

ÖĞRENCİLERİN ÖĞRENME YÖNETİM SİSTEMİ KULLANIM ALIŞKANLIKLARININ PANDAS İLE ANALİZİ

ANALYSIS OF STUDENTS' LEARNING MANAGEMENT SYSTEM USE HABITS WITH PANDAS

Tolga DEMİRHAN¹, İlkay DEMİRALAY²

ÖZ: İnternet bugün, eğitimden ticarete pek çok alanda yaşantımıza yön vermektedir. İnternetin bu kadar yaygın kullanılıyor olması ortaya çıkan bilgi büyüklüğünü etkileyerek web sunucu düzeyinde ziyaretçi analizleri yapılmasının gerekliliğini arttırmıştır. Bu sebeple günümüzde bu tür bilgi analizine yönelik ilgi ve çalışmalar artarak devam etmektedir.

Bu bağlamda yapılan çalışmada, okulumuz Öğrenme Yönetim Sistemi (ÖYS) sunucusu web erişim günlük (Log) dosyalarının PANDAS kütüphanesi ile analizinin yapılarak öğrencilerin, sistemi kullanım alışkanlıklarının araştırılması amaçlanmıştır. Çalışmada, tarama yöntemiyle elde edilen dosyalar üzerinde web kullanım madenciliği yöntemlerinin kullanıldığı ve istatistik bilgilerin çıkartılmasını sağlayan bir yazılım geliştirilmiştir. Geliştirilen yazılım ile öğrencilerin; Öğrenme Yönetim Sistemine erişim yolları, ziyareti gerçekleştirdiği ülkeler, ziyareti gerçekleştirdiği iller, ders bazlı erişim yoğunluğu, ders doküman indirme yoğunluğu, erişimde kullanılan donanım, işletim sistemi, web tarayıcı yazılımı, aylık erişim trafiği, haftalık erişim trafiği, günlük erişim trafiği, saatlik erişim trafiği konularında veriler elde edilmiştir. Elde edilen veriler üzerinden yapılan analizler neticesinde varılan sonuçlar araştırmacılar ile paylaşılmıştır.

Anahtar sözcükler: Öğrenme Yönetim Sistemi, Pandas, Log, Web Kullanım Madenciliği.

ABSTRACT: Today, the Internet is directing our lives in many fields from education to commerce. Since the Internet is being used extremely wide, the necessity of performing visitor analysis has increased at the level of the webserver by affecting the size of information. For this reason, attention and studies are increasingly going on directed to such information analysis in today's world.

In this context, it was aimed to investigate the system usage habits of the students by analyzing the web access log (Log) files of the Learning Management System (LMS) server of our school with the PANDAS library. In the study, a software has been developed that uses web mining methods on the files obtained by scanning method and provides statistical information to be extracted. With the developed software, students can; Access routes to the Learning Management System, countries visited, provinces visited, course-based access density, course-based document download density, hardware used for access, operating system, web browser software, monthly access traffic, weekly access traffic, daily access traffic, hourly access data on traffic were obtained. The results obtained as a result of the analyzes made on the obtained data were shared with the researchers.

Keywords: Learning Management System, Pandas, Log, Web Usage Mining

Bu makaleye atf vermek için:

Demirhan, T. ve Demiralay, İ. (2023). Öğrencilerin öğrenme yönetim sistemi kullanım alışkanlıklarının pandas ile analizi, *Trakya Eğitim Dergisi*, 13(3), 1736-1751.

Cite this article as:

Demirhan, T. & Demiralay, I. (2023). *Analysis of students' learning management system use habits with pandas*, *Trakya Journal of Education*, 13(3), 1736-1751.

EXTENDED ABSTRACT

Introduction

The Internet is used extensively in many areas from education to commerce. As a result, the amount of data generated and stored is increasing day by day (Uladi & Ari, 2023).

Web mining and data analysis techniques are used to make sense of these high-dimensional data (Yildiz, 2022). Pandas, one of the tools that can be used for data analysis, stands out as a powerful tool that accelerates and facilitates the analysis process by using SQL-like scripts (Bantilan, 2020).

¹ Öğr.Gör.Dr., Trakya Üniversitesi, Edirne/Türkiye, e-mail:tolgademirhan@trakya.edu.tr, ORCID: 0000-0001-9840-4457

² Öğr.Gör., Trakya Üniversitesi, Edirne/Türkiye, e-mail:ilkaydemiralay@trakya.edu.tr, ORCID: 0000-0002-0708-0627

The Learning Management System (LMS), which is the subject of the study, is a system that enables learners to watch and download course content and to make course-based assessment and evaluation. It is a necessity for system and student analysis to keep student mobility on these systems with the access logs collected by the server.

In the present study, web log files were obtained from the LMS server. With the software developed using Pandas, information such as activities on the system, temporal accesses, access types, courses and files accessed, country, province and district information, and hardware and software used in the connection were analyzed.

Web access log files are automatically generated by the server for each new day. Each interaction of the visitors with the system is added to these files as a new line. These unstructured data, which include information such as ip number, reference address, request realization time, added to the log files, are created out of the user's request.

It is very difficult to extract information from log files that contain large amounts of data due to their format. Therefore, a tool is needed to analyze the data. (Çınar & Şakir, 2016) In this study, Python programming language and Pandas library were used to analyze log files.

In the literature, it is seen that Pandas is widely used as a tool used in a certain part of the studies. In the present study, the Pandas library was used to cover all the processes from the beginning to the end of the analysis process. It is thought that the study will contribute to the literature by benefiting the researchers in the analysis of LMS log files, analysis processes with the Pandas library.

Method

The study was carried out within the framework of the scanning model, which is one of the scientific research models. In the study, Web usage mining process was applied in the web access log files obtained by scanning the fall semester of the 2019-2020 academic year on the server.

Findings

The following findings were obtained in the analysis:

- The address of <https://lms.trakya.edu.tr> is used at a rate of 99.72% to access the LMS.
- Computers are used with a rate of 95.59% to access the LMS.
- Windows operating system is used with a rate of 90.25% to access LMS.
- To view the LMS, Google Chrome is used with a rate of 34.66%, and Firefox is used with a rate of 23.75% in the second place.
- Participants from 12 different countries participated in the LMS.
- The most of participants was from Turkey with 98.46%.
- It was found that the LMS was accessed from 69 different provinces and the highest participation rate was 63.82%, followed by Ankara with 12.41%.
- It was found that the most frequently visited course was the Introduction to Programming course with a rate of 12.47%.
- It was found that the most downloaded course was the Introduction to Programming course with a rate of 21.34%.
- It was found that the most intense access to LMS was between 17:00-18:00 with 7.53%.
- It has been found that 94.20% of the days when access to the LMS is the busiest is Monday.
- It was found that the week with the highest number of access to LMS was the first week of January with a rate of 87.44%.
- It was found that the most intense month of accessing LMS was 90.34% in January.

Discussion and Conclusion

In this study, it is aimed to analyze the learning management system usage habits of distance education students with the Pandas library. In this context, a software has been developed in which web usage mining methods are used and statistical information is extracted on the web access log files of the 2019-2020 academic year fall semester belonging to the LMS server. The software has been developed in Python using the Pandas library.

As a result of the study, important information was obtained about our students' LMS usage habits. It has been seen that the flexibility and finalization performance of the Pandas library used with query scripts is satisfactory.

It is thought that the study will contribute to educational web usage mining studies. The next study is planned to be conducted on the effect of the pandemic process on the distance education students' LMS usage habits.

GİRİŞ

Bilindiği üzere internet, eğitimden ticarete pek çok alanda yoğun bir şekilde kullanılmaktadır. Buna bağlı olarak ortaya çıkan ve depolanan veri miktarı her geçen gün artmaktadır (Uladi & Ari, 2023). Bu veriler internet üzerinde kullanıcılar ile paylaşılan bilgiler dışında arka planda kullanıcı isteklerini cevaplayan sunucular tarafından tutulan ziyaret kaynaklı bilgiler şeklinde de bulunmaktadır.

Kurumların elinde bulunan, yüksek boyutlu bu verilerin anlamlandırılmadan depolanması veriye sahip olan bu kurumlar açısından herhangi bir fayda ifade etmez (Yıldız, 2022). Bu noktada web madenciliği, veri analiz teknikleri kullanılır. Veri analiz işlemi kısaca; işe uygun özelleştirilmiş yazılımlarla, eldeki verileri inceleme süreci olarak tanımlanabilir (Baykal, A., & Coşkun, C., 2009). Bu işlem ile depolanan verilerin analizi yapılarak, kuruma mevcut durumu hakkında anlamlı çıktılar verilerek kurumun gelecek için bir vizyon belirlemesini sağlayacak ipuçlarına sahip olması sağlanır. Veri analizi için kullanılacak pek çok araç olmakla birlikte Pandas Kütüphanesi, SQL benzeri betikler kullanarak verilerin yüklenmesi, işlenmesi ve analiz edilmesini sağlayan analiz sürecini hızlandırıp kolaylaştıran güçlü bir araç olarak öne çıkar (Bantilan, 2020).

