- multi-type data management,
- unstructured data searching,
- knowledge display
- data analysis based on semantic expansion and reduction.

Experiments conducted registered proposed new model to be powerful. It can execute in adaptation to various knowledge expression requirements of electric power big data.

4.1.2 Transmission

The transmission lines are typically at a high voltage, ranging from 765kV, 500kV, 345kV, 230kV and 138kV. They are commonly working in a three phase alternating current. When power generating stations are located far from the consumers, electricity has to travel a great distance, which is usual in today's power systems. To reduce the energy lost along a great distance transmission, high voltage is preferred in transmission.

There exists a proposal of a consumer partitioning methodology at the distribution level spite of probable disturbances and subsequent connectivity issues in the network induced by renewable energy resources [95]. K-means and a fuzzy logic engine is used on big data environment.

A study performed on purpose to review and synthesize the relevant literature about ecosystems in relevance to smart sustainable cities. Identifying and distilling the core enabling technologies of big data analytics are reviewed [96]. The key computational and analytical techniques and processes associated with the functioning of such ecosystems are illustrated. Moreover, to review the important works, important applications, challenges and open issues are highlighted. It is argued that big data analytics and context-aware computing are prerequisite technologies for the functioning of smart sustainable cities in the future.

A study is to mix two concepts, namely topical literature review and thematic analysis [97]. Provided a basis for urban researchers to draw on this analytical framework in future research. Framework designed will add additional depth to studies in the field of smart sustainable cities.

4.1.3 Distribution

The distribution stage is the stage of end customers, where a more accessible level of voltage is required. This requirement is satisfied by a step down transformer substation. The consumers are classified into three; sub-transmission customers at 26kV and 69kV, industrial customers at 13kV and 4kV and finally, residential customers at 120V and 240V.

An algorithm for distribution system parameter estimation is presented [98]. It is demonstrated using a real utility feeder model with advanced metering infrastructure (AMI) and PV micro-inverters. Illustration in Figure 4.10 shows the flows of big data for distribution system parameter estimation.

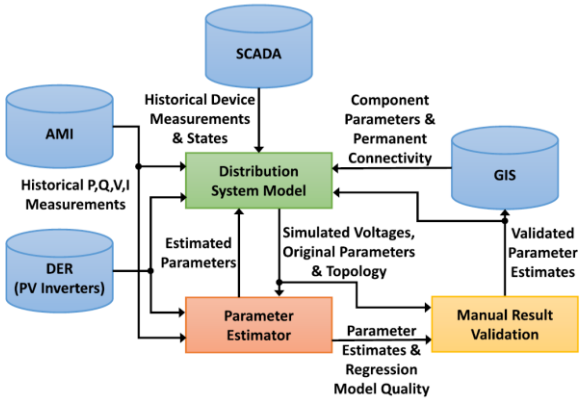


Figure 4.10 Big data for distribution system parameter estimation

A thesis provides a comprehensive overview of big data analytics applications in the energy sector [99]. The energy sector holds huge amount of data. Moreover the systems are getting more decentralized, customer demands are growing and regulatory pressure on mitigating climate change is larger than ever. These trends make it mandatory for big data analytics to be in adaptation to these changes.

The energy sector is divided into different parts according to their relevance to data sources: generation, transmission & distribution network, flexibility management and customer relations. The activities of electricity utilities and companies' usage of big data to create products & services were analyzed. Other sectors that are similar to energy in terms of data size and structure were investigated. Business models that are derived from these applications were identified.

At last, a business case appropriate for Iberdrola was elaborated and performed SWOT analysis methods on it. The results proven that optimal operations both for generation and transmission & distribution network can be performed when real-time visibility of assets combined with predictive analysis. Better management and enhanced productivity mitigates

operational costs. For renewable energy generation, uncertainties caused by weather conditions are reduced which brings better adoption of these technologies. It is seen that when flexibility management is applied to a system, more business opportunities are opened. Distributed energy resources are aggregated and managed by big data analytics in near-real time. Thus, services for residential, commercial and industrial end-users can be utilized by it, which leads to obtaining more DER assets in flexibility management programs. Utilities or aggregators can execute these operations. From viewpoint of customer relations, customer data is converted into energy insights to provide better engagement.

Iberdrola Efficiency Shop is discussed which is a business idea that converts smart meter data into energy advice. It is responsible to drive customers to purchase more efficient home appliances. The final case study showed the potential of big data analytics for future studies targeting business systems.

Some researches on subject classifies big data applications on power distribution systems into two, namely short term and long term applications [1]. Energy fraud detection, EV detection and Rooftop Solar Integration, Develop More Granular Load Forecast, Develop More Granular Renewable Generation Forecast, Visualization in Power Distribution Systems, State Estimation and Equipment Diagnosis are short term applications. Model Customer Consumption Behavior under Various Incentive and Pricing Structures, Transformation of Distribution System Planning Process and as long term applications.

4.2 Renewable Energy Systems

The sun and wind being theoretically endless resources by their nature, Renewable energy systems are accepted to be reliable. In order to operate a large photovoltaic (PV)-battery system independent, usually a sophisticated energy management strategy is determined also. Addition of a collected data by sensors managed with powerful technologies such as big data and analytics, the system is enabled to efficiently react to environmental factors and respond to consumers' requirements in real time. When the weather parameters concerned, the output of PV could be symmetric, supplying an asymmetric electricity demand. A smart adaptive switching module that includes a forecasting component which improves the symmetry between the PV output and daily load curve is proposed [100]. Scaling approach for smaller

off-grid systems developed. The aim of this scaling is to provide an accurate forecast of the PV output based on data collected from sensors. The methodology is based on sensor implementation in RES operation big data technologies used for data processing and analytics namely Support Vector Machine (SVM) Root Mean Square Error (RMSE), considerably improved prediction ability.

Energy internet, is an open framework for integrating equipment of generation, transmission, storage, consumption of energy. By energy internet global energy can be managed and controlled efficiently by information and communication technologies. A published paper focuses on the coordination of renewable and traditional energy sources [101]. This is a typical issue on energy connections. Conventional power system includes the utility company, the energy storage company, the microgrid, and electricity users. First, energy management problem is formulated as a Stackelberg game, which has three stages. Every player in the electricity market aims to maximize its individual pay. At the same time system reliability and satisfying users' electricity demands should be guaranteed. Backward induction method is used to solve problem. Next topic in paper is the big data-based power generation forecasting techniques, and introduction of wind power forecasting. The properties of the proposed energy management algorithm proved also including the existence and uniqueness of Nash equilibrium and Stackelberg equilibrium.

Electric power sequence can benefit from the application of big data techniques. A short overview of possible applications and challenges is presented [102]. As case study BigDataOcean, an H2020 funded project is described. It is shown how a data-driven approach improved the development of offshore renewable sources like wave energy. Value chain shown in Figure 4.11, demonstrates the followed methodology.

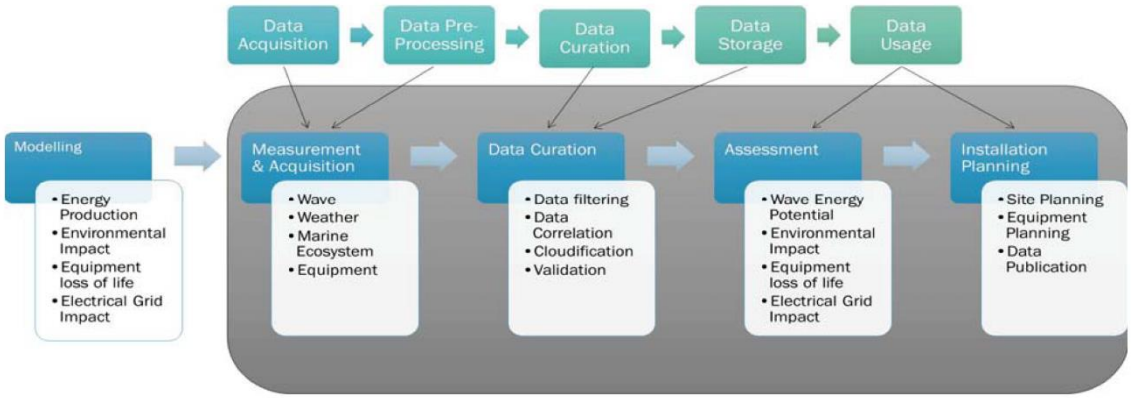


Figure 4.11 BigDataOcean wave power value chain

Energy internet is an open framework where every element in a power system is integrated with information and communication technologies. A chapter of a book about subject is reviewed [103]. Coordination of renewable and traditional energy is a typical issue on energy interconnection. Authors adopt game theory to big data for this purpose. Then formulate the energy management problem as a three-stage Stackelberg game to derive expressions of the optimal strategies. Next studied is big data-based power generation forecasting techniques. They introduce wind power forecasting. Simulations resulted microgrid to have more accurate prediction and consequently better energy management of wind power.

Parallel computing models are used to meet high computational demands. MapReduce is such a modal. On the other hand, big data analytics have become a main consumer of energy in data centers for the reason of these computational tasks. The granularity of task partitioning in each moldable job has a significant impact on energy efficiency. A paper presents formulation of a workflow mapping problem to minimize the dynamic energy consumption [104]. They design a pipeline-structured workflow based polynomial-time approximation scheme (FPTAS). Applied to both homogeneous clustered and heterogeneous clustered structures. The performance superiority of the proposed solution has better performance than recent examples.

Renewable energies development is investigated in Iran [105]. Techno-Econo-Socio-Environmental Multivariate Analysis (TESEMA) is an approach. Six annual hourly

consumption variables of Iran's power industry from 2011 to 2017, are predicted. For this prediction seven dynamic and intelligent models are used. According to the results DNN is the best model to predict demand load in Iran (RMSE=73.15%). Since DI varies in a wide range from 0 to 248.83 and the population density is significantly correlated with TESEMA variables (R2=91.86%) Results also show that the current centralistic policies should be revised in Iran. To reach sustainable development a four-cluster management strategy accompanied by smart monitoring can efficiently lead to sustainable development in Iran.

4.2.1 Solar Energy

It is detected that main parameters of poor quality need to be removed before processing PV plants data [106]. Then, all PV plants in a power system with single-PV-plant resolution is presented. Dataset analyzed is the official 237,588 monthly energy values obtained from 2005 to 2017 for the 1523 PV plants in the Canary Islands. Dataset is obtained from PV plants ranging from 0.53 kWp up to 9MWp. With simulation using their data they come up with a managed data having errors under % 3. Results shown in Figure 4.12.

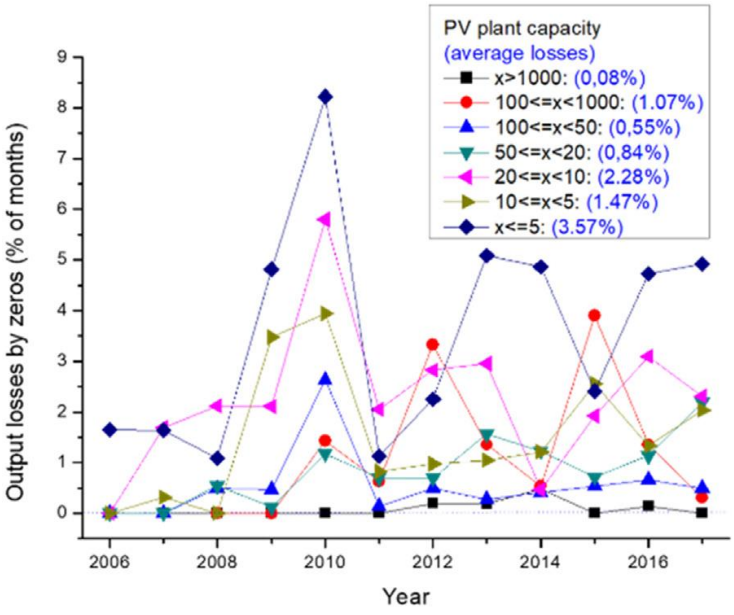


Figure 4.12 Output losses in the grid for different PV capacity ranges.

Solar power is the leading renewable electricity sources for future use and is expected to play a substantial role globally [107]. However, its natural variability still forms an important barrier to overcome. Machine learning and data mining techniques are used to forecast solar electricity generation, in order to reduce the impacts of its nature of being inconstant. A literature review is presented on big data models for solar photovoltaic electricity generation forecasts. They are compared by means of applicability accuracy of the technique, motivation behind each project proposal, the characteristics and quality of data used to address the problem.

4.2.2 Wind Energy

Because of its nature integration of wind generation to power system brings great challenges to the secure operation. Thus accurate wind power forecasting plays a key role for power systems including wind generation. A combined approach based on Extreme Learning Machine (ELM) and an error correction model is proposed for prediction [108]. ELM is used to forecast the short-term wind power. After, the ultra-short-term wind power forecasting is acquired. However it is seen that for short-term forecasting, the ELM doesn't perform well. The overall performance of NRMSE (Normalized Root Mean Square Error) forecasting for 66 days is 21.09 %. For the ultra-short term forecasting errors lie in range of [−10 MW, 10 MW]. This means a success of 5.76 %. The scatter of wind power versus wind speed of the wind farm is plotted in Figure 4.13 and error distribution is shown in Figure 4.14.

