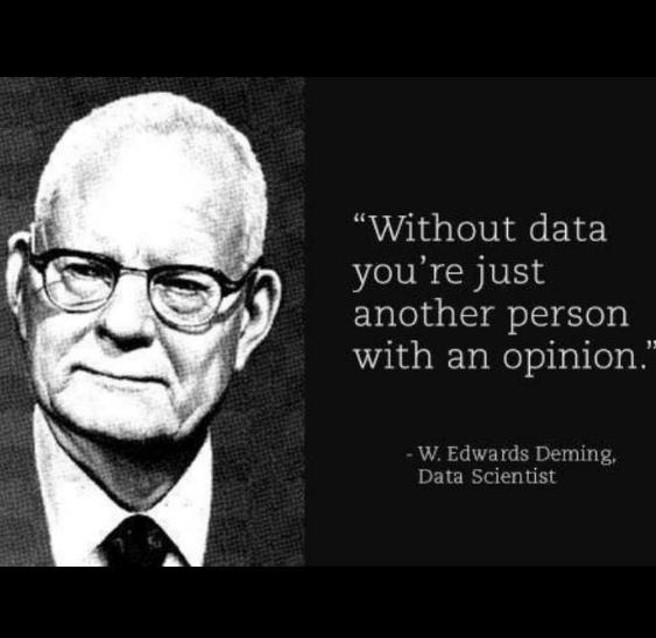
****

**Introduction**

**Data Analytics Engineering is a subfield of engineering. It is a science of collecting, analyzing and making use of raw data to get useful information in order to guide leaders and managers to make an accurate decision. Data analytics is a widely and deeply used subject nowadays, contributing a lot in all kinds of fields, such as finance, medical researching, pharmaceutical, banking, marketing and so on. This technique can keep people from being lost in a huge amount of data and lead people to find out useful information, analyze the information, explore trends and associations among what people are interested in and finally aim to make an optimal decision for future strategies. There are many subfields in data analytics engineering, for example, data mining, statistical analysis, cognitive computing development, machine learning and so on.**

**Writing Center Guide**



Agenda-setting theory describes the ability of the news media to influence the importance placed on the topics of the public agenda [1]. That means, through the media’s selection and display of the news, the repeated news can emphasize the news material so that the audience will consider the repeated news as important ones [1]. This ability to influence the salience of topics on the public agenda has come to be called the agenda-setting role of the news media [1]. This theory has been applied in the big data area: companies induce their consumers’ thought to lay the importance of their own merchandise so that people buy their products [1].



*Agenda——Setting Theory*

*Data Governance*



Governance refers to what decisions must be made to ensure effective management and use of IT (decision domains) and who makes the decisions (locus of accountability for decision-making) [2]. Data analytics is never a single subject or project. It is involved in diverse fields, so how to effectively use the data information to make an accurate decision is vital for projects and leaders. Data governance is exactly the process to help foster the dataset effectively and contribute a quick and correct decision [2]. Data governance includes many management aspects, such as data security, integrity, availability, quality, quantity, performance and so on so forth [2].



**Current pressing issues**

**in**

**the**

**field**

Privacy and Security Problem

As the emerging and development of technologies like Cloud Computing, cloud storage, analytics engines and social media, the security and privacy problem of big data is attracting a large number of people to pay attention to. For social media, like Facebook and Twitter, whose data is the most credible resource to do data mining so that information that can interest the subscribers can be launched timely and accurately. However, the privacy setting often limits the mining of data. For subscribers, this is good news, but for the companies, this is a limitation for their advertisement. In Cloud infrastructures, databases are often outsourced based on the well-known DaaS (Database as a Service) paradigm. This gives rise to very problematic security issues as query processing procedures may easily access sensitive data sets and determine privacy breaches.

Scalability Problems

The amount of data on big projects always increases at a high speed. However, most companies and organizations do not realize that their storage space cannot catch up with the speed of data increasing, so as the project goes through, the storage will be run out. Thus, it is very important for the organizer to realize the necessity of scaling up before doing a big data project. For example, on-premise Hadoop analytics, which depends on the physical environment and commodity servers. This will lead to scalability problems and the serious limit problem of storage.

