TAMAMEN RASTGELE EKSİK (MCAR) VERİLERDE VERİ MADENCİLİĞİ VE SINIFLANDIRMA ALGORİTMALARI

Editör
Dr. Ersin YILMAZ

Yazarlar:
Dr. Saygın DİLER
Doç. Dr. Yıldırım DEMİR

ISBN
978-625-375-173-9

Kitap Adı
Tamamen Rastgele Eksik (MCAR) Verilerde Veri Madenciliği ve Sınıflandırma Algoritmaları

Editör
Dr. Ersin YILMAZ
ORCID iD: 0000-0002-9871-4700

Yazarlar
Dr. Saygın DİLER
ORCID iD: 0000-0002-9056-412X
Doç. Dr. Yıldırım DEMİR
ORCID iD: 0000-0002-6350-8122

Yayın Koordinatörü
Yasin DİLMEN

Sayfa ve Kapak Tasarımı
Akademisyen Dizgi Ünitesi

Yayıncı Sertifika No
47518

Baskı ve Cilt
Vadi Matbaacılık

Bisac Code
TEC071000

DOI
10.37609/akya.3382

Kütüphane Kimlik Kartı
Diler, Saygın. Demir, Yıldırım.
Tamamen Rastgele Eksik (MCAR) Verilerde Veri Madenciliği ve Sınıflandırma Algoritmaları / Saygın Diler, Yıldırım Demir ; ed. Ersin Yılmaz.
Ankara : Akademisyen Yayınevi Kitabevi, 2024.
118 s. : şekil, çizelge. ; 135x210 mm.
Kaynakça var.
ISBN 9786253751739

GENEL DAĞITIM

Akademisyen Kitabevi A.Ş.

Halk Sokak 5 / A
Yenişehir / Ankara
Tel: 0312 431 16 33
siparis@akademisyen.com

www.akademisyen.com

ÖNSÖZ

Bu kitap, veri madenciliği alanında sınıflandırma algoritmalarının, veri setinde karşılaşılan eksikliklerin veri kalitesini nasıl etkilediğini ve bu eksikliklerin sınıflandırma performansına olan yansımalarını ele almaktadır. Aşağıdaki amaçlar doğrultusunda kitapta yapılan çalışmalar, veri madenciliği araştırmalarında eksik veri sorunu ile daha etkili başa çıkmanın yollarını sunmayı hedeflemektedir. Kitap, Dr. Saygın DİLER tarafından yazılan "*Veri Kalitesinin Bozulduğu Durumlarda Veri Madenciliği Sınıflandırma Algoritmalarının Performanslarının Karşılaştırılması*" başlıklı doktora tezine dayanmaktadır. Kitabın temel amacı, eksik veri yapılarının sınıflandırma algoritmaları üzerindeki etkilerini incelemek, bu algoritmaların performansını karşılaştırmak ve araştırmacılara eksik veri ile başa çıkma konusunda rehberlik etmektir. Bu doğrultuda kitap şu ana başlıklara odaklanmaktadır:

Eksik Verilerin Sınıflandırma Algoritmalarına Etkisini İncelemek: Eksik veriler, veri madenciliği süreçlerinde önemli bir sorun teşkil etmektedir. Bu kitap, tamamen rastgele eksik (MCAR) veri yapısında, veri setlerindeki eksikliklerin sınıflandırma performansını nasıl etkilediğini analiz etmektedir.

Farklı Sınıflandırma Algoritmalarının Performansını Karşılaştırmak: Naive Bayes, Lojistik Regresyon, K-En Yakın

Komşu (K-NN), Destek Vektör Makineleri (SVM) ve Aşırı Gradyan Arttırma (XGBoost) gibi popüler sınıflandırma algoritmaları incelenmiş ve eksik verilerle çalışırken gösterdikleri performans detaylı bir şekilde karşılaştırılmıştır.

Eksik Veri Oranlarının Etkisini Değerlendirmek: Eksik veri oranlarının (%5, %15 ve %30) sınıflandırma performansı üzerindeki etkileri değerlendirilerek, artan eksiklik oranlarının doğruluk, duyarlılık ve seçicilik gibi metrikler üzerindeki etkisi incelenmiştir.

Gerçek ve Simülasyon Verileri ile Uygulamalar Yapmak: Hem gerçek veri setleri hem de simülasyon yoluyla oluşturulmuş veri setleri üzerinde algoritmaların performansları karşılaştırılmıştır. Bu uygulamalar, farklı örneklem yapılarında algoritmaların dayanıklılığını ortaya koymuştur.

Veri madenciliği ve makine öğrenmesi, günümüz bilim ve teknolojisinin önemli alanları arasında yer almakta olup, bu alanlarda başarılı uygulamalar için veri kalitesinin yüksek olması büyük önem taşımaktadır. Eksik veriler, veri madenciliği algoritmalarının performansını olumsuz etkileyen temel sorunlardan biridir. Bu kitap, eksik veri problemiyle ilgili farkındalığı artırmayı ve araştırmacıların algoritma seçiminde daha bilinçli kararlar alabilmelerine yardımcı olmayı amaçlamaktadır. Eksik veri ile çalışan veri bilimciler ve araştırmacılar için faydalı bir kaynak olmayı hedefleyen bu çalışma, sınıflandırma algoritmalarının eksik veri karşısındaki performansını değerlendirmek üzere hazırlanmıştır.

