

Data Mining Framework for Nutrition Ranking

Nauman Aziz

Superior University, Pakistan, na.pitafi@gmail.com

Shabib Aftab

Virtual University of Pakistan, Pakistan, shabib.aftab@gmail.com

Abstract:

The goal of this research is to use the technology of Data Mining in a dataset for a ranking of three diets on the respondents and investigate such tools advantages and limitations such as large amount of manipulation before analysis replications with the same results from the analysis bias. Here we can see the ethical consequences of such programs in details.

Keywords—Data Science, Nutrition Assessment, Clustering Analysis

Introduction

Dietary nutrition tool is basically a software framework for the assessment and evaluation of different respondent's nutrition status. There is a classified database which is attached to this tools in order to investigate the group class and intake meal species.(H. M. Alzoubi, Alshurideh, et al., 2021; Koroušić Seljak et al., 2013) The operator of such tools should firstly select the specific group in which the respondent belong to, such as infant, children, pregnant woman, adult, people with illness, afterwards the gender, height, weight and other general conditions and the intake food class such as different kind of dietary conditions such as the amount within a certain intake time are specified.(M. A. Afifi, n.d.; Al Kurdi, n.d.; Alhamad et al., 2021; Batayneh et al., 2021; Hamadneh & Alshurideh, n.d.; Khan et al., 2021) The ultimate goal of this tool is to nutrition assessment on the patient in caring hospital. Anyway the tool can be applied for the healthy persons in schools, hotels, restaurants and other places as a reference model to help and structure them for a good eating habit.(Ali et al., 2021; Ghazal, Hasan, et al., 2021; Matloob et al., 2021; Nunes et al., 2015) The aim of this article is to use the Dataset on the between 78 respondents on different Data Mining Algorithms.(M. A. M. Afifi et al., 2020; Bennett, 1992) We would like also to visualize the results through different clustering and modeling into Heat Map Matrix for the better evaluation of the algorithm and understanding of the dataset. Therefore, the ultimate goal in this research is to finding out the possibility of Input variables age, height and weight and gender to rank the three Diet types included in the Dataset.(Al Shebli et al., 2021; Ali et al., 2021; Alnuaimi et al., 2021; Alzoubi, Vij, et al., 2021; Alzoubi & Aziz, 2021; Ghazal et al., 2020; Naqvi Rabab and Soomro, 2021)

Data clustering is simply the process of organizing (clustering) data into several clusters for easy, efficient and most effective use. A properly designed clustering algorithm makes data available for business use, in other words, it makes the data easy to retrieve and analyze for modeling. After importing the Dataset file named Project 2 - stcp-Rdataset-Diet.csv the first

visual records the 78 instances and 7 attributes. At the first glance it seemed that the number of instances for our evaluation is quite small.

DATASET DESCRIPTION

Since the Diet data analysis based on the ethical and subjective reasons behind the dietary status of the respondents it was more valuable to exactly watch the variable importance of the predictors of the analysis rather than just writing equations for gender, age or height and weight.(Alnazer et al., 2017; H. Alzoubi & Ahmed, 2019; H. M. Alzoubi et al., 2020) The following figure shows the distribution and some statistics on out raw dataset and we can see the normal distribution on the age, height and weight after 6 weeks.

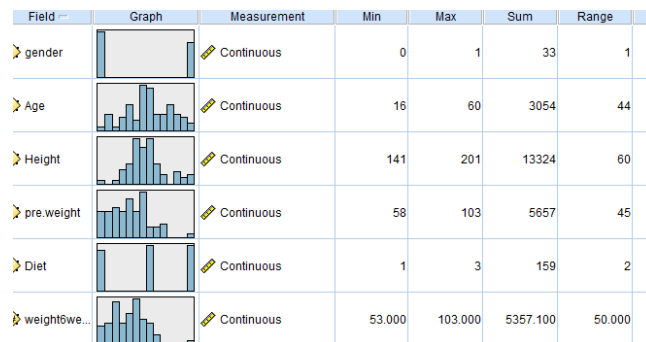


Figure 1 - Statistical overview of the Raw Datasets

Now it seems not only the number of the total respondents are enough but also the type and the size of the predictors and attributes are inefficient. It can be seen that there is a correlation between the pre-weight and weight after six weeks of diets. This can be seen on the next figure too. In our view it shouldn't be catastrophic about this since were experimenting the ethical and ranking phenomena behind dietary and not the statistics.

