

Dünya İnsan Bilimleri Dergisi

World Journal of Human Sciences

ISSN: 2717-6665

Ocak 2024

Cilt/Volume: 2024

Sayı/Issue: 1

İnsanların Gelir Düzeyleri ile Enflasyon Artışından Kaynaklı Harcama Alışkanlıklarının Veri Madenciliği Algoritmaları ile Analizi

Analysis Of People's Income Levels and Expenditure Habits Due to Increasing Inflation with Data Mining Algorithms

Mehlika Kocabaş Akay*
Muhammed Murat Çiftçi**
Meral Şahin Günkut***

Özet

Günümüzde, satın alma kararları veri madenciliğini aktif yönetebilen alt yapıların yönlendirmesi ile etkilenmektedir. Artık göz bebeklerinin hareket etmesi bile izlenerek veriler toplanmakta ve işlenmektedir bu da teknoloji sayesinde veri madenciliğinin önemini arttırmıştır. Teknolojinin ivmesi sayesinde insanların bilgisayar ortamında yaptığı işlemlerin sayısı her geçen gün katlanarak artmaktadır. Bu işlemler çokça veriyi oluşturmaktadır. Veriler veri tabanları veya veri ambarlarında anlamsız olarak depolanmaktadır. Yığın halde bulunan verileri çeşitli algoritmalar ile analiz edip işe yarar bilgiler üretmektedir. Özellikle satınalma kararlarını da etkilediği için firmalar için kritik öneme sahiptir. Bu çalışmada artan enflasyona göre insanların harcamalarındaki yönelim veri madenciliği kümeleme algoritmalarından K-

* **Sorumlu Yazar/Corresponding author:** Dr.Öğr.Üyesi, Kocaeli Üniversitesi, mehlika.kocabas@kocaeli.edu.tr, ORCID ID: 0000-0003-0564-4625.

** Uzman, Kocaeli Üniversitesi, ciftcimmurat23@gmail.com, ORCID ID: 0000-0002-7178-758X.

*** Uzman, Körfez Milangaz Hacer Demiröen Mesleki ve Teknik Anadolu Lisesi, meralmary@gmail.com, ORCID ID: 0000-0001-7023-0803.

Atıf/Citation: Kocabaş Akay, M., Çiftçi, M.M. & Şahin Günkut, M. (2024). İnsanların Gelir Düzeyleri ile Enflasyon Artışından Kaynaklı Harcama Alışkanlıklarının Veri Madenciliği Algoritmaları ile Analizi. *Dünya İnsan Bilimleri Dergisi*. 2024 (1), 32-47.

Geliş/Received: 21.02.2023

Kabul/Accepted: 25.12.2023

Means ile analiz edilmiştir. Bu analiz dijital ortamda hazırlanan form üzerinden katılımcılardan çeşitli bilgiler alınarak hazırlanmıştır. Çalışma kapsamında elde edilecek bulgular, içinde bulunulan enflasyon ortamında firmalara yol gösterici nitelik taşımaktadır.

Anahtar Kelimeler: Alım Gücü, Kümeleme, Sınıflandırma, Veri Madenciliği ve Modelleri.

Abstract

Nowadays, purchasing decisions are affected by the policies of companies that actively use data mining. It is now possible to collect and process data by even monitoring the movement of the pupils, increasing the impact of technology on decision-making. Thanks to the acceleration of technology, the number of transactions people perform on computers is increasing day by day. This multiprocessor creates the data. Data is stored meaninglessly in databases or data warehouses. Collected data is analyzed with various updates and produces useful information. It is especially critical for companies to make purchasing decisions. According to this increasing inflation, people's expenditures were analyzed with K-Means, one of the data mining clustering products. Various information is brought together through the digitalized form of this analysis. The findings to be obtained within the scope of the study show that they will guide companies in the current inflation environment.

Keywords: Data Mining And Models, Clustering, Classification, Purchasing Power.

Giriş

Sanayi devriminin başlangıcı sayılan Endüstri 0 noktasından günümüze teknoloji ivmelenerek hızla hayata dahil olmuştur. İnsanlar yaptıkları birçok işlemi dijitalde gerçekleştirmektedir. Verilerin dijitalde saklanmaya başlamasıyla birlikte verilerin saklanması ve işlenmesi daha kolay hale gelmiştir. Her geçen gün bu veriler katlanarak artmaya devam etmektedir. Bu büyük veriler veri tabanı adı verilen yapılarda saklanmaktadır. Veri tabanındaki veriler tek başına değersizdir. Kullanıcı için bir anlam ifade etmez. Bu veri tabanındaki veri sistematik bir şekilde belirli amaca yönelik analiz edilirse eğer çok değerli bilgilere ulaşılabilir. Veri madenciliği astronomi, biyoloji, finans, pazarlama, sigorta, sağlık gibi birçok alanda uygulanmaktadır (Özekes, 2003).