Çalışmaya konu olan Öğrenme Yönetim Sistemi (ÖYS) ise öğrenenlerin ders içeriklerini izlemesini, indirmesini ve ders bazlı ölçme ve değerlendirme yapılmasını sağlayan sistemdir. Bu sistemler üzerindeki öğrenci hareketliliğinin sunucu tarafından toplanan erişim günlükleri (log) tarafından saklanması, sistem ve öğrenci analizi noktasında yapılacak çalışmalar için bir gereklilik olmaktadır.

Çalışmada ÖYS sunucusu üzerinde tarama yapılarak belirlenen döneme ait log dosyaları elde edilmiştir. Geliştirilen yazılım bu dosyalar üzerinde kullanılarak, sistem üzerindeki hareketlilikler, zamansal erişimler, erişim türleri, erişilen ders ve dosyalar, bağlantı kurulan ülke, il ve ilçe bilgileri, bağlantı da kullanılan donanım ve yazılım gibi bilgiler çözümlenerek elde edilmiş ve analizler gerçekleştirilmiştir.

Çalışmanın ÖYS log dosyalarının analizi, Pandas kütüphanesi ile analiz işlemleri konulu çalışmalarda araştırmacılara fayda sağlayarak literatüre katkı sunacağı düşünülmektedir.

Log Dosyaları:

Bilişim dünyasında sistemler, üzerlerinde gerçekleşen olayları genellikle .log uzantılı olarak kullanılan log dosyası veya günlük dosyası diye anılan dosyalarda kayıt altına almaktadır. Bu dosyalar kullanıldıkları alanlara göre farklı biçimlerde olabilmektedir. (Baykara, Daş, & Tuna, 2016)

Özellikle internetin kullanıldığı Uzaktan Eğitimde, içeriklerin değerlendirilmesi ve öğrenci takibi bu log kayıtları ile mümkün olmaktadır. Eğitim ortamı geri bildirimleri ve öğrenci hareketliliğinin gözlemlenmesi, log dosyaları analiz edilerek yapılabilmektedir. (Aricı & Çiftçi, 2007)

Çalışmaya konu olan log dosyaları web sunucu üzerinde tutulan ziyaretçilerin siteye erişimleri ile ilgili bilgileri tutan web erişim log dosyalarıdır.

Web erişim log dosyaları, sunucu tarafından otomatik olarak her yeni gün için oluşturulur. Oluşturulan bu dosyalara, ziyaretçilerin sistem ile her bir etkileşimi yeni bir satır olarak eklenir. Log dosyalarına eklenen ip no, referans adresi, isteğin gerçekleştirme zamanı vb. bilgiler bulunan bu yapılandırılmamış veriler kullanıcının isteği dışında oluşur.

Bu bilgiler, kullanılan web sunucusuna, kullanılan log biçimine ve sunucu üzerinde yapılan ayarlamalara göre farklılık gösterir (Özseven & Düğenci, 2011).

Web sunucularında oluşturulabilecek log dosyaları W3C (World Wide Web Consortium), Apache, Amazon Elastic Load Balancing, HAProxy ve JSON tipinde olabilmektedir. Çalışmaya konu olan veri kaynağı W3C formatlı log dosyaları kullanılarak oluşturulmuştur. W3C, Notepad gibi basit metin düzenleyici programlarında görüntülenebilir düz metin dosyalarıdır. Oluşturulan Log dosyasındaki bir satırlık kayda ait bilgiler aşağıda verilerek tabloda açıklanmıştır.

Tablo 1.
W3C log dosya tipi

Tip	İçerik
W3C	2021-09-29 00:00:44 193.255.140.27 POST /Notification/GetNotificationCount - 443 system_admin:1:False 193.255.xxx.xxMozilla/5.0+(Windows+NT+6.3;+Win64;+x64)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/91.0.4472.106+Safari/537.36 https://lms.trakya.edu.tr/Home/Index 200 0 0 15

Tablo 2.
Log dosyası alan açıklamaları

Alan adı	Örnek	Açıklama
DATE	2021-09-29	İsteğin Gerçekleşme tarihi
TIME	00:00:44	İsteğin Gerçekleşme zamanı
S-İP	193.255.140.27	Sunucunun ip adresi
CS-METHOD	POST	İstemcinin istediği yöntem GET or POST vb..
CS-URI-STEM	/Notification/GetNotificationCount	action'ın hedefi, istemcinin gitmek istediği adres
CS-URI-QUERY	-	URI sorgusu sadece dinamik sayfalar için tanımlı olur sorgu varsa olur
S-PORT	443	Sunucun hizmet port numarası
CS-USERNAME	system_admin:1:False	Kimliği doğrulanan kişi bilgileri
C-İP	193.255.xxx.xx	İstemci ip adresi, istekte bulunan bilgisayarın ip adresi
CS(USER-AGENT)	Mozilla/5.0+(Windows+NT+6.3;+Win64;+x64)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/91.0.4472.106+Safari/537.36	İstemcinin kullandığı işletim sistemi ve tarayıcı türü
CS(REFERER)	https://lms.trakya.edu.tr/Home/Index	Ziyaretçinin son ziyaret ettiği site.
SC-STATUS	200	Sunucudan istemciye giden HTTP protokol durum yanıt kodlarıdır: 2xx, 3xx, 4xx
SC-SUBSTATUS	0	Alt Durum Hata kodu
SC-WIN32-STATUS	0	Windows durum kodu
TIME-TAKEN	15	Talep edilen işlemin milisaniye cinsinden gerçekleşme süresi

Büyük miktarda veri içeren log dosyalarından sahip oldukları biçimden dolayı bilgi çıkarımı yapmak oldukça zordur. Dolayısıyla verileri analiz etmek için bir araca gereksinim duyulmaktadır. (Çınar & Şakir, 2016)

Bu çalışmada log dosyalarını analiz etmek için Python programlama dili ve Pandas kütüphanesi kullanılmıştır.

Web Kullanım Madenciliği:

Web kullanım madenciliği kısaca web erişim log dosyalarından elde edilen veriler üzerinde çalışarak bilgi elde edilmesi süreci olarak tanımlanmaktadır. Web madenciliği yaklaşımlarından birisi olan Web kullanım madenciliği ilk 1997 yılında Cooley, Mobasher ve Srivastava tarafından, web sunucularından kullanıcı erişim örüntülerinin otomatik keşfi olarak tanımlanmıştır (Cooley, Mobasher, & Srivastava, 1997)

Log dosyalarında Web kullanım madenciliği yapılarak, ziyaretçi sayısı, ziyaret sıklığı, erişimde kullanılan yazılım ve donanımlar vb. istatistiksel sonuçlar yanında saldırı tespitleri, sayfa geçişleri ve

yapılandırılmış veriler üzerinde algoritmalar çalıştırılarak amaca uygun analizler yapılabilmektedir. Literatür incelendiğinde Web kullanım madenciliğinin gerçekleştirildiği aşağıdaki gibi benzer örnekler ile karşılaşılmaktadır;

Çınar, Çınar ve Bilge (2014), Web sunucusu günlüklerinin analizini WEKA programı ile yapmış ve elde ettikleri istatistiksel sonuçları, saldırıların tespit edilmesi, saldırı çeşitleri ve saldırıda kullanılan yöntemlere ilişkin bilgi sağlaması amacıyla kullanmıştır.

Gezer, Erol ve Gülsece, uluslararası akademik değişim programları ile ilgili bilgileri içeren web sunucu log dosyalarını, WUMprep ve WUM web madenciliği yazılımlarını kullanarak incelemiştir. İnceleme sonucunda sitede en çok kullanılan sayfalar ile en az ziyaret edilen sayfalar belirlenmiş, bağlantılar arasındaki geçişler keşfedilmiştir. (Gezer, Erol, & Gülsece, 2007)

Özseven ve Düğenci, web sunucu log dosyalarının analizi için “Log Analiz” isminde bir yazılım geliştirmiştir. Yazılım web sitesine ait çeşitli istatistik bilgileri çıkarmakta ve apriori algoritması ile birliktelik kurallarını keşfetmiştir. (Özseven & Düğenci, 2011)

Literatürde ÖYS log dosyaları üzerinde eğitim alanlı yapılan web kullanım madenciliği çalışmalarının yeterli sayıda olmadığı ve yapılan çalışmaların öğrencinin sistemi kullanımı ile başarısının ilişkilendirildiği (Özbay & Ersoy, 2017) veya sistemi kullanımı ile öğrenci başarısı tahminini yapmak (Yavuzarslan & Erol, 2022) amaçlı çalışmalar gerçekleştirildiği görülmektedir.