İçindekiler

KAYNAKLAR

Abar, H. (2020). XGBoost ve MARS yöntemleriyle altın fiyatlarının kestirimi. EKEV Akademi Dergisi, 83(0), 427-446.

Abbas, A. K., Al-haideri, N. A., Bashikh, A. A. (2019). Implementing artificial neural networks and support vector machines to predict lost circulation. Egyptian Journal of Petroleum, 28(4), 339–347.

Abe, S. (2005). Support vector machines for pattern classification. Springer-Verlag: Berlin, Germany.

Acuña, E., Rodriguez, C. (2004). The treatment of missing values and its effect on classifier accuracy. In D. Banks, F. R. McMorris, P. Arabie, W. Gaul (Eds.), Classification, clustering, and data mining applications (pp. 639-647). Springer-Verlag: Berlin, Germany.

Addin, O., Sapuan, S. M., Mahdi, E., Othman, M. (2007). A Naive-Bayes classifier for damage detection in engineering materials. Materials and Design, 28, 2379–2386.

Adje, E. A., Houndji, V. R., Dossou, M. (2022). Features analysis of internet traffic classification using interpretable machine learning models. IAES International Journal of Artificial Intelligence (IJ-AI), 11(3), 1175–1183.

Akbaş, U., Koğar, H. (2020). Nicel araştırmalarda kayıp veriler ve uç değerler. Pegem Akademi: Ankara, Türkiye.

Akinnuwesi, B. A., Macaulay, B. O., Aribisala, B. S. (2020). Breast cancer risk assessment and early diagnosis using Principal Component Analysis and support vector machine techniques. Informatics in Medicine Unlocked, 21, 1-13.

Akpınar, H. (2014). Data : veri madenciliği veri analizi. Papatya Bilim Yayınevi: İstanbul, Türkiye.

Akşehirli, Ö. Y., Ankaralı, H., Aydın, D., Saraçlı, Ö. (2013). Tıbbi tahminde alternatif bir yaklaşım: destek vektör makineleri. Türkiye Klinikleri Journal of Biostatistics, 5(1), 19–28.

Akyel, N., Seçkin, K. (2012). K-en yakın komşuluk algoritmasının hile dene-

timinde kullanımı. Muhasebe ve Vergi Uygulamarı Dergisi, 5(1), 21–40.
Albay, A. G., Doğan, Y. (2020). Veri madenciliği yaklaşımlarını kullanan yeni bir tarım takip sistemi. Avrupa Bilim ve Teknoloji Dergisi, Özel Sayı, 313–322.
Alcázar, Á., Jurado, J. M., Palacios-Morillo, A., de Pablos, F., Martín, M. J. (2012). Recognition of the geographical origin of beer based on support vector machines applied to chemical descriptors. Food Control, 23(1), 258–262.
Alkan, Ö., Demir, A. (2019). Tütün kullanımını bırakma başarısını etkileyen faktörlerin lojistik regresyon ile analizi. İktisadi ve İdari Bilimler Dergisi, 33(4), 1227–1244.
Alkhatib, K., Najadat, H., Hmeidi, I., Shatnawi, M. K. A. (2013). Stock price prediction using k-nearest neighbor algorithm. International Journal of Business, Humanities and Technology, 3(3), 32–44.
Alp, E. A. (2019). Lojistik regresyon analizi. In S. Alp, E. Öz (Eds.), Makine öğreniminde sınıflandırma yöntemleri ve R uygulamaları. Nobel Akademik Yayıncılık: Ankara, Türkiye.
Alpar, R. (2013). Çok değişkenli istatistiksel yöntemler. Detay Yayıncılık: Ankara, Türkiye.
Altunkaynak, A., Başakın, E. E., Kartal, E. (2020). Dalgacik k-en yakin komşuluk yöntemi ile hava kirliliğinin tahmini. Uludağ Üniversitesi Mühendislik Fakültesi Dergisi, 25(3), 1547–1556.
Anderson, J. A. (1979). Multivariate logistic compound. Biometrika, 66(1), 17–26.
Anderson, J. A. (1983). Robust inference using logistic models. Bulletin of International Statistical Institute, 48(0), 35–53.
Arumugam, P., Jose, P. (2018). Efficient decision tree based data selection and support vector machine classification. Materials Today: Proceedings, 5(1), 1679–1685.
Asraf, H. M., Nooritawati, M. T., Rizam, M. S. B. S. (2012). A comparative study in kernel-based Support Vector Machine of oil palm leaves nutrient disease. Procedia Engineering, 41(Iris), 1353-1359.
Asselman, A., Khaldi, M., Aammou, S. (2021). Enhancing the prediction of student performance based on the machine learning xgboost algorithm. Interactive Learning Environments, 0(0), 1–20.
Atasoy, N. A., Tabak, D. (2018). Destek vektör makineleri kullanarak yüz tanıma uygulaması geliştirilmesi. Engineering Sciences, 13(2), 119–127.
Avuçlu, E., Elen, A. (2019). Classification of cardiotocography records with Naïve Bayes. International Scientific and Vocational Studies Journal, 3(2), 105–110.
Ayan, B., Kuyumcu, B., Ceylan, B. (2019). Twitter üzerindeki islamofobik twitlerin duygu analizi ile tespiti. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 7(2), 495–502.
Ayhan, S., Erdoğmuş, Ş. (2014). Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi, 9(1), 175–201.