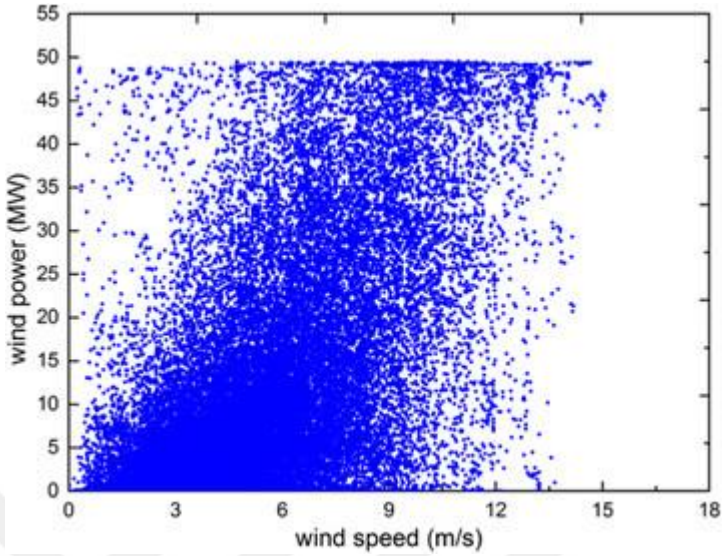


Figure 4.13 Scatter plot of wind power versus wind speed loads.

These results show that the ultra-short-term wind power forecasting accuracy is improved by using error correction in terms of normalized root mean squared error (NRMSE).

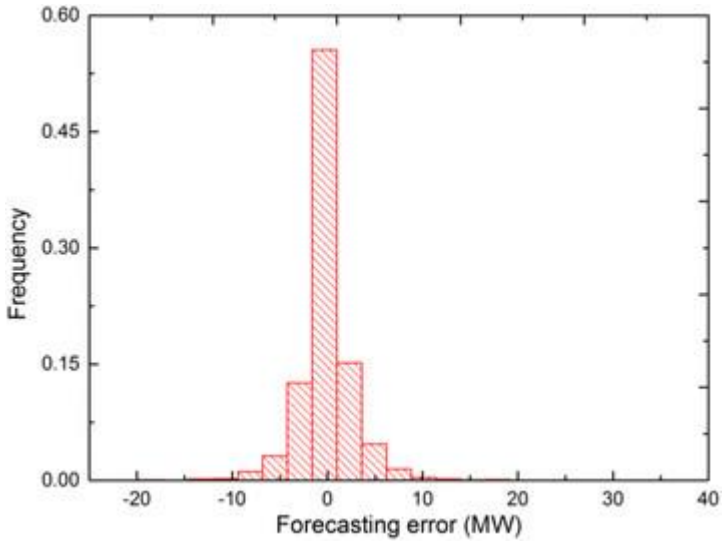


Figure 4.14 Ultra-short-term forecasting error distribution after error correction

A similar study is on a self-powered wearable sweat-lactate analyzer basing on sweat-evaporation-biosensing coupling effect [109]. This time aim is to build sports big data. The

outputting voltage increases with increasing sweat-lactate concentration, which can be regarded as biosensing signal. The biosensor can be integrated with wireless transmitter and capacitor, constructing a self-powered wearable sweat-lactate analyzer. Concentration of sweat-lactate can be detected on skin. Exercise physiological information is transmitted wirelessly to terminal. A new research direction for wearable exercise-analyzing system can be promoted on the way to build sports big data using bioenergy.

4.3 Demand Side Management and Demand Response

4.3.1 Demand Forecasting

For a prediction issue a combined algorithm of Map Reduce and neural network is used [110]. Clustering and nonlinear mapping of neural network is used, to solve the problem of nonlinear objective function approximation, after which neural network is applied to predict amount of fusion. As a neural network model multi layer feed forward network is used. The parallelism and real-time property of the algorithm is considered to achieve large-scale data sets in parallel computing, Reduce Map model is combined to system in order to realize the parallel processing of the algorithm. And finally, simulations conducted to verify proposed model.

People's behaviors influence power industry and big data technology is needed to improve for economy and reliability. After completion of Jiangsu Electricity Information Acquisition System the accumulation of historical information is started. These steps bring an opportunity for the study of the electricity big data. Electricity big data is the area of research of a paper [111]. Works on short-term load forecasting and corresponding results are discussed. It made a certain contribution for the promotion of big data technology in Jiangsu Electricity Power Company. Neural network training algorithm is shown in Figure 4.15.

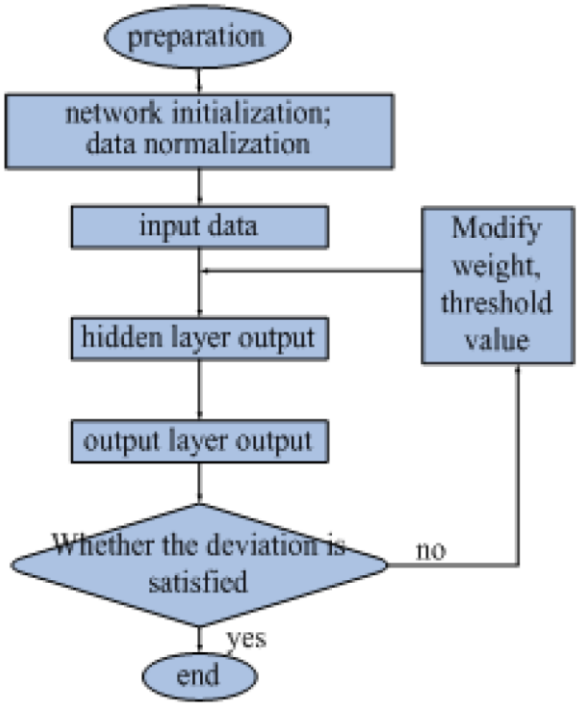


Figure 4.15 Neural network training

Construction of smart grid yields lots of renewable energy resources to be deployed in power system. Renewable energy deployed are wind solar or others. In this case, power system is loaded more than it is before which brings difficulties in short term forecasting. A paper proposes framework based on big data technologies. Cluster analysis was the first step of study which is done to classify daily load patterns. Next, step was determining critical influential factors by an association analysis [112]. Decision tree is applied to establish classification rules. Appropriate forecasting models are chosen for load patterns. The forecasted total system load is calculated as a sum of individual loads forecasting results. Comparing real load data with the forecasting results of proposed new framework shows that it is successful in short-term load forecasting within required limits.

Utilization of ‘Big Data’ is realized by using powerful algorithms combined with robust understanding of the data. Find out relevant factors, which are the variables is a key challenge in predictive analysis. In a paper, a case study is presented to come over this difficulty. Exogenous variables are identified [113]. This means, prediction of the German electricity

spot prices is tried. Least Absolute Shrinkage Selection Operation (LASSO) is used to choose stations. Improvement of forecasting accuracy up to 16.9% is achieved in terms of mean average error.

Recorded atmospheric events also provide big data which in a way supports scientific decision-making for industry and people's life [114]. A self-powered flexibly arranged gas monitoring system has been presented. The working principle of the new system is based on a new water-evaporation/gas-sensing coupling effect. A device evaporates rainwater and generates electricity which is used to power whole system. Since output voltage/current dependent on atmospheric conditions, device is acting as the gas sensing device. Naturally, a self-powered gas monitoring system can be fixed out of doors. It can detect the air quality of different locations. These records periodically sent to a database which in fact is a big data. An atmospheric environmental big data is developed and a research direction is opened here.

Ensemble models are used for forecasting big data time series. These ensemble models are presented [115]. Due to good results that achieved in previous big data applications, an ensemble is proposed to be composed of three methods. These are decision tree, gradient boosted trees and random forest. The performance of the proposed ensemble models is evaluated on an official data. This data is Spanish electricity consumption data for 10 years measured with a 10-minute frequency. From the results it is seen that both the dynamic and static ensembles performed well. The most accurate model was dynamic ensemble, which is shown in Figure 4.16, mean relative error (MRE) of which is calculated as 2%. Proposed ensembles are also evaluated on another data which is of solar power from Australia for two years measured with 30-min frequency.

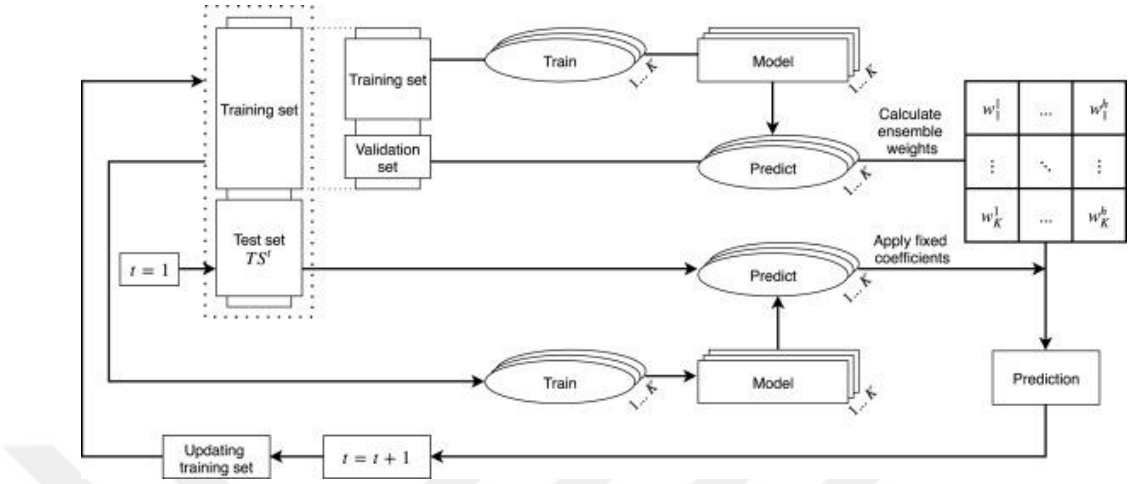


Figure 4.16 Workflow of the proposed dynamic ensemble

A reliable and modern electricity grid is supposed to be responsible, efficient and environmentally aware of energy consumption. It is presented that an intelligent data mining model can be used to analyze, forecast and visualize energy time series to uncover various temporal energy consumption patterns [116]. These patterns define the appliance usage in time period such as hour of the day, period of the day, weekday, week, month and season of the year. It is not trivial to determine the multiple relationships among different appliances usage from concurrent streams of data. Also, it is difficult to derive accurate relationships between interval-based events where multiple appliance usages persist for some duration. To overcome these problems, an unsupervised data clustering and frequent pattern mining analysis is proposed. Energy time series are considered and Bayesian network prediction is used for energy requirement forecasting. Many experiments are performed using real-world context-rich smart meter datasets. The accuracy results of identifying appliance usage patterns outperformed by Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP). Accuracy of 81.82%, 85.90%, 89.58% for 25%, 50% and 75% of the training data size respectively.

Short term load forecasting is highly important for transmission system planning and operations, It forms a base for planning, identify contingencies, calculate power flows, load switching and consumption market. As a solution a paper is published to present a model using Neo4j graph technology a big data NoSQL data store, processing with Time Series

method to forecast the load in short term basis [117]. A case study conducted for south part of Kosovo's power transmission system substation. The past load consumption data with weather's parameters, such as temperature, humidity and wind speed, captured from sensors every 15 minutes. A proposal about a model processing and storing those large amounts of data. The concept of path on graph with Cypher query language database are used to match the data for a specific time. Time Series method is used to predict load. Results has shown that the combination of using big data graph technology and Time Series is suitable to forecast the load prediction more accurately.

One of the main factors that affect Power systems is weather. Long-term planning perspective does not tell us enough for predicting weather impacts caused by short-term changes. A demonstration presented by utilizing the past weather data and climate change projections in a macro geographical area to predict future load patterns in a relatively micro geographical area [118]. Results show that temperature rising can have either positive or negative impact on the load, and the deviations shown to be quite large depending on the projected climate change data.

A study performs short term load forecasting using multi-variable linear regression (MLR) for big data [119]. In a power system total electric demand dynamically changes. It depends on temperature, humidity, wind speed, human nature, regular activities, events, and many other parameters. Accepted that enough historical and future input data with good accuracy are easily available whereas linear regression is a proven method, widely used for forecasting. However, method is slow for big data since it uses large size matrix operations. Paper published after study, linear regression is formulated for small number of variables with big data. For matrix operations Multi-core parallel processing is applied. This allow processing of unlimited historical big data and unlimited scenarios in acceptable time period. Mean absolute percent error is 3.99%. Actual and forecast during a 168-hour period is compared. Comparison results shown in Figure 4.17.

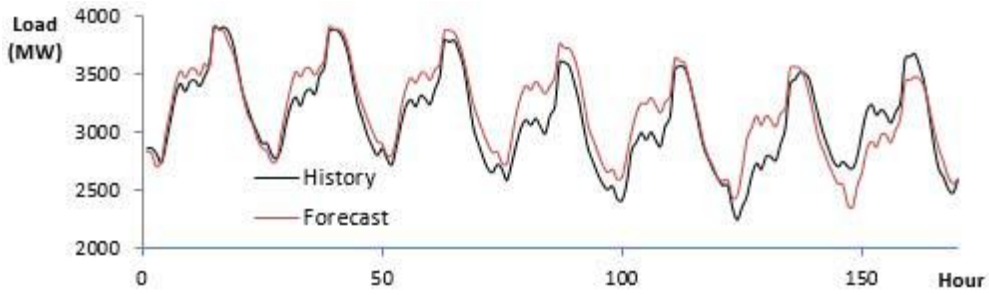


Figure 4.17 Comparison between actual and forecast loads.

Vehicular edge computing is essential for sake of vehicular networks. However, energy consumption and carbon pollution problems occurs in massive deployment of edge computing infrastructures. In order to solve this, a developed framework is presented. The name BEGIN (Big data enabled EnerGy-efficient vehIcular edge computiNg) is given to framework [120]. Overall design principle of BEGIN is described. Big data for the framework is classified into four categories and then detailed information is given about features and potential values of each category. A case study conducted to demonstrate the feasibility of BEGIN.