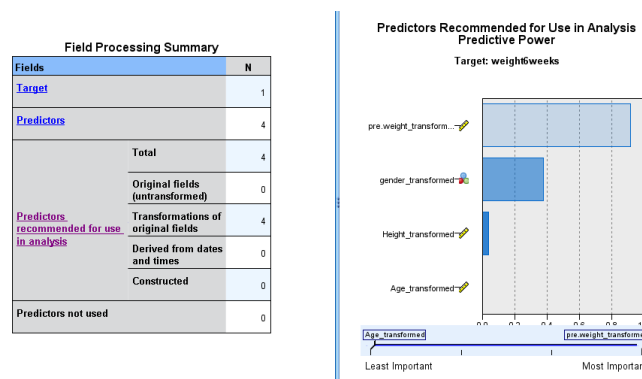


Figure 2- predictive power of the Dataset

Obviously the height, age and gender was not important attribute. We have reduced the attribute with a supervised clustering algorithms through different such as Kohonen, Two steps Cluster Node and K-Means where some of them resulted in different cluster number and size.

A Silhouette Ranker clustering selection is used to evaluate the clustering efficiency of the algorithm. This algorithm has no automatic way of normalizing the data interpretation. Preprocessing of the data is therefore a very important step when doing data classification using SPSS Modeler.

Use?	Graph	Model	Build Time (mins)	Silhouette	Number of Clusters	Smallest Cluster (N)
<input checked="" type="checkbox"/>		Two...	< 1	0.573	2	33
<input type="checkbox"/>		K-m...	< 1	0.484	5	2
<input type="checkbox"/>		Koh...	< 1	0.282	11	1

Smallest Cluster (N)	Smallest Cluster (%)	Largest Cluster (N)	Largest Cluster (%)	Smallest/Largest	Importance
33	43	43	56	0.767	0.749
2	2	22	28	0.091	1.0
1	1	14	17	0.071	0.995

Figure 3- Clustering Models used for Data Interpretation

The predictor importance result of the first and second clustering namely, TwoStep and K-Means are represented in figure 4. As it can be seen the results are in good quality of clustering.

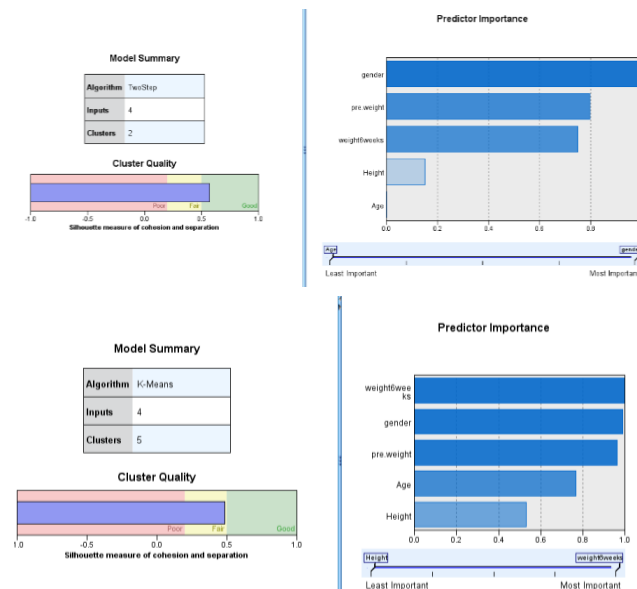


Figure4 - Graphical representation of clustering in accordance to the 6 weeks weight

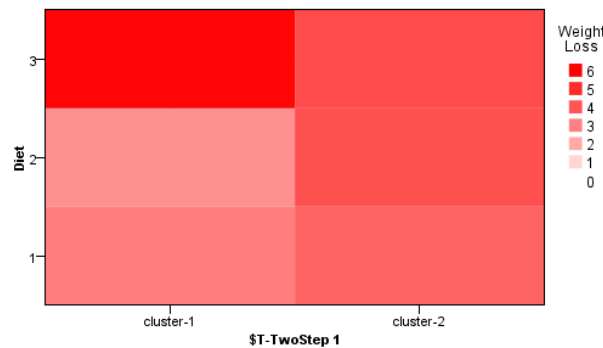
From the above results it is clear that in spite of better quality of the Two Step the K-Means clustering exhibits a better prediction power. In the two step method the Age and Height are not competent for the prediction. It should be furthermore noted that cluster 3 in K-Means are

outlier and can be excluded from other clusters. This result also confirm the K-Means can include the Age and Height into the model.