Web kullanım madenciliği temelde üç aşamadan meydana gelir. Birinci aşama ön işlem olarak anılan aşamadır. Bu aşama, veri temizleme, birleştirme, parçalama, silme, tanımlama işlemlerini içerir. İkinci aşama, veri madenciliği temeli oluşturulan uygulamaları içeren örüntünün keşfidir. Üçüncü aşama ise örüntünün analizidir. (Burçak, 2012)

Ön işlem aşamasında yer alan veri temizleme, web kullanım madenciliğindeki ilk adımdır. Bu adımda, log kayıtlarından yer alan ilgisiz veriler çıkarılır, dosya uzantıları düzeltilir, çoklu kayıtlar birleştirilir. Temizleme işleminden sonra kayıtlı kullanıcıların kullanıldığı uygulamalar için birleştirme işlemleri yapılabilir (Baykal & Coşkun, 2009). Veri ön işlem, doğru yapıldığında kaynak konularında tasarruf sağlamanın yanında sonuçların güvenilirliğini arttırmakta, değerlendirmeyi de kolaylaştırmakta olduğundan önemli bir aşamadır.

Web kullanım madenciliğinin ikinci aşamasında örüntü keşfi için sınıflandırma, istatistiksel analiz, kümeleme, ilişkilendirme kuralları gibi birçok yöntem kullanılabilir. (Burçak, 2012)

Üçüncü aşama olan örüntü analizinde ise örüntü keşfi aşamasında ortaya çıkan bilginin analiz edilmesini kapsar, bu bağlamda veri ve bilgi sorgulama ile görselleştirme işlemlerinden faydalanılır.

Veri Analizi ve Pandas:

Veri Analizi, özelleştirilmiş sistemler ve yazılımlarla, veri kümelerinin içerdikleri bilgiler hakkında sonuçlar çıkarma sürecidir. Veri analizi teknolojileri ve teknikleri, daha yerinde kararlar vermek için yaygın olarak kullanılmaktadır. (Enache, 2019)

Pandas bu teknolojilerden biri olmakla birlikte, Python ekosisteminde modern veri mühendisliği, analizi ve modellemesi için veri ile çalışan tüm alanlarda kullanılabilen, (Bantilan , 2020) açık kaynaklı bir Python kütüphanesidir. Matplotlib ve Seaborn kütüphaneleri ile birlikte Pandas, tablo verilerinin görsel analizi için çok çeşitli imkanlar sunar. Ayrıca Pandas, SQL benzeri sorgular kullanarak bu verileri yüklemeyi, işlemeyi ve analiz etmeyi çok kolaylaştırır (Kashnitsky, 2021)

Pandas'taki temel veri yapısı, seri, satır ve sütunlardan oluşan tablo şeklindeki DataFrame isimli bir yapıdır. Kullanıcıların DataFrame içeriklerini dönüştürmesine, filtrelemesine ve değiştirmesine izin veren çeşitli işlemler içerir. Ayrıca herhangi bir Python programında kullanılacak bir çerçeve olduğundan, döngüler, if-else dallanma veya modüller gibi prosedürel bir programlama dilinin avantajlarını, veri analizleri ve manipülasyonu avantajlarını birleştirir. (Hagedorn, Kläbe, & Sattler, 2021) Farklı alanlardaki veri kümeleriyle çalışmayı kolaylaştırmayı ve istatistiksel modellerin uygulanması için bir dizi temel yapı taşı sağlamayı amaçlar. (Mckinney, 2010)

Pandas, daha az kod satırı ile bir analiz modunu ifade etmek çok daha kolay olduğundan, hesaplamayı hızlandırmanın yanı sıra hızlı kodlama ve gelişmiş kod bakımına izin verir. Birleştirme, sıralama, gruplama ve yeni değişkenler oluşturma gibi işlemler paket içinde, derlenmiş ve optimize edilmiş makine kodunda gerçekleştirildiğinden, hız kazancı muazzam olabilir. (Ayer, Miguez, & Toby, 2014)

Pandas yüksek performanslı, kullanımı kolay veri yapıları ve analiz araçları sağlayan tamamen belgelenmiş, açık kaynaklı, BSD lisanslı bir kütüphanedir ve onu geliştirmek için çalışan güçlü bir geliştirici topluluğuna sahiptir. (Hummel, 2016)

Bugün Pandas kütüphanesi, verinin olduğu her alanda akla ilk gelen araçlardan biridir. Veri madenciliği konulu, Pandas kullanımının gerçekleştirildiği akademik çalışmalar incelendiğinde

araştırmacıların Pandas kütüphanesinden, analiz sürecinin belli başlı bölümlerinde yararlandığı görülür. Örneğin;

- Sodhi, Awasthi ve Sharma, Pandas kütüphanesini veri dosyasını okumak, depolamak ve ardından okunan verileri sütunlara ayırıp bir vektöre dönüştürmek yani veriyi yapılandırmak için kullanmıştır. (Sodhia , Awasthib , & Sharmac , 2019)

- Sumali, Pandas kütüphanesini veri setindeki önemli değişkenler arasındaki ilişkileri ortaya çıkarmak için verilerin basit bir şekilde sıralanması, listelenmesi gereken işlemlerde kullanmıştır. (Sumali, 2020)

- Patchaiammal, Sundar ve Thirumalaiselv, Pandas kütüphanesini verisetindeki verileri dosyadan okumak, boş kayıtları silmek gibi veri eksensli temel işlemlerde kullanmıştır. (Patchaiammal, Sundar, & Thirumalaiselv, 2021)

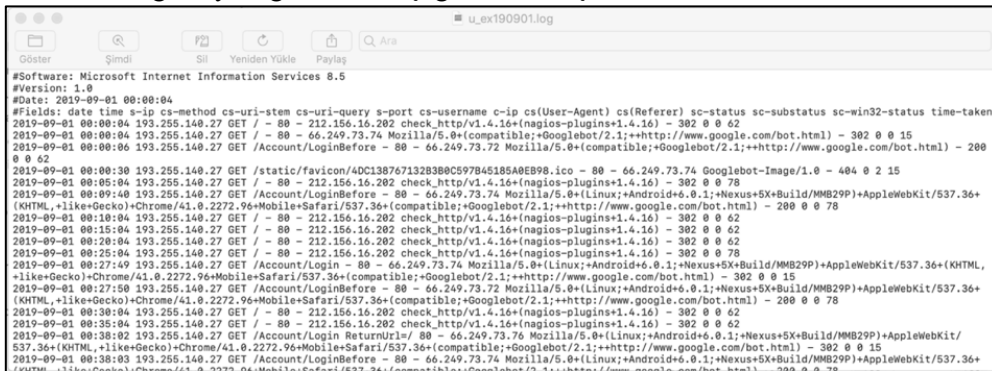
Yukarıda da bahsedildiği gibi Pandas'ın yaygın olarak çalışmaların belirli bir bölümünde başvurulan bir araç olarak kullanıldığı görülmektedir. Bu çalışmada ise Pandas kütüphanesi analiz sürecinin başından sonuna, yani log veri dosyalarını okumak, kayıtlar üzerinde manipülasyon yapmak, silmek, aramak, filtrelemek, farklı formatlarda kaydetmeye ve son olarak sonuçları görselleştirmeye kadar ki tüm süreçleri kapsayacak şekilde kullanılmıştır. Literatürde yer alan çalışmalarda log dosyaları üzerine yapılan çalışmaların verilerin tamamı yerine hedeflenen birkaç bilgi üzerine yoğunlaştığı, log dosyasında tutulan verilerin tamamını inceleyen bir çalışma olmadığı görülmüştür. Sunulan bu çalışmanın diğer çalışmalardan en önemli farkı çalışmada ÖYS sunucusunun log dosyalarının detaylı bir şekilde incelenmiş olması ve bu analiz sürecinin tüm aşamalarının Pandas kütüphanesi kullanılarak yapılmasıdır.