Azizoğlu, F., Ünsal, E. (2022). Missing IoT data prediction with machine learning techniques. El-Cezeri Fen ve Mühendislik Dergisi, 9(4), 1388–1397.

Baitharu, T. R., Pani, S. K. (2013). Effect of missing values on data classification. Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS), 4(2), 311–316.

Balaban, M. E., Kartal, E. (2015). Veri madenciliği ve makine öğrenmesi temel algoritmaları ve R dili ile uygulamaları. Çağlayan Kitapevi: Ankara, Türkiye.

Baoli, L., Shiwen, Y., Qin, L. (2003). An improved k-nearest neighbor algorithm. 20th International Conference on Computer Processing of Oriental Languages. Shenyang, China.

Batista, G. E. A. P. A., Monard, M. C. (2002). A study of k-nearest neighbour as an imputation method. In Abraham, A., Solar, J.R., Köppen, M. (Ed.), Frontiers in artificial intelligence and applications (Vol. 87, pp. 251–260). IOS Press: Santiago, Chile.

Batista, G., Silva, D. F. (2009). How k-nearest neighbor parameters affect its performance. Simposio Argentino deInteligencia Artificial, ASAI, Rosario, Arjantin.

Battineni, G., Chintalapudi, N., Amenta, F. (2019). Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). Informatics in Medicine Unlocked, 16, 1-8.

Bedell, Z. (2018). Support vector machines explained. Erişim tarihi: 4 Nisan 2022. https://medium.com/@zachary.bedell/support-vector-machines-explained-73f4ec363f13

Berkson, J. (1944). Application of the logistic function to bio-assay. Journal of the American Statistical Association, 39(227), 357–365.

Bhatia, N., Vandana (2010). Survey of nearest neighbor techniques. International Journal of Computer Science and Information Security, 8(2), 302–305.

Bhavsar, Y. B., Waghmare, K. C. (2013). Intrusion detection system using data mining technique: support vector machine. International Journal of Emerging Technology and Advanced Engineering, 3(3), 581–586.

Bircan, H. (2004). Lojistik regresyon analizi: tıp verileri üzerine bir uygulama. Kocaeli Üniversitesi Sosyal Bilimler Dergisi, 2, 185–208.

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Blomberg, L. C., Ruiz, D. D. A., (2013). Evaluating the influence of missing data on classification algorithms in data mining applications. Simpósio Brasileiro De Sistemas De Informação, SBSI, João Pessoa,Brezilya.

Boser, B. E., Guyon, I. M., Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT, Pittsburgh, USA.

Bramer, M. (2007). Principles of data mining. Springer-Verlag: Berlin, Germany.

Bridge, D. (2013). Classification: k nearest neighbours. Erişim tarihi: 6 Mart 2021. http://www.cs.ucc.ie/~dgb/courses/tai/notes/handout4. pdf

Brown, M. L., Kros, J. F. (2003). Data mining and the impact of missing data. Industrial Management & Data Systems, 103(8), 611–621.

Buciu, I., Kotropoulos, C., Pitas, I. (2006). Demonstrating the stability of support vector machines for classification. Signal Processing, 86(9), 2364–2380.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognation. Data Mining and Knowledge Discovery, 2(2), 121–167.

Buuren, S. V. (2012). Flexible imputation of missing data (1st Ed.). Chapman and Hall/CRC: London, UK.

Campbell, C., Yiming, Y. (2011). Learning with support vector machines (synthesis lectures on artificial intelligence and machine learning). Springer: Switzerland.

Cardona, T. A., Cudney, E. A. (2019). Predicting student retention using support vector machines. Procedia Manufacturing, 39(2019), 1827–1833.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189–215.

Chandrasekar, P., Qian, K. (2016). The ımpact of data preprocessing on the performance of naïve bayes classifier. 2016 IEEE 40th Annual Computer Software and Applications Conference, COMPSAC, Atlanta, USA.

Chen, H., Zhang, J., Xu, Y., Chen, B., Zhang, K. (2012). Performance comparison of artificial neural network and logistic regression model for differentiating lung nodules on CT scans. Expert Systems with Applications, 39(13), 11503–11509.

Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA.

Chen, H., Wang, N., Du, X., Mei, K., Zhou, Y., Cai, G. (2023). Classification prediction of breast cancer based on machine learning. Computational Intelligence and Neuroscience, 2023(0), 1–9.

Cherif, I. L., Kortebi, A. (2019). On using eXtreme Gradient Boosting (XGBoost) machine learning algorithm for home network traffic classification. 2019 Wireless Days, WD, Manchester, UK.

Cortes, C., Vapnik, V. N. (1995). Support vector networks. Machine Learning, 20, 273–297.

Cover, T. M., Hart, P. E. (1967). Nearest neighbor pattern classification. IEEE Trans Inf Theory, 13(1), 21–27.

Cox, D. R. (1969). Analysis of binary data. Champan and Hall: London, UK.

Cristianini, N., Taylor, J. S. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press: Cambridge, UK.

