A SMART DEVELOPMENT OF PREDICTION ANALYSIS-BASED DATA SCIENCE MODEL BY USING DEEP LEARNING APPROACH

R. Vilasini and M. Sutharsan

Department of Electronics and Communication Engineering, Muthayammal Engineering College, India

Abstract

The field of data analysis is one of the fastest growing fields. All large and small companies are now prioritizing the field of data analysis. It is possible to advance those industries using data analysis widely across all industries like Medicine, Agriculture, Architecture, Stock Market, Marketing, and many more. There are very few professionals in the market due to lack of sufficient knowledge about data science among the technicians. So as the company demand is high and the number of data analysts is low there are more opportunities for easy availability of jobs in this field. Data science is the study of the information we have. Usually the data contain useful and useless information. The field of data science involves the efficient extraction of useful information, recording, storing and analyzing data. In this paper, a smart model was proposed to enhance the data science model. The proposed model one step further and build useful data management-based on the analysis. This can keep track of the events that happened and predict what will happen next.

Keywords:

Prioritizing, Data Analysis, Sufficient Knowledge, Data Science, Information

1. INTRODUCTION

The main purpose of this data science is to explore millions of pieces of information and make some important conclusions using the knowledge that comes through it [1]. This means that all user information must be parsed and analyzed to obtain that general useful information. Data science is what is used to analyze that information. Many of us would have thought that data science and artificial intelligence were one and the same [2]. Wrong, the two are different. Data science is the study of the information we have (not only that but it's important) [3].

The field of artificial intelligence is the process of injecting artificial knowledge generated by programming languages into a tool and making it work as we wish [4]. The world is rapidly becoming digitally enchanted. We always go around with our hands. Millions of pieces of information are generated in the world every minute [5]. The use of the computer has come even in our brother's shop which owns the small box store, which stores information on the computer. So think about how much information the biggest companies have. The field of data science will continue to grow as information increases [6]. The fall of the field of data science as a whole is unlikely in this century, except for the possibility of successes and failures within the many tools used for data science [7]. Data science will not collapse without the biggest economic and technological downturn in the world. Data are useful references. They can be numbers, measurements of temperature, sound, light, pressure, altitude, etc., words, or other useful references [8]. These are called data because they are references or messages that need to be basically given if anything needs to be analyzed and analyzed. With this kind of data you can systematically analyze and get new results, ideas and findings [9]. It may come as a surprise to you that 90% of the data generated in the world so far has been generated in the last 2 years alone. Humans produce about 2.5 quintillion bytes of data per day. It is predicted that by 2025 humans will produce 463 Exabyte's of data per day. The power of data has grown to the point. Advertisements, suggestions and more appear on the screen-based on the data generated by users' web activity. The Internet world today is dominated by programs-based on data science with these free data. Our mobility in the digital world is determined by the data we produce; it is beginning to resonate in our real life. It is in this context that data science has evolved into something of great importance today. This is basically an information technology division. Data science is the process of compiling aggregate data, integrating it and extracting the required data from it for use, and creating specific new data for future operation [10].

Many people are confused as to whether data science is a part of computer science. In fact, data science belongs to computer science, but is different from computer science [11]. The two terms have similarities, but there are significant differences between the two. Computer science includes small areas of artificial intelligence, analysis, programming, natural language processing, machine learning, web development and many more. Data science is also a part of computer science, but requires more knowledge of mathematics and statistics. In other words, computer science handles programming software and hardware when dealing with data science analysis, programming, and statistics. So, if a computer scientist focuses on programming, statistics and analysis, he can become a data scientist.

Computer science is defined as the study of computer engineering, design and application in science and technology. The application of computer science includes various aspects and technical concepts such as networking, software, hardware and the Internet. Knowledge of computer science varies in different fields such as design, architecture, and manufacturing [12].

