

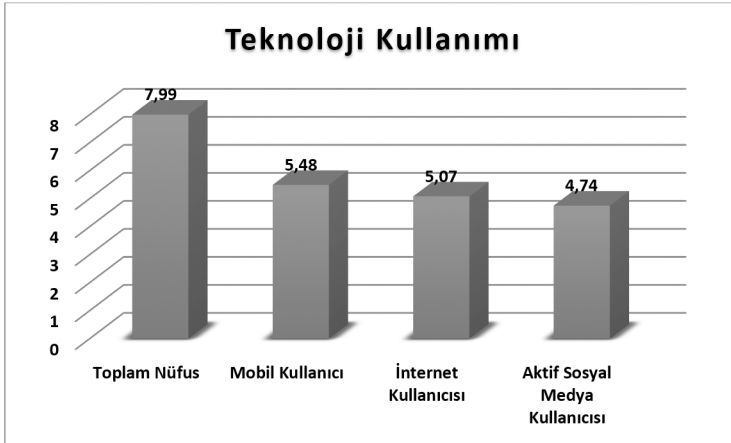
BÖLÜM 6

ÖZNİTELİK SEÇME YÖNTEMLERİ

Ali ŞENOL¹

GİRİŞ

Günümüz teknolojisinin geldiği nokta, mobil teknolojinin ve internet kullanımının yaygınlığı ve sosyal ağların sahip oldukları kullanıcı sayıları göz önünde bulundurulduğunda bilgisayar ortamına aktarılan/aktarılabilecek veri miktarının ne kadar büyük olduğu görülebilmektedir. Datareportal'ın Ekim 2022 verilerine göre 7.99 milyar olan Dünya nüfusunun 5.48 milyar kişisi yani; nüfusun %68.6'sı mobil teknoloji, 5.07 milyarı internet teknolojisi ve 4.74 milyarı aktif sosyal medya kullanıcısıdır. Dünya nüfusu %0.8 artarken, mobil teknoloji kullanıcı sayısı %3.2, internet kullanıcı sayısı %3.5 ve sosyal medya kullanıcı sayısı ise %4.2 artmaktadır. Aynı raporun Kasım 2022 istatistiki verilerine göre, insanlar internet ortamında günlük olarak yaklaşık 6 saat harcamaktadır ve *araştırma yapma* %58.4 ile en başta gelen amaç olurken; 2. sırada *arkadaş çevreleri ile etkileşimde olma* %54.2 ile gelmektedir [1]. Bütün bu istatistiki veriler bilgisayar ortamına aktarılan veri miktarı sayısı ve artış hızı hakkında net bir fikir vermektedir.



Şekil 1. İstatistiklere göre dünya genelinde teknoloji kullanımı

¹ Dr. Öğr. Üyesi Tarsus Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, alisenol@tarsus.edu.tr

SONUÇ VE DEĞERLENDİRMELER

Teknolojinin gelişmesi yaygınlaşması ve dönüşümü neticesinde bilgisayar ortamına aktarılmış olan veri miktarı günden güne artmaktadır. Bununla beraber, insanların bu devasa verilerden değerli bilgiye ulaşma talebi de artmaktadır. Ancak bu kadar büyük veri miktarlarını hızlı bir şekilde işleyebilecek ve yüksek doğrulukta sonuç verebilecek yaklaşımlar mevcut değildir. Bu nedenle öz nitelik seçme yöntemleri yüksek doğruluk elde etmek için yaygın olarak kullanılmaktadır. Öz nitelik seçme yöntemleri ile modellerin hızlı ve yüksek performansta sonuç üretebilmeleri sağlanabilmektedir.

Öz nitelik seçme yöntemlerinin başarılı sonuçlar vermesi bu alana olan talebi de arttırmaktadır. Bu nedenle literatürde çok sayıda yöntem ve yaklaşım önerilmiş olmasına rağmen yeni yaklaşımlar önerilmeye devam etmektedir. Özellikle her veri kümesinin kendine has özellikler barındırması, bu alanda farklı yöntemlerin önerilmesine olan talebi arttırmaktadır. Gelecekte de bu alanda çalışmaların devam edeceği öngörülmektedir.

KAYNAKLAR

- [1] S. Kemp., *Digital 2022: October Global Spyshot Report*. Erişim Tarihi: 11.15.2022, Web adresi: <https://datareportal.com/reports/digital-2022-october-global-statshot>
- [2] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective, *Neurocomputing*, vol. 300, 70-79, 2018.
- [3] F. Korn, B. U. Pagel, and C. Faloutsos, "On the "dimensionality curse" and the "self-similarity blessing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 1, 96-111, 2001.
- [4] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," *Applied Intelligence*, vol. 52, no. 4, 4543-4581, 2022.
- [5] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data Classification: Algorithms and Applications*, 37-64, 2014.
- [6] Kaya M., Bilge H.S. " A hybrid feature selection approach based on statistical and wrapper methods", *24th Signal Processing and Communication Application Conference, SIU 2016 - Proceedings*, , art. no. 7496186 , pp. 2101-2104, 2016.
- [7] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," *in 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1200-1205, 2015.
- [8] N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, "MIFS-ND: A mutual information-based feature selection method," *Expert Systems with Applications*, vol. 41, no. 14, 6371-6385, 2014.
- [9] U. M. Khaire, R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *the Journal of King Saud University - Computer and Information Sciences*, 1060-1073, 2022.
- [10] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data," *in Proceedings, Twentieth International Conference on Machine Learning*, T. Fawcett and N. Mishra, Eds., 856-863, 2003.
- [11] I. H. Witten, E. Frank, and M. A. Hall, *Data mining: practical machine learning tools and techniques, 3rd Edition*. Morgan Kaufmann, Elsevier, I-XXXIII, 2011.