**Current pressing issues**

**in**

**the**

**field**

**Research questions and methods**

The US health care system has been rapidly adopting Electronic Health Record (EHR) and Health Care Information Exchange (HIE), which help to accumulate large amounts of heterogeneous medical data from payers, providers and pharmaceutical departments in the healthcare industry [5]. These data can be analyzed to arrive at insights that can improve care levels [6] and medical quality [7]. However, due to the sheer size and complexity of such data, its analysis and practical application become a challenge. Because big data can process massive amounts of data and data at high speed, it has the potential to create significant value in healthcare by improving outcomes and reducing costs [8]. Studies have shown that it can improve the quality of care, improve operational efficiency, predict and plan responses to disease epidemics, and optimize health care spending at all levels [9]. Here, we explore how big data analytics has revolutionized the healthcare industry.

Healthcare System

**Research questions and methods**

Clinical informatics focuses on the application of information technology (IT) in the healthcare field. It includes activity-based research, analysis of the relationship between major patient diagnosis (MD) and potential cause of death (UCD), and data storage from EHR and HIE. The main contribution of big data lies in how EHR and HIE data are stored. Clinical real-time streaming data is stored using NoSQL databases, Hadoop and HBase databases due to its high-performance characteristics. Some research efforts have also studied and proposed several interactive methods for sharing medical data from multiple platforms.

HIE is used for effective information sharing between heterogeneous medical entities, thereby improving the quality of care provided. Janakiraman et al [10]. (2017) studied the use of HIE in the emergency department (ED) and found that the benefits of HIE increased with more information about patients, doctors, and previous interactions between them [10]. Yaraghi et al. (2014) Modeling HIE as a multi-platform [11]. Users self-service technology that evaluates models based on user-specific and network-specific factors. Another research institution is studying whether the health care reform model will lead to better patient-centered outcomes.

Big data technology has made the availability and analysis of large amounts of clinical data possible. Insights from data analysis can help medical professionals identify diseases

Symptoms and disease predictions and improvements are much better, ultimately leading to improved overall quality of care. Because data is very large in scale and complexity and often involves integrating clinical data from various platforms to learn more, data security is often compromised during clinical data analysis. Big data technology can solve this problem. Researchers have proposed several models and frameworks to effectively protect data privacy and efficiently handle the concurrent analysis of data sets..