Computer scientists analyze algorithms and study the performance of computer software and hardware. Key areas of computer science are computer systems, artificial intelligence and networks, human-computer communication, vision and graphics, and programming language, numerical analysis, biometrics, software engineering, and computer theory. Data science is the study of different types of data such as unstructured, semi-structured and structured data. Data can be in any available format and used to retrieve information contained therein. There are many techniques used in data science to study data. This includes data mining, data processing, data transfer and more. Data science focuses on exploiting data for prediction, analysis and understanding. Therefore, it emphasizes the effective correlation of data analysis results. Furthermore, data science prioritizes knowledge of optimization algorithms by managing the required trade between speed and accuracy. Computer science is the study of the performance of computers, while data science finds

meaning within large data. Computer science students learn advanced computing, which includes in-depth experience building database systems and enterprise-level application development [13].

Data science, on the other hand, learn about the analysis of large data sets, data visualization, data mining, efficient data management and forecast data analysis using mathematics and computer applications. Computer security is the development of technology in cyber security, software and intelligent systems. Although data science is-based on the skills required for data mining, it also clarifies the meanings of large data sets used in decision making in large companies and organizations.

Computer science is important because it is the main impetus for technological innovation today. However, data science is very important for an organization and its application requires experts in data mining and analysis. For Computer Science students, Application Developer, Computer Programmer, Computer Engineer, Database Developer, Database Engineer, Data Center Manager, IT Engineer, Software Engineer, System Programmer, Networker and Network. Let us first define computer science and data science separately.

2. RELATED WORKS

Data science is the collection of more than one function in a field. If you are a data scientist, you can use it to run data collection (or data mining), data analysis and modeling, as well as results that turn data into one-way information. This last bit of "data driven decision making" in the field of entrepreneurship. In our time, data science is considered a part of computer science or at least nowadays you can say that it is strongly related to it [1]. Using scientific methods to process data with algorithms and other systems, a data scientist works to turn data into information and to gain knowledge about something previously obscure or completely unknown [2]. To do this, a data scientist can use both structured and unstructured data. The exact requirements depend on the method used and the purpose of the tasks. Essentially, the data science job should move towards publishing opportunities, help make decisions and maximize the company's profitability in a way. You need to collect data, configure it, analyze it, and determine the best way to align sets and variables [4].

If you encounter "big data" issues, you need to organize the results in a way that makes sense right now and in the future. By combining all these historical sources with one kind of middleware you will come across different databases where you need to manage and manage a "source of gold" data. Data scientists also need to find ways to construct unstructured data and cut continuous data into bits so they can be used. Eliminates wasted data and ensures that useful data is optimally alienated.

Once that is done, the data scientist will process the data to find the data and trends. By doing so, it can "alert" the business of persistent problems or other harmful events if ignored or allow the company to earn more revenue if it is aware of that event in advance. The data scientist will create cases and explain the results to the shareholders and management. This can happen in the form of a visual presentation or other means [5]. The real-world practices of the data scientist and statistician often differ. For example, a data scientist might need to create and oversee databases that may be middleware solutions. From what I know, statisticians are rarely required to handle management information technology and software engineering tasks. They can use it, but not make or maintain them [7]. A few years ago terms like Business Intelligence (PI) or Management Information (MI) shared functions with current job descriptions of what a data scientist should do. There seems to be a password effect and exaggeration around the period of data science, especially in conjunction with artificial intelligence (AI) and machine learning (ML), but still innovative / disruptive work of data scientists. Simply put, a data scientist is an expert in teaching data science [6]. They use their technical and non-technical skills to distort complex data using various methods and algorithms to extract valid results. Therefore, management can make decisions with data.

3. PROPOSED MODEL

It is-based on methods of storing and analyzing data to extract valuable insights. The ultimate goal of data science is to extract information from any data, and that insight will enable the company to make better decisions. It will improve the growth and market share of a company. Most people have the misconception that data science is a programming language. It is a technology that requires three main components: first element domain knowledge, second element statistics and probability, and third element programming skills.