These big data brings opportunities for development of new energy services and ensures better energy management and security. These advantages targeted sensor-based energy forecasting as a research topic in buildings. A paper investigates sensor-based forecasting in an especially difficult issue, namely consumptions caused by the hosted events which has large variations [121]. Neural networks (NN) and support vector regression (SVR) is two machine-learning approaches. These were used in cooperation with three periodic data. Data recording periods are daily, hourly, and 15 minutes. The approach has been applied to a large utility located in Ontario, Canada. Results show that daily data intervals resulted in higher consumption prediction accuracy than hourly or 15-min readings. This is explained by the inability of the hourly and 15-min models to capture random variations. Machine learning techniques are compared also. It is seen that the NN model achieved better than the SVR with daily data unlike with hourly and 15-min data whereas two technique achieved closer results.

Nowadays, Cloud Computing (CC) paradigm becomes insufficient for the storage of big data (BD) produced by Internet of Things (IoT). A new platform is needed to meet this

requirement. That is "Fog Computing". Fog Computing proposes most functions of data processing to be performed out from the cloud, required that a reliable and efficient communication system exists to get a robust, affordable and secure power supply through Smart Grids (SGs). For a SG application computational requirements can be done by Fog Computing (FC) model. In order SGs to have Fog-supported nodes, a study introduced SmartLocalGrid (SLG) as a communication method between micro-grids [122]. FC model is considered efficient for SLGs and SGs, but it has some constraints as well. Security and reliability are two of them. In a work on subject, fog platform, technical support, security, big data analysis and utilization issues are searched in detail according to current FC architecture. In FC, the fog node transports big data to the remote cloud, in order to perform in-depth data analysis.

In the future of renewable energy, weather and climate dynamics will be bounded to energy systems. So as to enable modeling such a future energy system, a data-driven renewable electric power system is described [123]. Generation and demand information and transmission network model is combined together by a dataset. Conventional generators generates with their special technical and economic characteristics. Weather forecasts realize demand for renewable energy generation for a period of 3 years. Future European energy system is taken as base for forecasting. Completeness and resolution of the dataset, lead to the evaluation, scaling analysis and satisfaction of wealth of proposals.

Traditional power load forecasting and evaluation methods for power transmission and transformation equipment become insufficient with rapid development of the power grid. A framework is proposed for elaborating the method to use built big data platform in the power grid industry to predict the level of power load and assess the status of power transmission and transformation equipment [124]. The proposed big data methods for forecasting and assessment on power grid revealed satisfactory results in practical application.

Machine Learning (ML) is an effective tool for predictions based on past history. ML algorithms operating principle is to build a model from training input examples which makes predictions or decisions for future. ML integrated with big data would be a candidate scheme to build effective predictive systems or to solve complex problems. An electricity generation

forecasting system proposed to predict the amount of power required for electricity consumption in United States [125]. For analytics, data collected on power management in the past 20 years. A ML model is applied to the system for prediction. Test results show that the system can predict the required power generation in ratio of 99%, which means a success of prediction. It is indicated that the integration of ML with big data improves efficiency in forecasting techniques.

A paper focuses on a data correlation framework using innovative issues [126]. Big data driven analytics achieved for improved in distribution networks. Data sources extended and their relation between network model are described. Lastly, Improved data analysis advantages are discussed before giving several examples.

4.3.2 Price Prediction

Another concept that proves usefulness of smart grid is ability of price forecasting. It is said that this makes smart grid cost efficient [127].When the huge size of price data is considered, methods for price forecasting may be difficult to handle. Besides, there should exist some redundancy from feature and integrated infrastructure in some cases may not be sufficient for forecasting electricity price. For these reasons a modal is developed. Specifically, three modules are integrated in the proposed model. After merging of Random Forest (RF) and Relief-F algorithm, a hybrid feature selector based on Grey Correlation Analysis (GCA) is proposed to get rid of redundancy. To reduce complexity caused by dimensionality in feature extraction process, an integration of Kernel function and Principle Component Analysis (KPCA) is used. At the end of process, for forecasting, differential evolution (DE) based Support Vector Machine (SVM) classifier is used. It is claimed to have superior performance than other methods. Comparison of model with other benchmarks is shown in Figure 4.18.

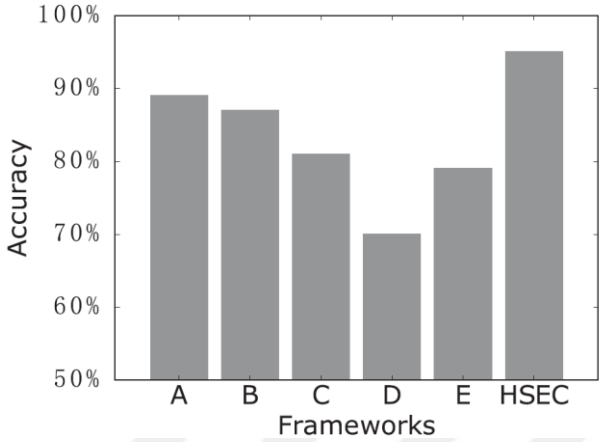


Figure 4.18 Comparison on accuracy among HSEC and benchmark frameworks.

By using an intelligent big data platform, some other researches are done also. There are some considerations on analytics of an extremely large dataset of smart grid electricity price and load, which is very difficult to process with conventional computational models [128]. A deep learning based model is proposed for the forecast of price and demand on big data using deeper Long Short-Term Memory (LSTM). Due to adaptive and automatic feature learning of DNNs, processing of big data is easier with LSTM, compared to purely data driven methods. The proposed model is evaluated using a well-known real electricity markets' data.

4.4 Energy Management

4.4.1 Residential, Commercial & Industrial Consumption

Smart Home is defined a house to be able to generate energetic values such as electricity consumption, temperature or luminosity. Main aim of some research is to extract information from raw data [129]. This information is imperative to improve energy efficiency management. An document in literature stands for presentation of a system which uses various Machine Learning approaches to get knowledge about the consumer’s behaviors. Presented system is able to generate collaborative recommendations and consumption predictions, whereas it helps user to consume better, by sharing recommendations. Moreover, by using

consumption values, the system learns to identify devices, after which device is enabled to be anticipated.

Electricity demand heavily influenced by temperature. In some cases, psychology is in consideration when factors of electricity demand is searched. An article is about adaptation of "recency effect" to energy system, which is a term of psychology, to represent the fact that the electricity demand is affected by the temperatures of the preceding hours. Usually load forecasting achieved due to variables from lagged hourly temperatures and moving average temperatures. A comprehensive study is conducted to model the big data driven recency effect [130]. How many lagged hourly temperatures or moving average temperatures are needed to be able to capture the recency effect? This question is tried to be answered with advantages of modern computing power. A modal with the recency effect is demonstrated. Its performance for forecasting the load series at the top level is by 18% to 21%. Then the recency effect in order to customize load forecasting models at the bottom level of a geographic hierarchy is modeled and performance was 12% to 15% on average. Two findings are listed. These are;

- The naive models are not useful for benchmark purposes in load forecasting at aggregated level due to their lack of accuracy.

- Slicing the data into 24 pieces to develop one model for each hour is not necessarily better than building one interaction regression model using all 24 hours together.

Heating and cooling in buildings is directly affecting energy efficiency measures and implementing local policies for energy planning. Knowledge of features and performance of existing systems is processed for establishing realistic energy savings strategies. A work presented an analysis on building heating systems of Lombardy, the most populated Italian region [131]. A dataset of almost 2.9 million of heating systems. Several aspects are considered. Type of heating systems, thermal power, fuels, age, nominal and measured efficiency are some of them.

Heating, ventilation and air conditioning systems are considered systems to improve the energy saving while supporting occupants' comfort. Optimal control of ventilation systems is

use to be subject of many studies. These approaches could not be efficiently performed under every contexts. An intelligent approach is introduced that selects the most appropriate control among three of existing strategies [132]. The proposed approach is implemented using IoT devices and recent Big-data technologies for real-time monitoring and data processing. A case study is also deployed in a test site for real testing. Experiment results show the effectiveness adaptive control approaches for ventilation systems control. Topology of system is shown in Figure 4.19.

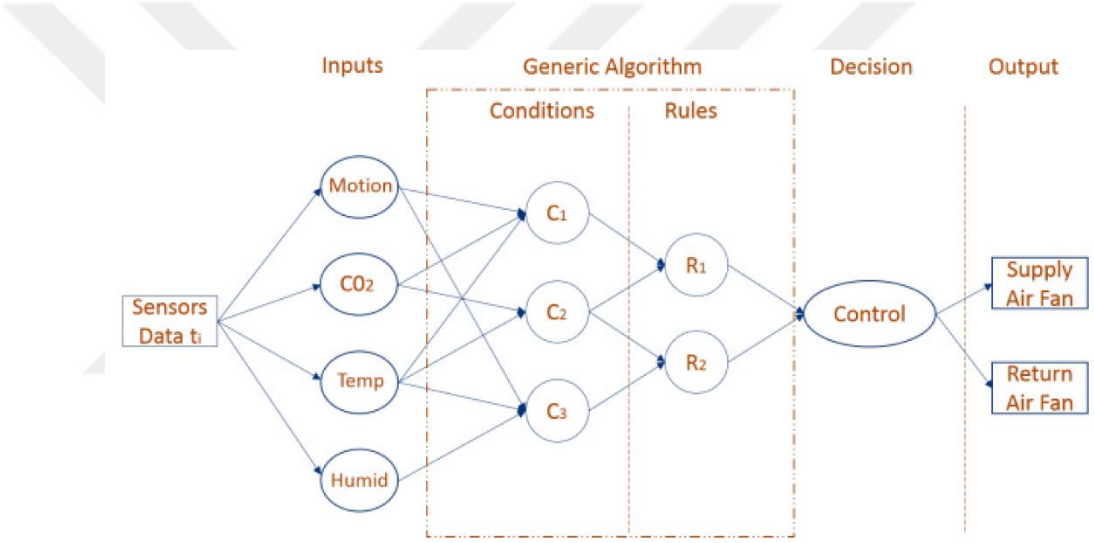


Figure 4.19 The data monitoring/processing topology

In recent years, residential and commercial building energy consumption has risen continuously. This increase mainly due to their HVAC systems. The quantity and quality of the energy used in buildings is affected by expected energy loads, transportation, and storage. One other parameter that influence consumption is behavior of consumers. Technology of today can accurately monitor, collect, and store the huge amount of data involved in this process to analyze and calculate such data in meaningful ways. Today, a great deal of attention and interest is spent to the use of data science techniques to increase energy efficiency. A review paper is established how data science is used to solve the most difficult problems faced in the field of energy management [133].

4.4.2 Objectives

4.4.2.1 Optimization

On-line solution optimization problems is crucial for a reliable, secure and economic smart grid operation. Reducing the effect of perturbations or adopting the grid state to new load conditions, data acquired by sensors should be continuously analyzed to identify proper control actions. A case based reasoning extracting useful knowledge from historical data is suggested to overcome these high number of computations. A solution is presented aimed at searching k most similar cases based on a paradigm which they call MapReduce paradigm [134]. It allows to distribute data and jobs over a compute cluster of commodity hardware. Experiments are done under different conditions, such as the dataset, size, the number of available nodes/jobs, the block size.

Increasingly customized manufacturing causes unpredictable stricter requirements on sustainability. Hence it becomes challenging to achieve energy efficient optimization for machining processes in shop floors. This optimization problem is tried to be solved by a novel Cyber Physical System (CPS) integrated with big data [135]. Figure 4.20 shows the structure. The innovations and characteristics of the presented system include four aspects:

(1) Designation of "scheduling, monitoring/learning, rescheduling" sequences to enhance system adaptability during manufacturing.

(2) Developing an energy model, which is enabled by CPS, big data analytics and intelligent learning algorithms to support energy efficient optimization over manufacturing lifecycles

(3) Application an effective evolutional algorithm based on Fruit Fly Optimization (FFO) to generate an adaptive energy efficient schedule, and improve schedule when there are significantly varying working conditions

(4) Deployment of the system into European machining companies to verify capabilities.

As results, around 40% energy saving and 30% productivity improvement have been obtained in the companies which demonstrates the effectiveness and great potential applicability of the system in practice.

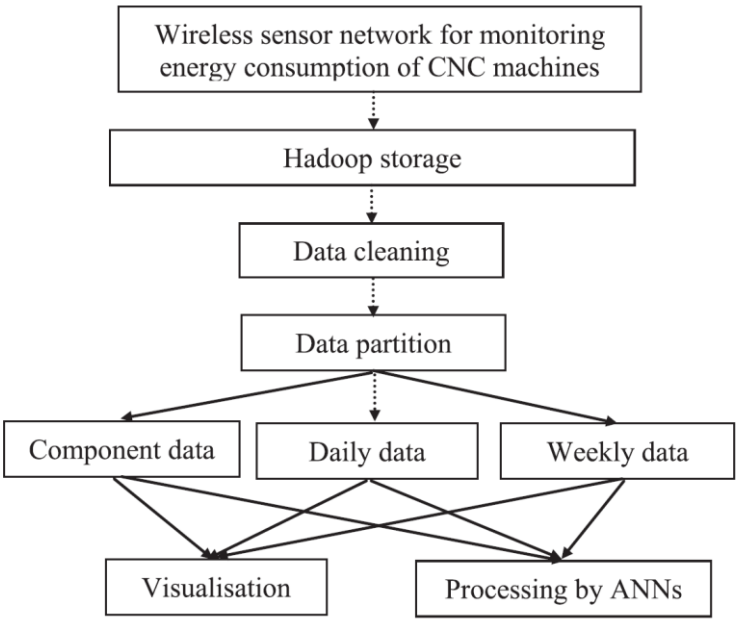


Figure 4.20 Design of big data structure

Obtaining Low response time in big data stream computing environments requires optimization of framework, which is achieved by energy efficient resource scheduling. A study proposes a real-time resource scheduling [136]. It name is given as Re-Stream. Proposed scheduling is used by a framework which setup a mathematical correlation between energy consumption, response time and resource utilization. By this it obtains the conditions to meet required efficiency and response time. Distributed stream computing theories are used to model a graph. Such a methodology is to calculate the energy consumption of a resource allocation. Afterwards, the Re-Stream is allocated tasks by utilizing an energy-efficient critical path scheduling mechanism subject to the architectural requirements. In this way scheduling is optimized by reallocating the critical vertices on the critical path of a data stream graph. Minimization of response time and system fluctuations also provided. Experiments conducted results as to demonstrate that the Re-Stream improves energy efficiency. The Re-

Stream provides an issue of intelligent combination of increased efficiency and decreased response time.