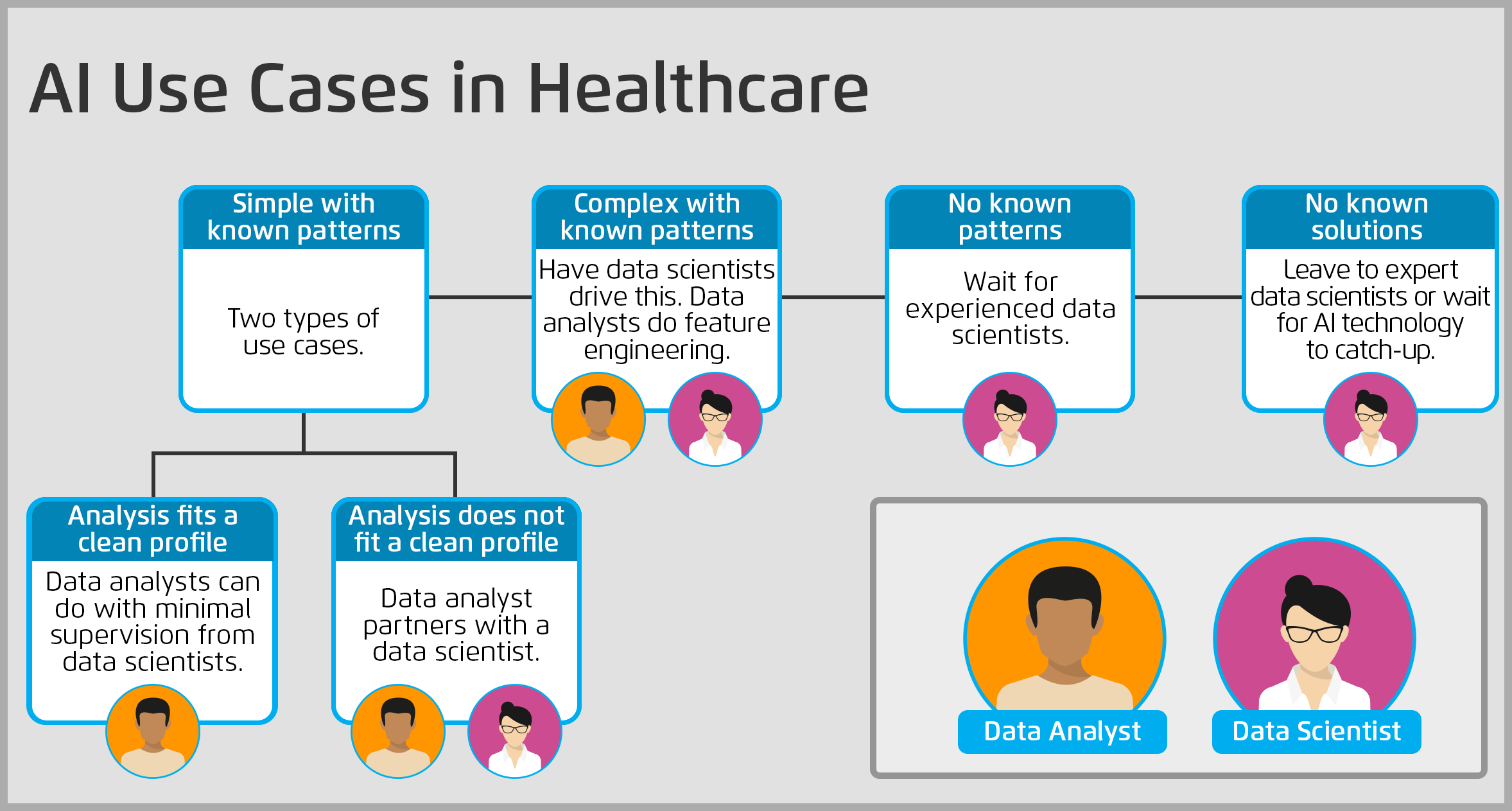
Healthcare Information Exchange

With the advancement of imaging technology, EHR is often accompanied by high-quality medical images. Studying clinical data and analyzing such images will help to better diagnose and predict disease more accurately in the future. Medical Image Informatics focuses on using large data tools and techniques to process images for meaningful insights. Similarly, picture archiving and communication systems (PACS) are also highly advantageous for the medical community as these medical images can be used to improve decisions regarding patient care and readmissions [12]. It discusses how data can be integrated with PACS when PACS's Medical Digital Imaging and Communications (DICOM) object repository and database system are transferred to the cloud [12]. Since the use of big data analytics to analyze large numbers of high-quality clinical images generates rich, spatially oriented information at the cellular and subcellular levels, systems such as Hadoop-GIS are cost-effective Parallel systems are being developed to help manage advanced spatial queries.

Medical Image Informatics

**Research questions and methods**

Recent research has also used big data technology to analyze the content of social media as a means of monitoring infectious diseases and monitoring the occurrence of diseases worldwide. Big data analytics tools are used in social media communications to detect mood patterns associated with depression to identify people with depression from users. Health IT infrastructure such as the US Veterans Health Administration (VHA) promotes improved quality of care by providing structured clinical data in EHR and unstructured data such as doctor notes [13].



Health Management

**Research questions and methods**



Credibility means it is transparent enough for us to know where the data come from, what kinds of method do we use to analyze the data and what the associated results are according to the data. Credibility is a subjective concept, and credible evidence has a different explanation in a different field. When defining the definition of credible evidence the following should be in consideration. When is the evidence came up and who put it forward? Who is the audience that receives the evidence? What is this evidence for and which field does the evidence refer to? Whether the evidence is viewed as credible or not? All these aspects should be considered in order to judge whether it is credible evidence. The value of credible evidence is huge, especially in research, which lay much attention on evidence. Every sentence shown in the research paper is on behalf of the author’s knowledge, and even quality. Credible evidence is the foundation of a good research paper. Great oaks from little acorns grow, and credible evidence is just the groundwork of the tall building. It is the body of knowledge, analytical results of data, research findings that can lead comprehensively to answering questions, by giving us directions for practice and policy, to learn and create opinions around a particular issue.