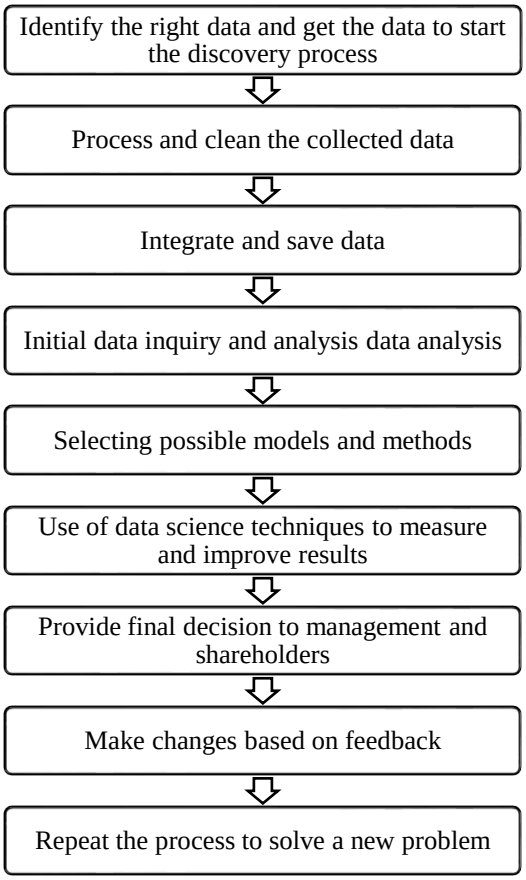

Fig.1. Proposed model flow

Data scientists need to work closely with management and stakeholders to determine how they can use data to achieve their

goals and achieve those goals. Based on those goals, they develop algorithms, forecast models and data modeling processes to extract, collect, store and analyze data and share insights with peers. The proposed model flow shown in Fig.1.

Step 1: Identify the right data and get the data to start the discovery process.

Step 2: Process and clean the collected data

Step 3: Integrate and save data

Step 4: Initial data inquiry and analysis data analysis

Step 5: Selecting possible models and methods

Step 6: Use of data science techniques to measure and improve results

Step 7: Provide final decision to management and shareholders

Step 8: Make changes-based on feedback

Step 9: Repeat the process to solve a new problem

The business world is in dire need of smart and creative individuals who can dive deep into the ocean of data and help businesses and economies grow. Data management is considered to be one of the most important tasks for various industry organizations and various sectors to successfully carry out their operations.

Companies are found to be constantly storing a lot of data through their operations. Data mining plays a leading role in converting such stored information into useful information.

Data mining is the process of identifying discrepancies found within large data sets, as well as unwanted data, in order to identify the best data set. The most common roles in data science are as follows shown:

- **Data Scientists**: They design processes to create predictive models and algorithms and perform custom analysis to obtain insights from data.
- **Data analysts** use large data sets to analyze and manipulate and identify trends in making strategic business decisions.
- **Data Engineers**: The main task of data engineers is to learn, synchronize and organize data from different sources and transfer it to databases.

A wide range of techniques are used here, which enable the company to utilize this information to increase revenue, reduce costs, improve customer relationships and reduce risks. The main objective of data mining is to "extract information (with intelligent methods) from a database and to provide relevant information for management decisions to transform the information into an understandable structure for use."

Data processing is the process of capturing large amounts of data in order to identify the ins and outs of that data. Nowadays, the demand for data sector is increasing very fast, which has also increased the demand for data analysts and data scientists.

Using this technique, we analyze the data and then convert that data into meaningful information. It helps the business to make accurate and better decisions in a company. Data processing helps to improve smart market results, conduct accurate campaigns, make predictions and much more. With the help of data processing, we can analyze customer behaviors and their insights. This leads to great success and data-based business.

4. RESULTS AND DISCUSSION

The proposed prediction analysis-based data science (PADS) model was compared with the existing Surgical data science model (SDSM), Smart city data science model (SCSM), Data science and prediction (DSP) and data-driven decision making (DDDM)

4.1 DEMAND COLLECTION MANAGEMENT

Data processing projects begin with the collection and understanding of need. Data mining analysts or users define the amount of demand with a vendor business perspective. Once the scope is defined, we move on to the next step.

When competition is tight in a market with a high load, the answers are often within your consumer data. Telecommunications, media and technology companies can use analytical models to understand mountains of customer data, predict customer behavior, and deliver more targeted and relevant campaigns.