With increasing energy demand, electricity grids are needed to be integrated with renewable energy sources. Usage of sustainable energy resources is also advantageous in reduction of carbon emission and dependency on fossil fuels. However, renewable sources has intermittent and unpredictable generations. Because of this, balancing supply and demand for integrated systems is a much more complicated scheme than it is for classical energy systems. For traditional systems it was enough to adopt supply side to energy demand, but for a sustainable energy system the demand side should be managed in a better way to be in compatible with available energy supply.

First part of a thesis focused on energy consumption behavior, which is called as demand analytics [137]. Unlike previous systems that obtain consumption monthly, contemporary systems by the help of metering devices obtain it almost in real-time. Versatile customer segmentation framework is developed which track changes continuously before determining key characteristics that define clusters. Segmenting the households into several clusters can be a base for predicting household-level consumption. Forecasting the energy consumption of each cluster, and then aggregating partial forecasts. Success of this strategy may depend on number of clusters and size of customer base.

In thesis a modeling approach is developed for reduction of uncertainty of future demand. Methods like simulation, bootstrapping, ensemble are inefficient methods, but they are used in previous work. In this study, prediction intervals are constructed by using the time-varying conditional mean and variance of future demand. Collecting customer energy data has an infringement risk of privacy. This study uses symbolic representations of smart meter data on which learning algorithms is applied which isolate accurate data to processed data.

Second part focuses on methods for encouragement of customers to change their energy consumption. Demand Response (DR) stands for maintaining grid stability. Previous studies about DR targeted high consumption customers like industrial companies. This work studied DR application on residential customers. Affects of DR baselines, which means estimation of

customers consumption in the absence of a DR event on customer participation is studied. A strategy is proposed to meet demand with supply by intelligent agents on customer side.

4.4.2.2 Efficiency

An optimized smart electricity grid faciliates power flow in terms of economic efficiency, reliability and sustainability. This is a two-way flow between suppliers and consumers. In this situation consumers and producers take active role in market and is called the dynamic energy management. A smart grid turns users' participation to power cost reduction. This infrastructure permits the consumers and the micro-energy producers to take a more active role in the electricity market and the dynamic energy management (DEM). The most important challenge in a smart grid (SG) is how to take advantage of the users' participation in order to reduce the cost of power. In order to achieve this intelligent methods and solutions for the real-time exploitation of large volumes of data generated by the vast amount of smart meters is required. A research on subject aimed to highlight the big data issues and challenges faced by the DEM employed in SG network and to provide a brief description of the most commonly used data processing methods used [57].

Optimization of energy usage in big data processing platforms is imperative [138]. Energy efficiency has a closed relation with the utilization of computing resources on a big data processing platform. Consequently, an efficient resource allocation strategy for data processing tasks improves the platform's resources utilization. A study on subject proposes an allocation according to a task-related Best Resource Ratio (BRR for short), such as "cpu, disk, network in the ratio of1:2:4__, rather than the resource's quantity, such as "cpu = 1 GHz, network = 20MB/s". Best Resource Ratio (BRR) of data processing tasks is estimated to design a resource ratio based approach including a task scheduling algorithm and resource allocation algorithm. At the end, energy efficiency is improved by 10%.

A service-oriented architecture is presented to support big data analytics for energy management task for buildings [139]. Supporting different types of analytics and decision-making operations, this architecture is integrating different technologies such as fog and cloud computing Different kinds of decision making processes are considered to manage energy efficiency for its buildings. Layers and components of architecture is shown in Figure 4.21.

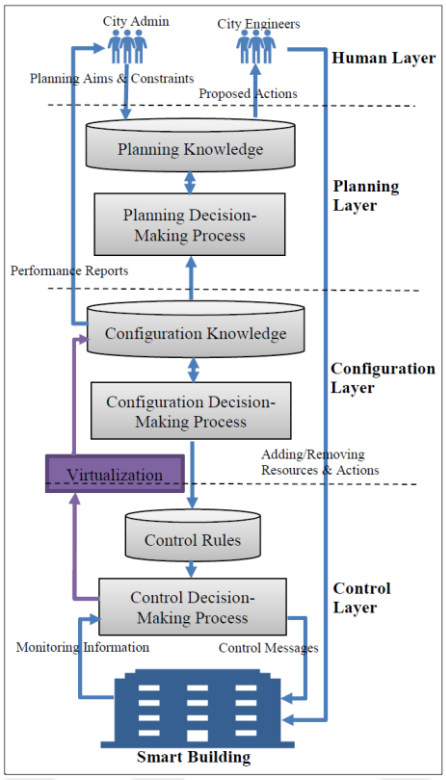


Figure 4.21 Decision-making layers for smart buildings in a smart city.

4.4.2.2.1 Consumption Efficiency

For power systems data management has grown in importance because extracting a significant value out of a huge data has a significant role to make different decisions. An article published to review the concept of big data, methods used for data analytics and challenges with big data in the field of energy [140]. Area of study is building energy across the construction sector. Collection of big amount of data, with various big data analysis techniques can help to enhance energy efficiency in construction process.

Big data mining approach is used to evaluate energy efficiency of China industry sectors by finding the key points for understanding the resources usage patterns, identifying the problematic companies, and establishing good energy consumption practices [141]. K-means clustering and multi-dimensional association rules algorithm is used, to analyze the characteristics of regional energy consumption in different industries and companies. This

provides a reference point to identify the industries and companies in terms of bad consumption practices and environmental performance. The output of research is a big data analytics platform and the results generated by analytics techniques applied specifically to solve regional energy efficiency problems.

Efficiency of not only big data collection in terms of equipment status, maintenance and grid operation in power systems, also data mining is an important research topics for big data application in smart grid. The application of big data technology in fast image recognition of transmission towers are discussed [142]. Recognition is obtained using fixed-wing Unmanned Aerial Vehicle (UAV) by large range tilt photography. A method that using fast region-based convolutional neural networks (Rcnn) convolutional architecture for fast feature embedding (Caffe) is used. To get deep learning, the image characteristics of the tower is extracted, the tower model is sequenced, and quick recognization of transmission tower image is achieved. The case study shows that efficiency improved around 14.2%.

Requirements of global competition and energy saving requirements, it is needed to seek out opportunities to reduce energy waste and overall cost. A data-driven stochastic manufacturing system modeling method is proposed [143]. Identification and prediction of energy saving opportunities is the main goal of study. A real-time distributed feedback production control policy is established. Policy integrates the current and predicted system performance and is to improve the overall profit and energy efficiency. A case study held to demonstrate the effectivity of proposed policy.

Excessive energy consumption is a big problem of supercomputers. Minimizing this energy consumption energy-conscious technologies are required, both hardware level and software level. Marcher is a heterogeneous high performance computing infrastructure which is used in big data analytics. The Marcher system is consists of Intel Xeon CPUs, Intel Many Integrated Cores (Xeon Phi), Nvidia GPUs, power-aware memory systems and hybrid storage with Hard Disk Drives (HDDs) and Solid State Disks (SSDs). These primary computing components endows user easy-to-use tools and interfaces to obtain decomposed and fine-grained power consumption data. The design of the Marcher system is presented in a paper.

Usage of Marcher power measurement tools in various projects is shown [144]. Figure 4.22 shows job scheduler design of system.

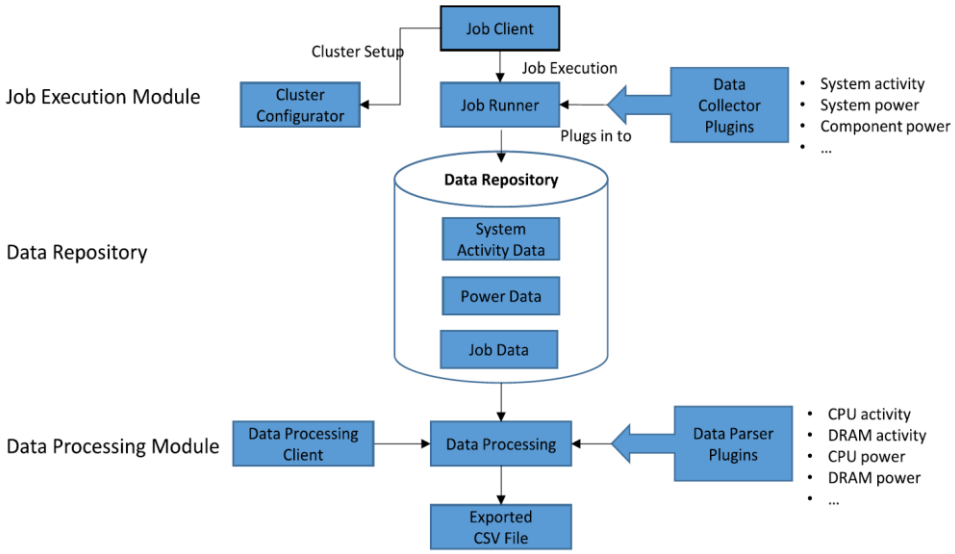


Figure 4.22 Workflow of Job Scheduling

Household energy consumption behavior is the main factor that influence energy efficiency and promote energy conservation. So understanding and changing behavior is important for managing energy systems. Penetration of conventional and emerging information and communication technologies (ICTs) digitized energy systems. This digitization is mostly advantageous analyze and understand individuals' energy consumption behavior. Proposal of the interdisciplinary research of energy framework is held [145]. This is study is in the area of energy social science, social informatics and energy informatics. Household energy consumption behavior is presented in different dimensions and paradigms. Time, user and spatial dimensions are household energy consumption behavior dimensions. Two major research paradigms are economic paradigm and the behavior-oriented paradigm.

Efficient usage of energy is significant to block energy consumption growth without compromising on the customers' requirements. Improving efficiency requires studying on usage patterns and practices. Granular energy usage data collected by smart energy grids data is generated in large volumes and in various formats which makes it big data.

A thesis published in literature focuses on big data tools and techniques such as MapReduce and Apache Hadoop ecosystem tools [146]. Collecting, processing and analyzing energy data to generate information that can be used to improve efficiency.

Studying energy efficiency to formulate the use cases is another aim of study. Author's analysis was performed on two data sets. The first one is hourly data of electricity consumption of different buildings. This data is analyzed to explore the seasonal and daily usage trends and to classify buildings on the basis of energy efficiency. This analysis is also used to build a model for examining inefficient buildings separately. Second data set is a set from device level electricity consumptions of different home appliances used in an apartment. This data is used to evaluate prediction models to forecast future consumption.

High financial cost is a big problem to be solved for large-scale big data processing systems. Various type of hardware and struggle between concurrent loads increases execution time and energy need. Detection and handling are two phases in reduction of struggle. In the detection phase, slow tasks are classified as stragglers, where slow means having speed under average execution time. Handling is performed after detection by the speculative execution technique. This technique involves copying the detected straggler and launching in parallel with the original straggler. Hence, reduction of straggler's execution time achieved.

Improvement of speculative execution technique is studied many times before, but only a few of them considered efficiency of proposed solution. An experimental study is conducted to completely understand the effects of this technique to performance and energy consumption of big data processing systems [147]. It is observed that current straggler mitigation techniques are not efficient. Consequently, a new energy-driven straggler detection mechanism is proposed, which is implemented into Hadoop and is shown to have higher energy efficiency and a new speculative copy allocation method is presented. Speculative copy allocation method differs from alternatives since it takes the impact of resource heterogeneity into consideration.

At last part, an energy efficient straggler handling mechanism, which provides more resource availability is introduced. This resource is used for launching speculative copies and adopting

a dynamic resource reservation approach. It is seen after study that a trace-driven simulation improves energy efficiency.

4.4.2.2.2 Cost Efficiency

Cost efficiency is a big task to come through. Usage of big data analytics to forecast electricity price is another area of big data searches [148]. There are some obstacles on the way to this kind of a solution. First of all there is a huge price data to process. There would be unavoidably redundancy from feature selection and a lack of integrated infrastructure for coordinating the procedures in electricity price. To solve this a modal proposal implemented by integrating three modules is done. Random Forest (RF) and Relief-F algorithms merged on Grey Correlation Analysis (GCA) to be a hybrid feature selector. An integration of Kernel function and Principle Component Analysis (KPCA) is used for feature extraction process to realize the dimensionality reduction. And to forecast price classification, a differential evolution (DE) based Support Vector Machine (SVM) classifier is suggested. Layer structure is shown in Figure 4.23.

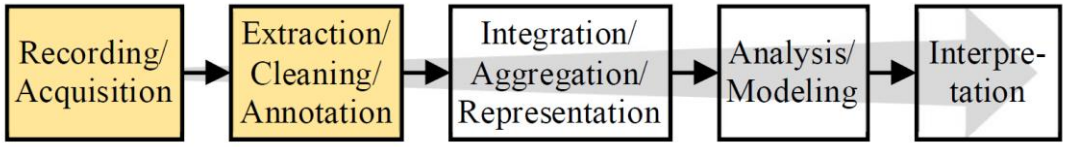


Figure 4.23 Big-data pipeline

Two-way flow of power is valid for smart electricity grid. Data exchanged to and from suppliers and consumers in order to optimize power flow. Economic efficiency, reliability and sustainability are key words for optimization. The most important challenge in a smart grid (SG) is the cost reduction of power, whereas effective energy management heavily depends on forecasting load and renewable production. Thus, data analytics, high performance computing, efficient data management, and cloud computing techniques are critical towards the optimized operation of SGs. One of the researches about subject done to highlight big data issues and challenges faced by DEM which is a must in SG networks [149]. A brief description about most common data processing methods is also provided.

