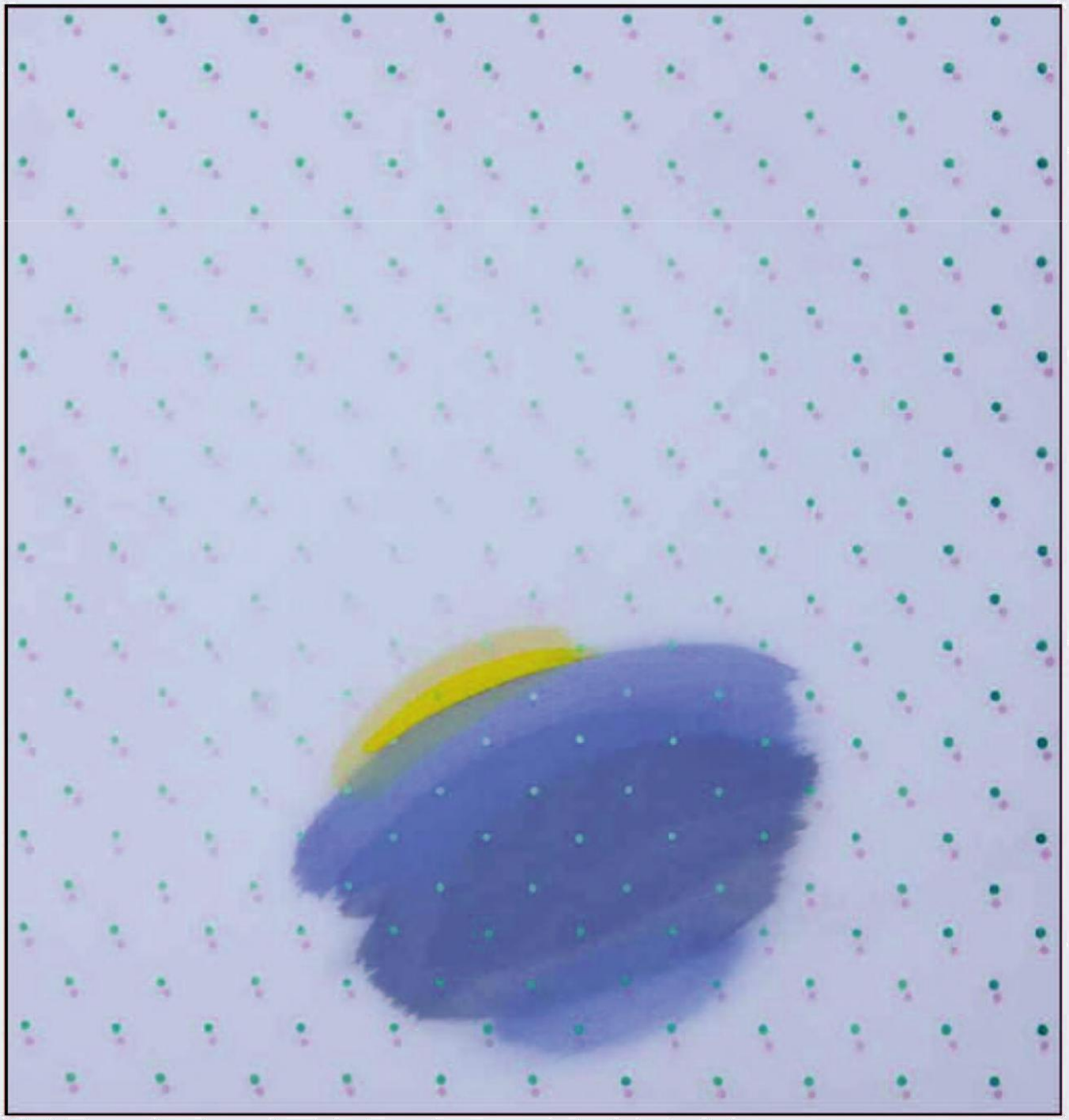
SEEJSD

SOUTH EAST EUROPEAN JOURNAL OF SUSTAINABLE DEVELOPMENT

Vol. 3 (2/2019)