DATA MODELING

For the modeling of two clustering's, the derived node from the clustering 2 Steps imported to the dataset as a new filed. In addition to the variable diet type the new filed is also applied as splitting variable. Afterwards two variable height and age are excluded from the model and the Auto numeric node has been used to build 8 models. In the following figure the variable types are indicated . Furthermore, the difference between the Weight after 6 weeks and pre weight are calculated and created in the calculation. (The new created variable is called Weight Loss)(Bennett & Mangasarian, n.d.; Chapman et al., 2000)

Model	Build Time (mins)	Correlation	No. Fields Used	Relative Error
KNN Algor...	< 1	0.591	2	0.67
Neural Ne...	< 1	0.276	2	0.928
Regressi...	< 1	0.160	2	0.975
Generaliz...	< 1	0.160	2	0.975
SVM 1	< 1	0.072	2	0.998



Field	Measurement	Values
gender	Flag	1/0
Age	Continuous	[16,60]
Height	Continuous	[141,201]
pre.weight	Continuous	[58,103]
Diet	Nominal	1,2,3
weight6weeks	Continuous	[53.0,103.0]
\$T-TwoStep 1	Nominal	cluster-1,c...
Weight Loss	Continuous	[-2.100000...

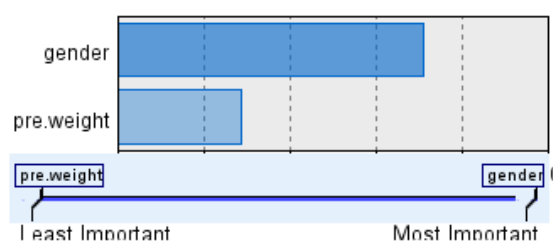
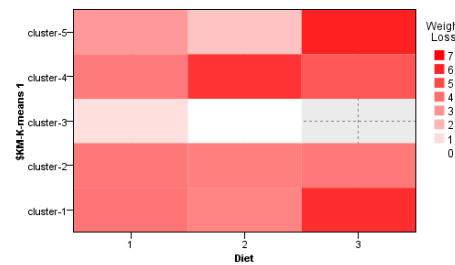
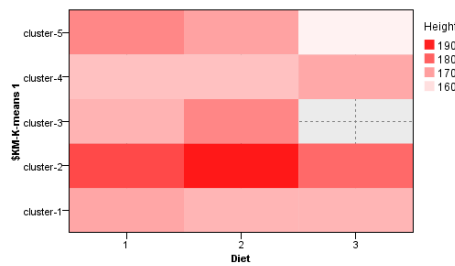
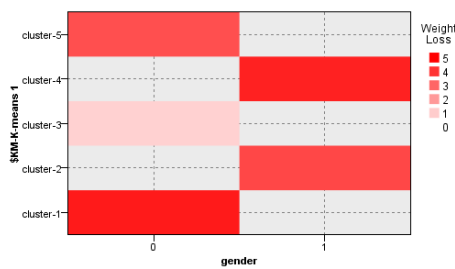
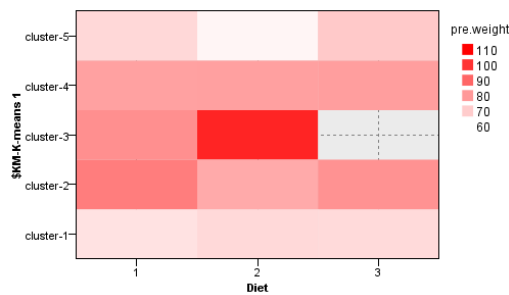