Ekonomik koşullar insanların satın alma kararlarını doğrudan etkilemektedir. Dünyanın enflasyon ve krizlere yenik düştüğü dönemlerde satın alma alışkanlıklarının değişiklik göstermesi olasıdır. Enflasyon ortamlarında gelir düzeyleri arasında oluşan farklılıkların da alışveriş alışkanlıklarını etkileyeceği öngörülmektedir.

Bu çalışma kapsamında, enflasyon ortamında gelir düzeyi farklılıklarının alışveriş alışkanlıklarını etkileyip etkilemediği veri madenciliği yöntemleri kullanılarak araştırılmıştır. Çalışmada gelir gruplarına göre alışkanlıkların değişip değişmediği araştırıldığı için veri madenciliği yöntemlerinden kümelemeye yarayan K-Means algoritması kullanılmıştır. IBM SPSS Modeler Subscription paket programı ile veriler modellenmiş ve K-Means algoritması kullanılarak yapılan modelleme güvenilir çıkmıştır.

Veriler anket yöntemi toplanmıştır. Çevrimiçi tarama yöntemleri kullanılarak gerçekleştirilen anket herkes tarafından erişilecek şekilde tasarlanmış ve çeşitli sosyal medya araçları ile yayılması sağlanmıştır. Bu nedenle katılım gönüllülük esaslı olup rassaldır. Anketi 234 kişi cevaplamıştır. Katılımcıların %59'u erkek, %41'i kadındır. Katılımcıların %88,5'i genç nüfustan (18-25 yaş aralığı) oluşmaktadır. Katılımcıların %20'si tam zamanlı olarak yüksek gelir ile çalışmaktadır.

Ticari hayat ancak alışverişin aktif olması ile mümkündür ve tüm girişimlerde amaç para kazanmaktır. Bu nedenle hedef kitlenin dönemselsel davranışları kritik öneme sahiptir. Bu çalışma farklı sektörlerle hizmet edebilme özelliğindedir ve özelleştirilerek kullanılması için bir örnek teşkil etmektedir. Dört bölümden oluşmaktadır, birinci bölümde veri madenciliği tanıtılmış ve algoritmalarından bahsedilmiştir. İkinci kısımda detaylı bir literatür taramasına yer verilmiştir. Literatür taraması yöntemin uygulamalarının çeşitlendirilmesi üzerine kurgulanmıştır. Üçüncü kısımda çalışmanın malzeme ve metotlarından bahsedilmiştir. Son olarak ise sonuç ve öneriler verilmiştir.

1. Veri Madenciliği

1.1. Veri Madenciliği Nedir

Depolanan verilerin artarak veri yığını oluşturması ve her geçen gün daha karmaşık hale gelmiştir. Bu veriler arasındaki ilişkiyi analiz etmek için geleneksel yöntemler yetersiz kalmıştır. Yeni yöntemlerin ortaya çıkması ihtiyacı oluşturmuştur. İstatistik, bilgisayar veri tabanı teknolojileri ve diğer disiplinlerin de birleşmesiyle veri madenciliği ortaya çıkmıştır (Özbay,2015:262-272).

Yığın halindeki verileri analiz edip anlamlı ve yararlı bilgiler haline dönüştüren veri madenciliği, bilgisayar programlarının kullanılarak anlamlı örüntülerin ortaya çıkarılması işlemidir. Büyük miktardaki verinin analiz edilip daha önceden bilinmeyen bir bilginin ortaya çıkarılması sürecidir. Veri madenciliği aşağıdaki resim 1.1 de olduğu gibi çeşitli disiplinlerini içerisinde barındırır (Özbay, 2015: 262-272).



Şekil 1: Veri Madenciliğini Oluşturan Yapılar

Veri madenciliğinin amacı önceden bilinmeyen verilerin analizler sonucunda tahminlenmesi ve ona uygun hareket edilmesini sağlamaktır. Mevcuttaki yapıların eksikliklerinin, aksaklıklarının tespit edip giderilmesi ve daha verimli çalışmak için gerekenlerin yapılmasını sağlamaktır. Bu yüzden çokça geniş uygulama alanı bulunmaktadır.