MATERYAL VE METOT

Çalışma, bilimsel araştırma modellerinden biri olan tarama modeli çerçevesinde yürütülmüştür. Tarama modelleri, mevcut veya geçmişte gerçekleşmiş bir durumu var olduğu biçimiyle betimlemeyi amaçlayan araştırma yaklaşımıdır. (Karasar, 2015) Çalışmada sunucu üzerinde 2019-2020 Eğitim Öğretim yılı güz dönemi taranarak elde edilen web erişim günlük (Log) dosyalarında Web kullanım madenciliği süreci uygulanmıştır. Bu bağlamda elde edilen dosyalardaki veriler öncelikle hazırlanmıştır. Veri hazırlama, veri analizi sürecinde hayati bir adımdır, veri ön işlemeyi ve veri işlemeyi içerir. Ön işleme, orijinal ham verileri veri analizi için kullanılabilir hale getirmek için temizlemeyi, entegre etmeyi, dönüştürmeyi ve küçültmeyi amaçlarken, veri işleme ise önceden işlenmiş veri kümesini işlenebilecek bir veri formatına dönüştürme sürecini içerir. (Stančin & Jović, 2019)

Veri Hazırlama

Çalışmaya konu olan veri seti Uzaktan Eğitim veren Trakya Üniversitesi Tunca Meslek Yüksekokulunun Öğrenme Yönetim Sistemi (ÖYS) sunucusuna ait web erişim günlük (Log) dosyaları kullanılarak oluşturulmuştur. Log dosyaları 2019-2020 Eğitim Öğretim yılı güz dönemini kapsar. Kaynak olarak kullanılacak log dosyası görünümü aşağıda verilmiştir.



```
#Software: Microsoft Internet Information Services 8.5
#Version: 1.0
#Date: 2019-09-01 00:00:04
#Fields: date time s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip cs(User-Agent) sc-status sc-substatus sc-win32-status time-taken
2019-09-01 00:00:04 193.255.148.27 GET / - 80 - 212.156.16.202 check_http/v1.4.16+(nagios-plugins1.4.16) - 302 0 0 62
2019-09-01 00:00:04 193.255.148.27 GET / - 80 - 66.249.73.74 Mozilla/5.0+(compatible);+Googlebot/2.1;+http://www.google.com/bot.html) - 302 0 0 15
2019-09-01 00:00:06 193.255.148.27 GET /Account/LoginBefore - 80 - 66.249.73.72 Mozilla/5.0+(compatible);+Googlebot/2.1;+http://www.google.com/bot.html) - 200 0 0 62
2019-09-01 00:00:30 193.255.148.27 GET /static/favicon/4DC1387671328380C597845185A0E998.ico - 80 - 66.249.73.74 Googlebot-Image/1.0 - 404 0 2 15
2019-09-01 00:05:40 193.255.148.27 GET /Account/LoginBefore - 80 - 66.249.73.74 Mozilla/5.0+(Linux;+Android;6.0.1;+Nexus+5X+Build/MMB29P)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/41.0.2272.96+Mobile+Safari/537.36+(compatible);+Googlebot/2.1;+http://www.google.com/bot.html) - 200 0 0 78
2019-09-01 00:10:04 193.255.148.27 GET / - 80 - 212.156.16.202 check_http/v1.4.16+(nagios-plugins1.4.16) - 302 0 0 62
2019-09-01 00:15:04 193.255.148.27 GET / - 80 - 212.156.16.202 check_http/v1.4.16+(nagios-plugins1.4.16) - 302 0 0 62
2019-09-01 00:20:04 193.255.148.27 GET / - 80 - 212.156.16.202 check_http/v1.4.16+(nagios-plugins1.4.16) - 302 0 0 62
2019-09-01 00:25:04 193.255.148.27 GET / - 80 - 212.156.16.202 check_http/v1.4.16+(nagios-plugins1.4.16) - 302 0 0 78
2019-09-01 00:27:49 193.255.148.27 GET /Account/Login - 80 - 66.249.73.74 Mozilla/5.0+(Linux;+Android;6.0.1;+Nexus+5X+Build/MMB29P)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/41.0.2272.96+Mobile+Safari/537.36+(compatible);+Googlebot/2.1;+http://www.google.com/bot.html) - 302 0 0 15
2019-09-01 00:27:50 193.255.148.27 GET /Account/LoginBefore - 80 - 66.249.73.72 Mozilla/5.0+(Linux;+Android;6.0.1;+Nexus+5X+Build/MMB29P)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/41.0.2272.96+Mobile+Safari/537.36+(compatible);+Googlebot/2.1;+http://www.google.com/bot.html) - 200 0 0 78
2019-09-01 00:30:04 193.255.148.27 GET / - 80 - 212.156.16.202 check_http/v1.4.16+(nagios-plugins1.4.16) - 302 0 0 62
2019-09-01 00:35:04 193.255.148.27 GET / - 80 - 212.156.16.202 check_http/v1.4.16+(nagios-plugins1.4.16) - 302 0 0 62
2019-09-01 00:38:02 193.255.148.27 GET /Account/LoginReturnUrl/+ / 80 - 66.249.73.76 Mozilla/5.0+(Linux;+Android;6.0.1;+Nexus+5X+Build/MMB29P)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/41.0.2272.96+Mobile+Safari/537.36+(compatible);+Googlebot/2.1;+http://www.google.com/bot.html) - 302 0 0 15
2019-09-01 00:38:03 193.255.148.27 GET /Account/LoginBefore - 80 - 66.249.73.74 Mozilla/5.0+(Linux;+Android;6.0.1;+Nexus+5X+Build/MMB29P)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/41.0.2272.96+Mobile+Safari/537.36+(compatible);+Googlebot/2.1;+http://www.google.com/bot.html) - 200 0 0 78
```

Şekil 1. Log Dosyası

Veri Setini Oluşturma ve Temizlik:

Günlük olarak oluşturulan log dosyalarının içeriği düz metin formatlı satırlardan oluşmaktadır. Anlamlı verilere ulaşmak ve ilişkisel veritabanlarında da olduğu gibi satır ve sütunlardan oluşan iki boyutlu bir veri seti oluşturmak için pandas kütüphanesi kullanılarak aşağıdaki adımlar gerçekleştirilmiştir;

a) Log dosyalarının her biri tek tek açılarak bilgi olan satırlar okunarak bir dizide toplanmıştır.

b) Dizideki satırlar incelendiğinde her bir kaydın, sahip olduğu özelliklerin boşluklar ile ayrılarak verildiği tespit edilmiştir. Bu bağlamda kayıtları temsil eden tüm metinler boşluklarından parçalara ayrılmıştır.

c) log dosyasından alınan başlıklar kullanılarak DataFrame'e dönüşümü gerçekleştirilmiştir. Elde edilen bilgiler csv uzantılı bir dosyaya kaydedilmiştir. Oluşan bu dosyanın büyüklüğü 2.7 GB olup 8.899.938 satırdan/kayıttan oluşmaktadır.

d) Parçalama işleminden sonra dizi içerisinde her bir satırın 15 özelliğe sahip olması beklenmektedir. Bu koşula uymayan kayıtlar diziden çıkarılmıştır. Koşula uymayan kayıt sayısı 84 adettir.

e) http protokol yanıt kodu 200 dışında olan kayıtlar DataFrame'den çıkarılmıştır. Veri setinden çıkarılan kayıt sayısı 2.211.880 tür.

f) Okulun öğrencisi olmayan diğer uzaktan eğitim öğrencilerine ait 5.270.556 adet kayıt veri setinden çıkarılmıştır. Son durumda veri setinde 1.417.502 kayıt kalmıştır. Bu kayıtlar bir .csv dosyasına aktarıldığında .csv dosyasının boyutu 466.4MB olmuştur.

Veriyi Düzenleme:

Veriyi düzenleme, oluşturulan DataFrame içerisindeki bilgileri kendi özelinde ele almak için yapılan işlemleri barındıran bölümdür. Bu bölümde sırasıyla;

a) DataFrame içerisinde Date/Tarih bilgisini içeren sütun öncelikle date formatına dönüştürülüp yıl, ay, gün bilgilerini barındıracak şekilde ayrı sütunlara çözümlenerek aktarılmıştır.

date	Yıl	Ay	Gün	time
2019-09-03	2019	Eylül	3	09:09:54
2019-09-03	2019	Eylül	3	09:09:56
2019-09-04	2019	Eylül	4	14:40:29
2019-09-04	2019	Eylül	4	14:40:29
2019-09-06	2019	Eylül	6	07:13:12

Şekil 2. Date çözümlenerek düzenlenmiş dataframe görünümü

b) IPstack apisi kullanılarak öğrencinin bağlantı adresi bilgileri ülke, il, ilçe olarak çözümlenmiştir. Çözümlenen bilgilerin her biri DataFrame'e ayrı bir sütun olarak eklenmiştir. IPstack apisini kullanımı ile ilgili detaylar için <https://ipstack.com/documentation> adresi ziyaret edilebilir.

c-ip	Ülke	Şehir	İlçe
88.247.███	'Turkey'	'Istanbul'	'Istanbul'
176.███	'Turkey'	'Ankara'	'Ankara'
88.24.███	'Turkey'	'Istanbul'	'Maltepe'
176.237.███	'Turkey'	'Ankara'	'Ankara'
195.21.███	'Turkey'	'Istanbul'	'Maltepe'

Şekil 3. İp çözümlenerek düzenlenmiş dataframe görünümü

c) DataFrame içerisinde bulunan cs(User-Agent) sütunu, ziyaretçinin sisteme nasıl bir cihaz, hangi işletim sistemi ve hangi web tarayıcı uygulaması ile erişimde bulunduğu bilgisini tutmaktadır. Kullanılan user_agents isimli kütüphane ile bu bilgiler alınıp çözümlenerek DataFrame içerisinde ayrı sütunlar olarak saklanmıştır.

cs(User-Agent)	Device	OS	Browser
Mozilla/5.0+(Windows+NT+6.3;+Win64;+x64;+rv:69.0)+Gecko/20100101+Firefox/69.0	PC	Windows	Firefox
Mozilla/5.0+(Windows+NT+10.0;+Win64;+x64)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/70.0.3538.102+Safari/537.36+Edge/18.18362	PC	Windows	Edge
Mozilla/5.0+(Windows+NT+10.0)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/70.0.3538.102+Safari/537.36+Edge/18.18362	PC	Windows	Edge
Mozilla/5.0+(Windows+NT+10.0;+WOW64)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/78.0.3904.108+YaBrowser/19.12.1.229+Yowser/2.5+Safari/537.36	PC	Windows	Yandex Browser
Mozilla/5.0+(Windows+NT+10.0;+Win64;+x64)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/77.0.3865.120+Safari/537.36	PC	Windows	Chrome

Şekil 4. Cs(user-agent) çözümlenerek düzenlenmiş dataframe görünümü

Veri Analizi

Aşağıda, çalışmaya konu olan veriler üzerinde kullanılan Pandas metotları kısa betikler ve açıklamaları ile birlikte verilmiştir.