What is credible evidence

**Professional Trade Journals**

1. **The Analytics Dispatch**
2. **Data Elixir**
3. **Data Science Weekly**
4. **Analytics Vidhya**
5. **Machine Learning Engineer**

**DATABASE USED FREQUENTLY**

1. **NTIS (National Technical Information Service)**
2. **ASTM Standards & Engineering Digital Library**
3. **SpringerLink**
4. **IEEE Xplore**
5. **Safari Books Online**

Scholarly Journals

1. **Artificial Intelligence**
2. **Big Data**
3. **Big Data Research**
4. **Computational Statistics & Data Analysis by International Association for Statistical Computing (IASC)**
5. **Foundation and trends of Machine Learning**

**Genres in the field of data analysis**

The main genres of popular articles, trade articles, short reports, academic articles, technical reports, and case studies are mainly used. Depending on the specific context, a type of article structure may vary differently.

1) Popular articles use easy-to-understand language and informal tone to make readers understand the science.

2) Trade articles informally talk about some ideas, surveys and analysis in an industry. They may contain some professional vocabulary and knowledge, but may be difficult for non-professionals to understand.

3) Short reports are usually written by researchers to briefly describe and present their research, experiments, progress reports or meetings. Their audience is usually a person who is interested in the professional or interested in the report and has relevant knowledge.

4) The technical report focuses on a detailed description of the research. People working in this field with relevant needs will read them.

5) Academic articles are the final reports of large-scale projects and research to be published, with certain unique articles.

**Genres in the field of data analysis**

In particular, academic articles are most helpful to data analysts, who can enable data analysis personnel to understand novel solutions in the field of data science. Ordinary data analysts can improve their academic skills and writing skills.

Academic articles are very important to everyone involved in data analysis. The genre consists of five elements, namely abstracts, introductions, methods, results, discussions and references.

1) Abstract is actually a brief introduction to the entire article and its background.

2) Introduction will further introduce the current trends, objectives and related research of the study.

3) Method part is the main body of the academic article, detailing the methods used in the study.

4) Results section shows the completion of the proposed method and a related assessment of the results.

5) Discussion section will not only recapitulate the research object, urgency and purpose, but also summarize the methods and main findings. In addition, it will discuss the relevance of research, including meaning and future work.

6) Reference section should list the references used in all the articles in order to clarify copyright and facilitate access.

**Differences Between Academic Articles and Technical/Professional Tests**

As the introduction of academic articles, scholarly articles have a more formal writing style and are very rigorous. Although sometimes a particular study modifies a given template based on context, it cannot delete these six elements. However, the focus of the technical article is on how to clearly articulate the research results and the specific experimental details. In addition, technical articles do not have to write every refence as well as academic articles. Moreover, the linguistic style and tone of academic articles are more rigorous and formal than technical articles.

Reference Website

1 <http://data.stats.gov.cn/>

On the basis of the "China Statistics Database" created by the National Bureau of Statistics in 2008, a new version of the statistical database was established in 2013. Here, you can not only query the time series data of the main indicators in the professional fields surveyed by the National Bureau of Statistics, but also create personalized statistical charts according to individual needs; you can browse not only the statistical yearbook data carrying history, but also the use. A modern, visual statistical product.

2 <http://www.cnnic.net.cn/>

China Internet Network Information Center (CNNIC) is a management and service organization established on June 3, 1997 with the approval of the competent national authorities. It exercises the responsibility of the National Internet Network Information Center.

3 <https://figshare.com/>

Figshare is an online data knowledge base based on cloud computing technology. Researchers can save and share their research results, including data, datasets, images, videos, posters and codes, so that their research, even preliminary research, exposure and get approval. Since Figshare follows the open data principle, users have free access to data and uploads. Therefore, researchers can publish their data through Figshare, and all entries on Figshare have a DOI for easy reference.

