

Loku, Lindita; Fetaji, Bekim; Krstev, Aleksandar; Fetaji, Majlinda; Zdravev, Zoran

**Conference Paper**

## Analyses of Diabetes Data and Its Data Analytics Perspectives of Usage in the Health System in Kosovo

**Provided in Cooperation with:**

IRENET - Society for Advancing Innovation and Research in Economy, Zagreb

*Suggested Citation:* Loku, Lindita; Fetaji, Bekim; Krstev, Aleksandar; Fetaji, Majlinda; Zdravev, Zoran (2019) : Analyses of Diabetes Data and Its Data Analytics Perspectives of Usage in the Health System in Kosovo, In: Proceedings of the ENTRENOVA - ENTERprise REsearch InNOVation Conference, Rovinj, Croatia, 12-14 September 2019, IRENET - Society for Advancing Innovation and Research in Economy, Zagreb, pp. 17-24

This Version is available at:

<https://hdl.handle.net/10419/207659>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-nc/4.0/>

# Analyses of Diabetes Data and Its Data Analytics Perspectives of Usage in the Health System in Kosovo

Lindita Loku

Computer Sciences, University Goce Delcev Stip, Macedonia

Bekim Fetaji

Faculty of Informatics, University Mother Teresa, Skopje, Macedonia

Aleksandar Krstev

Computer Sciences, University Goce Delcev Stip, Macedonia

Majlinda Fetaji

Computer Sciences, South East European University, Tetovo, Macedonia

Zoran Zdravev

Computer Sciences, University Goce Delcev Stip, Macedonia

## Abstract

The focus of the research study is to investigate and analyses the current data processing and analytics issues in health focusing in diabetes data order to make sense of the data and use it to improve the health system. Within the investigation of the set of data from the ministry of health of Kosovo realised several types of analyses using different software tools. The current main challenge is to efficiently translate science into modern medicine that is limited by our capacity to process and understand these data. So, it is obviously needed to devise new mathematical as well as computational model with the ability to analyse Data. This will help the clinicians to retrieve useful information and then accurately diagnose and treat patients to improve patient outcomes. Scientist as well as medicine workers should become more aware and understand the value of Data analytics in providing valuable insights. Data, derived by patients and consumers, also requires analytics to become actionable. Based on the above results and test of homogeneity we assume that the first group Insuline M will in the future retain higher rate with a standard deviation of .176 compared with the first group Insuline R. This gives as opportunity to predict the management of diabetes for the next years where we can conclude that group 1 will continue to prevail and require more special treatment compared with group 2. Insights are provided as well as arguments and discussed the benefits from the study.

**Keywords:** diabetes, data processing, machine learning, computational model, data analytics

**JEL classification:** A31

## Introduction

Predictive analytics which is help to healthcare organizations to analyse data collected on the past and predict the possibilities of future occurrences in order to enable proper decisions of their patient [2]. Devising such predictive model can help healthcare to understand better and bring decisions based on informed data and with this to become more effective and efficient in solving health issues. Predictive

analytics recently has been proven as very efficient in improving safety and performance of patient outcomes.

The modernization of the healthcare industry's focuses on processing massive health records, and accesses those for analysis and with this it has potential to substantially increase the likelihood of eliminating illnesses. The unstructured nature of Big Data that is collected, it is required to structure and normalise their size into smaller quantum to arrive to possible solution.