Temel olarak veri madenciliği bilgisayar destekli bir bilgi çözümlene işlemidir. Veri madenciliği, eldeki verilerden çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasıdır. Bu işlem kümeleme, veri özetleme, değişikliklerin analizi, sapmaların tespiti gibi belirli sayıda teknik yaklaşımları içerir. Başka bir deyişle, veri madenciliği, verilerin içerisindeki örüntülerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların keşfedilmesidir. Veri madenciliğinde amaç, kolaylıkla mantıksal kurallara ya da görsel sunumlara çevrilebilecek nitel modellerin çıkarılmasıdır. Bu bağlamda, veri madenciliği insan merkezlidir ve bazen insan – bilgisayar ara

yüzü birleştirilir. Veri madenciliği sahası, istatistik, makine bilgisi, veri tabanları ve yüksek performanslı işlem gibi temelleri de içerir (Özby, 2015: 262-272).

Veri madenciliği tanımlarından da anlaşılacağı gibi istatistik, makine öğrenmesi, veri tabanı yönetimi, görselleştirme gibi alanlardan faydalanan disiplinler arası bir alandır. İşlenmemiş veriden, son kullanıcının kolayca anlayıp karar alma sürecine dahil edebileceği bilgiyi oluşturana kadar geçen tüm süreci kapsayan bir yöntem olmasından, hipotez doğrulamaya yönelik değil yeni, gizli örüntüler bulmaya yönelik bir alan olmasından ve çok çeşitli teknikleri aynı uygulama içinde kullanabilmeye olanak sağlamasından dolayı veri madenciliği kullanıcılarına kendisini oluşturan makine öğrenmesi, istatistik matematik gibi yöntemlerden daha farklı bir perspektif suna (Özçınar, 2006).

Örgütler az veya çok milyarlarca byte uzunluğunda, kullanıcıları, süre giden işleri, hizmet ve ürünleri, çalışanları vb. hakkında veri üretirler. Ancak bu veriler bilgisayar sistemlerinin içinde her geçen gün ulaşılması daha da zor bir hal alarak gömülür gider. Uzmanlar ise, mevcut verilerden elde edilmiş, işlenmiş ve depolanmış sadece küçük bir parça veriye dayanarak değerlendirme yapabilmektedirler. Örgüt bünyesinde toplanan verilerin yalnız raporlama amaçlı kullanılması, yöneticilere yeterince karar desteği sağlanamamasına neden olmaktadır (Savaş, S., Topaloğlu, N., & Yılmaz, M., 2012: 1-23).

1.2. Veri Madenciliği Süreci

1. Veri temizleme
2. Veri bütünleştirme
3. Veri indirgeme
4. Veri dönüştürme
5. Veri madenciliği algoritmasının uygulanması
6. Analiz sonuçlarının değerlendirilmesi (Coşlu, 2013: 23-25)

Veri temizleme: Veri ambarında yer alan hatalı ve tutarsız veriler gürültü olarak adlandırılır. Bu gürültülü verilerin, eksik verilerin, atılabilir. Kayıp veriler yerine diğer verilerin ortalaması girilebilir.

Veri bütünleştirme: Farklı veri tabanlarında yer alan veriler aynı biçime dönüştürülerek analize hazır hale getirilmesidir. Örnek verecek olursak bir veri tabanında erkek olan cinsiyet diğer veri tabanında e ya da 0 ile 1 rakamlarıyla temsil edilebilir.

Veri indirgeme: Eğer analizin sonucunun değişmeyeceğine inanılıyorsa yüksek boyuttaki verinin belirli bir kısmının alınmasıdır.

Veri dönüştürme: Verilerin kullanılacak algoritmaya uygun içeriğini koruyarak

dönüştürülmesi işlemidir.

Veri madenciliği algoritmasının uygulanma: Veri hazır hale getirildikten sonra analiz programı yardımıyla uygun algoritma seçimi verilere uygulanır.

Analiz sonuçlarının değerlendirilmesi: Veriler analiz edildikten sonra sonuçlar düzenlenilip ilgili yerlere sunulur.

Veri Madenciliği Yöntemleri

1. Sınıflandırma
2. Kümeleme
3. Birliktelik kuralı

Veri madenciliği yöntemleri tahmin edici ve tanımlayıcı olmak üzere iki yönlüdür. Sınıflandırmada amaç sonucu tahmin etmektir. Kümeleme ile birliktelik kurallarında ise amaç karar verme aşamasında rehberlik edecek bilginin tanımlanmasıdır (Özbay,2015:262-272).

1.3. Sınıflandırma

Sınıflandırma veri madenciliğinin en çok kullanılan yöntemlerindedir. Veri tabanından belirli bir kısmı kullanılarak kurallar oluşturulur. Bu kurallar kullanılarak ortaya çıkacak durumlarda nasıl karar verileceği belirlenir (Osman,2019).