Skopje, North Macedonia

SEEJSD

SOUTH EAST EUROPEAN JOURNAL OF SUSTAINABLE DEVELOPMENT

Vol.3 (2/2019)

Publisher: Mother Teresa University in Skopje, North Macedonia

South East European Journal of Sustainable Development
ISSN (print) 2545-4463
ISSN (online) 2545-4471
is published twice a year.
Account No. 160016267778815

723019 - 45
Tax No. 4080016561272
Bank: Narodna Banka RM

Managing Editors: Sani Demiri, PhD
Olga Popovska, PhD
Technical Editing/Layout: Besnik Hamiti
Editorial Office: South East European Journal of Sustainable Development

Mother Teresa University in Skopje,
North Macedonia
12 UdarnaBrigada 2a/ VII, Skopje,
North Macedonia
Phone: +389 2 3161 004
E-mail: seejsd@unt.edu.mk
Web: www.seejsd.unt.edu.mk

The publication of the Journal is supported by:

Ministry of Culture of North Macedonia

Contents

EDITOR'S MESSAGE

SEEJSD is again on a board. This time, in front of you is its fourth edition. It takes some time to receive, review, consolidate and publish 19 articles from vast number of disciplines which somehow are closely related to sustainable development. In fact, sustainable development is one of the most important challenges of modern society interconnected with climate changes and a newcomer - ***green deal***.

In that sense, it is difficult to manage a journal like SEEJSD where every single topic in the articles which pretend to be published are from very heterogeneous fields and can be linked to sustainable development as top human priority. It was the main reason for Editorial board to undertake a measure of so called clustering of received and approved articles for publishing in this issue of SEEJSD.

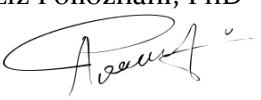
Namely, we decided to split this issue in two parts, having in mind that still there will be overlapping in the published manuscripts from both parts. So, part A is containing articles from technical, biotechnical and natural sciences, while part B is composed by the articles from humanities, health and social sciences. Doing this, the Editorial board is hoping that in such way it will be more convenient for SEEJSD more easily in enter in the process on international indexation of scientific journals and publications.

Regarding the disciplines, in this issue of SEEJSD in Part A the majority of the articles are from the field of information and communication technologies, more popular as ICT (5), followed by the ones from the environmental sciences (3) ending with the technical sciences (2). In Part B, there are mostly articles from an economy and humanities (4 of each) and one from health sciences. Nevertheless, we divided the SEEJSD in two parts; the articles in each part in some cases are overlapping in term of disciplines which implies multidisciplinary and interdisciplinary approach in most of the studies. SEEJSD is welcoming that reality as precondition for better understanding in wider publicity.

Related to the type of manuscripts, most of the manuscripts are original scientific papers, and few review papers and few professional articles. This composition of this issue is expressing its openness for different types of research activities and the promotion of their findings and recommendation.

In this occasion, let me express my optimistic expectation that in the new format, SEEJSD will continue to be attractive media for publishing of the scientific articles with appropriate quality which will accelerate the process of international indexation. At the end I would like to thank to the authors and coauthors who publish their manuscripts in this issue of SEEJSD and to the Editorial board members for their extraordinary efforts to finalize the completion of this issue of SEEJSD.

Editor in Chief,
Prof. Aziz Pollozhani, PhD



Application of Data Analytics with Python Programming In Healthcare

Majlinda Fetaji[†]
South East European University, Computer Sciences, Tetovo, North Macedonia, m.fetaji@seeu.edu.mk

Lindita Loku
Mother Teresa University, Technical Faculty, Skopje, North Macedonia, lindita.loku@unt.edu.mk

Bekim Fetaji
Mother Teresa University, Technical Faculty, Skopje, North Macedonia, Bekim.fetaji@unt.edu.mk

Aleksandar Krsteski
University Goce Delcev, Shtip, Faculty of Computer Sciences, Aleksandar.Krsteski@ugd.edu.mk

Zoran Zdravev
University Goce Delcev, Shtip, Faculty of Computer Sciences, Zoran.Zdravev@ugd.edu.mk