- [12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification, 2nd Edition*. Wiley, I-XX, 2001.
- [13] M. Dash and Y.-S. Ong, "RELIEF-C: Efficient Feature Selection for Clustering over Noisy Data," *23th International Conference on Tools for Artificial Intelligence (ICTAI)*, 2011. Available: <https://doi.org/10.1109/ICTAI.2011.135> <http://doi.ieeecomputersociety.org/10.1109/ICTAI.2011.135>
- [14] M. Robnik-Sikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1-2, 23-69, 2003.
- [15] J. Tang, S. Alelyani, and H. Liu, "Feature Selection for Classification: A Review," in *Data Classification: Algorithms and Applications*, 37-64, 2014.
- [16] D. M. Witten and R. Tibshirani, "A Framework for Feature Selection in Clustering," *Journal of the American Statistical Association*, vol. 105, no. 490, 713-726, 2010.
- [17] A. U. Ahmad and A. Starkey, "Application of feature selection methods for automated clustering analysis: a review on synthetic datasets," *Neural Comput. Appl.*, vol. 29, no. 7, 317-328, 2018.
- [18] M. A. F. A. Fida, T. Ahmad, and M. Ntahobari, "Variance Threshold as Early Screening to Boruta Feature Selection for Intrusion Detection System," in *2021 13th International Conference on Information & Communication Technology and System (ICTS)*, 46-50, 2021.
- [19] A. J. Ferreira and M. A. T. Figueiredo, "Efficient feature selection filters for high-dimensional data," *Pattern Recognition Letters*, vol. 33, no. 13, 1794-1804, 2012.
- [20] J. M. Peña, "Learning Gaussian Graphical Models of Gene Networks with False Discovery Rate Control," in *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, Berlin, Heidelberg, 165-176: Springer Berlin Heidelberg, 2008.
- [21] R. Peck and J. L. Devore, *Statistics: The exploration & analysis of data*. Cengage Learning, 2011.
- [22] G. Forman, "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," *Journal of Machine Learning Research*, vol. 3, 1289-1305, 2003.
- [23] S. Bahassine, A. Madani, M. Al-Sarem, and M. Kissi, "Feature selection using an improved Chi-square for Arabic text classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 2, 225-231, 2020.
- [24] M. Dash, K. Choi, P. Scheuermann, and L. Huan, "Feature selection for clustering - a filter solution," in *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, 115-122, 2002.
- [25] X. Deng, Y. Li, J. Weng, and J. Zhang, "Feature selection for text classification: A review," *Multi-media Tools and Applications*, vol. 78, no. 3, pp. 3797-3816, 2019.
- [26] A. Wang, H. Liu, and G. Chen, "Chaotic Harmony Search based Multi-objective Feature Selection for Classification of Gene Expression Profiles," in *2021 IEEE 9th International Conference on Bioinformatics and Computational Biology (ICBCB)*, 107-112, 2021.
- [27] F. Özyurt, "Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures," *The Journal of Supercomputing*, vol. 76, no. 11, 8413-8431, 2020.
- [28] E. S. M. El-Kenawy, A. Ibrahim, S. Mirjalili, M. M. Eid, and S. E. Hussein, "Novel Feature Selection and Voting Classifier Algorithms for COVID-19 Classification in CT Images," *IEEE Access*, vol. 8, 179317-179335, 2020.
- [29] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artificial Intelligence Review*, vol. 53, no. 2, 907-948, 2020.
- [30] D. Han and J. Kim, "Unified Simultaneous Clustering and Feature Selection for Unlabeled and Labeled Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 12, 6083-6098, 2018.
- [31] M. Z. Asghar, A. Khan, F. Khan, and F. M. Kundi, "RIFT: A Rule Induction Framework for Twitter Sentiment Analysis," *Arabian Journal for Science and Engineering*, vol. 43, no. 2, 857-877, 2018.
- [32] R. Jensen, C. Cornelis, and Q. Shen, "Hybrid fuzzy-rough rule induction and feature selection," in *2009 IEEE International Conference on Fuzzy Systems*, 1151-1156, 2009.

- [33] A. S. N. Huda and S. Taib, "Suitable features selection for monitoring thermal condition of electrical equipment using infrared thermography," *Infrared Physics & Technology*, vol. 61, 184-191, 2013.
- [34] X. Jin, E. W. M. Ma, L. L. Cheng, and M. Pecht, "Health Monitoring of Cooling Fans Based on Mahalanobis Distance With mRMR Feature Selection," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, 2222-2229, 2012.
- [35] V. Nasir and J. Cool, "Intelligent wood machining monitoring using vibration signals combined with self-organizing maps for automatic feature selection," *The International Journal of Advanced Manufacturing Technology*, vol. 108, no. 5, 1811-1825, 2020.
- [36] UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set. Erişim Tarihi: 1 Aralık 2022, Web adresi: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>