Sınıflandırmada kullanılan başlıca teknikler aşağıdaki gibidir (Özekes,2003).

1. Karar ağaçları
2. Yapay sinir ağları
3. Genetik algoritmalar
4. K – en yakın komşu
5. Bellek temelli nedenleme
6. Naive – Bayes

En çok kullanılan teknik karar ağacıdır. Yorumlanmasının rahat olması, kolayca entegre edilip güvenli sonuçlar vermesi iyi olmasının nedenlerindedir. Adından da anlaşılacağı üzere karar ağacı ağaca benzer bir şemadır. Dğümler, dallar ve yapraklardan oluşur (Han, J.-Kamber, M., Morgan Kaufmann Publishers,2000). Eğer bir dalın ucunda sınıflandırma gerçekleşiyorsa yaprak gerçekleşmiyorsa düğüm oluşmaktadır. Karar ağacı oluşturmak için birçok yöntem geliştirilmiştir. Bunlardan temel olanları aşağıdaki gibidir.

1. Entropiye dayalı algoritmalar
2. Sınıflandırma ve resesyon araçları
3. Bellek tabanlı sınıflandırma modelleri

1.4. Kümeleme

Kümeleme, veriyi sınıflara veya kümelere ayırma işlemidir (Han E.-Kumar V, 1999: 68-75). Aynı kümelerin elemanları birbirine benzerden farklı kümelerin elemanlarından farklıdır. Sınıflama modelinde olan veri sınıfları kümelemede yoktur. Sınıfı olmayan verilerde kümelemede dahil edilir. Bu yüzden bazı uygulamalarda kümeleme yapıldıktan sonra sınıflandırma algoritmaları kullanılmaktadır. Örnek verecek olursak marketlerdeki müşterilerin gruplanması örneği verilebilir. Veri tabanlarındaki verilerin her geçen gün artmasına orantılı olarak kümelemede önem kazanmaktadır. Literatürde pek çok kümeleme algoritması bulunmaktadır. Genel olarak kümeleme yöntemleri aşağıdaki gibi sınıflandırılabilir (Özekes, 2003)

1. Bölme yöntemleri
2. Hiyerarşik yöntemler
3. Yoğunluk tabanlı yöntemler
4. Izgara tabanlı yöntemler
5. Model tabanlı yöntemler

Bölme yöntemleri, n adet nesneyi k adet kümeye bölmeye yardımcı olur. En iyi bilinen ve en çok kullanılan kümeleme bölme yöntemi K – Means yöntemidir. Bu yöntemde ilk olarak n adet nesneden k adet nesne seçer. Bu nesnelerin her biri bir kümenin merkez noktasını temsil eder. Bu merkez noktaya yakınlıklarına göre nesnelere kümelere ayrılır. Bu işlem tüm nesnelere kümelere ayrılıncaya kadar devam eder.

1.5. Birliktelik Kuralları

Veri tabanı içerisindeki verileri analiz edip bu verilerin birbirleriyle olan ilişkisini inceleyen ve aynı zamanda gerçekleştirebileceklerini ortaya koyan yöntemdir. Pazarlama alanında çokça kullanılır. Bunlara aynı zamanda pazar sepet analizi de denmektedir. Örneğin marketten elma, armut alan kişilerin yüzde kaç olasılıkla yeni bir ürün alabileceği sonucunu ortaya koyar. Müşterilerin segmentlere ayrılmasını ve onlara özel teklifler sunulması için bilgiler üretir.

En sık kullanılan algoritması Apriori algoritmasıdır. Destek ve güven ölçütlerini belirlemek için gerekli formüller kullanılıp bu değerler verilerden elde edilir. Elde edilen destek aralığında bize seçilenin yanında güven oranı değeri kadar seçilecek olanının bilgisini verir.

2. Literatür İncelemesi

Çalışmanın bu kısmında literatürde yer alan veri madenciliği çalışmaları incelenmiş ve farklı alanlarda kullanımının mümkün olduğu ve yöntemlerin kapsayıcı olduğuna değinilmiştir.

Özekeş, veri madenciliği tekniklerini eğitim alanında kullanmıştır. Eğitim alanında öğrencilerin verileri analiz edilip başarılı ve başarısız olmalarındaki faktörlerin tespiti, çalışma ortamlarındaki aksaklıkların tespiti ve aralarındaki örüntü durumların incelenmesi veri madenciliği ile yapılmaktadır.