ABSTRACT

The goal of this research is to investigate how using data analytics and programming python platform can discover additional value from health information used in health care. Dominant part of the Healthcare information is frequently unstructured, exists in storehouses and dwells in imaging frameworks, clinical solution notes, patient information and so forth coordinating these heterogeneous information and calculating it in to progress investigation is basic to improve medical services results. The incorporation of health sector data analytics provides stakeholders with fresh perspectives that can advance personalized treatment, optimize patient outcomes and prevent excessive costs. Until now, medical healthcare has not completely gotten a handle on the possible advantages to be picked up from information examination. The developing medical care industry is creating an enormous volume of valuable information on tolerant socioeconomics, therapy plans, instalment, and protection inclusion pulling in the consideration of clinicians and researchers the same. Lately, various companion explored articles have tended to various components of information science application in medical care. Be that as it may, the absence of a thorough and methodical account roused us to build a study examinations on this field. This examination study characterizes information investigation and its qualities, remarks on its favourable circumstances and difficulties in medical care. Information examination gives new systematic open doors as well as faces parcel of difficulties. The research study begins from investigating the patient information. While picking the stage, a few rules like accessibility, usability, adaptability, level of security and coherence ought to be thought of. Analyses of all challenges of data analytics are analyzed and insights are represented and discussed and argumented.

CCS

Applied computing > Life and medical sciences >Health informatics

KEYWORDS

data analytics, healthcare, comparative analyses, data analytics platform

1 Introduction

Computer scientists and health-care providers may learn from one another when it comes to understanding the value of data and data analytics. There is an urgent need to develop and integrate new , mathematical, visualization, and computational models with the ability to analyze Data in order to retrieve useful information to aid clinicians in accurately diagnosing and treating patients to improve healthcare. According to [1] data science is a multi-disciplinary field that uses data analytics and different algorithms and systems to extract knowledge and insights from structured and unstructured data.. Data, derived by patients and consumers, also requires analytics to become actionable.

In the recent period, there is a growing research interest on the concept of Machine Learning that is used in data science. This approach deals with the creation of techniques and algorithms that facilitate thecomputers to acquire knowledge and procure intelligence that relies on the previous experience. Machine Learning represents a member of AI (Artificial Intelligence) and is much associated with statistics. Here, the system would be capable of recognizing and understanding the data related to the input, such that it could make predictions and decisions by depending on that data.

In this context, the learning process begins with the data collection by several means, from multiple resources. The subsequent step is data preparation which implies a data pre-processing method to address the data-associated issues and to decrease the space dimensionality by eliminating the unnecessary data. As the volume of data utilized for learning remains huge, it is problematic for the system to proceed with the decision making. In such scenario, algorithms are devised by employing logic, probability, statistics, and certain control theory for analyzing the data and retrieving it from the earlier experiences.

One of the most important issues in healthcare recently is Diabetes. Diabetes represents a well-known metabolic disease that could adversely affect the complete body system. Usually, type 2 diabetes onset occurs in the middle age and rarely in the old age. However, diabetes incidences are also identified in children. Diabetes is driven by multiple etiologic factors such as sedentary lifestyle, food habits, body weight, and genetic susceptibility. An undiagnosed diabetes could cause the levels of blood sugar to become excess. This condition is known as hyperglycemia and this could cause complications such as cardiac stroke, diabetic foot ulcer, neuropathy, nephropathy, and retinopathy. Hence, diabetes detection at the earliest stage is central to enhance the patient related QOL (quality of life) and life expectancy enhancement (Kaur & Kumari, 2018).

2 LITERATURE REVIEW

In order to identify articles that employ data analytics and its application in healthcare especially in diabetes extensive efforts were made. Several databases were searched: the extensively used in biomedical sciences, PubMed, the IEEE digital library, ACM digital library, the DBLP Computer Science Bibliography, containing more than 3.4 million journal articles, conference papers, and other publications on computer science.

Machine learning could be also used in several areas such as traffic management, prediction of disease, robotics, gaming, face tagging and identifying, filtering of email, ranking of web page and in search engine. Among these, the prediction of disease has good implications for the clinicians (Kaur & Kumari, 2018).