Table.1. Demand collection management

Files	SDSM	SCSM	DSP	DDDM	PADS
100	78.31	92.76	82.53	89.56	88.27
200	76.65	89.90	83.37	89.15	89.37
300	74.99	87.04	84.21	88.74	90.47
400	73.33	84.18	85.05	88.33	91.57
500	71.67	81.32	85.89	87.92	92.67
600	70.01	78.46	86.73	87.51	93.77
700	68.35	75.60	87.57	87.10	94.87

4.2 DATA ANALYSIS MANAGEMENT

At this stage, data mining experts collect, evaluate and explore the need or plan. Experts understand problems and challenges and turn them into metadata. At this point, data mining statistics are used to identify and modify data formats. With analytical knowledge, insurance companies can solve problems related to fraud, compliance, risk management and customer shortages. Companies use data mining techniques to use products in commercial mining to more effectively determine pricing and to find new ways to deliver products in line with existing customer standards.

Table.2. Data analysis management

Files	SDSM	SCSM	DSP	DDDM	PADS
100	76.20	88.04	90.66	82.66	88.44
200	80.78	86.90	92.80	79.42	88.49
300	81.28	86.02	91.23	80.14	88.53
400	81.12	84.82	89.61	80.27	88.56
500	80.38	83.17	87.81	79.00	88.56
600	82.56	82.24	87.76	78.36	88.61
700	83.43	81.06	86.87	77.71	88.64

4.3 DATA MINING

Data mining experts translate data into meaningful information for the modeling step. They use the nucleus process - extraction, transformation and loading. They are also responsible for creating new data properties. Various tools are used here to render data in structured form without changing the meaning of the data sets. With integrated, data-based perspectives, educators can predict students' performance before they set foot in the classroom - and develop intervention strategies to keep them confident. Data processing allows educators to access student data, predict achievement levels, and point out groups of students or groups of students who need extra attention.

Table.3. Data arrangements management

Files	SDSM	SCSM	DSP	DDDM	PADS
100	79.77	80.20	83.26	84.34	95.62
200	80.10	81.70	83.85	86.21	96.66
300	81.44	82.81	84.83	87.04	96.79
400	82.58	83.19	86.04	87.95	97.75
500	83.63	84.20	87.18	88.87	97.32
600	84.34	85.13	88.29	90.20	98.56
700	85.64	86.13	88.99	91.07	98.67

4.4 MODELING MANAGEMENT

Data professionals put their best tools into this process as it plays an important role in the complete processing of data. All modeling methods are used to filter the data appropriately. Modeling and evaluation are followed simultaneously to verify the relevant steps and parameters. The final result quality is proven after the final modeling is completed. Requirements such as early detection of problems, quality assurance and investment in brand equity need to align distribution plans with forecasts. Manufacturers can predict wear and tear on production assets, which can increase time and keep production sequence on schedule.

Table.4. Modeling management

Files	SDSM	SCSM	DSP	DDDM	PADS
100	63.60	84.92	81.53	83.35	95.43
200	63.93	86.42	82.12	85.22	96.47
300	65.27	87.53	83.10	86.05	96.60
400	66.41	87.91	84.31	86.96	97.56
500	67.46	88.92	85.45	87.88	97.13
600	68.17	89.85	86.56	89.21	98.37
700	69.47	90.85	87.26	90.08	98.48

4.5 EVALUATION MANAGEMENT

This is the filtration process after successful modeling. If the result is not satisfactory, it will be redone to the sample. In the end, the demand is re-checked with the seller so no point is missed. Data mining experts ultimately determine the absolute outcome. Automated algorithms help banks understand their customer base and the billions of transactions at the heart of the financial system. Data processing enables financial services companies to gain a better view of market risks, quickly detect fraud, manage regulatory compliance obligations, and obtain optimal returns on their marketing investments.

Table.5. Evaluation management

Files	SDSM	SCSM	DSP	DDDM	PADS
100	65.71	94.48	73.56	83.06	97.03
200	65.38	92.98	72.97	81.19	96.02
300	64.04	91.87	71.99	80.36	95.86
400	62.90	91.49	70.78	79.45	94.90
500	61.85	90.48	69.64	78.53	95.33
600	61.14	89.55	68.53	77.20	94.13
700	59.84	88.55	67.83	76.12	93.97

4.6 DEPLOYMENT MANAGEMENT

This is the final stage of the whole process. Experts provide data to vendors in the form of spreadsheets or maps. Large customer databases contain hidden customer intelligence that can help improve relationships, improve marketing campaigns and predict sales. With more accurate data models, retailers can deliver more targeted campaigns - and find the offer that has the biggest impact on the customer.