Savaş, çalışmasında Türkiye’de veri madenciliğini karar destek sistemi olarak kullanan alanların tespiti için çalışılmış ve mühendislik alanında, tıp alanında, bankacılık ve borsa alanında, eğitim alanında, ticari alanda, telekomünikasyon alanında olmak üzere 6 alanda incelenmiştir. Kurumların farklı amaçlar doğrultusunda veri madenciliğinin farklı özelliklerinden faydalandıkları ve kimisinin gruplama kimisinin tahminlemede kullandığını belirtmiştir.

Coşlu 2013 yılındaki çalışmasında veri madenciliğinin süreçleri olan veri temizleme, veri bütünleştirme, veri indirgeme, veri dönüştürme, veri madenciliği algoritmasının uygulanması, sonuçlar ve sunumlar açıklanmıştır. Karar ağacı sınıflandırma yöntemi kullanılarak gelir, statü, risk durumu üzerinden örnek çalışma yapılmıştır. Birliktelik kurallarından apriori algoritması açıklanıp örnek verilmiştir.

Koyuncugil ve Özgülbaş (2009), bu çalışmada veri madenciliğinin tıp ve sağlık alanında kullanımı ve uygulamaları konusu işlenmiştir. Verinin oluşum süreci ve veri ambarlarında kullanılmak üzere depolanması, veri tabanı gibi konuları açıklanmıştır. Veri madenciliği ile istatistik arasındaki ilişki açıklanmıştır. Tıp ve sağlık sektöründe elektronik hasta dosyalarının oluşturulmasında, kronik hastalıklar için erken uyarı sinyallerinin tespitinde, laboratuvar testleri için hata ve suistimal tespitinde, sağlıkta kalitenim geliştirilmesinde, yolsuzluk ve fatura suistimallerinin tespitinde, maliyet minimizasyonunda, finansal riskin belirlenmesi için erken uyarı sisteminin geliştirilmesinde, yönetsel karar destek sistemlerinin geliştirilmesi gibi konularda veri madenciliği yöntemlerinden faydalanılabileceği söylenmiştir.

Yurdakul (2015), bu çalışmada lise öğrencilerinin performansları anket verileri ışığında veri madenciliği yöntem ve teknikleri ile analiz edilip incelenmiştir. Öğrenci performansını etkileyen yöneticiden kaynaklı nedenler, öğretmenlerden kaynaklı nedenler, aileden kaynaklı nedenler, çevreden kaynaklı nedenler, akademik kaygıdan kaynaklı nedenler, bireysel nedenlerden bahsedilmiştir. Çalışmada kullanılan çok katmanlı algılayıcı algoritmalar, k en yakın komşu algoritması, J48 algoritması, JRIP algoritması ve kullanılan program tanıtılıp açıklanmıştır.

Özkan ve Boran (2014:59-82), bu çalışmada veri madenciliğinin finansal kararlarda kullanımı konusu incelenmiştir. Finansal tablolarda analiz yapmanın amacı karar alma

mekanizmasının somut verilerle desteklenmesi ve karar almanın kolaylaştırılması olduğu söylenmiştir. Çalışmada kümeleme algoritmalarından OPTICS, DBscan, K – Means, EM algoritmaları kullanılmıştır. Ağırlıklı olarak K – Means ile EM algoritmaları karşılaştırılarak kullanıldığı belirtilmiştir.

Rygielski, Wang ve Yen (2002:483-502), bu çalışmada müşteri ilişkileri yönetimi için veri madenciliği yöntem ve tekniklerinden faydalanılmıştır. Geleneksel pazarlama modellerinde sürekli müşteri kitlesini artırmanın yüksek maliyetlere neden olduğuna değinmiştir. Bunun yerine var olan müşterilerin ihtiyaçlarını tespit edip satış yapmanın önemi vurgulanmıştır. Veri madenciliği yöntemlerinden sınıflandırma ve regresyon analizinin tahminlerde kullanıldığı, birliktelik analizinin davranışı tanımlamak için kullanıldığı söylenmiştir. Mevsimsel trend ürünlerin satışındaki problem CHAID ve sinir ağları yöntemleri açıklanıp kullanılarak sonuçlar ve çözümler üretilmiştir.

Özçınar (2006), bu çalışmada KPSS sonuçlarının veri madenciliği yöntemi kullanılarak tahmin edilmesinden bahsedilmiştir. Çalışmada kullanılan yapay sinir ağları ve regresyon analizi detaylı olarak anlatılmıştır. Elde edilen iki farklı algoritma sonuçları karşılaştırılmıştır. KPSS’de soru çıkan dersler, bu derslere ait not ortalamaları, puanlara ilişkin veriler alınarak analizin gerçekleştirildiği söylenmiştir. SPSS ve WEKA programlarının kullanıldığı söylenmiştir. Analizler sonrası elde edilen sonuçlar tablolarla ifade edilmiştir.