For instance, predictive analytics employs a machine learning strategy for predicting the unknown or future outcomes. Predictive analytics has been explored widely in health care especially in diabetes. By applying predictive analytics in diabetic care, diabetes related diagnosis, prediction, self-management, and prevention could be possible base on the surveys.

From the past research, there are two important predictive analytic types. These are unsupervised learning and supervised learning. Unsupervised learning will not employ any earlier known findings for training its models. It relies on employing descriptive statics. It recognizes groups or clusters. On the other hand, supervised learning represents a method of building predictive models employing historical data set and generates predictive findings. Examples include time-series analysis, regression, and classification (Jayanthi, Babu & Rao, 2018).

Additionally, predictive model classification is of nine kinds such as logistic regression, naive Bayes and linear regression, classification and decision trees, business rules, natural language processing (NLP), support vector machines (SVMs), machine learning, and neural networks (NNs). However, predictive analytics employs regression models based on the existing data for predicting the majority of outcomes in the medical field. With regard to diabetes, a multi stage adjustment model is believed to be applied. This has low rate of misclassification rate and could predict which individuals is susceptible to acquire diabetes. Say, researchers use KoGES dataset to build this model (Jayanthi, Babu & Rao, 2018).

Researchers devised the physiological model that could help in predicting the level of blood glucose thirty-minutes in advance by employing the data of five patients by training with the physiological characteristics. This assisted in giving reliable outcomes than that contributed by the physicians. In the similar

context, another predictive model is a graph model based on a sparse factor. The researchers could not only acquire get a forecast of the complications but also could detect the hidden connection between the complications specific to diabetes and the test types of a lab.

In an investigation, every algorithm was implemented by employing C++ program with features like 4 GB of memory, Intel Core i7 2.66 GHz, and Mac running Mac OS X. The data set employed for the trial was gathered from a geriatric hospital. The data set consists of one year old data related to 35,530 patients, 181,934 medical records, 1945 kinds of lab tests specific to Mac running Mac OS X

and 65% of data was selected for model training and the remaining for the testing. The proposed model was thought to address knowledge skewness and sparseness. In the similar context, a hybrid model was devised for predicting if the diagnosed patient could acquire diabetes within a period of five years or not. For this purpose, the tool employed was WEKA and the specific data set used was that of PIMA Indian population with diabetes.

This model attained an accuracy of 93% (Jayanthi, Babu & Rao, 2018).The authors have adopted the process in devising the predictive model where they initially did dataset pre-processing, and then calculating the values of F-score related to chosen characteristics with increased F-score as the discriminative characteristics.

Temurtas et al. (2009) employed two separate neural networks for expressing that would produce the precise classifier to predict diabetes. The two models of neural network are probabilistic neural network and multilayer neural network. The dataset carries the diabetic data of Pima Indians with 769 samples in two classes. Among these, researchers used 575 samples for the purpose of training and 190 were employed for testing.

Divya et al. (2014) devised a prediction model as per a H-TSVM (Hybrid-Twin Support Vector Machine) to predict whether or not a novel patient can suffer from diabetes. They employed the Pima dataset for carrying out the experiment. Here, 'kernel function' served as a unique factor to keep the proposed method distinct from the others. The classifier was able to give a 88% accuracy.In a study, Ahmed (2016) proposed a predicting model which classifies the treatment plans specific to type 2 diabetes into three categories such as medication, diet and insulin.. The dataset employed for devising the model was specific to the clinic centre named 'JABER ABN ABU ALIZ' that carries tree eighteen medical records. The model was devised employing WEKA tool by using the J48 classifier and it has induced 71% accuracy.

Yet in another trial by Devi et al. (2016), the study team devised a prediction model for predicting various disease types a diabetic patient could develop. For devising the data set model a 3 year period was spent in gathering the data from a hospital containing details of 740 patients as well as 31 attributes. Following the deletion of outliers by employing DBOD (distance based outlier detection), the pre processed data was provided as the logistic regression model input and this was constructed by employing the Bipolar Sigmoid Function. This, in turn, was evaluated by employing the function of Neuro based Weight Activation. The model induced 91% accuracy in the prediction.