Cunningham, P., Delany, S. J. (2007). K-neighbor classifiers. Mult Classif Syst, 34(8), 1–17.

Çırak, G., Çokluk, Ö. (2013). Yükseköğretimde öğrenci başarılarının sınıflandırılmasında yapay sinir ağları ve lojistik regresyon yöntemlerinin kullanılması. Mediterranean Journal of Humanities, 3(2), 71–79.

Çiftçi, C., Çağlar, A. (2014). Ailelerin sosyo-ekonomik özelliklerinin öğrenci başarısı üzerindeki etkisi: fakirlik kader midir? International Journal of Human Sciences./ Uluslararası İnsan Bilimleri Dergisi, 11(2), 155–175.

Çinicioğlu, E. N., Atalay, M., Yorulmaz, H. (2016). Trafik kazaları analizi için bayes ağları modeli bayesian network model for analysis of traffic accidents. Bilişim Teknolojileri *Dergisi,* 6(2), 41–52.

Dai, X., Wang, N., Wang, W. (2019). Application of machine learning in BGP anomaly detection. Journal of Physics: Conference Series, 1176(3), 1-12.

Davidson, I., Tayi, G. (2009). Data preparation using data quality matrices for classification mining. European Journal of Operational Research, 197(2), 764-772.

Demir, Y., Keskin, S. (2021). Examination of OECD countries for the presence of livestock by non-metric multidimensional scaling. Livestock Studies, 61(2), 46-54.

Desdhanty, V. S., Rustam, Z. (2021). Liver cancer classification using random forest and extreme gradient boosting (XGBoost) with genetic algorithm as feature selection. 2021 International Conference on Decision Aid Sciences and Application, DASA, Sakhir, Bahreyn.

Diler, S., Demir, Y. (2023). Sağdan sansürlü veriler için veri madenciliği algoritmaları performanslarının karşılaştırılması. İstatistik Araştırma Dergisi, 13(1), 34-47.

Diler, S., Demir, Y. (2024). Çoklu Doğrusal Bağlantı Olması Durumunda Veri Madenciliği Algoritmaları Performanslarının Karşılaştırılması. Nicel Bilimler Dergisi, 6(1), 40-67. https://doi.org/10.51541/nicel.1371834.

Dilki, G. (2020). Makine öğrenmesi algoritmalarının sınıflama problemleri üzerinden karşılaştırılması: satış tahmini. Press Academia Procedia, 12(1), 82–83.

Dilki, G., Başar, Ö. D. (2020). İşletmelerin iflas tahmininde k – en yakın komşu algoritması üzerinden uzaklık ölçütlerinin karşılaştırılması. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 19(38), 224–233.

Doger, Ş., Kurgun, A. (2021). Şarap üretiminde veri kalitesine ilişkin eksik veri sorunlarının derin öğrenme ile çözülmesi: üretici çekişmeci ağlarla bir uygulama. Uluslararası Güncel Turizm Araştırmaları Dergisi, 5(1), 99–111.

Donders, A. R. T., Heijden, G. J. M. G., Stijnen, T., Moons, K. G. M. (2006). Review: a gentle introduction to imputation of missing values. Journal of Clinical Epidemiology, 59(10), 1087–1091.

Dong, J. X., Krzyzak, A., Suen, C. Y. (2005). An improved handwritten Chinese character recognition system using support vector machine. Pattern Recognition Letters, 26(12), 1849–1856.

Dong, J. X., Krzyzak, A., Suen, C. Y. (2005). Fast SVM training algorithm with decomposition on very large data sets. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(4), 603–618.

Dong, Z., Li, X., Luan, F., Ding, J., Zhang, D. (2023). Point and interval prediction of the effective length of hot-rolled plates based on IBES-XGBoost. Measurement, 214(0), 112857.

Dragomir, E. G. (2010). Air quality ındex prediction using k-nearest neighbor technique. Bulletin Of Pg University Of Ploiesti Series Mathematics Informatics Physics, 62(1), 103–108.

Dumais, S., Platt, J., Heckerman, D., Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. CIKM '98: Proceedings of the Seventh İnternational Conference On Information And Knowledge Management, Bethesda, USA.

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., Tabona, O. (2021). A survey on missing data in machine learning. Journal of Big Data, 8(140), 1–37.

Enders, C. K. (2010). Applied missing data analysis. Guilford Press: New York, USA.

Ercan, U., Irmak, S. (2018). Lojistik regresyon analizi kullanılarak kanatlı hayvan eti tüketimini etkileyen faktörlerin belirlenmesi. Aksaray Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi, 10(2), 9–18.

Ercan, U. (2022). Havayolu taşımacılığında müşteri memnuniyetinin topluluk öğrenmesi yöntemleri ile belirlenmesi. Alanya Akademik Bakış, 6(3), 2763–2774.

Erpolat, S., Öz, E. (2010). Kanser verilerinin sınıflandırılmasında yapay sinir ağları ile destek vektör makinelerinin karşılaştırılması. İstanbul Aydın Üniversitesi Dergisi, 2(5), 71-83.

Fix, E., Hodges, J. L. (1951). Discriminatory analysis. nonparametric discrimination; consistency properties. Report Number 4, Project Number 21-49-004, USAF of Aviation Medicine, Randolph Field, Texas.