Cengiz ve Topsakal (2014:1-11), bu çalışmada veri madenciliği kullanılarak beton basınç dayanımının belirlenmesi incelenmiştir. Betondan beklenen en temel özellik basınç dayanımı olduğu söylenmiştir. Bu yüzden çeşitli veriler veri madenciliği yöntemleri ile analiz edilip anlamlı bilgiler edinilmesi amaçlanmıştır. Gerekli deneyler sonucunda girdi verileri elde edilmiştir. Bu girdiler karışım, doz, boyut, zaman özelliklerine göre hazırlanmıştır. Veriler WEKA programına yüklenip korelasyon katsayısı en yüksek olan dokuz algoritma belirlenmiştir. Bu dokuz algoritmadan elde edilen sonuçlar tablo halinde sunulmuştur. En yüksek korelasyon katsayısına sahip algoritma K – Star algoritması olarak belirlenmiştir. Elde edilen sonuçlara göre büyükten küçüğe sırasıyla çimento dozajı, numune boyutu, numune alma zamanı, karıştırma metodu olarak tespit edilmiştir.

Literatür taraması ile veri madenciliğinin birbirinden çok farklı alanda farklı şekillerde kullanıldığı gösterilmiştir. Çalışma özelinde yapılan literatür bilgileri ise uygulama ve yöntem anlatımı kısımlarında değinilmiştir.

3. Malzeme ve Yöntem

3.1. Çalışmanın Amacı

Bu çalışmanın amacı, her geçen zamanda artan enflasyona karşı insanların gelir düzeyleri arasındaki ilişkiye göre harcama alışkanlıklarının nasıl değiştiğini ortaya koymaktır. Bunun için rastgele yapılan ankette insanların aldıkları maaşlar, enflasyonun etkisi, harcamalarının eğiliminin nasıl değiştiği ile alakalı veriler toplanmıştır. İnsanlarda artan enflasyonun ve maaş zamlarının etkisi ile tüketim davranışlarının nasıl etkilendiği araştırılmıştır.

3.2. Çalışmanın Sınırları ve Verilerin Toplanması

Çalışmada kullanılmak üzere toplanan veriler form hazırlanarak sosyal platformlar aracılığı ile duyurulmuştur. Verileri dolduran kişiler sosyal platformlardan ilgili anket sayfasına yönlendirilmiştir.

Uygulamada kullanılan verilerimiz 10 Aralık – 20 Aralık 2022 tarihleri arasında 233 kişinin katılımı ile katılımı ile gerçekleştirilmiştir. Forma giriş yaptıktan sonra kişilerin tek seferlik cevap gönderebileceği şekilde ayarlama yapılmıştır. Ankette insanlara 12 farklı soru yöneltilmiştir. 4 soruda katılımcılara aralık belirtilerek sorular sorulmuştur. Bu cevapların karşılıkları tablo 2, tablo 3, tablo 4 ve tablo 5’ te verilmiştir. Soruların tamamı çoktan seçmelidir. Yorumla açık uçlu soru bulunmamaktadır. Ankette yöneltilen sorular ekler kısmında tablo 1’ de gösterilmiştir.

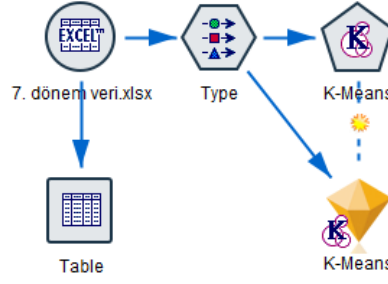
Anket verileri MS Excel dosyası olarak alınmıştır. Gerekli düzenlemeler yapılarak IBM SPSS Modeler Subscription programına aktarılarak analizi yapılmıştır.

3.3. Çalışmanın Yöntemi ve Örneklemi

Çalışmada kullanılan K – Means algoritması kümeleme algoritmalarından birisidir. Elimizdeki verileri belirlemiş olduğumuz k adetteki kümeye ayırır. Aynı kümedeki veriler birbirleriyle benzer niteliğe sahiptir.