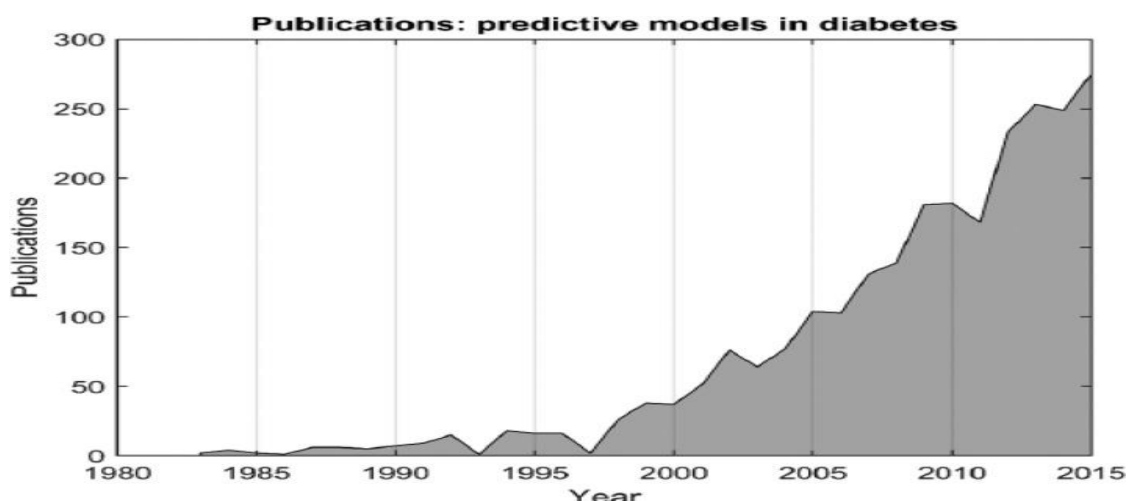
With the advancement of the technology most of medical records and information is being stored and processed electronically. This brings a lot of opportunities in the era of "big health care data" for different predictions and exploring different aspects of health care previously humanly impossible.

Devised different predictive models have the potential to improve the chances for healing of patients, use different kinds of patient information and output prognostic results in a clinical setting [2]. According to [1] "This could be used for clinical decision support, disease surveillance, and population health management to improve patient care"

One of such illness that can be predicted is Diabetes, that is considered as one of the top priorities in medical science. Today more and more people are dying from Diabetes than from Cancer and other illnesses. "Diabetes is a very serious disease that can lead to a large number of very serious long-term complications such as blindness, amputation and heart disease if not treated properly in time" [5] [6] [7]. Also, the very early stages of type 2 diabetes are asymptomatic, and therefore many patients can be undiagnosed for years [2]. "Treatment, especially with insulin, is not without adverse effects such as risk of hypoglycaemia and weight gain" [4]. Fortunately, recently we are being witnessing rapid growing awareness of the possibilities in the field of Data Science by using available information for predicting diabetes. The number of published articles has risen every year, from 5 publications in 1990 to about 300 in 2015, as illustrated in Figure 1 below.

Figure 1

Predictive Models in diabetes - Number of publications index by PubMed with keywords "predictive AND model AND diabetes."



Note: The 2015 count is extrapolated based on the number from May 27, 2015.

Source: Authors' work

## Literature Review

Predictive Analytics involves different statistical models as well as analytical techniques that are used for devising different models that can predict many future occurrences [2].

As discussed by [6] "almost any statistical regression model can be used as a predictive model". Generally, according to [2] "there are 2 kinds of models: parametric and nonparametric. Parametric models can make assumptions regarding the underlying data distribution, whereas nonparametric models (and semiparametric models) make fewer or no assumptions about the underlying distribution. The most common approach is to use a regression model for prediction". This often also involves the use of classic statistical methods to devise the predictive model based on some statistical technique [5]. "These models often utilize a broad range of methods involving machine learning and pattern recognition, among others, [3], [4] and they are often, but not always, limited to classification tree, neural network, k-nearest neighbour" [7].

## Results and findings from research data from Kosovo

Our study is built on numerous studies from the reviewed published literature and then used data from Ministry of Health of Kosovo.