Fletcher, T. (2008). Support vector machines explained. Erişim tarihi: 20 Mart 2021. https://static1.squarespace.com/static/58851af9eb-bd1a30e98f-b283/t/58902fbae4fcb5398aeb7505/1485844411772/SVM+Explained.pdf

Frank, E., Hall, B., Pfahringer, B. (2003). Locally weighted naive bayes. Proceedings of Conference on Uncertainty in Artificial Intelligence, Akapulko, Meksika.

Friedjungová, M., Jiřina, M., Vašata, D. (2019). Missing features reconstruction and its impact on classification accuracy. International Conference on Computational Science, ICCS, Faro, Portekiz.

Friedman, N., Geiger, D., Goldszmidt, M. (1997). Bayesian network classifiers. Machine Learning, 29(2), 131–163.

Gamgam, H., Altunkaynak, B. (2017). SPSS uygulamalı regresyon analizi (2. Basım). Seçkin Kitapevi: Ankara, Türkiye.

García-Laencina, P. J., Sancho-Gómez, J.-L., Figueiras-Vidal, A. R. (2010). Pattern classification with missing data: a review. Neural Computing and Applications, 19(2), 263–282.

Gharamaleki, P. S., Seyedarabi, H., Branch, A. (2015). Face recognition using Eigen faces, PCA and support vector machines. European Journal of Applied Engineering and Scientific Research, 4(3), 24–30.

Ghazvini, K., Yousefi, M., Firoozeh, F., Mansouri, S. (2019). Predictors of tuberculosis using a logistic regression model. Reviews in Clinical Medicine, 6(3), 108-112.

Girginer, N., Cankuş, B. (2008). Tramvay yolcu memnuniyetinin lojistik regresyon analiziyle ölçülmesi: estram örneği. Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 15(1), 181–193.

Govindarajan, M., Chandrasekaran, R. (2010). Evaluation of k-nearest neighbor classifier performance for direct marketing. Expert Systems With Applications, 37(1), 253-258.

Göker, H., Tekedere, H. (2019). Çocukluk çağı dikkat eksikliği ve hiperaktivite bozukluğunun öngörülmesine yönelik dinamik uzman sistem tasarımı. Bilişim Teknolojileri Dergisi, 12(1), 33–41.

Granik, M., Mesyura, V. (2017). Fake news detection using naive bayes classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering, UKRCON, Kiev, Ukrayna.

Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K. (2003). KNN model-based approach in classification. In R. Meersman, Z. Tari, D. C. Schmidt (Eds.), On the move to meaningful internet systems 2003: CoopIS, DOA, and ODBASE (pp. 986–996). Springer-Verlag: Berlin, Germany.

Gül, E., Kalyoncu, M. (2020). Ağır vasıta hava kompresörü piston segmanı aşınması durumlarında k-en yakın komşu algoritmasının sınıflandırma performansının incelenmesi. Avrupa Bilim ve Teknoloji Dergisi, Özel Sayı, 78–90.

Gülcü, Ş. (2019). Bilgisayar ağ güvenliğinde naive bayes algoritmasının kullanımı. Kilis 7 Aralık Üniversitesi Fen ve Mühendislik Dergisi, 3(1), 1-15.

Güner, N., Çomak, E. (2011). Mühendislik öğrencilerinin matematik i derslerindeki başarısının destek vektör makineleri kullanılarak tahmin edilmesi. Pamukkale University Journal of Engineering Sciences, 17(2), 87–96.

Güner, Z. B. (2014). Veri madenciliğinde cart ve lojistik regresyon analizinin yeri: ilaç provizyon sistemi verileri üzerinde örnek bir uygulama. Sosyal Güvenlik Uzmanları Derneği Sosyal Güvence Dergisi, 6, 53–99.

Haltaş, A., Alkan, A. (2009). İmmunohistokimyasal boyalar ile tiroid tümörü teşhisinde naive bayes algoritması kullanılması. XVI. Akademik Bilişim Konferansı, Mersin, Türkiye.

Han, J., Kamber, M., Pei, J. (2012). Data mining concepts and techniques (Third Edition). Morgan Kaufman Publishers: Massachusetts, USA.

Harrington, P. (2012). Machine learning in action. Manning Publications: New York, USA.

Hosmer, D. W., Lemeshov, S., Sturdivant, R. X. (2013). Applied logistic regression (Third Edition). John Wiley & Sons, Inc: New Jersey, USA.

Hsu, C., Lin, C. (2002). A comparison of model selection methods for multi-class support vector machines. IEEE Transactions on Neural Networks, 13(2), 415–425.

Hu, L.-Y., Huang, M.-W., Ke, S.-W., Tsai, C.-F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus, 5(1304), 1–9.

Huang, J., Shao, X., Wechsler, H. (2002). Face pose discrimination using support vector machines (SVM). Fourteenth International Conference on Pattern Recognition, Brisbane, Avusturalya.

İlarslan, K. (2016). K-en yakın komsu (knn) algoritması ile hisse senedi fiyatlarının tahmin edilmesi: Bist'den örnek bir uygulama. The Journal of Academic Social Sciences, 4(30), 375–392.