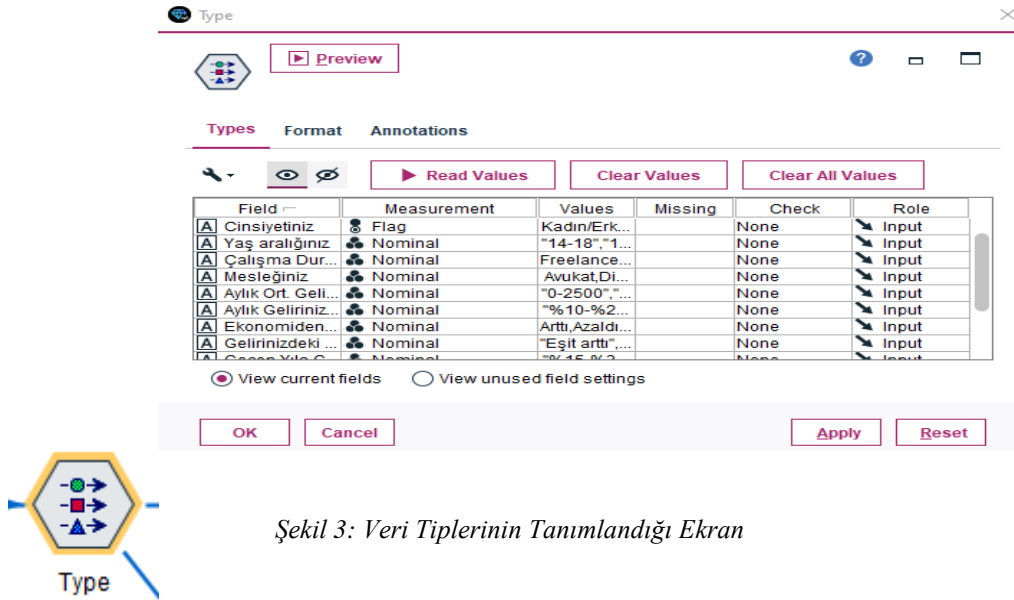
Anket sonucunda elde verilen IBM SPSS programına yüklenerek sonuçlar çıkarılmıştır. Bu sonuçlar K – Means algoritması uygulanarak analiz edilmiştir. Analizde algoritmadan verileri 5 farklı kümeye ayırması istenmiştir.

Aşağıdaki şekil 2 ‘de olduğu gibi model kurulmuştur. Formdan alınan veriler hazırlama aşamasında geçirildikten sonra Excel dosyasına aktarılmıştır. Kaynak olarak Excel dosyası kullanılmıştır.



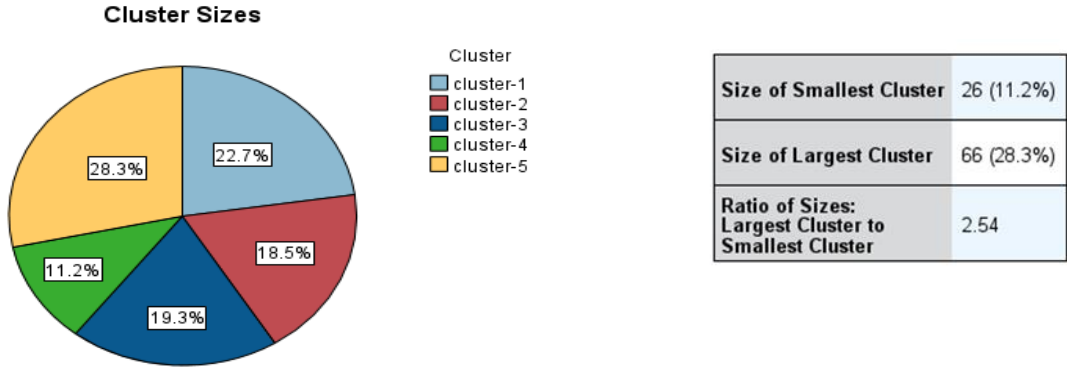
Şekil 2 SPSS'te Kurulan Yapının Görseli

Şekil 3' te olduğu gibi verilerin hepsi K – Means algoritmasına ve Programa uygun hale getirilmiştir. Veri tipi olarak hepsi Input olacak şekilde girilmiştir.