Figure 5- Results of estimation with twosteps clustering of weight Loss

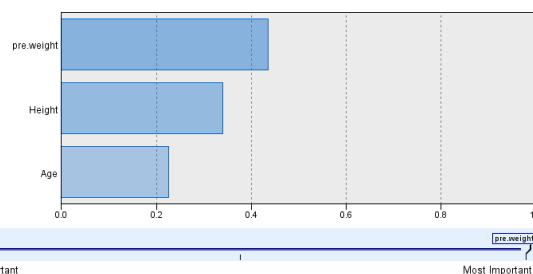
RESULTS

Results and error rates for the build model are shown. Accordingly, the Diet three is better than 2 and 1 in all clusters but it is still difficult to generalize this between diet 1 and 2. Since the relative error is high, and only gender is considered and still age and height were excluded this model is not efficient. Finally yet importantly, the effect of gender on this model outperforms the pre.weight variable.

It is therefore important to test the other clustering.



Predictor Importance
Target: Weight Loss



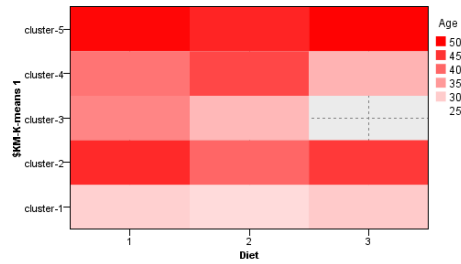
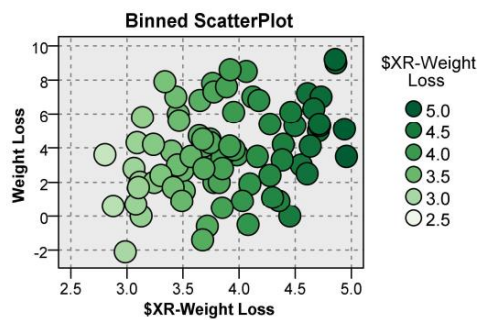
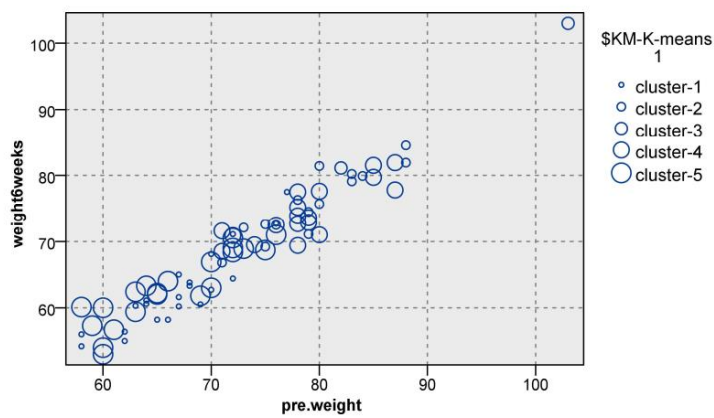


Figure 6 - Results of estimation with K-Means clustering of weight Loss

In spite of problem for the clustering with variable age, preweight and height can be distinguished from the above graph. It is also notable to check the comparison between diet 2 and 1 are clearer. However, the diet 3 outperform the other one with a good distance.

SPSS Modeller Bivariate Analysis)



Results for output field Weight Loss Overall Results Comparing \$ E-Weight Loss with Weight Loss

Minimum Error	-4.906
Maximum Error	5.704
Mean Error	-0.0
Mean Absolute Error	1.809

Standard Deviation	2.306
Linear Correlation	0.428
Occurrences	78

Output field Weight Loss, splitting by field Diet = 1

Minimum Error	-3.829
Maximum Error	5.704
Mean Error	0.0
Mean Absolute Error	1.597
Standard Deviation	2.191
Linear Correlation	0.209
Occurrences	24

Comparing \$ E-Weight Loss with Weight Diet = 2

Minimum Error	-4.906
Maximum Error	4.236
Mean Error	0.0
Mean Absolute Error	2.011
Standard Deviation	2.499
Linear Correlation	0.14
Occurrences	27