İşleyen, Ş., İnan, S. (2019). Eksik veri tiplerinde (MAR, MCAR, MNAR) tamamlama algoritmalarının parametre tahmin analizi. Akademisyen Kitapevi: Ankara, Türkiye.

İyit, N., Genç, A. (2005). Lojistik regresyon analizi yardımıyla denekte menopoz evresine geçişe ilişkin bir sınıflandırma modelinin elde edilmesi. Selçuk Üniversitesi Fen Fakültesi Fen Dergisi, 1(25), 19–28.

Jebril, N. A., Al-Zoubi, H. R., Abu Al-Haija, Q. (2018). Recognition of handwritten arabic characters using histograms of oriented gradient (HOG). Pattern Recognition and Image Analysis, 28(2), 321–345.

Josephus, B. O., Nawir, A. H., Wijaya, E., Moniaga, J. V., Ohyver, M. (2021). Predict mortality in patients infected with covid-19 virus based on observed characteristics of the patient using logistic regression. Procedia Computer Science, 179(0), 871–877.

Karatay, S., Algahani, M. (2021). 1999 Marmara depremi ve güneş tutulmasının naive bayes sınıflayıcısı ile istatistiksel analizi. European Journal of Science and Technology, 23, 643–648.

Kartal, E., Balaban, M. E. (2019). Destek vektör makineleri: teori ve R dili ile bir uygulama. In M. E. Balaban, E. Kartal (Eds.), Veri madenciliği ve makine öğrenmesi temel kavramlar, algoritmalar, uygulmalar (pp. 207-241). Çağlayan Kitapevi: İstanbul, Türkiye.

Kartal, C. (2020). Destek vektör makineleri ile borsa endekslerinin tahmini. İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 9(2), 1394–1418.

Kaya, D., Türk, M., Kaya, T. (2018). Examining the effect of dimension reduction on eeg signals by k-nearest neighbors algorithm. El-Cezeri Fen ve Mühendislik Dergisi, 5(2), 591–595.

Kaya, D. (2019). Alt uzay knn eritmato-skuamöz hastalık türlerinin sınıflandırılması. Fırat Üniversistesi Mühendislik Bilim Dergisi, 31(2), 583–587.

Kemalbay, G., Alkış, B. N. (2020). Borsa endeks hareket yönünün çoklu lojistik regresyon ve k-en yakın komşu algoritması ile tahmini. Pamukkale Üniversitesi Mehindislik Bilimleri Dergisi, 27(4), 556-569.

Khan, M., Ding, Q., Perrizo, W. (2002). k-nearest neighbor classification on spatial data streams using p-trees. In M.-S. Chen, P. S. Yu, B. Liu (Eds.), Advances in knowledge discovery and data mining (pp. 517–528). Springer Berlin Heidelberg.

Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., Özçift, A. (2016). KNN algoritması ve R dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. Marmara Fen Bilimleri Dergisi, 28(3), 89–94.

Kleinbaum, D. G., Klein, M., Pryor, E. R. (2002). Logistic regression a self learning text (2nd Edition). Springer-Verlag: Berlin, Germany.

Kökver, Y., Barışçı, N., Çiftçi, A., Emekçi, Y. (2014). Hipertansiyone etki eden faktörlerin veri madenciliği yöntemleri ile belirlenmesi. Engineering Sciences, 9(2), 15–25.

Kumar, R., Geetha, S. (2020). Malware classification using XGboost-Gradient boosted decision tree. Advances in Science, Technology and Engineering Systems Journal, 5(5), 536–549.

Kutlugün, M. A., Çakır, M. Y., Kiani, F. (2017). Yapay sinir ağları ve k-en yakın komşu algoritmalarının birlikte çalışma tekniği (ensemble) ile metin türü tanıma. 22. Türkiye'de İnternet Konferansı, İstanbul, Türkiye.

Lewis, N. D. (2017). Machine learning made easy with R: An intuitive step by step blueprint for beginners. CreateSpace Independent Publishing Platform: Carolina, USA.

Li, X., Cervantes, J., Yu, W. (2010). A novel SVM classification method for large data sets. 2010 IEEE International Conference on Granular Computing, GrC 2010, California, USA.

Lin, W. C., Tsai, C.-F. (2020). Missing value imputation: a review and analysis of the literature (2006–2017). Artificial Intelligence Review, 53(2), 1487–1509.

Little, R. J. A., Rubin, D. B. (1987). Statistical analysis with missing data. Wiley: New York, USA.

McCallum, A., Nigam, K. (1998). A comparison of event models for naive bayes text classification. learning for text categorization. 1998 Association for the Advancement of Artificial IntelligenceWorkshop, AAAI, Madison, USA.

McNamara, J. M., Green, R. F., Olsson, O. (2006). Bayes' Theorem and ıts applications in animal behaviour. Oikos, 112(2), 243–251.

Metin, S. (2021). OECD endüstriyel üretim verilerinde bulunan kayıp verilerin knn yöntemi ile tahmini. Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi, 9(4), 955–967.

Metlek, S., Kayaalp, K. (2020). Derin öğrenme ve destek vektör makineleri ile görüntüden cinsiyet tahmini. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 8, 2208–2228.

Mucherino, A., Papajorgji, P. J., Paradalos, P. M. (2009). Data mining in agriculture. Springer: Dordrecht, Hollanda.