Metot: pandas.read_csv()

Açıklama: pandas_read_csv() metodu ile belirlenen bir csv dosyasına erişilip okunabilir. Örneğin pandas.read_csv('/Users/tdr/Desktop/dosyam.csv') betiği ile bilgisayarın masaüstünde ki dosyam isimli csv dosyasına erişilip kod betikleri içerisinde DataFrame olarak kullanılmaktadır.

Metot: pandas.DataFrame()

Açıklama: pandas.DataFrame() metodu ile bir dizi, seri kullanıcı tarafından belirlenen sütun başlıkları kullanılarak DataFrame'e dönüştürülebilmektedir.

Örneğin pandas.DataFrame(columns=["Ülke","Şehir"]) betiği ile hiçbir kayda sahip olmayan, Ülke ve Şehir başlıklı sütunlara sahip boş bir DataFrame oluşturulmuştur.

Metot: .to_csv()

Açıklama: .to_csv() metodu ile sahip olunan bir DataFrame, .csv dosyası olarak kaydedilebilir. Örneğin DF.to_csv (r'/Users/tdr/Desktop/dosyam.csv', index = False, header=True) betiği ile bilgisayarın masaüstüne dosyam isminde bir csv dosyası DataFrame'in sahip olduğu başlıkları kullanılarak, index numaraları olmadan kaydedilmektedir.

Metot: pandas.set_option()

Açıklama: pandas.set_option() metodu ile DataFrame'de gösterilecek satır, sütun sayısı, boyutları, sayısal ifadelerin biçimi düzenlenebilir. Bunlardan çalışmada kullanılanlar aşağıda örneklendirilmiştir.

pandas.set_option("display.max_columns", 20) betiği ile DataFrame'de gösterilecek sütun sayısının en fazla 20 olacağı ayarlanmıştır.

pandas.set_option("display.max_rows", 500) betiği ile DataFrame'de gösterilecek satır sayısının en fazla 500 olacağı ayarlanmıştır.

pandas.set_option("precision", 3) betiği ile DataFrame'de gösterilecek virgüllü değerlerin sadece virgülden sonra 3 basamağının görüntülenmesi ayarlanmıştır.

pandas.set_option('max_colwidth', 200) betiği ile DataFrame'de gösterilecek sütunların max sütun genişliğinin 200 olması sağlanmıştır.

pandas.options.display.float_format = '{:.2f}%'.format ile ondalık lı sayıların % sembolü ile görüntülenmesi sağlanmıştır.

Metot: pandas.to_datetime()

Açıklama: pandas.to_datetime() metodu ile DataFrame'in sahip olduğu sütunlarda barınan bilginin türü datetime türüne dönüştürülür. Örneğin DF["date"]= pandas.to_datetime(DF["date"]) betiği ile DF DataFrame'inde bulunan date isimli sütunun biçimi datetime türüne dönüştürülmüştür.

Metot: .drop_duplicates()

Açıklama: Drop_duplicates() metodu kullanılarak DataFrame içerisinde tekrarlı kayıtların DataFrame içerisinden tamamen silinmesi veya tercihe göre ilk veya son kaydın bırakılması diğerinin silinmesi sağlanabilir. Örneğin DF.drop_duplicates(subset=["date","time","cs-username"]) betiği ile DF isimli

DataFrame 'de date, time ve cs-username sütunlarında bulunan değerlerin birlikte aynı olduğu diğer satırlar bulunarak bunlardan ilkinin korunması, diğerlerinin silinmesi sağlanmıştır. Metodun sahip olduğu keep özelliği ile tekrarlı kayıtlardan hangisinin tutulacağı belirlenebilir. Eğer boş bırakılırsa varsayılan keep='first' olarak çalışır. Keep='last' kullanımı ile tekrar eden kayıtlardan sonuncu kaydın dataframe de tutulması sağlanır, önceki kayıtlar silinir. Keep özelliğine Keep=False değeri atanırsa tekrarlı kayıtların hepsinin DataFrame' den çıkarılması sağlanır.

Metot: .sort_values()

Açıklama: Sort_values metodu ile DataFrame içerisindeki satırların belirtilen sütun veya sütunlardaki bilginin türüne göre şöyle ki metinler için "A" dan "Z" ye, sayısal bilgiler için küçükten büyüğe veya metinler için "Z" den "A" ya, sayısal bilgiler için büyükten küçüğe sıralanması sağlanabilir. Aşağıdaki Örnekte DF.sort_values(["date","time"],ascending=True) betiği ile DF isimli DataFrame'deki kayıtların öncelikle date sütunundaki, daha sonra time sütunundaki bilgiler dikkate alınarak küçükten büyüğe sıralanması sağlanmaktadır.

Metot: .groupby()

Açıklama: groupby() metodu, bir DataFrame'de bulunan tekrarlı değerleri kategorik olarak gruplayarak, tekrarsız göstermek için kullanılmaktadır. Bu metodun devamını yazılan .sum(), .avg() vb.. hesaplama metotları ile birlikte grup içindeki kayıtların sahip olduğu sayısal ifadelerin toplamı, ortalaması vb.. hesaplamalar yapılabilir ya da group içindeki eleman sayısını bulmak için size() metodu kullanılabilir. Örneğin DF.groupby(["Ay","Hafta"]).size() betiği ile DF içerisindeki kayıtların Ay ve Hafta bilgileri bakımında gruplanması ve grup içerisindeki kayıtların sayısının gösterilmesi sağlanmıştır.

Metot: .reset_index(drop=True)

Açıklama: reset_index(drop=True) metodu ile DataFrame'de bulunan her bir kaydın sahip olduğu index değerinin tekrar oluşturulması sağlanır. DataFrame üzerinde yapılan silme, ekleme, filtreleme veya gruplama işlemlerinden sonra reset_index(drop=True) eklenerek kayıtların index numaralarının 0'dan başlayarak tekrar yazılması sağlanabilir. Ayrıca Reset_index metodunun sahip olduğu name özelliği ile hesaplanan sayısal verilerin bulunacağı sütuna bir isim atanabilir. Aşağıdaki DF.groupby(["Ülke","Şehir"]).size().reset_index(name="Tekrar Sayısı") betiği ile DF isimli DataFrame de Ülke'nin şehir bilgileri gruplanarak bu şehir bilgilerinin DataFrame içerisindeki tekrar sayıları "Tekrar Sayısı" isimli bir sütunun içerisinde görüntülenecek şekilde ayarlanmıştır.

Metot: .str.startswith()

Açıklama: .str.startswith() metodu kullanılarak DataFrame de işaret edilen bir sütunda parantez içerisinde verilen metin ile başlayan metinlere sahip kayıtların filtrelenmesi sağlanmaktadır. Örneğin DF[DF["cs-username"].str.startswith("A")] betiği ile DF isimli DataFrame içerisinde cs-username sütununda "A" ile başlayan bilgiye sahip olan kayıtların filtrelenmesi sağlanmıştır.

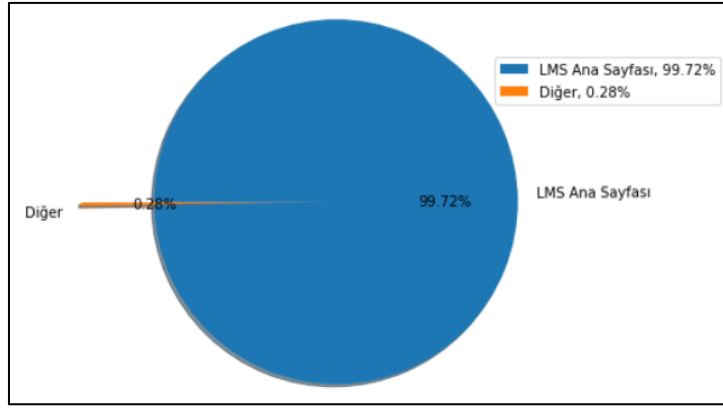
Metot: .str.extract()

Açıklama: .str.extract() metodu ile DataFrame'de belirtilen sütundaki bilginin istenen kriterlere uyan kısmı parantez içerisinde ifade edilerek hızlı ve kolay bir şekilde alınabilir. Örneğin DF["name"]=DF["cs-username"].str.extract('([A-Za-z]+\.\.?)') betiği ile cs-username sütununda bilginin başından noktaya kadar olan tüm metinler alınıp DF isimli DataFrame içerisinde yeni oluşturulan name sütununa eklenecektir.