Table 1

Analysis of Insulin Factor Over Time Period of 1 year

Description	Insulina Humane (Gensulin M)			Insulina Humane (Gensulin R)		
Orders over 2016	The first six months 2016	The second sixth month of 2016		The first six months 2016	The second month of 2016	sixth
	0	190,820		0	22,965	
Total orders 2015	190,820			22,965		
Contracted amount for the year (2015-2018)	726,000			85,080		
Percentage of movement in relation to the contracted value of 2016	380.46% (280.46% decrease below the contracted value of 2016)			370.47% (270.47% decrease below the contracted value of 2016)		
Orders and planning throughout 2017	Orders, the first six months of 2017	Planning of the second sixth month 2017		Orders, the first six months of 2017	Planning of the second sixth month 2017	the sixth month
	49,920	85,540		10,000	16,838	
Total	135,460			26,838		

<b>Contracted amount for the year (2016-2018)</b>	<b>726,000</b>		<b>85,080</b>	
<b>Percentage of Growth / Decrease in the ratio: 2017 through 2016</b>	29.01% decrease below the contracted value of 2016		116.86% (16.86% increase over the commissioned value of 2017)	
<b>Planning throughout 2018</b>	The first six months 2018	The second sixth month of 2018	The first six months 2018	The second sixth month of 2018
	85,540	64,464	8,622	5,774
<b>Total Planning 2018</b>	132,490		14,396	
<b>Contracted amount for the year (2016-2018)</b>	<b>726,000</b>		<b>85,080</b>	
<b>Percentage of Growth / Decrease in the ratio: 2018 through 2017</b>	2.19% decrease below the contracted value of 2017		46.36% decrease below the contracted value of 2017	

Source: Authors' work

From Table 1 we can see the change within one year for Gensulin M and Gensulin R and the influence on the health system they have.

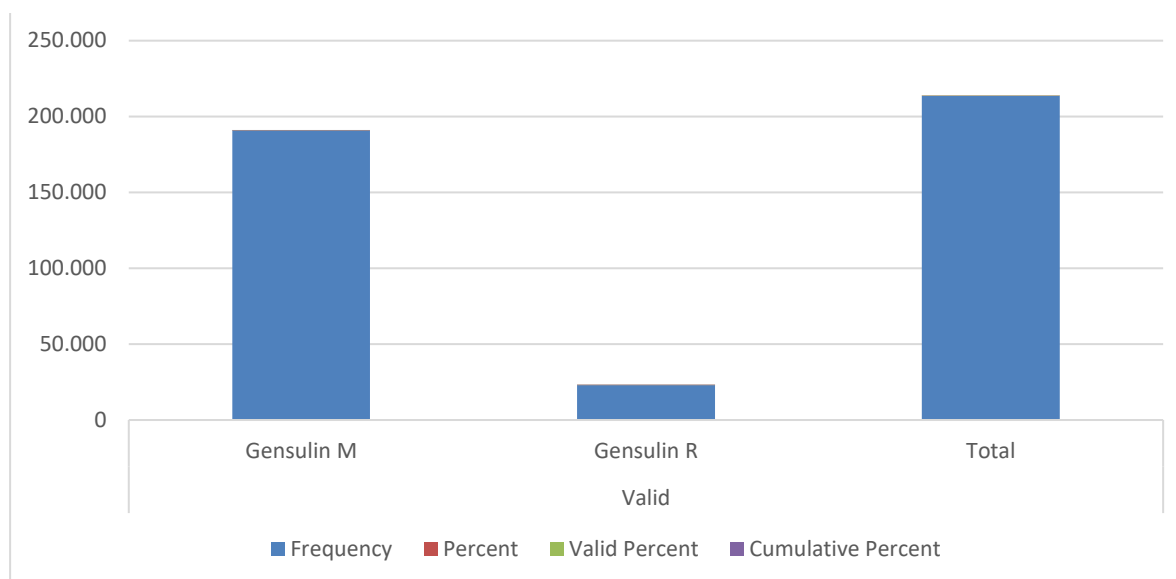
Table 2  
Data Analyses of Patients

		<b>Type of Gensuline</b>			
		Frequency	Percent	Valid Percent	Cumulative Percent
<b>Valid</b>	Gensulin M	190,820	89.25789929	89.2	89
	Gensulin R	22,965	10.74210071	10.7	11
	Total	213,785	100.0	100.0	

Source: Authors' work

From Table 2 we can see the frequency of patients that use Gensulin M is much higher than the patients with Gensulin R. This influences the health system to prepare and be more engaged to patients with Gensulin M needs.