Manufacturing optimization is another common usage of are of big data driven approach. Developing of an innovative big data enabled Intelligent Immune System (I2S) is an example of this [150]. This development endows system monitoring, analyzing and optimization machining processes over lifecycles in order to achieve energy efficient manufacturing. Two major functions are stands to be basic for system; Artificial Neural Networks (ANNs)-based algorithm for identifying abnormal electricity consumption and abnormal electricity consumption for optimization. The applicability validation of the system is done by multiple industrial trials in European factories. It is seen that approximately 30% energy saving and over 50% productivity improvement have been achieved.

A Smart Manufacturing (SM) system should be efficient in power consumption. Capabilities of handling high volume data, processing with high velocity and extracting useful information are required to catch efficiency. Big data analytics uses machine learning and predictive analytics together to make correct decisions. The objective of a document in literature is to present big data analytics modeling in the metal cutting industry [151]. Three studied context listed;

- identification of manufacturing data
- design of a functional architecture for deriving analytic models
- design of an analytic model to predict a sustainability performance especially power consumption, using the big data infrastructure.

For tests a prototype system has been developed which uses open platform solutions such as MapReduce, Hadoop Distributed File System (HDFS). A machine-learning tool is implemented also. In addition cause-effect relationships are determined.

It is often presented easy and low-cost intervention applications to alter individual behaviors for determining policy of behavioral insights. However, there is limited insight into the cost effectiveness of these attempts. On the other hand, there has been little focus on the intersection of large datasets and behavioral insights. A paper presents a conceptual overview on the intersection of big data and behavioral knowledge in the renewable energy sector [152]. Due to technological trajectories, inserting behavioral insights into the energy instrument mix is complex. Large behavioral datasets that can inform can be used not only for policy

implementation, but also policy design and evaluation planning. Findings enables author to introduce future research streams of government capacity in combination of behavioral insights and data. Also compatibility of this information with existing policy instruments is seen.

4.4.3 Dynamic Energy Management

Dynamic energy management (DEM) is to manage load at demand side in an effective way based on an innovative approach.

An integrated system having smart end-use devices and distributed energy resources, an advanced control system and communications capabilities enabled system to managed dynamically. This brings incorporation of the conventional energy management principles represented in demand side management, demand response, and distributed energy resource programs. A framework which simultaneously addresses permanent energy savings, permanent demand reductions, and temporary peak load reductions is merged into the system. This measures are simultaneously implemented which promotes system to be more than a conventional energy use management, eliminating any inherent inefficiencies. Illustration in Figure 4.24 shows building blocks of the dynamic energy management concept.

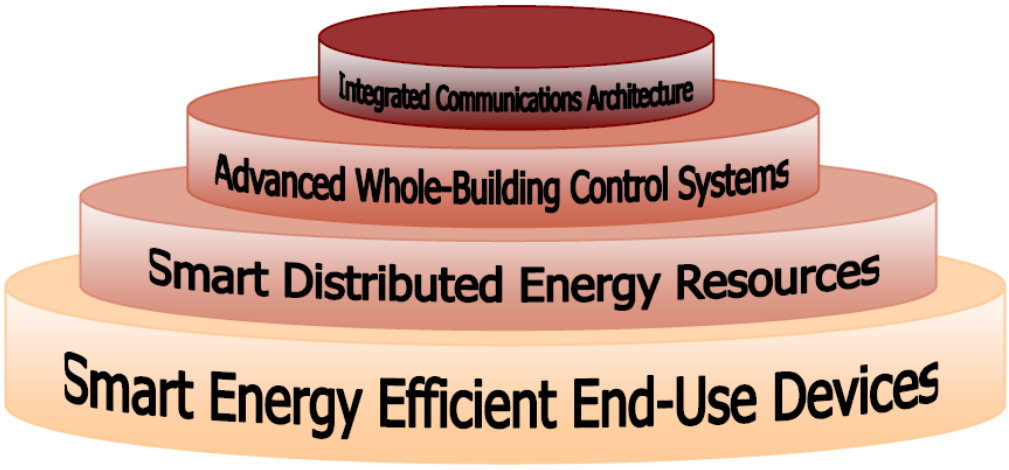


Figure 4.24 Building blocks of dynamic energy management

Reliable and Feasible micro energy generated due to Dynamic Energy Management (DEM) and the electricity market including consumers and suppliers. Smart grid aims to reduce the cost of power by the help of their enhanced operations such as data analytics, high performance estimation, adequate data network management and cloud computing. Some studies focuses on the issues in big data and difficulties of it experienced by the DEM handled in Smart Grid [153]. Data processing techniques that are mostly implemented are explained in details, a brief description of the most commonly used data processing methods is given and recommended proposes a directional future research in the field.

4.5 Power Quality

Intelligent energy networks (IENs) concept based heavily on data analysis. An article published to review and discuss the application of data analysis methods processing on energy big data required for clustering, demand forecasting, energy generation optimization, energy pricing, monitoring and diagnostics issues [154]. pattern recognition, machine learning, data mining, statistics methods are examples of data analysis technologies . Article identifies that there is a lack of existing methods of big data analysis so as to meet requirements for processing the data produced by IENs.

By an article a preliminary and systematic study on the field of energy informatics is presented [155].A comprehensive discussion of the fundamentals and standardization of energy informatics is also included. Moreover, several key open issues are identified.

An edition is interested in studies that examines big data and power in details [156]. Articles from various disciplines who analyze the fields agriculture, education, consumer policy are collected. What big data is and how it relates to power is discussed.

4.6 Fault Diagnosis and Asset Management

A study about big data based energy systems is presented [157] to show potential of energy big data and basics of smart energy management. The sources and characteristics of energy big data is discussed. A process model is proposed. Four major points of view is considered;

- Power generation side management
- Microgrid and renewable energy management
- Asset management and collaborative operation
- Demand side management (DSM).

Nowadays, big data have become a key concept, hence researchers are forced to enhance the existing data mining techniques. They aim to come over challenges of the evolved nature of data and to develop new analytic techniques. A study done and documented which provide a detailed comprehensive analysis and discussion of the data mining techniques [158]. Main data mining tasks is shown in Figure 4.25. Study also investigates that could not be used in big data analytics. Considers what their disadvantage is. Finally, a presentation is done for discussion of the application of big data analytics in the field of renewable energy studies.

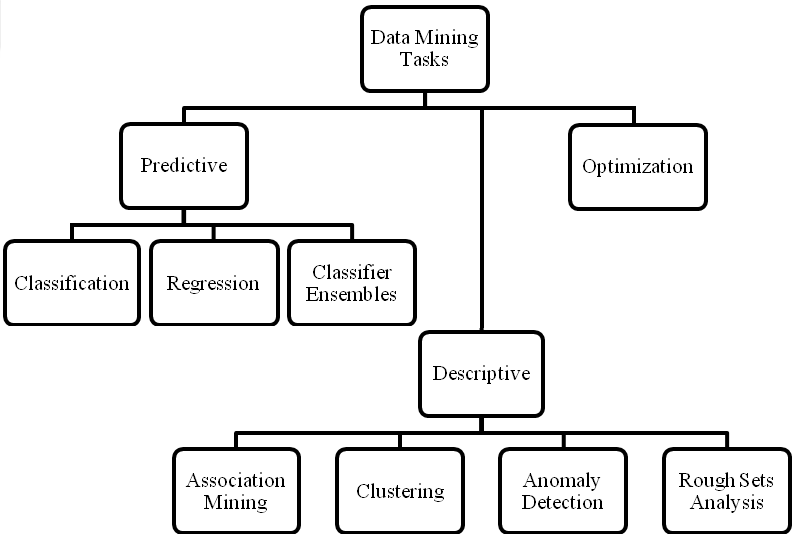


Figure 4.25 Main data mining tasks

4.7 Smart and Sustainable Cities

A study performed a big data analysis on a set of energy consumption data from 2012 up to thesis date for 15000 buildings in Stockholm [159]. Outcomes of a big data analysis wanted to be demonstrated.

KNIME is a big data analysis tool, which is used in this study to find patterns and inefficiencies. Initiating a retrofitting plan for the city to counteract these inefficiencies is another purpose of study. Defined numbers of energy saving scenarios implies estimation of the potential for increased efficiency and the resulting methodology can be used by other cities. Moreover, weather circumstances, building location and building types and their impact is studied. After basic concepts are defined overview of previous studies given.

After understanding and describing the dataset statistically, the annual fluctuations in energy consumption is discussed. Authors clustered consumer groups due to total consumption. Next step was definition of a number of energy saving scenarios that are used to estimate the maximal potential for energy savings, regardless of actual measures, financial or temporal aspects. A more realistic retrofitting plan is proposed to explore the feasibility of Stockholm's Climate Action Plan for 2012-2015. The analysis ended with a spatial regression that determines the influence of wind velocity and n in different parts of Stockholm.

At the end conclusions show that the potential for energy savings in the studied data set can go up to 59% or 4.6 TWh. Savings are validated financially to be 6% using investment parameters. However, these savings quickly reduced because of a high sensitivity on the input parameters. Clustering analysis did not end with anticipated results, but they can be used as a tool and may be beneficial and feasible to applied investments towards groups of buildings.

Exploring energy performance of building sector in the city of Stockholm is a subject of a thesis in literature [160]. Building energy database is the data platform to study on. Analyzing data collected from 20,000 buildings in Stockholm region, Study involves exploring the relation between building performance and different factors on building energy consumption. Clustering method is used to identify buildings with different energy consumptions. Thesis consists of a brief introduction of several key concepts about energy consumption in buildings, building energy declaration and big data, background information. Data processing, descriptive analysis, regression analysis, clustering analysis and energy saving potential analysis methods are used throughout thesis. Data is firstly processed in MS Excel then reorganized in MS Access. As for the data analysis process, IBM SPSS is used for the descriptive analysis and graphical representation. After definition of performance indicators,

the descriptive analysis presents the energy consumption and composition for different building classifications. Result of this process also give the application details of different systems in various building types. Five different independent variables is analyzed by a linear regression model. Clustering analysis is performed on studied buildings and aimed to enhance efficiency of inefficient group of buildings. By this study it is proved that although clustering method is quite useful in the big data analysis, some parameters of clustering needs to be adjusted in order to achieve better results. Energy saving potential calculation is done for studied buildings. After calculations results show that the maximal potential for energy savings in buildings is estimated at 43% (2.35 TWh) for residential buildings and 54% (1.68 TWh) for non-residential premises.

4.8 Electrical vehicles

Power industry promotes the transformations to achieve a strong and excellent strategic objective [161]. It proposed the system construction. Building the system is not only the upgrading and transformation of power facilities, but also a comprehensive and in-depth innovation of power operation mode and business model. Firstly, it makes information and communication technologies fuse rapidly with power production and enterprise management with unprecedented breadth and depth, at the same time it makes the services coordinate better and the business flow and information flow tend to be consistent; Power industry's data will further enrich and expand from the aspects of timeliness. Second, taking full advantage of these data based on actual grid, such as in-depth analysis of it and digging deeper worthy, can provide a lot of high value-added services. These value-added services will help power companies refine operational guidance, improve the management level and operational efficiency, and generate a lot of innovative businesses.

Battery technology is known to be a problem for electric vehicles, since chemical reactions inside it are difficult to monitor directly. To ensure the secure and stable operation of the battery management system, a precise mathematical model should be established and this is the subject of some searches [162]. Generally raw data is not suitable for machine learning and a data cleaning method is put forward, after which establishment of a lithium-ion battery model based on deep learning algorithm is held. The error of model based on different

algorithms is compared and proved that data cleaning method achieves good results. Dataset is obtained from electric buses. The mean absolute percentage error is calculated to be stayed under 4%.

Having a strategy of energy recovery, a numerical simulation for pure electric vehicle is achieved [163]. Big data calculation techniques used on motor loss model. Braking energy of the two-motor dual-axis drive electric vehicle is aimed to be recovered. The efficiency numerical simulation model is established. Under Different speed and braking conditions, the optimization motor braking torque distribution model was established. Thus, the distribution rule of the power optimization for the front and rear electric mechanism was obtained. Constraint condition assumption was Economic Commission of Europe (ECE) braking safety regulation. A new strategic numerical simulation for regenerative braking torque distribution was developed. The results obtained from numerical simulations show that under the proposed strategy, the average utilization efficiency of the motor system is increased by 3.24%. Relation between automobile braking force with the driver's driving behavior and energy recovery efficiency is analyzed.

Power quality of large-scale electric vehicles (EVs) applied into the future smart grid may be a problem in some cases. Besides, Vehicle-to-grid (V2G) technology is a candidate to solve these probable problems. Prediction of schedulable capacity (SC) of EVs is the first step to use this advantages. An article published for proposing two different time scales schedulable capacity forecasting (SCF) methods [164]. These are real-time SCF method and a day head SCF method. With tests parallel decision tree regression algorithm (PDTR) and parallel random forest regression algorithm (PRFR) are compared and seen that PDTR is faster and better than PRFR. Procedure of processing is shown in Figure 4.26.