Diet = 3 Comparing \$ E-Weight Loss with Weight Loss

Minimum Error	-4.231
Maximum Error	4.339
Mean Error	-0.0
Mean Absolute Error	1.796
Standard Deviation	2.293

Analysis **Standard Deviation** 2.293

Linear Correlation	0.289
Occurrences	27

Analysis Number of clusters: 5

Iteration	Error
1	0.247
2	0.134
3	0.116
4	0.024
5	0.037
6	0.048
7	0.027
8	0.0

DISCUSSION

To ensure the quality of documenting the results, CRISP-DM (Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology) has been used in this report to analysis the data. The whole processes of CRISP-DM are Data understanding, Data preparation, Modeling, Evaluation and Deployment.(Alshurideh et al., 2020; H. Alzoubi et al., 2020; Jevons et al., 2010; Rehman et al., 2021)

As mentioned strategical reflections and giving the solution for the nutrition evaluation tool is the subject of this research. We have experienced on vast methods of data mining for the dietary dataset to answer the questions of research. Are there any main factors effecting the quality of the information out of the 87 attributes? We have deduced from the results and finding and comparing the pair of data that supposed to have effect on the weight Loss but still some points and comparison remain unclear such as age and height. Most models confirm the pre weight as an explanatory variable for the ranking of the dietary nutrition.(H. M. Alzoubi & Yanamandra, 2020; Dr. Deepak Kalra Dr. Mohammed A. M. Afifi, 2020; Joghee et al., n.d.; Mehmood & Al-Gasaymeh, 2019; Sahi et al., 2017; Vallaster & von Wallpach, 2013)

We have searched the Internet and see the reflection of our work from the websites of some nutrition tools developed for the improvement of the eating habit. The ultimate goal of this tool is to nutrition assessment on the patient in caring hospital. Anyway the tool can be applied for the healthy persons in schools, hotels, restaurants and other places as a reference model to help and structure them for a good eating habit.(Ghazal et al., 2013; Ghazal, Alshurideh, et al., 2021; Ghazal, Said, et al., 2021; Hasan et al., 2021)