Figure 2  
Data Analyses of Patients



Source: Authors' work

In total, there are 213,785 patients and 22,965 of them (10.7%) are from the Insuline R group and 190,820 patients (89.2%) are from the Insuline M group. All of the answers are valid rate (100%).

Table 3  
Data Analyses of Frequencies

		Gensulin M	Gensulin R
<b>N</b>	Valid	89	11
	Missing	0	0
<b>Mean</b>		8.54	4.23
<b>Std. Error of Mean</b>		.280	.140
<b>Median</b>		89.25	10.74
<b>Mode</b>		10	1.5
<b>Std. Deviation</b>		1.945	.921
<b>Variance</b>		4.784	.757

Source: Authors' work

The mean values for the group 1 are higher than group 2. (Group 1 mean = 89.25, Group 2 mean = 10.74) This shows that group 1 students were affected by current status prevalence of Gensuline M. This shows that data set is close to the normal distribution with a little skewness to the left.

Table 3  
Test of Homogeneity of Variances

Test of Homogeneity of Variances			
Grade Average			
Levene Statistic	df1	df2	Sig.
.892	1	78	.176

Source: Authors' work

To test the Homogeneity of Variances, we apply Homogeneity test as above. This shows that our P value (.476) is bigger than the standard a level (.05) so we retain the

null hypothesis (has difference) for the assumption of homogeneity of variance and conclude that there is a significant difference between the two groups' variances. Based on the above results and test of homogeneity we assume that the first group Insuline M will in the future retain higher rate with a standard deviation of .176 compared with the first group Insuline R.

## Conclusion and future work

Within the research study much effort has been put into developing predictive model for use in the management of diabetes and its complications.

Based on the study results and test of homogeneity we assume that the first group Insuline M will in the future retain higher rate with a standard deviation of .176 compared with the first group Insuline R. This means that the health system and hospitals should expect and to prepare and be more engaged to patients with Gensulin M needs. This gives an opportunity to predict the management of diabetes for the next years where we can conclude that group 1 will continue to prevail and require more special treatment compared with group 2.

However, this model has not been implemented and the clinical impact has not been yet investigated and remains as future work.

Although evidence from implementation is lacking, we argue that this predictive model has the potential to transform the way health care providers use sophisticated technologies and much insight may be gained and more informed decisions made by drawing on the large amount of electronically stored clinical data. Scientists and health-care providers may learn from one another when it comes to understanding the value of Data analytics.

Data, derived by patients and consumers, also requires analytics to become actionable.

## References

1. Aljumah, A. A., Ahamad, M. G., Siddiqui, M. K. (2013), "Application of data mining: Diabetes health care in young and old patients", *Journal of King Saud University-Computer and Information Sciences*, Vol. 25, No. 2, pp. 127-136.
2. Bhat, V. H., Rao, P. G., Krishna, S., Shenoy, P. D., Venugopal, K. R., Patnaik, L. M. (2011), "An efficient framework for prediction in healthcare data using soft computing techniques", in the *Proceedings of the First International Conference on Advances in Computing and Communications*, Kochi, India, Springer, pp. 522-532.
3. Bhat, V. H., Rao, P. G., Shenoy, P. D. (2009), "An Efficient Prediction Model for Diabetic Database Using Soft Computing Techniques," in the *Proceedings of the 12th International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, Delhi, India, Springer, pp. 328-335.
4. Mishra, N., Silakari, S. (2012), "Predictive analytics: A survey, trends, applications, opportunities & challenges", *International Journal of Computer Science and Information Technologies*, Vol. 3, No. 3, pp. 4434-4438.
5. Rajesh, K., Sangeetha, V. (2012), "Application of data mining methods and techniques for diabetes diagnosis", *International Journal of Engineering and Innovative Technology (IJEIT)*, Vol. 2, No. 3, pp. 224-229.
6. Sabibullah, M., Shanmugasundaram, V., Priya, R. (2013), "Diabetes patient's risk through soft computing model", *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, Vol. 2, No. 6, pp. 60-65.
7. Sadhana, S. S., Shetty, S. (2014), "Analysis of diabetic data set using hive and r", *International Journal of Emerging Technology and Advanced Engineering*, Vol. 4, No. 7, pp. 626-629.