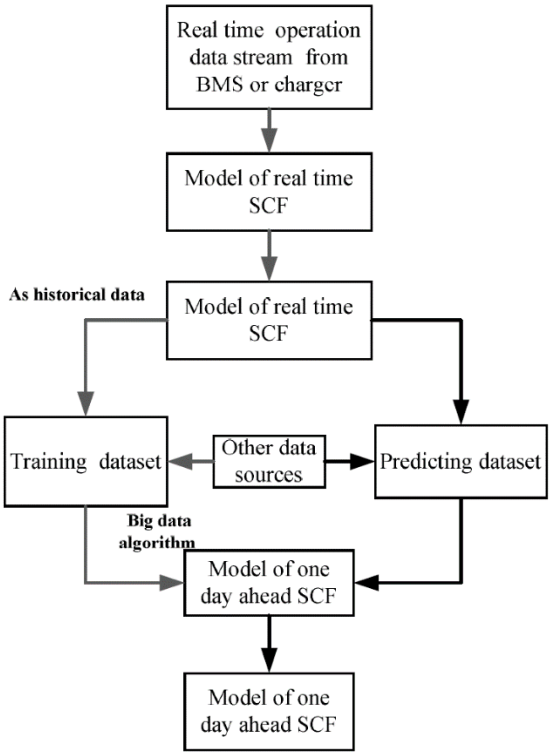


Figure 4.26 Data processing procedure

A study done upon a forecasting big data driven model to estimate electric vehicle charging demand [165]. Unlike previous studies real-world traffic distribution data and weather conditions are considered. Experimental data of historical traffic and weather is of South Korea. During forecasting processes a cluster analysis used to classify traffic patterns, a relational analysis used to identify influential factors. To establish a classification criteria decision tree is used. Charging starting time and the initial state-of-charge of a battery are variables in this study. Case studies held during each weekday and weekend. Summer and winter seasons are also considered. These results were presented to show the different charging load profiles of electric vehicles.

4.9 Challenges and future perspectives

Still big data can be used to reveal demands, automatic decision making, forecasting and etc. However, big data technology applied to power system is thought to be in its infancy stage.

On the way to setup advanced systems there inevitably would be challenges in the future. Today, these difficulties are analyzed and classified to be a base for future researches [166].

Data integration and storage: Classical data analysis usually considers data collected from single source, it is essential to integrate dataset collected from multisource, which has different standards, formats, and representations. Although some storage systems are designed to collect big data, it still needs to be developed and revised in order to accommodate big data collected from power grid.

Data processing: Fault detection and oscillation detection are urgent applications for a smart grid for which the reaction time should be minimized. Even in the cloud system which is able to provide fast computation service, a complicated algorithm with a massive amount of data is not so easy to process.

Data compression: Data compression technique is not applicable in Wide Area Monitoring System (WAMS). In order to achieve required high compression ratio to detect disturbances, some special compression methods are also needed.

Data visualization: Visual graphs and charts are much more effective to warn operators with abnormal values of voltage or frequency. Finding and representing the correlations or trends between multi-source data may become a big challenge. Other challenges lie in visualization algorithms, information extraction & presentation and image synthesis technology [167].

Data security: Smart grids are interconnected in a SCADA system but cyber-attack prevention is not in the considerations of a SCADA system design. A power system could be threatened by meta-data spoofing, wrapping and phishing attacks. Interoperation through APIs prevent grid versus these risks [168].

Data privacy: On the customer side, risk is the increase in private data as a result of smart meters data collection for measurements of household energy consuming. Private data leakage would be a disaster [169].

Another survey paper investigates four challenges related to big data and the autonomous grid operations [30].

- The technical readiness level of cloud computing services
- Limitations of wireless telecommunication technology
- Smart meter related privacy issues
- The intrinsic uncertainty in data analytics

Listed challenges indicate current cloud computing performance and wireless telecommunication abilities is not yet sufficient to automatically control power systems. Furthermore, technical solutions have to be developed to satisfy consumer privacy.

A study discusses about future challenges in managing electric utilities. These utilities have been collecting data in large amounts but they are almost not utilized because they are huge in amount and also there is uncertainty [170]. Future directions on implementing data analytics is shown in Figure 4.27. Challenges tried to be answered by pathways and guidelines to make the current management practices for the future.

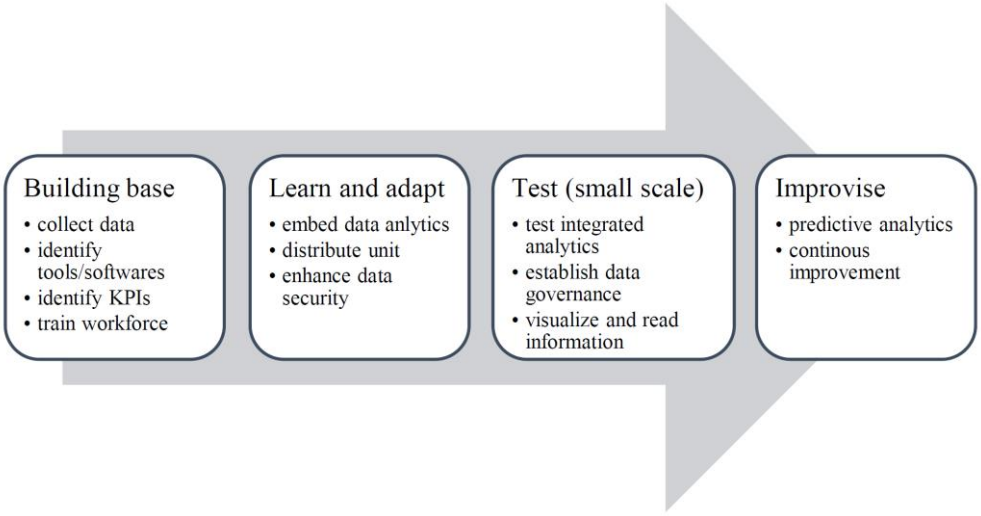


Figure 4.27 Future directions on implementing data analytics

CHAPTER 5

CONCLUSIONS

In this thesis, big data applications in smart grid which is the intelligent center members of electrical distribution systems have been reviewed and discussed. The large scale data which contains valuable information are collected from smart meters installed in the power systems, electricity market, meteorological system and other systems related to electrical distribution system. Given applications in this paper processes big data to solve physical problems, eliminate fault risks, forecast future realizations, compensate external parameters that affect transmission and demand oriented to every state of distribution systems like operation, maintenance, supply/demand projection and protection.

Advanced information and communication technology and developed big data processing techniques are used in concerned systems. Desired features are extracted from raw information and they are used as a feedback critical information again for the system. This process goes on to obtain an effective, cost reduced enhanced system.

Despite developed technological abilities of information technology, performing fast and efficient data analytics on a huge data is still a challenging business. Sometimes it takes much effort to provide cyber security and to satisfy privacy principles. In such a multi-parameter systems there would be difficulties for identifying consumers' behavior analysis. Data analytics in smart grids is a complicated area of work. It involves information technologies, artificial intelligence and mathematics, electrical engineering, meteorological science and some others. So, multidisciplinary teamwork is needed to design and implement big data based electrical distribution systems of future.

REFERENCES

[1] Yu, N., Shah, S., Johnson, R., Sherick, R., Hong, M., Loparo, K., *Big Data Analytics in Power Distribution Systems,* IEEE Power & Energy Society Innovative Smart Grid Technologies Conference*,* 2016.

[2] *Energy Big Data Analytics Unlocking the benefits of Smart metering and Smart Grid Technologies*, DIGITALEUROPE, 2015.

[3] Lai, C. S., Lai, L. L., *Application of Big Data in Smart Grid*, IEEE International Conference on Systems, Man, and Cybernetics, 2015.

[4] *Big data in the cloud: Converging technologies*, Intel IT Center, 2015.

[5] Bitzer, B. , *Application of Cloud Computing for Power Systems*, Trivent Publishing, 2015.

[6] http://en.wikipedia.org/wiki/Quantum_computing, 2015. [Online].

[7] Finogeev, A., Parygin, D., Finogeev, A., *The convergence computing model for big sensor data mining and knowledge discovery*, Human-centric Computing and Information Sciences, 2017.

[8] B. Parhami, Energy Implications of Big Data, in *Encyclopedia of Big Data Technologies*, S. Sakr and A. Zomaya, Eds., Springer, 2019.

[9] Koomey, J. G., Berard, S., Sanchez, M., Wong, H., *Implications of historical trends in the electrical efficiency of computing*, IEEE Ann Hist Comput, vol. 33, no. 3, pp. 46-54.

[10] Zhang, Y., Huang, T., Bompard, E. F., *Big data analytics in smart grids a review*, Energy Informatics, pp. 1-8, 2018.

[11] V. Marinakis, H. Doukas, J. Tsapelas, S. Mouzakitis, Á. Sicilia, L. Madrazo and S. Sgouridis, *From big data to smart energy services: An application for intelligent energy management*, *Future Generation Computer Systems*, 2018.

[12] Zhang, Y., Ma, S., Yang, H., Lv, J., Liu, Y., *A big data driven analytical framework for energy-intensive manufacturing industries*, Journal of Cleaner Production, vol. 197, pp. 57-72, 2018.

[13] Chou, S.-C., Yang, C.-T., Fuu-Cheng, , Chang, J. C.-H., *The Implementation of a Data-Accessing Platform Built from Big Data Warehouse of Electric Loads*, 2018 42nd IEEE International Conference on Computer Software & Applications, 2018.

[14] Al-Ali, A., Zualkernan, I. A., Rashid, M., Gupta, R., AliKarar, M., *A smart home energy management system using IoT and big data analytics approach*, IEEE Transactions on Consumer Electronics, vol. 63, 2017.

[15] Ku, T.-Y., Park, .-K., Choi, H., *IoT energy management platform for microgrid*, IEEE 7th International Conference on Power and Energy Systems (ICPES), 2017.

[16] Guilan , W., Guoliang, ., Hongshan, ., Hongyang, ., *Real-time big data technologies of energy internet platform,* IEEE International Conference on Power System Technology (POWERCON), 2016.

[17] Wang, K., Li, H., Feng , Y., Tian, G., *Big Data Analytics for System Stability Evaluation Strategy in the Energy Internet,* IEEE Transactions on Industrial Informatics, vol. 13, 2017.

[18] N. K. M. S. Anish Jindal, *A unified framework for big data acquisition, storage, and analytics for demand response management in smart cities*, Future Generation Computer Systems, 2018.

[19] Zhou, W. , *Research on the Application of Energy Internet Big Data in Integrated Energy Market*, Energy and Power Engineering, vol. 9, pp. 329-335.

[20] Alahakoon, D., Yu, X., *Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey,* IEEE Transactions on Industrial Informatics, vol. 12, 2016.

[21] Liu, X., Nielsen, P. S., *Streamlining Smart Meter Data Analytics,* Proceedings of the 10th Conference on Sustainable Development of Energy, Water and Environment Systems, 2015.

[22] Wilcoxa, T., Jinb, N., Flachc, P., Thumim, J., *A Big Data platform for smart meter data analytics*, Computers in Industry, vol. 105, pp. 250-259, 2019.

[23] Yao, H.-w., Wang, X.-w., Wu, L.-s., Jiang, D., Liang, D., *Prediction Method for Smart Meter Life Based On Big Data*, Procedia Engineering, vol. 211, pp. 1111-1114, 2018.

[24] Jaradata, M., Jarrah, M., Bousselhamb, A., Jararweha, Y., Al-Ayyouba, M., *The Internet of Energy: Smart Sensor Networks and Big Data Management for Smart Grid*, Procedia Computer Science, vol. 56, pp. 592-597, 2015.

[25] Ahsan, U. , *Distributed Architecture for Managing Big Data in Smart Grid*, Electronic Systems Engineering University of Regina, 2018.

[26] Sulaiman,S. M., Jeyanthy, A., *Big data analytics of smart meter data using Adaptive Neuro Fuzzy Inference System (ANFIS)* in 2016 International Conference on Emerging Technological Trends (ICETT), 2017.

[27] Sun, M. , *Big Data Analytics in Power Systems*, Imperial College London, 2016.

[28] K. Hawes, NIST Cloud Computing Program-NCCP, 2010.

[29] Taieb, S. B., Huser, R., Hyndman, R. J., Genton, M. G., *Forecasting Uncertainty in Electricity Smart Meter Data by Boosting Additive Quantile Regression*, IEEE Trans. smart grid, vol. 7, p. 2448, 2016.

[30] Subhani, S., Gibescu, M., Kling, W. L., *Autonomous control of distributed energy resources via wireless machine-to-machine communication; a survey of big data challenges*, IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), 2015.

[31] Banadaki, A. D., Taufik, , Feliachi, A., *Big Data Analytics in a Day-Ahead Electricity Price Forecasting Using TensorFlow in Restructured Power Systems*, International Conference on Computational Science and Computational Intelligence Las Vegas, Nevada, USA, 2018.

[32] Z. Q. BoEmail author, X. N. L. Q. P. W. Y. H. Y. a. F. Q. Z. , Developments of power system protection and control Protection and Control of Modern Power Systems, 2019.

[33] Goyal, V., Movva, R., Kunadharaju, A., *A Secure Cloud Computing based Framework for Big Data Information management of smart Grids,* International Journal of Advanced Research in Computer and Communication Engineering.

[34] Zeng, S. , *An Intrusion Detection System Based on Big Data for Power System*, International Symposium on Advances in Electrical, Electronics and Computer Engineering (ISAEECE 2016), 2016.

[35] Kezunovic, M. Obradovic, Z. Dokic, T. Zhang, B. Stojanovic, J. DehghanianP. and ChenP.-C. , Predicting Spatiotemporal Impacts of Weather on Power Systems Using Big Data Science., in Data Science and Big Data: An Environment of Computational Intelligence, 2017, pp. 265-299.

[36] Wang, Y., Zhang, L., A Combined Fault Diagnosis Method for Power Transformer in Big Data Environment. Hindawi, vol. 2017, 2017.

[37] Ebrahimy, R., Pourmirza, Z., Cyber-interdependency in Smart Energy Systems. In Proceedings of the 3rd International Conference on Information Systems Security and Privacy (ICISSP 2017), pp. 529-537, 2017.