Reference Forum

<http://bbs.pinggu.org/index.php?gid=148>

This is the China University Economic Forum, which mainly involves exchange of statistical software related to econometrics, statistics, data mining, statistical software, and data.

Reference Journal

1) Journal of Business and Economics Statistics

2) Journal of the Royal Statistical Society: Series B

3) Annals of Statistics

Tutorial

1 The Info Guides of data collection: https://dsc.gmu.edu/

2 use Mason ID to take the free online LinkedIn Learning courses, such as R, Python, etc. https://lil.gmu.edu/

# **Reference:**

|  |  |
| --- | --- |
| [1] | A. R. Maxwell McCombs, “News influence on our pictures of the world,” 2002. |
| [2] | V. &. B. C. V. Khatri, “Designing data governance,” *Communications of the ACM*, 2010. |
| [3] | A. Cuzzocrea, “Privacy and security of big data: current challenges and future research perspectives,” *In Proceedings of the First International Workshop on Privacy and Secuirty of Big Data* , 2014, pp. 45-47. |
| [4] | D. Kerby, “ https://dzone.com/articles/top-5-big-data-integration-challenges-be-prepared,” 18 4 2018. |
| [5] | T. K. S. S. A. &. S. C. Rajapakshe, “Sustainability planning for healthcare information exchanges with supplier rebate program,” 2018. |
| [6] | J. &. R. C. K. Sun, “Big data analytics for healthcare,” *In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1525-1525. |
| [7] | P. K. B. K. D. &. K. S. V. Groves, “The'big data'revolution in healthcare: Accelerating value and innovation,” 2016. |
| [8] | J. B.-L. G. W. &. A. T. A. Roski, “Creating value in health care through big data: opportunities and policy implications,” *Health affairs*, 2014, pp. 1115-1122. |
| [9] | R. B. R. S. A. &. V. R. Nambiar, “ A look at challenges and opportunities of big data analytics in healthcare,” *In 2013 IEEE international conference on Big Data* , 2013, pp. 17-22. |
| [10] | R. P. E. D. E. &. K. S. Janakiraman, “The effects of health information exchange access on healthcare quality and efficiency: An empirical investigation. Mays Business School Research Paper,” 2017. |
| [11] | N. D. A. Y. S. R. G. R. D. &. R. R. Yaraghi, “Health information exchange as a multisided platform: adoption, usage, and practice involvement in service co-production.,” *Information Systems Research,* pp. 1-18, 2014. |
| [12] | L. A. B. C. C. &. O. J. L. Silva, “A PACS archive architecture supported on cloud services,” *International journal of computer assisted radiology and surgery,* pp. 349-358, 2012. |
| [13] | J. F. J. K. E. K. S. P. L. K. R. M. &. P. J. B. Kupersmith, “Advancing Evidence-Based Care For Diabetes: Lessons From The Veterans Health Administration: A highly regarded EHR system is but one contributor to the quality transformation of the VHA since the mid-1990s,” *Health Affairs,* p. 26, 2007. |
| [14] | S. T. P. Doherty, “https://www.journals.elsevier.com/artificial-intelligence/,” 2019. |
| [15] | P. Zoran Obradovic, “https://home.liebertpub.com/publications/big-data/611/overview,” |
| [16] | Z. W. T. Palpanas, “Big Data Research”. |
| [17] | E. K. B. P. A.M. Colubi, “Computational Statistics & Data Analysis”. |
| [18] | B. Michael Jordan and University of California, “https://www.nowpublishers.com/MAL,” |
| [19] | “https://mode.com/analytics-dispatch/archive/,” 2019. [联机]. |
| [20] | “https://dataelixir.com/,” |
| [21] | “https://www.datascienceweekly.org/,” |
| [22] | “https://www.analyticsvidhya.com/,” |
| [23] | “udacity.com/course/machine-learning-engineer-nanodegree--nd009t,” 2019. |