## About the authors

Lindita Loku is a Doctoral student at the Faculty of Computer Sciences at Computer Sciences, University Goce Delcev Stip. Her main research interests are Statistical data analysis, data analytics, virtual learning environments, and closely related fields. Published 2 scientific papers in international conferences and 1 international journal. The author can be contacted at [lindita.loku@student.ugd.edu.mk](mailto:lindita.loku@student.ugd.edu.mk).

Bekim Fetaji is Full Professor of Informatics at University Mother Teresa (UMT) - Skopje. He is currently Vice Rector for Science & Research at UMT. Former head of Research group in Programming and software Engineering and Formals specifications. Previously Dean of Computer Science Faculty and before that vice-dean for academic issues in computer Sciences at South East European University. Vice President of Macedonian Board of Accreditation of Higher Education in Macedonia. Received his PhD in Computer Sciences at the Faculty of Computer Sciences in Graz University of Technology with the dissertation thesis "E-learning indicators- a multidimensional model for designing and developing e-learning software solutions". He received his Master Degree in Oxford Brookes University, Oxford, UK with his master thesis entitled: "Agile approach in software engineering web applications" Main research interests are in software engineering, programming, technology enhanced education, data processing, and closely related fields. Participated in several project teams within different programs such as Tempus, Erasmus and other national and international research projects. Published more than 100 scientific papers in international conferences and more than 25 international journals. The author can be contacted at [bekim.fetaji@unt.edu.mk](mailto:bekim.fetaji@unt.edu.mk).

Aleksandar Krstev is an Associate Professor at the Faculty of Computer Sciences at Computer Sciences, University Goce Delcev Stip. He received his PhD in Computer Sciences at the same University in PhD in 2010. His main research interests are: data analytics, data science, web applications, service-oriented architecture. Published more than 40 scientific papers in international conferences and more than 10 international journals. The author can be contacted at [aleksandar.krstev@ugd.edu.mk](mailto:aleksandar.krstev@ugd.edu.mk).

Majlinda Fetaji is an Associate Professor at the Faculty of Computer Sciences at South East European University -SEEU. She received her PhD in Computer Sciences at the Faculty of Contemporary Sciences and Technologies at South East European University -SEEU with the dissertation thesis "MAI instructional model, MLUAT testing methodology and TBMLM methodology framework for developing mobile learning software solutions". Her main research interests are algorithms and data structures, programming, e-learning, m-learning, virtual learning environments, and closely related fields. Awarded "Researcher of the year 2008" from the Macedonian Academy of Sciences for her research work in mobile learning and Technology Enhanced Education. Participated in different projects in Tempus, Erasmus and other national and international research projects. Head of the Quality Team of the Faculty of Computer Sciences. Published more than 80 scientific papers in international conferences and more than 20 international journals. The author can be contacted at [m.fetaji@seeu.edu.mk](mailto:m.fetaji@seeu.edu.mk).

Zoran Zdravev is a Full Professor at the Faculty of Computer Sciences at Computer Sciences, University Goce Delcev Stip. He received his PhD in Computer Sciences at in 2001. His main research interests are: e-learning, semantic enabled architectures,



mobile and ubiquitous applications, e-business and e-government services. Published more than 60 scientific papers in international conferences and more than 15 international journals. The author can be contacted at [zoran.zdravev@ugd.edu.mk](mailto:zoran.zdravev@ugd.edu.mk).