[38] Wang , B., Fang , ., Wang , ., Liu , ., Liu, ., Power System Transient Stability Assessment Based on Big Data and the Core Vector Machine IEEE Transactions on Smart Grid, vol. 7, 2016.

[39] Gu, Y., Jiang, H., Zhang, Y., Zhang, J. J., Gao, T., Muljadi, E., Knowledge Discovery for Smart Grid Operation, Control, and Situation Awareness –A Big Data Visualization Platform. North American Power Symposium (NAPS), 2016.

[40] Kezunovic, M., Dokic, T., Chen, P. C., Malbasa, V., Improved Transmission Line Fault Location Using Automated Correlation of Big Data from Lightning Strikes and Fault-induced Traveling Waves. HICCS – Hawaii International Conference on System Science Kauai, Hawaii, 2015.

[41] Khan, M., Li, M., Ashton, P., Taylor, G., Liu, J., Big Data Analytics on PMU Measurements. 2014.

[42] Buffat, R., Froemelt, A., Heeren, N., Raubal, M., Hellweg, S., Big data GIS analysis for novel approaches in building stock modelling. Applied Energy, vol. 208, pp. 277-290, 2017.

[43] GU, Y., Renewable Energy Integration in Distribution System – Synchrophasor Sensor Based Big Data Analysis, Visualization, and System Operation. Faculty of School of Engineering and Computer Science University of Denver, 2018.

[44] Ketter, W., Peters, M., Collins, J., Gupta, A., Competitive Benchmarking: An IS Research Approach to Address Wicked Problems with Big Data and Analytics MIS Quarterly ERIM Report Series, 2018.

[45] YAN, Y., SHENG, G., QIU, R. C., JIANG, X., Big Data Modeling and Analysis for Power Transmission Equipment: A Novel Random Matrix Theoretical Approach. IEEE Access / Special Section On Data-Driven Monitoring, Fault Diagnosis And Control Of Cyber-Physical Systems, vol. 6, 2018.

[46] Akhavan-Hejazi, H., Mohsenian-Rad, H., Power systems big data analytics: An assessment of paradigm shift barriers and prospects. Energy Reports 4, pp. 91-100, 2018.

[47] Perçuku, A., Minkovska, ., Stoyanova, L., Mining optimum models of generating solar power based on big data analysis. Procedia Computer Science, vol. 141, pp. 167-174, 2018.

[48] Zaman, I., Pazouk, K., Norman, R., Younessi, S., Coleman, S., Challenges and Opportunities of Big Data Analytics for Upcoming Regulations and Future Transformation of the Shipping Industry. Procedia Engineering, vol. 194, pp. 537-544, 2017.

[49] Herman, J., Herman, H., Mathews, M. J., Vosloo, J. C., Using big data for insights into sustainable energy consumption in industrial and mining sectors. Journal of Cleaner Production, vol. 197, pp. 1352-1364, 2018.

[50] Yuan, L., Chen, H., Tian1, B., Gong, J., The Application Analysis of Data Mining-Based Energy Internet. Advances in Computer Science Research, vol. 86, 2018.

[51] Zhang, D., Qiu, R. C., Research on big data applications in Global Energy Interconnection. Global Energy Interconnection, vol. 1, pp. 352-357, 2018.

[52] Wang, L., Shang, L., Ma, M., Ma, Z., Fault Diagnosis and Trace Method of Power System Based on Big Data Platform IOP Conf. Series: Materials Science and Engineering 394, 2018.

[53] Wang, G.-p., Wen, X.-j., Research on Electric Power Monitoring System Based on Wireless Big Data Platform. Journal of Computers, vol. 29, pp. 110-117, 2018.

[54] Wen, L., Zhou, K., Yang, S., Li, L., Compression of smart meter big data: A survey. Renewable and Sustainable Energy Reviews, vol. 91, pp. 59-69, 2018.

[55] Veiga, J., Enes, J., Expósito, R. R., Touriño, J., BDEv 3.0: Energy efficiency and microarchitectural characterization of Big Data processing frameworks. Future Generation Computer Systems, vol. 86, pp. 565-561, 2018.

[56] Perçuku, A., Minkovska, D., Stoyanova, L., Modeling and Processing Big Data of Power Transmission Grid Substation Using Neo4j. Procedia Computer Science, vol. 113, pp. 9-16, 2017.

[57] Wu, W., Lin, W., Hsu, C.-H., He, L., Energy-efficient hadoop for big data analytics and computing: A systematic review and research insights. Future Generation Computer Systems, vol. 86, pp. 1351-1367, 2018.

[58] Kirihara, K., Big Data In Power Systems: A Statistical Approach On Synchrophasor Application, University of Illinois at Urbana-Champaign, 2016.

[59] Perrons, R. K., W, J. J., Data as an asset: What the oil and gas sector can learn from other industries about Big Data. Energy Policy, vol. 81, pp. 117-121, 2015.

[60] Yu, L., Zhao, Y., Tang, L., Yang, Z., Online big data-driven oil consumption forecasting with Google trends. International Journal of Forecasting, vol. 35, pp. 213-223, 2019.

[61] Yang, Z., Ciwei, G., Zhao, M., Utilizing Big Data to Explore The Running Opportunity of Powerto-Gas in Energy System. Energy Procedia, vol. 158, 2019.

[62] Perrotta, F., Parry, T., Neves, L. C., Buckland, T., Mesgarpour, M., Verification of the HDM-4 fuel consumption model using a Big data approach: A UK case study. Transportation Research Part D: Transport and Environment, vol. 67, pp. 109-118, 2019.

[63] Park, E. , Positive or negative? Public perceptions of nuclear energy in South Korea: Evidence from Big Data. Nuclear Engineering and Technology, vol. 51, pp. 626-630, 2019.

[64] Roh, S. , Big Data Analysis of Public Acceptance of Nuclear Power in Korea. Nuclear Engineering and Technology, 2018.

[65] Hua, T., Zheng, M., Tana, J., Zhu, L., Miao, W., Intelligent photovoltaic monitoring based on solar irradiance big data and wireless sensor networks. Ad Hoc Networks, vol. 35, pp. 127-136, 2015.

[66] H. M. N. M. F. L. S. C. Konstantinos J. Chalvatzis, Sustainable resource allocation for power generation: The role of big data in enabling interindustry architectural innovation, Technological Forecasting and Social Change, 2018.

[67] https://www.smartgrid.gov/the_smart_grid/smart_grid.html, 2019. [Online].

[68] Feng, G. Big Data Application in Power Systems., R. Arghandeh and Y. Zhou, Eds., Elsevier, 2018.

[69] Liu, Q., Cui, L., Chen, H., Key technologies and applications of internet of things. Computer Science, vol. 6, no. 37, 2010.

[70] Ukil, A., Zivanovic, R., Automated Analysis of Power Systems Disturbance Records: Smart Grid Big Data Perspective IEEE Innovative Smart Grid Technologies, 2015.

[71] He, X., Ai, Q., Qiu, R. C., Huang, W., Piao, L., Liu, H., A Big Data Architecture Design for Smart Grids Based on Random Matrix Theory. IEEE Transactions On Smart Grid, vol. 8, 2018.

[72] Chen, J., Chen, H. Y., Wen, M., Sun, S., Industrial power demand forecasting based on big data technology orienting to energy internet: A case study of Hunan Province IOP Conference Series: Earth and Environmental Science 188, 2018.

[73] Bishnu P. Bhattarai, S. P. , Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions 2019.

[74] Guo, Y., Yang, Z., Feng, S., Hu, J., Complex Power System Status Monitoring and Evaluation Using Big Data Platform and Machine Learning Algorithms: A Review and a Case Study Hindawi Complexity, vol. 2018.

[75] Shyam, R., Bharathi Ganesh, H. B., Sachin Kumar, S., Poornachandran, P., Soman, K. P., Apache Spark a Big Data Analytics Platform for Smart Grid. Procedia Technology, vol. 21, pp. 171-178, 2016.

[76] C. Chelmis, J. Kolte and V. K.Prasanna, Big data analytics for demand response: Clustering over space and time., in IEEE International Conference on Big Data, 2015.

[77] Benhaddou, D., Mohamed, R. A., Achahbar, O., Khalil, N., Rachidi, T., Al Assaf, M., BIG DATA PROCESSING FOR SMART GRIDS. IADIS International Journal on Computer Science & Information Systems, vol. 10, pp. 32-46, 2015.

[78] Chou, J.-S., Ngo, N.-T., Smart grid data analytics framework for increasing energy savings in residential buildings. Automation in Construction, vol. 72, pp. 247-257, 2016.

[79] Guo, Y., Feng, S., Li, K., Mo, W., Liu, Y., Wang, Y., Big Data Processing and Analysis Platform for Condition Monitoring of Electric Power System. UKACC 11th International Conference on Control Belfast, 2016.

[80] Troiano , L., Vaccaro , ., Vitelli, . C., On-line smart grids optimization by case-based reasoning on big data. IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS), 2016.

[81] Guo, L., Yan, H., Hao, Y., Chen, Y., Comprehensive analysis and evaluation of big data for main transformer equipment based on PCA and Apriority. IOP Conf. Series: Earth and Environmental Science 108, 2018.

[82] Zhang, J. , Application and Challenges of Big Data and Cloud Computing in Power Industry International Symposium on Communication Engineering & Computer Science (CECS 2018), 2018.

[83] Zhanga, Y., Hea, Y., Li, S., Li, X., Chena, Z., Dinga, Y., Guoa, M., Key Technologies Research of Urban Power Grid AVC System with High Load Density Based on Big Data. Energy Procedia, vol. 152, pp. 1200-1205, 2018.

[84] Kaur, D., Aujla, . S., Kumar, ., Zom, . Y., Tensor-Based Big Data Management Scheme for Dimensionality Reduction Problem in Smart Grid Systems: SDN Perspective IEEE Transactions on Knowledge and Data Engineering, vol. 30, 2018.

[85] Sagiroglu, S., Terzi, R., Canbay, Y., Colak, I., Big Data Issues in Smart Grid Systems. IEEE International Conference on Renewable Energy Research and Applications (ICRERA), 2016.

[86] Jiang, H., Wang, K., Wang, Y., Gao, M., Energy Big Data: A Survey, IEEE Access Special Section On Theoretical Foundations For Big Data Applications: Challenges And Opportunities, 2016.

[87] Schuelke-Leech, B.-A., Barry, B., Muratori, M., Yurkovich, B. J., Big Data issues and opportunities for electric utilities Renewable and Sustainable Energy Reviews, vol. 52, pp. 937-947, 2015.

[88] Fusco, F., Fischer, U., Lonij, V., Pompey, P., Fiot, J.-B., Chen, B., Gkoufas, Y., Sinn, M., Data Management System for Energy Analytics and its Application to Forecasting EDBT/ICDT Workshops, 2016.

[89] Simonov, M., Li, H., Chicco, G., Gathering Process Data in Low-Voltage Systems by Enhanced Event-Driven Metering. IEEE Systems Journal, vol. 11, 2017.

[90] Ahsan , U., Bais, ., Distributed big data management in smart grid. 26th Wireless and Optical Communication Conference (WOCC), 2017.

[91] Munshi, A. A., Mohamed, Y. A. -R. I., Big data framework for analytics in smart grids. Electric Power Systems Research, vol. 151, pp. 369-380, 2018.

[92] Daki, H., Hannani, A. E., Aqqal, A., Haidine, A., Dahbi, A., Big Data management in smart grid: concepts, requirements and implementation Journal of Big Data, 2017.

[93] Dang-Ha, T.-H., Olssony, R., Wang, H., The Role of Big Data on Smart Grid Transition. 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), 2017.

[94] Huang, Y., Zhou, X., Knowledge model for electric power big data based on ontology and semantic web. CSEE Journal Of Power And Energy Systems, vol. 1, 2015.

[95] Nasiakou, A., Alamaniotis, ., Tsoukala, L. H., Power distribution network partitioning in big data environment using k-means and fuzzy logic. Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion, 2016.

[96] Bibri, S. E., Krogstie, J., The core enabling technologies of big data analytics and context-aware computing for smart sustainable cities: a review and synthesis. Journal of Big Data, 2017.

[97] EliasBibri, S. , The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. Sustainable Cities and Society, vol. 38, pp. 230-253, 2018.

[98] Peppanen, J., Reno , . J., Broderick , . J., Grijalva, ., Distribution System Model Calibration With Big Data From AMI and PV Inverters. IEEE Transactions On Smart Grid, vol. 7, 2016.

[99] Kıranlar, o. , Big Data Applications in the Energy Sector. 2017.

[100] Preda, S., Oprea, S.-V., Bâra, A., (Velicanu), A. B., PV Forecasting Using Support Vector Machine Learning in a Big Data Analytics Context. Symmetri 2018, vol. 10, p. 748, 2018.

[101] Zhou, Z., Xiong, F., Huang, B., Xu, C., Jiao, R., Game-Theoretical Energy Management for Energy Internet With Big Data-Based Renewable Power Forecasting. IEEE Access, vol. 5, 2017.

[102] Amaro, N., Pina, J. M., Big Data in Power Systems Leveraging grid optimization and wave energy integration. 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC), 2017.

[103] F. X. C. X. a. R. J. Zhenyu Zhou, Energy Management in Microgrids A Combination of Game Theory and Big Data-Based Wind Power Forecasting.

[104] Shu, T., Wu, C. Q., Energy-efficient Mapping of Large-scale Workflows Under Deadline Constraints in Big Data Computing Systems. Future Generation Computer Systems, 2017.

[105] Ifaei, P., Karbassi, A., Lee, S., Yoo, C., A Renewable Energies-Assisted Sustainable Development Plan For Iran Using Techno-Econo-Socio-Environmental Multivariate Analysis And Big Data. Energy Conversion and Management, vol. 153, pp. 257-277, 2017.