Mukherjee, S., Sharma, N. (2012). Intrusion detection using Naive Bayes classifier with feature reduction. Procedia Technology, 4, 119–128.

Mulla, G. A. A., Demir, Y., Hassan, M. (2021). Combination of PCA with SMOTE oversampling for classification of high-dimensional ımbalanced data. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 10(3), 858–869.

Muralidharana, V., Sugumaranc, V. (2012). A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. Applied Soft Computing, 12, 2023–2029.

Naik, V. A., Desai, A. A. (2017). Online handwritten Gujarati character recognition using SVM, MLP, and KNN. 2017 8th International Conference on Computing, Communication and Networking Technologies, ICCCNT, Delhi, Hindistan.

Nemade, V., Fegade, V. (2023). Machine learning techniques for breast cancer prediction. Procedia Computer Science, 218(0), 1314–1320.

Noaman, H. M., Elmougy, S., Ghoneim, A., Hamza, T. (2010). Naive Bayes classifier based arabic document categorization. 2010 The 7th International Conference on Informatics and Systems, INFOS, Kahire, Mısır.

Novakovic, J. (2010). The impact of feature selection on the accuracy of Naive Bayes classifier. 18th Telecommunication Forum, TELFOR, Belgrad, Sırbistan.

Onan, A. (2017). Türkçe twitter mesajlarında gizli dirichlet tahsisine dayalı duygu analizi. 19. Akademik Bilişim Konferansı, Aksaray, Türkiye.

Ou, T., Liu, J., Liu, F., Chen, W., Qin, J. (2023). Coupling of XGBoost ensemble methods and discrete element modelling in predicting autogenous grinding mill throughput. Powder Technology, 422(0), 1-12.

Öz, E. (2019). Destek vektör makineleri. In S. Alp, E. Öz (Ed.), Makine öğreniminde sınıflandırma yöntemleri ve R uygulamaları (pp. 167-189). Nobel Akademik Yayıncılık: Ankara, Türkiye.

Özdamar, K. (2018). Paket programlar ile istatistiksel veri analizi-2. Nisan Kitapevi: Eskişehir, Türkiye.

Özdamar, K. (2019). Paket programları ile istatistiksel veri analizi-1 (11. Baskı). Nisan Kitapevi: Eskişehir, Türkiye.

Özdemir, A. K., Tolun, S., Demirci, E. (2011). Endeks getirisi yönünün ikili sınıflandırma yöntemiyle tahmin edilmesi: İMKB 100 endeksi örneği. Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 4(2), 45–59.

Özkan, Y. (2008). Veri Madenciliği Yöntemleri. Papatya Yayınevi: İstanbul, Türkiye.

Pande, S., Khamparia, A., Gupta, D. (2023). Feature selection and comparison of classification algorithms for wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing, 14(3), 1977–1989.

Phillips, P. J. (1999). Support vector machines applied to face recognition. Advances in Neural Information Processing Systems, NIPS, Massachusetts, USA.

Raheja, J. L., Mishra, A., Chaudhary, A. (2016). Indian sign language recognition using SVM. Pattern Recognition and Image Analysis, 26(2), 434–441.

Rençber, Ö. F. (2018). Sınıflandırma problemlerinde çoklu lojistik regresyon, yapay sinir ağı ve ANFIS yöntemlerin karşılaştırılması: insani gelişmişlik endeksi üzerine uygulama. Gazi Kitapevi: Ankara, Türkiye.

Roiger, R., Geatz, M. (2003). Data mining; a tutorial based primer. Addison Wesley: Boston, USA.

Sadri, J., Suen, C. Y., Bui, T. D. (2003). Application of support vector machines for recognition of handwritten arabic/persian digits. Second Iranian Conference on Machine Vision and Image Processing & Applications, MVIP, Tahran, İran.

Salmi, N., Rustam, Z. (2019). Naïve Bayes classifier models for predicting the colon cancer. 9th Annual Basic Science International Conference 2019, BaSIC, Malang, Endonezya.

Selimoğlu, M., Yılmaz, A. (2021). Kredi kartı dolandırıcılık tespitinin makine öğrenmesi yöntemleri ile tahmin edilmesi. Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 13(2), 28–33.

Ser, G. (2012). Application of multiple ımputation method for missing data estimation. Gazi University Journal of Science, 25(4), 869–873.

Shmaglit, L., Khryashchev, V. (2013). Gender classification of human face images based on adaptive features and support vector machines. Optical Memory and Neural Networks (Information Optics), 22(4), 228–235.

Silahtaroğlu, G. (2013). Veri Madenciliği Kavram ve Algoritmaları. Papatya Yayınevi: İstanbul, Türkiye.

Silverman, B., Jones, M. (1989). An important contribution to nonparametric discriminant analysis and density estimation: commentary on Fix and Hodges (1951). International Statistical Review, 57(3), 233–238.

Solmaz, R., Günay, M., Alkan, A. (2014). Fonksiyonel tiroit hastalığı tanısında naive bayes sınıflandırıcının kullanılması. XVI. Akademik Bilişim Konferansı, Mersin, Türkiye.

Sopharak, A., Nwe, K. T., Moe, Y. A., Dailey, M. N., Uyyanonvara, B. (2008). Automatic exudate detection with a Naive Bayes classifier. International Conference on Embedded Systems And Intelligent Technology, Bangkok, Tayland.