REFERENCES

1. Afifi, M. A. (n.d.). The Impact of Deploying the Internet of Things and How Will It Change Our Lives. Retrieved September 29, 2021, from www.solidstatetechnology.us
2. Afifi, M. A. M., Kalra, D., & Ghazal, T. M. (2020). The Role of Training in Determining Citizen-Consumer Attitudes Towards the Use of e-Government.
3. Al Kurdi, B. (n.d.). THE INTERPLAY AMONG HRM PRACTICES, JOB SATISFACTION AND INTENTION TO LEAVE: AN EMPIRICAL INVESTIGATION.
4. Al Shebli, K., Said, R. A., Taleb, N., Ghazal, T. M., Alshurideh, M. T., & Alzoubi, H. M. (2021). RTA's Employees' Perceptions Toward the Efficiency of Artificial Intelligence and Big Data Utilization in Providing Smart Services to the Residents of Dubai. 573–585. https://doi.org/10.1007/978-3-030-76346-6_51
5. Alhamad, A. Q. M., Akour, I., Alshurideh, M., Al-Hamad, A. Q., Kurdi, B. Al, & Alzoubi, H. (2021). Predicting the intention to use google glass: A comparative approach using machine learning models and PLS-SEM. *International Journal of Data and Network Science*, 5(3), 311–320. <https://doi.org/10.5267/IJDNS.2021.6.002>
6. Ali, N., Ahmed, A., Anum, L., Ghazal, T. M., Abbas, S., Khan, M. A., Alzoubi, H. M., & Ahmad, M. (2021). Modelling supply chain information collaboration empowered with machine learning technique. *Intelligent Automation and Soft Computing*, 30(1), 243–257. <https://doi.org/10.32604/IASC.2021.018983>
7. Alnazer, N. N., Alnuaimi, M. A., & Alzoubi, H. M. (2017). Analysing the appropriate cognitive styles and its effect on strategic innovation in Jordanian universities. *International Journal of Business Excellence*, 13(1), 127–140. <https://doi.org/10.1504/IJBEX.2017.085799>
8. Alnuaimi, M., Alzoubi, H. M., Ajelat, D., & Alzoubi, A. A. (2021). Towards intelligent organisations: An empirical investigation of learning orientation's role in technical innovation. *International Journal of Innovation and Learning*, 29(2), 207–221. <https://doi.org/10.1504/IJIL.2021.112996>
9. Alshurideh, M., Gasaymeh, A., Ahmed, G., Alzoubi, H., & Kurd, B. Al. (2020). Loyalty program effectiveness: Theoretical reviews and practical proofs. *Uncertain Supply Chain Management*, 8(3), 599–612. <https://doi.org/10.5267/J.USCM.2020.2.003>
10. Alzoubi, H., & Ahmed, G. (2019). Do TQM practices improve organisational success? A case study of electronics industry in the UAE. *International Journal of Economics and Business Research*, 17(4), 459–472. <https://doi.org/10.1504/IJEER.2019.099975>
11. Alzoubi, H., Alshurideh, M., Kurdi, B. Al, & Inairat, M. (2020). Do perceived service value, quality, price fairness and service recovery shape customer satisfaction and delight? A practical study in the service telecommunication context. *Uncertain Supply Chain Management*, 8(3), 579–588. <https://doi.org/10.5267/J.USCM.2020.2.005>
12. Alzoubi, H. M., Ahmed, G., Al-Gasaymeh, A., & Al Kurdi, B. (2020). Empirical study on sustainable supply chain strategies and its impact on competitive priorities: The mediating role of supply chain collaboration. *Management Science Letters*, 10(3), 703–708. <https://doi.org/10.5267/J.MSL.2019.9.008>
13. Alzoubi, H. M., Alshurideh, M., & Ghazal, T. M. (2021). Integrating BLE Beacon Technology with Intelligent Information Systems IIS for Operations' Performance: A Managerial Perspective. 527–538. https://doi.org/10.1007/978-3-030-76346-6_48
14. Alzoubi, H. M., & Aziz, R. (2021). Does Emotional Intelligence Contribute to Quality of Strategic Decisions? The Mediating Role of Open Innovation. *Journal of Open Innovation: Technology, Market, and Complexity* 2021, Vol. 7, Page 130, 7(2), 130. <https://doi.org/10.3390/JOITMC7020130>
15. Alzoubi, H. M., Vij, M., Vij, A., & Hanaysha, J. R. (2021). What leads guests to satisfaction and loyalty in UAE five-star hotels? AHP analysis to service quality dimensions. *Enlightening Tourism*, 11(1), 102–135. <https://doi.org/10.33776/ET.V11I1.5056>
16. Alzoubi, H. M., & Yanamandra, R. (2020). Investigating the mediating role of information sharing strategy on agile supply chain. *Uncertain Supply Chain Management*, 8(2), 273–284. <https://doi.org/10.5267/J.USCM.2019.12.004>
17. Batayneh, R. M. Al, Taleb, N. M., Said, R. A., Alshurideh, M. T., Ghazal, T. M., & Alzoubi, H. M. (2021). IT Governance Framework and Smart Services Integration for Future Development of Dubai