[106] Guerrero-Lemus, R., Cañadillas-Ramallo, D., Reindl, T., Valle-Feijóo, J. M., A Simple Big Data Methodology And Analysis Of The Specific Yield Of All PV Powerplants In A Power System Over A Long Time Period, Renewable and Sustainable Energy Reviews, vol. 107, pp. 123-132, 2019.

[107] Viscondi, G. d. F., Alves-Souza, S. N., A Systematic Literature Review on big data for solar photovoltaic electricity generation forecasting. Sustainable Energy Technologies and Assessments, vol. 31, pp. 54-63, 2019.

[108] Li, Z., Ye, L., Zhao, Y., Song, X., Teng, J., Jin, J., Short-Term Wind Power Prediction Based On Extreme Learning Machine With Error Correction, Prot Control Mod Power Syst (2016) 1: 1. https://doi.org/10.1186/s41601-016-0016-y.

[109] Guan, H., Zhong, T., He, H., Zhao, T., Xing, L., Zhang, Y., Xue, X., A self-powered wearable sweat-evaporation-biosensing analyzer for building sports big data. Nano Energy, no. 59, pp. 754-761, 2019.

[110] Yan, L., Yu, S., Gang, L., Weiqiang, L., Power Big Data Fusion Prediction. Computer Technology and Application, vol. 7, pp. 165-171, 2016.

[111] Wei, H., Qian, M., Chao, F., Zheng, X., Cong, J., Chunlin, Z., Short Term Load Forecasting Research Based on Electricity Big Data. 4th International Conference on Advanced Materials and Information Technology Processing (AMITP 2016), 2016.

[112] Zhang, P., Wu, X., Wang, X., Bi, S., Short-Term Load Forecasting Based on Big Data Technologies. CSEE JOURNAL OF POWER AND ENERGY SYSTEMS, vol. 1, 2015.

[113] Chalvatzis, K. J., Malekpoor, H., Mishra, N., Lettice, F., Choudhary, S., Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests, Journal of Decision Systems, 2015.

[114] Zhong, T., Guan, H., Dai, Y., He, H., Xing, L., Zhang, Y., Xue, X., A self-powered flexibly-arranged gas monitoring system with evaporating rainwater as fuel for building atmosphere big data. Nano Energy, vol. 60, pp. 52-60, 2019.

[115] Galicia, A., Talavera-Llames, R., Troncoso, A., Koprinska, I., Martínez-Álvarez, F., Multi-step forecasting for big data time series based on ensemble learning. Knowledge-Based Systems, vol. 163, pp. 830-841, 2019.

[116] Singh, S., Yassine, A., Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting. 2018.

[117] Perçuku, A., Minkovska, D., Stoyanova, L., Big Data And Time Series Use In Short Term Load Forecasting In Power Transmission System Procedia Computer Science, vol. 41, pp. 167-174, 2018.

[118] Chen, P.-C., Kezunovic, M., Load Consumption Prediction Utilizing Historical Weather Data and Climate Change Projections. International Conference on Intelligent Systems Applications to Power (ISAP), 2017.

[119] Saber, A. Y., Alam, A. K. M. R., Short term load forecasting using multiple linear regression for big data. 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2018.

[120] Zhou, Z., Yu, H., Xu, C., Chang, Z., Mumtaz, S., Rodriguez, J., BEGIN: Big Data Enabled Energy-Efficient Vehicular Edge Computing. IEEE Communications Magazine, vol. 56, 2018.

[121] Grolinger, K., L'Heureux, A., Capretz, M. A. M., Seewald, L., Energy Forecasting for Event Venues: Big Data and Prediction Accuracy. Energy and Buildings, vol. 112, pp. 222-223, 2016.

[122] Naranjo, P. G. V., Shojafary, M., Vaca-Cardenasz, L., Canali, C., Lancellottiy, R., Baccarelli, E., Big Data Over SmartGrid-A Fog Computing Perspective. Proc. SOFTCOM Workshop, 2016.

[123] Jensen, T. V., Pinson, P., Data Descriptor: RE-Europe, a large-scale dataset for modeling a highly renewable European electricity system. SCIENTIFIC DATA, 2017.

[124] Hu, B., Chuan Pang, L. W. ,. H. C., Mao, C., Big Data Management and Application Research in Power Load Forecasting and Power Transmission and Transformation Equipment Evaluation. Bo Hu et al 2018 J. Phys.: Conf. Ser., 2018.

[125] Rahman, M. N., Esmailpour, A., Zhao, J., Machine Learning with Big Data An Efficient Electricity Generation Forecasting System. Big Data Research, vol. 5, pp. 9-15, 2016.

[126] Chen, P.-C., Dokic, T., Kezunovic, M., The use of big data for outage management in distribution. CIRED Workshop Rome, 2014.

[127] Wang, K., Xu, C., Zhang, Y., Guo, S., Zomaya, A. Y., Robust Big Data Analytics for Electricity Price Forecasting in the Smart Grid. IEEE Transactions on Big Data, vol. 5, no. 1, 2019.

[128] Mujeeb, S., Javaid, N., Khalid, R., Nazeer, O., Sha, I., Khan, M., Big Data Analytics for Price and Load Forecasting in Smart Grids. the 13th International Conference on Broad-Band Wireless Computing, Communication and Applications (BWCCA-2018), 2018.

[129] Fernández, M. R., García, A. C., Alonso, I. G., Casanova, E. Z., Using the Big Data generated by the Smart Home to improve energy efficiency management. Energy Efficiency, vol. 9, pp. 249-260, 2016.

[130] Wang, P., Liu, B., Hong, T., Electric load forecasting with recency effect: A big data approach. International Journal of Forecasting, vol. 32, pp. 585-597, 2016.

[131] Noussan, M., OrcID, , Nastasi, B., Data Analysis of Heating Systems for Buildings—A Tool for Energy Planning, Policies and Systems Simulation. Energies 2018, 2018.

[132] Lachhab, F., Bakhouya, M., Ouladsine, R., Essaaidi, M., Towards an Intelligent Approach for Ventilation Systems Control using IoT and Big Data Technologies. Procedia Computer Science, vol. 130, pp. 926-931, 2018.

[133] Molina-Solana, M., Rosa, M., Ruiza, M. D., Gómez-Romeroa, J., MartinBautistaa, M., Data science for building energy management: A review. Renewable and Sustainable Energy Reviews, pp. 598-609, 2017.

[134] Batra, N., Amarjeet, S., Whitehouse, K., Neighbourhood NILM: A Big-data Approach to Household Energy Disaggregation. Computer Science > Machine Learning, 2015.

[135] Liang, Y. C., Lu, X., Li, W. D., Wang, S., Cyber Physical System and Big Data enabled energy efficient machining optimisation. Journal of Cleaner Production, vol. 187, pp. 46-62, 2018.

[136] Sun, D., Zhang, G., Yang, S., Zheng, W., Khan, S. U., Li, K., Re-Stream: Real-time and energy-efficient resource scheduling in big data stream computing environments. Information Sciences, pp. 92-112, 2015.

[137] WIJAYA, T. K. , Pervasive Data Analytics for Sustainable Energy Systems. EPFL - Ecole Polytechnique F ed erale de Lausanne, 2015.

[138] Song, J., Ma, Z., Thomas, R., Yu, G., Energy efficiency optimization in big data processing platform by improving resources utilization. Sustainable Computing: Informatics and Systems, vol. 21, pp. 80-89, 2019.

[139] Mohamed, N., Al-Jaroodi, J., Jawhar, I., Service-Oriented Big Data Analytics for Improving Buildings Energy Management in Smart Cities. 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC), 2018.

[140] Koseleva, N., Ropaite, G., Big data in building energy efficiency: understanding of big data and main challenges. Procedia Engineering, vol. 172, pp. 544-549, 2017.

[141] Liu, G., Yang, J., Hao, Y., Yan, Z., Big data-informed energy efficiency assessment of China industry sectors based on K-means clustering. Journal of Cleaner Production, vol. 183, pp. 304-314, 2018.

[142] Hu, Z., He, T., Zeng, Y., Luo, X., Wang, J., Huang, S., Liang, J., Sun, Q., Xu, H., Lin, B., Fast image recognition of transmission tower based on big data. Protection and Control of Modern Power Systems, p. 3:15, 2018.

[143] Jing, Z., Chang, Q., Arinez, J., Xiao, G., Data-driven modeling and real-time distributed control for energy ef fi cient manufacturing systems. Energy, vol. 147, pp. 247-257, 2017.

[144] Zong, Z., Ge, R., Gu, Q., Marcher: A Heterogeneous System Supporting Energy-Aware High Performance Computing and Big Data Analytics. Big Data Research, vol. 8, pp. 27-38, 2017.

[145] Zhou, K., Yan, S., Understanding household energy consumption behavior: The contribution of energy big data analytics. Renewable and Sustainable Energy Reviews, vol. 56, pp. 810-819, 2016.

[146] Ahmed, H. , Applying Big Data Analytics for Energy. Aalto University School of Science, 2014.

[147] Phan, T.-D. , Energy-efficient Straggler Mitigation for Big Data Applications on the Clouds. 2018.

[148] Flick, D., Kuschicke, F., Schweikert, M., Thiele, T., Herrmann, C., Ascertainment of Energy Consumption Information in the Age of Industrial Big Data. Procedia CIRP, vol. 72, pp. 202-208, 2018.

[149] Diamantoulakis, P. D., Kapinas, V. M., Karagiannidis, G. K., Big Data Analytics for Dynamic Energy Management in Smart Grids. Big Data Research, vol. 2, pp. 94-101, 2016.

[150] Wang, S., Liang, Y. C., Li, W. D., Cai, X. T., Big Data enabled Intelligent Immune System for energy efficient manufacturing management. Journal of Cleaner Production, vol. 195, pp. 507-520, 2018.

[151] Shina, S.-J., Wooa, J., Rachuria, S., Predictive analytics model for power consumption in manufacturing. 21st CIRP Conference on Life Cycle Engineering, 2014.

[152] Giest, S., Mukherjee, I., Behavioral instruments in renewable energy and the role of big data: A policy perspective. Energy Policy, vol. 123, pp. 360-366, 2018.

[153] Lydia, E. L., Kumar, B. P., Ramya, D., Generation of dynamic energy management using data mining techniques basing on big data analytics isssues in smart grids. International Journal of Engineering & Technology, pp. 85-89, 2018.

[154] Ma, Z., Xie, J., Li, H., Sun, Q., Si, Z., Zhang, J., Guo, J., The Role of Data Analysis in the Development of Intelligent Energy Networks. IEEE Network, vol. 31, 2017.

[155] Huang, B., Bai, X., Zhou, Z., Cui, Q., Zhu, D., Hu, R., Energy informatics: Fundamentals and standardization. ICT Express, vol. 3, pp. 76-80.

[156] Ulbricht, L., Big data: big power shifts?, Internet Policy Review Journal On Internet Regulation, vol. 5, 2016.

[157] Zhou, K., Chao, F., Yang, S., Big data driven smart energy management: From big data to big insights. Renewable and Sustainable Energy Reviews, vol. 56, pp. 215-225, 2016.

[158] Fawzy, D., Moussa, S., Badr, N., The Evolution of Data Mining Techniques to Big Data Analytics: An Extensive Study with Application to Renewable Energy Data Analytics. Asian Journal of Applied Sciences, vol. 4, 2016.

[159] B. v. d. Heijde, Big Data Analytics towards a Retrofitting Plan for the City of stockholm., Industrial Ecology Royal Institute Of Technology, 2014.

[160] Y. MA, Big Data Analytics of City Wide Building Energy Declarations., Industrial Ecology Royal Institute Of Technology, 2014.

[161] Chen, C., Yang, D., Wang, S., Yang, D., Big Data Applications in Power Industry. International Conference on Information Sciences, Machinery, Materials and Energy (ICISMME 2015), 2015.

[162] Li, S., Li, J., He, H., Wang, H., Lithium-ion battery modeling based on Big Data. Energy Procedia, vol. 159, pp. 168-173, 2019.

[163] Xiong, H., Zhu, X., Zhang, R., Energy Recovery Strategy Numerical Simulation for Dual Axle Drive Pure Electric Vehicle Based on Motor Loss Model and Big Data Calculation. Hindawi, vol. 2018, 2018.

[164] Mao, M., Wang, Y., Yue, Y., Chang, L., Multi-time scale forecast for schedulable capacity of EVs based on big data and machine learning. IEEE Energy Conversion Congress and Exposition (ECCE), 2017.

[165] Arias, M. B., Bae, S., Electric vehicle charging demand forecasting model based on big data technologies. Applied Energy, vol. 183, pp. 327-339, 2016.

[166] Tu, C., He, X., Shuai, Z., Jiang, F., Big data issues in smart grid – A review. Renewable and Sustainable Energy Reviews, vol. 79, pp. 1099-1107, 2017.

[167] Kennedy, S. J. , Transforming big data into knowledge: experimental techniques in dynamic visualization. Massachusetts Institute of Technology, 2012.

[168] Cuzzocrea, A. , Privacy and security of big data: current challenges and future research perspectives. Proceedings of the international workshop on privacy & security of big data, vol. 45, p. 7, 2014.

[169] Bertino, E. , Big data – security and privacy. Proceedings of the IEEE international congress on big data, 2015.

[170] Khuntia, S. R., Rueda, J. L., van der Meijden, M. A. M. M., Smart Asset Management for Electric Utilities: Big Data and Future. 12th World Congress on Engineering Asset Management (WCEAM), 2017.

CIRRICULUM VITAE

PERSONAL INFORMATION

Name Surname : Okan Şen

Date of Birth : 04.11.1971

Phone : +90 532 371 25 29

E-mail : okansen1@gmail.com

EDUCATION

High School : Antalya Anatolian Highschool

Bachelor : METU Electrical-Electronics Engineering Department

TOPICS OF INTEREST

- Power Systems