Some workers developed FNC a tool which could be employed for a diabetes diagnosis. This model was devised by introducing three methods Case based reasoning, neural network and fuzzy logic, with with the details of two hundred patients who possess sixteen attributes of input attributes. The study team applied Matlab to implement neural networks and fuzzy logic, and applied MyCBR plug-in to implement Case based Reasoning. Following the collection of results from three methods, the research team applied rule oriented algorithm to every technique for enhancing the accuracy. Finally, the reliable accuracy was sought for case based reasoning (Thirugnanam et al., 2016).

Osman et al. (2017) devised a KSVM hybrid model. This model has chief criteria a selection algorithm which makes it distinct from various approaches. Data set specific to PIMA was used for the trials and findings were collected. It was demonstrated that the results specific to diagnosis employing K-SVM are 99.75, 99.79, and 99.82 for learning trials, and 99.92 for the testing trials.

Anand (2015) devised a prediction model which helps in predicting if an individual would acquire diabetes by relying on the activities specific to daily lifestyle. Here, for constructing the prediction model, data set specific to PIMA diabetes was applied and the machine learning classifier Classification and Regression Trees (CART) was used. This proposed model was thought to give a 75% accuracy.

Jakhmola (2015) devised a prediction model which helps in predicting if an individual acquired diabetic condition or not. For achieving that, dataset specific to PIMA diabetes was employed. In the method proposed, earlier regulated binning technique was applied and then multiple regressions was employed for enhancing the model accuracy. Following the introduction of every method a 78% accuracy was attained.

A team led by Aljarullah (2011) devised a type 2 diabetes diagnostic tree model. They employed the data set specific to Pima Indian diabetes. The techniques specific to pre-processing are those based on numerical discretization, handling missing values, recognition and selection for enhancing the data quality. The study team applied J48 decision tree classifier and Weka tool for building the decision tree model which produced a 79% accuracy.

Jahani and Mahdavi (2016) devised a prediction model by employing the neural networks for classifying and diagnosing the diabetes specific onset and its progression. They employed data specific to 550 patients from a diabetes center. Initially, they provided training and examined the neural networks with variety of neurons and observed a neural network with various neurons that induced highest accuracy.