BULGULAR

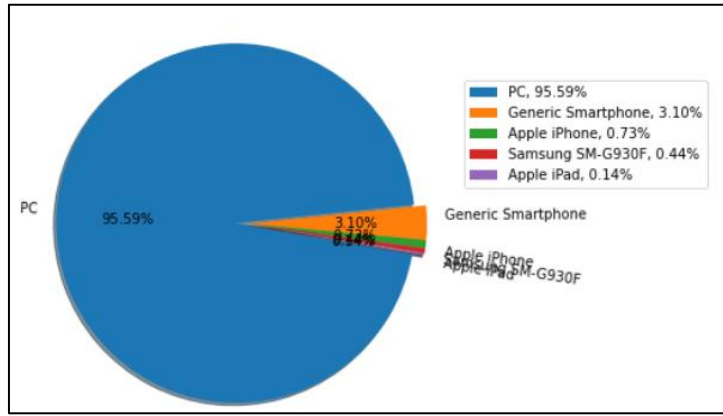
Çalışmada elde edilen bulgular, maddeler halinde aşağıda verilmiştir;

- ÖYS'ne erişim için arama motorları veya benzer alternatifler yerine ÖYS ana sayfası olan <https://lms.trakya.edu.tr> adresinin %99,72 oranında daha çok tercih edildiği bulunmuştur.



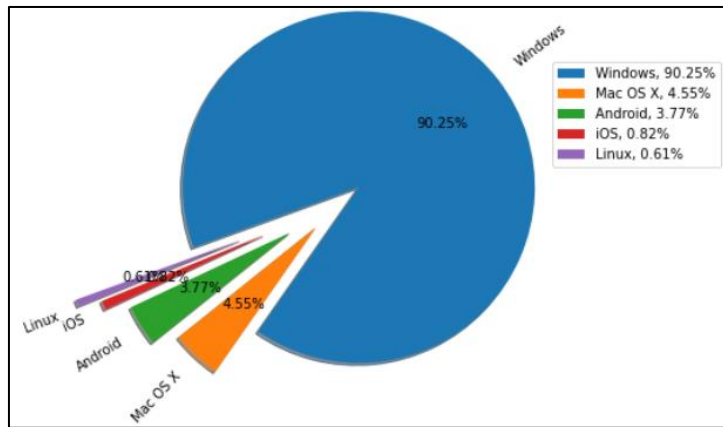
Grafik 1. Erişim yolu

- ÖYS'ne erişim için tercih edilen cihazlar araştırıldığında %95,59 oran ile bilgisayarın erişimde en çok tercih edilen cihaz olduğu onun hemen arkasından %3,10 oranı ile mobil cihazların yer aldığı bulunmuştur.



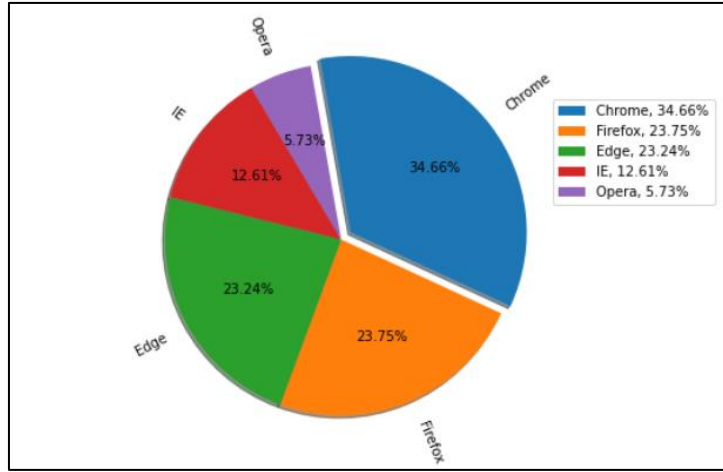
Grafik 2. Erişimde kullanılan cihazlar

- Erişimde kullanılan cihazların sahip oldukları işletim sistemleri araştırıldığında bu cihazların %90,25 oranı ile Windows işletim sistemine sahip oldukları ikinci sırada ise %4,55 ile Mac OSX olduğu bulunmuştur.



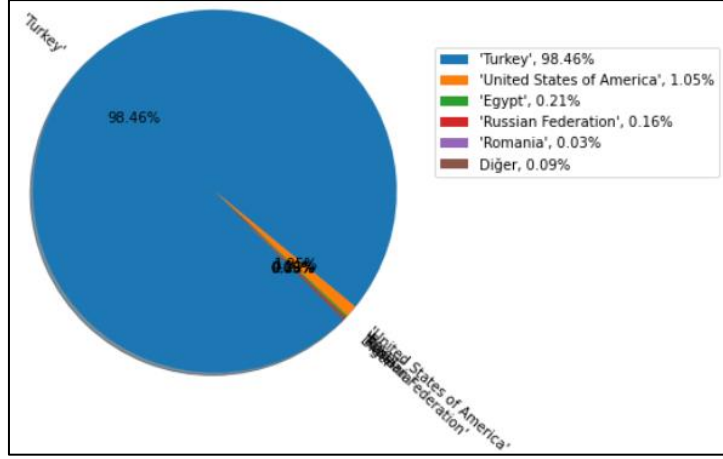
Grafik 3. Erişimde kullanılan işletim sistemi

- ÖYS'ni görüntülemek için kullanılan web görüntüleme programları araştırıldığında %34,66 oran ile Google Chrome'un diğer uygulamalara göre daha çok tercih edildiği bu listenin ikinci sırasında ise %23,75 oranı ile Firefox programının olduğu bulunmuştur.



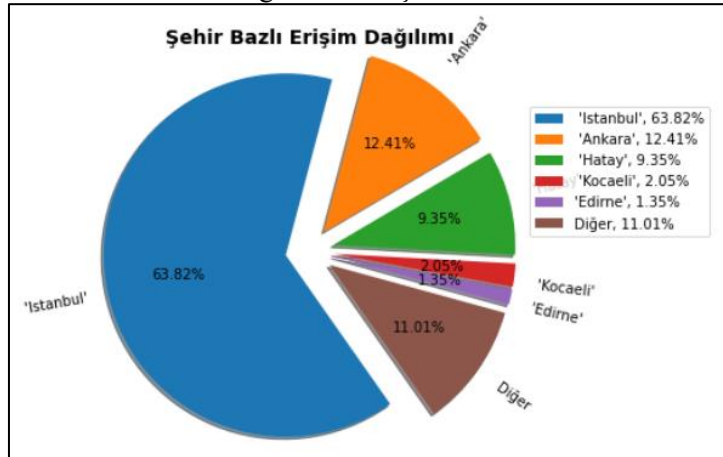
Grafik 4. Erişimde kullanılan web browser

- ÖYS'ne ülke odaklı erişimler araştırıldığında 12 farklı ülkeden katılımın olduğu en yoğun katılımın %98,46 ile Türkiye'den sonrasında ise Amerika, Mısır, Rusya, Romanya'da olduğu bulunmuştur.



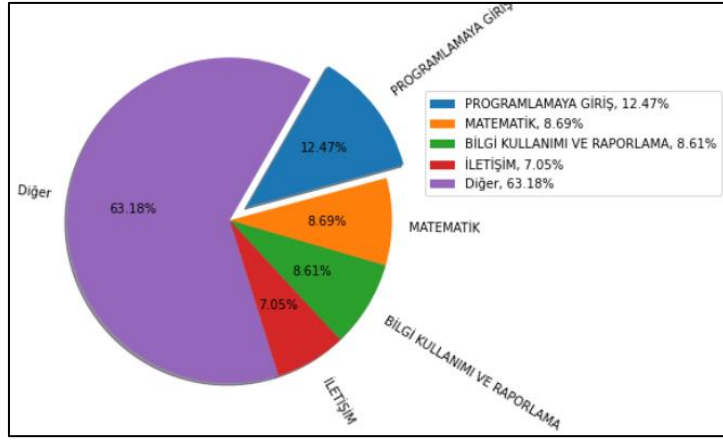
Grafik 5. Ülke bazlı erişim

- ÖYS'ne il odaklı erişimler araştırıldığında erişimin 69 farklı ilden yapıldığı ve en yoğun katılımın %63,82 ile İstanbul sonrasında %12,41 ile Ankara olduğu bulunmuştur.