Stoean, C., Stoean, R. (2014). Evolutionary support vector machines and their application for classification. Springer International Publishing: New York, USA.

Syaliman, K. U., Nababan, E. B., Sitompul, O. S. (2018). Improving the accuracy of k-nearest neighbor using local mean based and distance weight. 2nd International Conference on Computing and Applied Informatics, ICCAI, Medan, Endonezya.

Şamkar, H., Güven, G. (2019). Özel markalı ürünleri satın alma sıklığının sıralı lojistik regresyon analizi ile incelenmesi. Kırklareli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 8(1), 79–96.

Şimşek, D. Ö. (2018). Tria*j Sistemlerine genel bakış ve Türkiye'de acil servis başvurularını etkileyen faktörleri lojistik regresyon ile belirlenmesi. Sosyal Güvenlik U*zmanları Derneği Sosyal Güvence Dergisi, 7(13), 84–115.

Tabachnick, B. G., Fidell, L. S. (2013). Using multivariate statistics (Sixth Edit). Pearson Education, Inc.: London, UK.

Taş, E., Gökçe, B. (2017). Yumurtaların çevrimiçi bir Destek Vektör Makinesi kullanılarak sınıflandırılması. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 17, 914–921.

Taşcı, E., Onan, A. (2016). K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. 18. Akademik Bilişim Konferansı, Aydın, Türkiye.

Tayyar, N., Tekin, S. (2013). İMKB-100 endeksinin destek vektör makineleri ile günlük, haftalık ve aylık veriler kullanarak tahmin edilmesi. AİBÜ Sosyal Bilimler Enstitüsü Dergisi, 13(1), 189–2017.

Ting, S. L., Ip, W. H., Tsang, A. H. C. (2011). Is Naïve Bayes a good classifier for document classification? In International Journal of Software Engineering and its Applications, 5(3), 37–46.

Türkmen, H. İ., Kurt, Z., Karslıgil, M. E. (2006). Destek vektör makinesi yöntemi ile yüz güzelliği kararı. 2006 IEEE 14th Signal Processing and Communications Applications, Antalya, Türkiye.

Uğuz, S. (2019). Makine öğrenmesi teorik yönleri ve python uygulamaları (1. Basım). Nobel Akademik Yayıncılık: Ankara, Türkiye.

Uludağ, O., Gürsoy, A. (2020). On the financial situation analysis with knn and naive bayes classification algorithms. Journal of the Institute of Science and Technology, 10(4), 2881–2888.

Ünver, H. M., Kökver, Y., Çiftçi, A. (2020). Hipertansiyon tahmini için temel bileşen analizinin k ullanımı. Uluslararsı Mühendislik Araştırma ve Geliştirme Dergisi, 12(3), 42–51.

Üstüner, M., Abdikan, S., Bilgin, G., Şanlı, F. B. (2020). Hafif gradyan artırma makineleri ile tarımsal ürünlerin sınıflandırılması. Türk Uzaktan Algılama ve CBS Dergisi, 1(2), 97–105.

Verma, R., Krishan, K., Rani, D., Kumar, A., Sharma, V., Shrestha, R., Kanchan, T. (2020). Estimation of sex in forensic examinations using logistic regression and likelihood ratios. Forensic Science International: Reports, 2(0), 1–6.

Wang, Z., Shao, Y. H., Bai, L., Li, C. N., Liu, L. M., Deng, N. Y. (2018). Insensitive stochastic gradient twin support vector machines for large scale problems. Information Sciences, 462, 114–131.

Yan, Z., Chen, H., Dong, X., Zhou, K., Xu, Z. (2022). Research on prediction of multi-class theft crimes by an optimized decomposition and fusion method based on XGBoost. Expert Systems with Applications, 207, 117943.

Yayar, R., Daşçı, A. N. (2020). Özel sağlık sigortası talebini etkileyen faktörlerin ikili lojistik regresyon yöntemiyle analizi: İstanbul örneği. Sosyal Güvenlik Dergisi, 10(1), 19–40.

Yazıcı, M. (2018). Kredi risk analizlerinde diskriminant analizi, lojistik regresyon ve yapay sinir ağlarının karşılaştırılması. Maliye ve Finans Yazıları, 1(109), 91–106.

Yin, S., Yin, J. (2016). Tuning kernel parameters for SVM based on expected square distance ratio. Information Sciences, 370, 92–102.

Zhang, W., Gaoa, F. (2011). An improvement to Naive Bayes for text classification. Procedia Engineering, 15, 2160–2164.

Zhang, Y., Wang, Y., Xu, J., Zhu, B., Chen, X., Ding, X. ve Li, Y. (2022). Comparison of prediction models for acute kidney ınjury among patients with hepatobiliary malignancies based on XGBoost and LASSO-Logistic algorithms. International Journal of General Medicine, 14(0), 1325-1335.

Zhu, J., Ge, Z., Song, Z., Gao, F. (2018). Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. Annual Reviews in Control, 46(1), 107–133.

Zontul, C., Hayta, E., Zontul, M., Taş, A., Siliğ, Y. (2017). Destek vektör makineleri ile fibromiyalji sendromu sınıflaması. Acta Infologica, 1(2), 92–98.