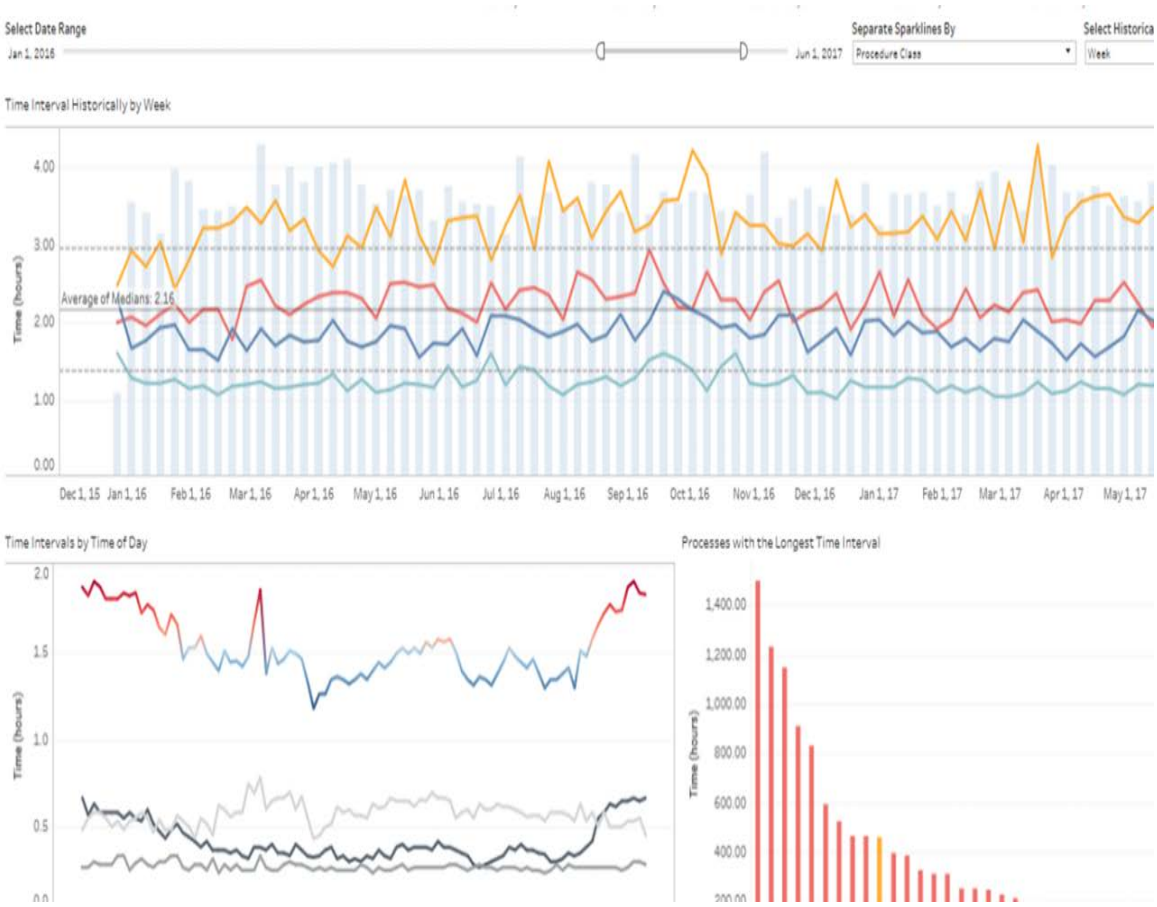


Figure 1. Data Analytics in Healthcare Trends

3 Conclusion

The main purpose of the research study was to investigate data analytics and its applications in healthcare. Primarily the focus was on diabetes as one of the biggest silent killers of patients. In conclusion, the survey done on various models specific to the area of diabetes is clear that data science has evolved well in predictive analysis with regard to prediction and diagnosis. This has furnished insights on the efficacy associated with the construction of clinical prediction models particularly in developing nations. Although, there seem to be a few gaps that need to be addressed in areas like plans specific to type 1 diabetes treatment, prediction model

implementation optimizations, using larger dataset. So, predictive analytics appears to gain much reputation with regard to the modern technology Big data. It has implications to go beyond the level of data mining.

This application has made feasible the exploitation of a huge quantity of available medical data is with regard to disease, signs and symptoms, etiology, and their impact on health, altogether.

An evidence based practice could also help in better streamlining the clinical research when applying the data science with a special emphasis on predictive analysis. Recommendations
As such and according to the results of the study, some recommendations can be given:

- The data analysts must raise their seriousity in developing secure and serious data analytics applications.
- The data science analyses applications should be user friendly and usable, so that will increase the clients' satisfaction.
- By having serious dissemination and presentation of the data science application and providing sufficient training, can raise the confidentiality towards the use of the data science software tools.

REFERENCES

[1] Ahmed, T.M. (2016). Developing a predicted model for diabetes type 2 treatment plans by using data mining. J Theor Appl Inf Technol, 90(2),181–7.

[2] AlJarullah, A.A. (2011). Decision tree discovery for the diagnosis of type II diabetes. In: International conference on innovations in information technology. New York: IEEE; 2011.

[3] Devi, M.N, et al. Developing a modified logistic regression model for diabetes mellitus and identifying the important factors of type II DM. Indian J Sci Technol, 9(4).

[4] Jahani M., & Mahdavi, M. (2016). Comparison of predictive models for the early diagnosis of diabetes. Healthc Inform Res, 22(2),95–100.

[5] Jayanthi, N., Babu, B.V. & Rao, N.S. (2017).Survey on clinical prediction models for diabetes prediction. J Big Data (4), 26.

[6] Kaur,H., & Kumari, V. (2018).Predictive modelling and analytics for diabetes using a machine learning approach. Retrieved from https://www.sciencedirect.com/science/article/pii/S221083271830365X

[7] Osman AH, et al. (2017).Diabetes disease diagnosis method based on feature extraction using K-SVM. Int J Adv Comput Sci Appl, 8(1).

[8] Thirugnanam, M, et al.(2016). Hybrid tool for diagnosis of diabetes. IIOAB J, 7(5).