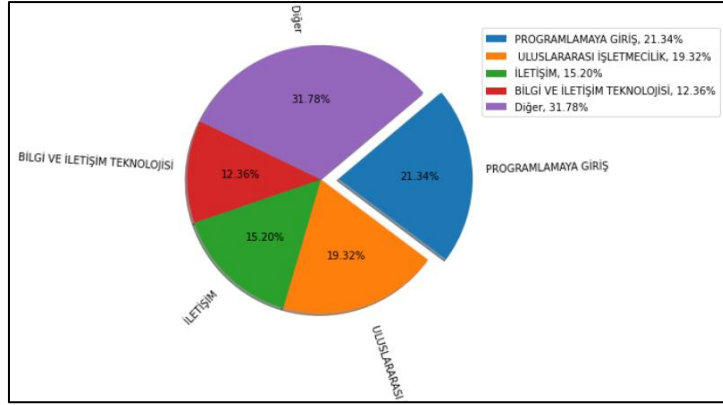
Grafik 6. Şehir bazlı erişim

- Derslerin erişim yoğunluğu araştırıldığında en yoğun ziyaret edilen dersin %12,47 oranı ile Programlamaya Giriş dersi olduğu ve hemen arkasından ikinci sırada %8.69 oranı ile Matematik dersinin olduğu bulunmuştur.



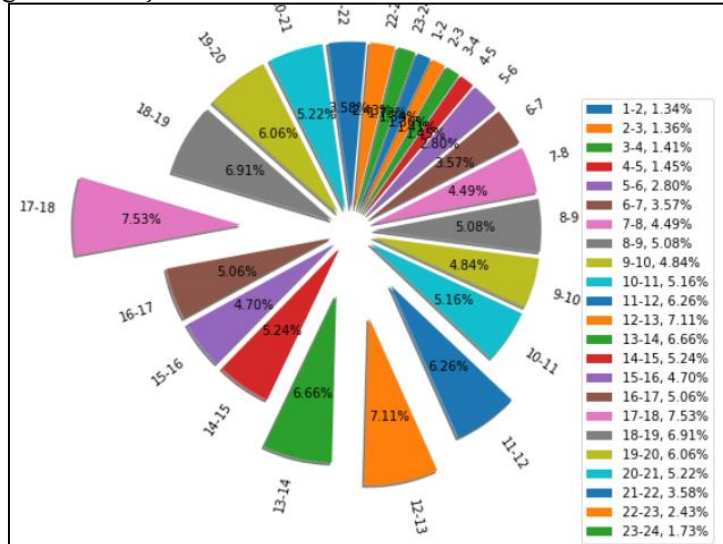
Grafik 7. Ders bazlı ziyaretçi yoğunluğu

- Derslerin içeriklerinin indirilmesi yoğunluğu araştırıldığında birinci sınıf derslerinin daha yoğun ziyaret edilip içeriklerinin indirildiği bulunmuştur. En çok içerik indirilen dersin %21,34 oranı ile Programlamaya Giriş dersi olduğu bulunmuştur.



Grafik 8. Ders bazlı doküman indirme yoğunluğu

- Öğrencilerin ÖYS'ne erişimleri 24 saatlik zaman diliminde incelendiğinde en yoğun erişimin %7,53 ile 17:00-18:00 arasında olduğu, ikinci sırada %7.11 oranı 12:00-13:00 üçüncü sırada ise %6,64 oranı ile 13:00-14:00 saatleri olduğu bulunmuştur.



Grafik 9. Saat bazlı tekil ziyaretçi yoğunluğu

- Öğrencilerin ÖYS'ne erişimleri günlük dilimde incelendiğinde öğrencilerin %94,20'sinin sisteme pazartesi günü erişmeyi tercih ettiği, en az ziyaretin yapıldığı günün ise %27,54 ile cuma günü olduğu bulunmuştur.

Tablo 3.
Günlük bazlı öğrenci katılım yoğunluğu

Gün	Yüzde
Pazartesi	%94.20
Salı	%33.82
Çarşamba	%30.43
Perşembe	%34.30
Cuma	%27.54
Cumartesi	%28.99
Pazar	%34.30

Tablo 4.
Günlük bazlı genel internet trafiği yoğunluk dağılımı

Gün	Yüzde
Pazartesi	%28.35
Salı	%14.10
Çarşamba	%11.29
Perşembe	%11.53
Cuma	%10.05
Cumartesi	%10.75
Pazar	%13.94

• Öğrencilerin ÖYS'ne erişimleri haftalık bazda incelendiğinde, okulumuz öğrencilerinin %87,44'nün Ocak ayının ilk haftasında sisteme eriştiği, ayrıca bu haftanın ilgili dönemin en yoğun erişiminin gerçekleştiği hafta olduğu bulunmuştur.

Tablo 5.
Haftalık öğrenci katılım yoğunluğu

Ay	Hafta	Yüzde
Ekim	42	%29.47
Kasım	45	%33.33
Aralık	51	%23.67
Ocak	1	%87.44

Tablo 6.
Hafta bazlı internet trafiği yoğunluk dağılımı

Ay	Hafta	Yüzde
Ekim	42	%8.41
Kasım	45	%9.11
Kasım	46	%6.85
Ocak	1	%19.63

• Öğrencilerin ÖYS'ne erişimleri aylık dilimde incelendiğinde, okulumuz öğrencilerinin Ocak ayında %90,34'nün sisteme erişim sağladığı bulunmuştur. En düşük erişim sayısı Eylül ayında olmuştur. Bu dönem içinde Kasım ayındaki erişim yoğunluğunun diğer aylara göre daha fazla olduğu bulunmuştur.

Tablo 7.

Aylık öğrenci katılım yoğunluğu

Ay	Yüzde
Eylül	%25.12
Ekim	%43.96
Kasım	%43.00
Aralık	%34.78
Ocak	%90.34

Tablo 8.

Ay bazlı internet trafiği yoğunluk dağılımı

Ay	Yüzde
Eylül	%10.67
Ekim	%22.98
Kasım	%27.88
Aralık	%18.85
Ocak	%19.63

SONUÇ VE TARTIŞMA

Bu çalışmada uzaktan eğitim öğrencilerinin öğrenme yönetim sistemi kullanım alışkanlıklarının pandas kütüphanesi ile analizi amaçlanmıştır. Bu bağlamda ÖYS sunucusuna ait 2019-2020 Eğitim Öğretim yılı güz dönemi web erişim günlük dosyaları üzerinde web kullanım madenciliği yöntemlerinin kullanıldığı ve istatistiki bilgiler çıkartılmasını sağlayan bir yazılım geliştirilmiştir. Yazılım Python dilinde Pandas kütüphanesinden faydalanılarak geliştirilmiştir. Geliştirilen bu yazılım ile öğrencilerin; Öğrenme Yönetim Sistemine erişim şekilleri, ziyareti gerçekleştirdiği ülkeler, ziyareti gerçekleştirdiği iller, ders erişim yoğunluğu, doküman indirme yoğunluğu, kullanmış olduğu işletim sistemleri, kullanmış olduğu tarayıcılar, kullanmış olduğu cihazlar, aylık erişim yoğunluğu, haftalık erişim yoğunluğu, günlük erişim yoğunluğu, saatlik erişim yoğunluğu konularında veriler elde edilmiştir. Elde edilen veriler üzerinden yapılan analizler neticesinde aşağıdaki sonuçlara ulaşılmıştır.

ÖYS'ne erişimin yoğun olarak ana sayfa üzerinden gerçekleştirilmesi web adresinin akılda kalıcı bir adres olarak seçildiğini doğrulamıştır.

Öğrencilerin ÖYS'ne yoğun olarak bilgisayar üzerinden erişmesi, mobil cihaz kullanımının daha düşük bir oranda kalması ÖYS'nin mobil uygulama ile desteklenmesi halinde öğrencilerin zaman ve mekan kısıtlamasından bağımsız olarak sistemi daha fazla kullanabileceği sonucunu düşündürmüştür.

Öğrencilerin ÖYS'ne erişimde yaygın olarak Windows işletim sistemini kullanması ülkemizde bu işletim sisteminin daha yaygın olarak kullanılmasından kaynaklandığını düşündürmüştür.

Erişimde yaygın olarak kullanıldığı belirlenen Google Chrome ve Firefox web görüntüleme programlarının öne çıkmalarının sebebinin sahip oldukları performans ve eklentilerinin rakiplerine göre daha üstün olmasından kaynaklandığı düşünülmektedir.

Öğrencilerin kayıtlarda belirttikleri adresleriyle örtüşmeyecek şekilde ÖYS'ne farklı ülkelerden erişim sağlanmaları internet bağlantılarında VPN sistemlerini kullandıklarını düşündürmektedir.

En yoğun takip edilen dersin bölümün ana derslerinden ve birinci sınıf dersi olan Programlamaya Giriş dersi olması, ayrıca bu derse ait dokümanların da sistemden en çok indirilen doküman olması öğrencilerin bölüm derslerine ilgilerinin daha yoğun olduğunu düşündürmüştür.