Table.6. Deployment management

Files	SDSM	SCSM	DSP	DDDM	PADS
100	68.01	76.78	70.16	80.32	97.94
200	67.68	75.28	69.57	78.45	96.90
300	66.34	74.17	68.59	77.62	96.77
400	65.20	73.79	67.38	76.71	95.81
500	64.15	72.78	66.24	75.79	96.24
600	63.44	71.85	65.13	74.46	95.00
700	62.14	70.85	64.43	73.59	94.89

5. CONCLUSION

It should be noted that it takes more time to get accurate information from the data. Therefore, after growing your business quickly, there is a need to make accurate and quick decisions so that you can take advantage of the opportunities available at the right time. In this technology-based world data mining is a fast-growing industry. Nowadays everyone has to use their data in the right way and in the right approach to get useful and accurate information.

Data aggregation is a type of data and information mining process where data is searched, collected and presented in a data-based, concise format to achieve specific business objectives or processes and / or conduct human analysis. Data integration is a component of Business Intelligence (BI) solutions.

Data integration software search databases can find relevant search query data and current data findings in a concise format and be useful and useful to the end user or application. Data integration typically operates on large data or data marts that do not provide much information value overall. The main applications of the Data Internet are the collection, use and dissemination of existing and available data on the Web.

REFERENCES

[1] L. Maier Hein, M. Eisenmann and S. Speidel, "Surgical Data Science from Concepts Toward Clinical Translation", *Medical Image Analysis*, Vol. 76, pp. 1-13, 2022.

[2] I.H. Sarker, "Smart City Data Science: Towards Data-Driven Smart Cities with Open Research Issues", *Internet of Things*, Vol. 19, pp. 1-14, 2022.

[3] H. Han and S. Trimi, "Towards a Data Science Platform for Improving SME Collaboration through Industry 4.0 Technologies", *Technological Forecasting and Social Change*, Vol. 174, pp. 1-17, 2022.

[4] H.S.Z. Kazmi, S.O. Shim and Y.B. Zikria, "Congestion Avoidance and Fault Detection in WSNs using Data Science Techniques", *Transactions on Emerging Telecommunications Technologies*, Vol. 33, No. 3, pp. 1-12, 2022.

[5] A. Rairikar, V. Kulkarni, V. Sabale, H. Kale and A. Lamgunde, "Heart Disease Prediction using Data Mining Techniques", *Proceedings of International Conference on Intelligence Computing and Control*, pp. 1-8, 2018.

[6] F.A.M. Al-Yarimi, and M.H.M. Bamashmos, "Feature Optimization by Discrete Weights for Heart Disease Prediction using Supervised Learning", *Soft Computing*, Vol. 25, pp. 1821-1831, 2021.

[7] M.D. Samad, A. Ulloa and G.J. Wehner, "Predicting Survival from Large Echocardiography and Electronic Health Record Datasets: Optimization with Machine Learning", *JACC Cardiovascular Imaging*, Vol 12, No. 4, pp. 681-689, 2019.

[8] M. Anoopkumar and A.M.J. Md Zubair Rahman, "A Review on Data Mining techniques and Factors used in Educational Data Mining to Predict Student Amelioration", *Proceedings of International Conference on Data Mining*, pp. 122-133, 2016.

[9] R.B. Kulkarni, "Appraisal Management System using Data mining Classification Technique", *International Journal on Computer Applications*, Vol. 135, No. 12, pp. 45-50, 2016.

[10] V. Tanvi Sharma, "Performance Analysis of Data Mining Classification Techniques on Public Health Care Data", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4, No. 6, pp. 1-14, 2016.

[11] V. Dhar, "Data Science and Prediction", *Communications of the ACM*, Vol. 56, No. 12, pp. 64-73, 2013.

[12] F. Provost and T. Fawcett, "Data Science and its Relationship to Big Data and Data-Driven Decision Making", *Big Data*, Vol. 1, No. 1, pp. 51-59, 2013.

[13] M. Swan, "The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery", *Big Data*, Vol. 1, No. 2, pp. 85-99, 2013.