- Infrastructure Utilizing AI and Big Data, Its Reflection on the Citizens Standard of Living. AICV.
18. Bennett, K. P. (1992). Decision Tree Construction Via Linear Programming. <https://minds.wisconsin.edu/handle/1793/59564>
 19. Bennett, K. P., & Mangasarian, O. L. (n.d.). ROBUST LINEAR PROGRAMMING DISCRIMINATION OF TWO LINEARLY INSEPARABLE SETS.
 20. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. Undefined.
 21. Dr. Deepak Kalra Dr. Mohammed A. M. Afifi, T. M. G. (2020). Integration of Collaboration Systems in Hospitality Management as a Comprehensive Solution. *International Journal of Advanced Science and Technology*, 29(8s), 3155–3173. <http://sersc.org/journals/index.php/IJAST/article/view/16386>
 22. Ghazal, T. M., Afifi, M. A. M., & Kalra, D. (2020). Data Mining and Exploration: A Comparison Study among Data Mining Techniques on Iris Data Set.
 23. Ghazal, T. M., Alshurideh, M. T., & Alzoubi, H. M. (2021). Blockchain-Enabled Internet of Things (IoT) Platforms for Pharmaceutical and Biomedical Research. AICV.
 24. Ghazal, T. M., Hasan, M. K., Alshurideh, M. T., Alzoubi, H. M., Ahmad, M., Akbar, S. S., Kurdi, B. H. Al, & Akour, I. A. (2021). IoT for Smart Cities: Machine Learning Approaches in Smart Healthcare - A Review. *Future Internet*, 13, 218.
 25. Ghazal, T. M., Said, R. A., & Taleb, N. M. (2021). Internet of vehicles and autonomous systems with AI for medical things. *Soft Computing*, 1–13.
 26. Ghazal, T. M., Soomro, T. R., & Shaalan, K. (2013). Integration of Project Management Maturity (PMM) Based on Capability Maturity Model Integration (CMMI). *European Journal of Scientific Research*, 99.
 27. Hamadneh, S., & Alshurideh, M. (n.d.). AN INVESTIGATION OF THE ROLE OF SUPPLY CHAIN VISIBILITY INTO THE SCOTTISH BLOOD SUPPLY CHAIN.
 28. Hasan, M., Ghazal, T., Alkhaifah, A., Abu Bakar, K., Omidvar, A., Nafi, N., & Agbinya, J. I. (2021). Fischer Linear Discrimination and Quadratic Discrimination Analysis based Data Mining Technique for Internet of Things Framework for Healthcare. *Front. Public Health*.
 29. Jevons, C., Veloutsou, C., Christodoulides, G., & de Chernatony, L. (2010). Introduction: Thought leadership in brand management. *Journal of Business Research*, 63(11), 1111–1112. <https://doi.org/10.1016/J.JBUSRES.2009.10.005>
 30. Joghee, S., Alzoubi, H. M., & Dubey, A. R. (n.d.). Decisions Effectiveness Of FDI Investment Biases At Real Estate Industry: Empirical Evidence From Dubai Smart City Projects. Retrieved September 30, 2021, from www.ijstr.org
 31. Khan, Q. T. A., Ghazal, T. M., Abbas, S., Khan, W. A., Khan, M. A., Said, R. A., Ahmad, M., Asif, M., & Qura-Tul-Ain Khan Taher M. Ghazal, S. A. W. A. K. M. A. K. R. A. S. M. A. M. A. (2021). No Title. 30(2), 539–552. <http://www.techscience.com/iasc/v30n2/44033>
 32. Koroušić Seljak, B., Stibilj, V., Pograjc, L., Mis, N. F., & Benedik, E. (2013). Food composition databases for effective quality nutritional care. *Food Chemistry*, 140(3). <https://doi.org/10.1016/j.foodchem.2013.02.061>
 33. Matloob, F., Ghazal, T. M., Taleb, N. M., Aftab, S., Ahmad, M., Khan, M. A., Abbas, S., & Soomro, T. R. (2021). Software Defect Prediction Using Ensemble Learning: A Systematic Literature Review. *IEEE Access*, 9, 98754–98771.
 34. Mehmood, T., & Al-Gasaymeh, A. (2019). SCHUMPETERIAN ENTREPRENEURSHIP THEORY: EVOLUTION AND RELEVANCE. *Academy of Entrepreneurship Journal*, 25(4).
 35. Naqvi Rababand Soomro, T. R. and A. H. M. and G. T. M. and A. M. T. (2021). The Nexus Between Big Data and Decision-Making: A Study of Big Data Techniques and Technologies. In A. and T. P. J. and B. L. and G. S. and A. A. T. and S. E. and B. D. Hassanien Aboul Ellaand Haqiq (Ed.), *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021)* (pp. 838–853). Springer International Publishing.
 36. Nunes, C. A., Alvarenga, V. O., de Souza Sant'Ana, A., Santos, J. S., & Granato, D. (2015). The use of statistical software in food science and technology: Advantages, limitations and misuses. *Food Research International*, 75, 270–280. <https://doi.org/10.1016/J.FOODRES.2015.06.011>
 37. Rehman, E., Khan, M. A., Soomro, T. R., Taleb, N., Afifi, M. A., & Ghazal, T. M. (2021). Using