Öğrencilerin ÖYS'ne en yoğun olarak eriştikleri dönemlerin sınav dönemleri olması, öğrencilerin içerikleri indirmek ve tekrar etmek için sınavlara yakın tarihleri tercih ettiklerini göstermiştir. Erişimde en yoğun günün pazartesi olması yeni ders içeriklerinin yayınlandığı gün olmasından kaynaklandığı düşünülmektedir. Arıcı ve Çiftçi'nin (2007) yaptığı çalışmada, öğrencilerin ÖYS sistemine katılım yoğunluğunun dönem sonu sınavlarını kapsayan ay ve haftalarda daha yoğun olduğunu bulmuşlardır. Buldukları bu sonuçlar bizim çalışmamızla örtüşmektedir. Günlük bazda ziyaretçi trafiği üzerine yaptıkları analizde ise en yoğun günün Pazar, trafiğin en az olduğu günün ise Cuma olduğunu bulmuşlardır. (Arıcı & Çiftçi, 2007) Bu sonuçlardan ziyaretçi trafiğinin en az olduğu gün bizim çalışmamız ile örtüşmekte iken en yoğun ziyaret trafiğinin olduğu gün bilgisi farklılık göstermektedir. Bunun nedeni yukarıda da bahsedildiği gibi dersin yeni içeriklerinin pazartesi yayınlanıyor olmasıdır. Çalışmamızda ziyaretçi trafiğinin en yoğun olduğu ikinci gün ise Pazardır.

Öğrencilerin sisteme erişimleri öğrenci bazlı, 24 saatlik zaman dilimi aralığında incelendiğinde ÖYS'nin 6-9, 11-14, 17-20 saatleri arasında yoğun kullanıldığı görülür. Bu durum bize, öğrencilerimizin yaklaşık bir kısmının çalışıyor olabileceğini ve bu saatleri ders çalışarak kendileri için verimli hale getirdiklerini düşündürmektedir.

Çalışmanın sonucunda, öğrencilerimizin ÖYS'ni kullanım alışkanlıkları hakkında önemli bilgiler elde edilmiştir. Çalışmanın başından, ham verinin elde edilmesinden, analizine kadar olan tüm süreçte, kullanılan Pandas kütüphanesinin, sorgu betikleri ile sağladığı esneklik ve sonuçlandırma performansının tatmin edici olduğu görülmüştür. Pandas, veri ile çalışan, programlama geçmişi olan veya programlamaya meraklı araştırmacılar tarafından öğrenilmesi ve kullanılması kolay güçlü bir kütüphanedir.

Çalışmanın eğitim alanlı web kullanım madenciliği çalışmalarına katkı sağlayacağı düşünülmektedir. Bir sonraki çalışmanın, pandemi sürecinin uzaktan öğretim öğrencilerin ÖYS'ni kullanım alışkanlıklarına etkisi üzerine yapılması planlanmaktadır.

KAYNAKÇA

- Arıcı, N., & Çiftçi, S. (2007, Ekim). Bilgisayar Destekli Uzaktan Eğitimde Öğrenci Log Kayıtlarının İncelenerek Analiz Edilmesi. *Natural and Applied Science Computer Engineering*.
- Ayer, V., Miguez, S., & Toby, B. (2014, Aralık 18). Why scientists should learn to program in Python. *Powder Diffr.*, Vol. 29, No. S2, s. 48-64.
- Bantilan, N. (2020, Ocak). Pandera: Statistical Data Validation of Pandas Dataframes. *Proc of the 19th Python in Science Conf. (SCIPY 2020)*, (s. 116-124).
- Baykal, A., & Coşkun, C. (2009). Web Madenciliği Teknikleri. *Akademik Bilişim'09-XI. Akademik Bilişim Konferansı*, (s. 797-800). Şanlıurfa, Bildirileri Harran Üniversitesi.
- Baykara, M., Daş, R., & Tuna, G. (2016). Web Sunucu Erişim Kütüklerinden Web Ataklarının Tespitine Yönelik Web Tabanlı Log Analiz Platformu. *Fırat Üniv. Müh. Bil. Dergisi Science and Eng. J of Fırat Univ.*, 291-302.
- Burçak, K. (2012). Kırıkkale Üniversitesi Web Sitesinin Kullanıcı Örüntülerinin Web Madenciliği ile Analizi, Yüksek Lisans Tezi. Kırıkkale.
- Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web Mining: Information and Pattern Discovery on the World Wide Web. *9th IEEE International Conference on Tools with Artificial Intelligence*, (s. 558-567).
- Çınar, I., & Şakir, H. (2016, Mayıs). Web Madenciliği Yöntemleri ile Web Loglarının İstatistiksel Analizi ve Saldırı Tespiti. *Bilge Bilişim Teknolojileri Dergisi*, Cilt: 9, SAYI: 2 ISSN: 1307-9697 / 2147-0715, s. 125- 135.
- Çınar, I., Çınar, M. S., & Bilge, H. Ş. (2014). Web Sunucu Loglarının Web Madenciliği Yöntemleri ile Analizi Akademik. *Bilişim'14- XVI. Akademik Bilişim Konferansı Bildirileri Mersin Üniversitesi*, (s. 417-427). Mersin.
- Enache, M. C. (2019). Data Analysis with Pandas. *Annals of "Dunarea de Jos" University of Galati Fascicle I. Economics and Applied Informatics Years XXV – no2/2019 ISSN-L 1584-0409 ISSN-Online 2344-441X www.eia.feaa.ugal.ro*, s. 69-74.
- Gezer, M., Erol, C., & Gülseçe, S. (2007). Bir web sayfasının web madenciliği ile analizi. *Akademik Bilişim 2007, Dumlupınar Üniversitesi. Kütahya*.
- Hagedorn, S., Kläbe, S., & Sattler, K.-U. (2021, Kasım 25). Putting Pandas in a Box. http://cidrdb.org: http://cidrdb.org/cidr2021/papers/cidr2021_paper07.pdf adresinden alındı
- Hummel, J. (2016, Ekim 12). Gadfly: A pandas-based Framework for Analyzing GADGET Simulation Data doi:10.1088/1538-3873/128/969/114503. *Publications of the Astronomical Society of the Pacific*, 128:114503 (6pp), s. 1-6.
- Karasar, N. (2015). Bilimsel araştırma yöntemi. Ankara: Nobel Akademik Yayıncılık.

- Kashnitsky, Y. (2021, Kasım 25). Topic 1. Exploratory Data Analysis with Pandas. <https://www.kaggle.com/>: . <https://www.kaggle.com/kashnitsky/topic-1-exploratory-data-analysis-with-pandas> adresinden alındı
- Mckinney, W. (2010). Data Structures For Statistical Computing In Python. Proc of the 9th Python in Science Conf. (SCIPY 2010), (s. 56-61).
- Özbay, Ö., & Ersoy, H. (2017). Öğrenme Yönetim Sistemi Üzerindeki Öğrenci Hareketliliğinin Veri Madenciliği Yöntemleriyle Analizi. Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 37(2), 523-558.
- Özseven, T., & Düğenci, M. (2011). Malatya Log Preprocessing: Web Kullanım Madenciliği Ön İşlem Aşaması Uygulama Yazılımı. Akademik Bilişim'11- XIII. Akademik Bilişim Konferansı, (s. 143-150). Malatya.
- Patchaiammal, P., Sundar, G., & Thirumalaiselv, R. (2021). A Large-Scale Study Of Fault Feature Extraction From Github Repository Using Data Science Techniques. Turkish Journal of Computer and Mathematics Education Vol.12 No.10, 2092-2103.
- Sodhia, P., Awasthib, N., & Sharmac, V. (2019). Introduction to Machine Learning and Its Basic Application in Python. Proceedings of 10th International Conference on Digital Strategies for Organizational Success.
- Stančin, I., & Jovic, A. (2019). An overview and comparison of free Python libraries for data mining and big data analysis. 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), (s. 20-24). Opatija, Croatia.
- Sumali, M. (2020). College Cost-Benefit Analysis Using Linear Regression Analysis, Pandas, and Seaborn. National Medicines Symposium 2020, (s. 122-157).
- Uladi, A. İ. & Arı, E. S. (2023). Büyük Veri, Büyük Veri Analizi ve Uygulama Alanları. Yönetim Bilişim Sistemleri Dergisi, 9 (1) , 1-14 . Retrieved from <https://dergipark.org.tr/tr/pub/ybs/issue/78203/1034655>
- Yavuzarslan, M., & Erol, Ç. (2022). Öğrenme Yönetim Sistemi Log Kayıtlarının Akademik Başarı Tahmininde Kullanılması. Bilişim TeknolojiLeri Dergisi, 15(5).
- Yıldız, A. (2022). Büyük Veri'nin V'leri ve Veri Analitiği. Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Yıl 2022 Sayı 51: Özel sayı 1 , 361-378 . DOI: 10.30794/pausbed.1117208