- Blockchain to Ensure Trust between Donor Agencies and NGOs in Under-Developed Countries. *Computers*, 10(8). <https://doi.org/10.3390/computers10080098>
38. Sahi, G. K., Sehgal, S., & Sharma, R. (2017). Predicting Customers Recommendation from Co-creation of Value, Customization and Relational Value: [Http://Dx.Doi.Org/10.1177/0256090916686680](http://Dx.Doi.Org/10.1177/0256090916686680), 42(1), 19–35. <https://doi.org/10.1177/0256090916686680>
 39. Vallaster, C., & von Wallpach, S. (2013). An online discursive inquiry into the social dynamics of multi-stakeholder brand meaning co-creation. *Journal of Business Research*, 66(9), 1505–1515. <https://doi.org/10.1016/J.JBUSRES.2012.09.012>
 40. Ghazal, T. M., Noreen, S., Said, R. A., Khan, M. A., Siddiqui, S. Y. et al. (2022). Energy Demand Forecasting Using Fused Machine Learning Approaches. *Intelligent Automation & Soft Computing*, 31(1), 539–553.
 41. Ghazal, T.M. Internet of Things with Artificial Intelligence for Health Care Security. *Arab J Sci Eng* (2021). <https://doi.org/10.1007/s13369-021-06083-8>
 42. Aslam, M. S., Ghazal, T. M., Fatima, A., Said, R. A., Abbas, S. et al. (2021). Energy-Efficiency Model for Residential Buildings Using Supervised Machine Learning Algorithm. *Intelligent Automation & Soft Computing*, 30(3), 881–888.
 43. Ghazal, T. M., Hussain, M. Z., Said, R. A., Nadeem, A., Hasan, M. K. et al. (2021). Performances of K-Means Clustering Algorithm with Different Distance Metrics. *Intelligent Automation & Soft Computing*, 30(2), 735–742.
 44. Khan, Q., Ghazal, T. M., Abbas, S., Khan, W. A., Khan, M. A. et al. (2021). Modeling Habit Patterns Using Conditional Reflexes in Agency. *Intelligent Automation & Soft Computing*, 30(2), 539–552.
 45. Rehman, E., Khan, M. A., Soomro, T. R., Taleb, N., Afifi, M. A., & Ghazal, T. M. (2021). Using Blockchain to Ensure Trust between Donor Agencies and NGOs in Under-Developed Countries. *Computers*, 10(8), 98. doi:10.3390/computers10080098
 46. Ghazal, T.M. Positioning of UAV Base Stations Using 5G and Beyond Networks for IoMT Applications. *Arab J Sci Eng* (2021). <https://doi.org/10.1007/s13369-021-05985-x>
 47. Ghazal, T.M., Said, R.A. & Taleb, N. Internet of vehicles and autonomous systems with AI for medical things. *Soft Comput* (2021). <https://doi.org/10.1007/s00500-021-06035-2>
 48. Matloob, Faseeha & Ghazal, Taher & Taleb, Nasser & Aftab, Shabib & Ahmad, Munir & Khan, Muhammad & Abbas, Sagheer & Soomro, Tariq. (2021). Software Defect Prediction Using Ensemble Learning: A Systematic Literature Review. *IEEE Access*. 9. 98754-98771. 10.1109/ACCESS.2021.3095559.
 49. T. M. Ghazal, M. Anam, M. K. Hasan, M. Hussain, M. S. Farooq et al., "Hep-pred: hepatitis c staging prediction using fine gaussian svm," *Computers, Materials & Continua*, vol. 69, no.1, pp. 191–203, 2021.
 50. Ghazal, T. M., Kalra, D., & Afifi, M. A. (2021). The Impact of Deploying the Internet of Things and How Will It Change Our Lives. *Solid State Technology*, 64(2).
 51. Taher M. Ghazal, Mohammed Kamrul Hasan, Rosilah Hasan, Shayla Islam, Siti Norul Huda Sheikh Abdullah, Mohammed A.M. Afifi, & Deepak Karla. (2020). Security Vulnerabilities, Attacks, Threats and the Proposed Countermeasures for the Internet of Things Applications *Solid State Technology*, 63(1), 1566-1574.