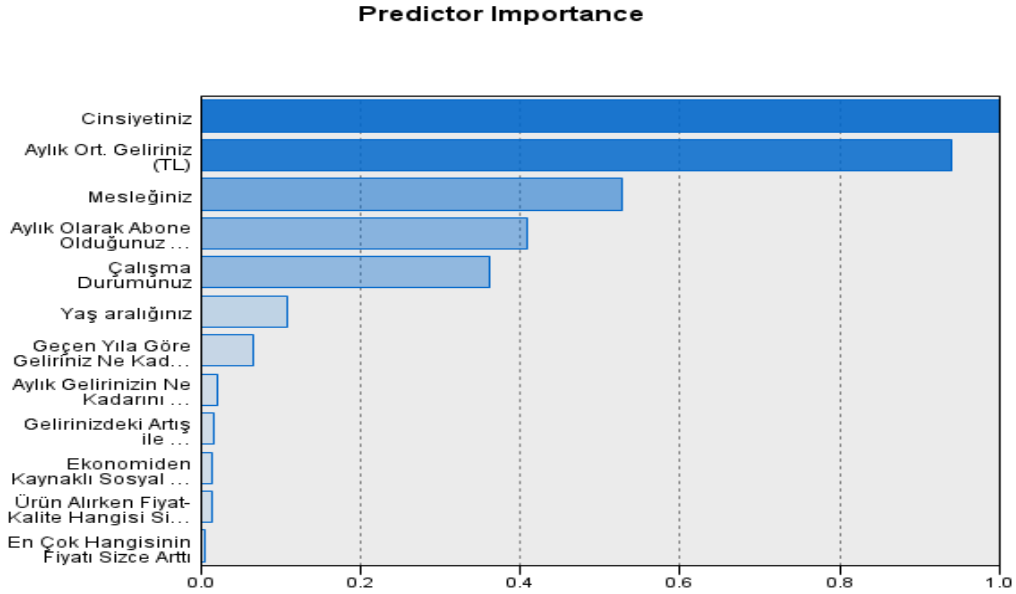
Şekil 3: Veri Tiplerinin Tanımlandığı Ekran

Uygulamada yapılan analizlerin sonucunda istenilen 5 adet kümeye ulaşılmıştır. Bu kümeler sırası ile %22,7, %18,5, %19,3, %11,2, %28,3 büyüklüğe sahiptir. En küçük kümeyi %11,2 oranla dördüncü küme almaktadır. En büyük kümeyi %28,3 oranla beşinci küme oluşturmaktadır. En büyük kümenin en küçük kümeye oranı 2,54' tür.



Şekil 4: Kümelerin Büyüklüklerinin Görseli

Aşağıdaki şekil 5’teki grafikte olduğu gibi en önemli 2 verimizin cinsiyet ve aylık ortalama gelir olduğunu görmekteyiz. Bunlardan sonra gelen diğer önemli veriler ve önem düzeyleri de resimde gösterilmiştir.

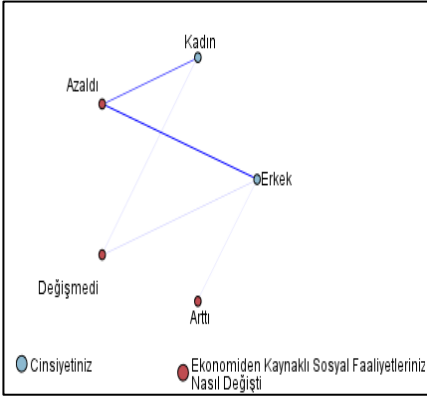


Şekil 5: En Önemli Verilerin Grafiği

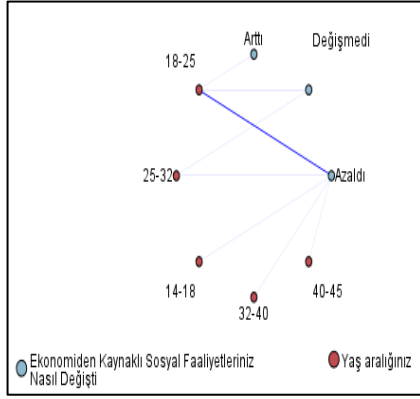
Verilerin analizi sonrasında oluşan kümeler ekler kısmında şekil 15, şekil 16, şekil 17, şekil 18, şekil 19’da olduğu gibidir. En büyük kümeyi %28,3 oranla beşinci küme oluşturmaktadır. Bu kümedeki kişilerin cinsiyeti erkektir. Aylık ortalama gelirleri 0 – 2500 TL arasındadır ve meslekleri öğrencidir. Enflasyondan dolayı aylık sürekli aboneliklerini iptal etmişlerdir. Bu kişiler aktif olarak çalışmıyordur. 18 – 25 yaş aralığındaki bireylerden oluşmaktadırlar. Geçen yıla oranla gelirleri 0- %15 arasında artış göstermiştir. Aylık gelirlerinden 0 - %10 aralığında biriktirebilmektedirler. Gelirlerine oranla giderlerindeki artışın daha yüksek olduğunu düşünmektedirler. Ekonomiden kaynaklı sosyal faaliyetlerini azaltan

kişilerdir. Aldıkları ürünleri değerlendirirken ihtiyaçlarını karşılayan en uygun fiyatlı ürünü tercih etmektedirler. En çok gıda fiyatlarının arttığını düşünmektedirler.

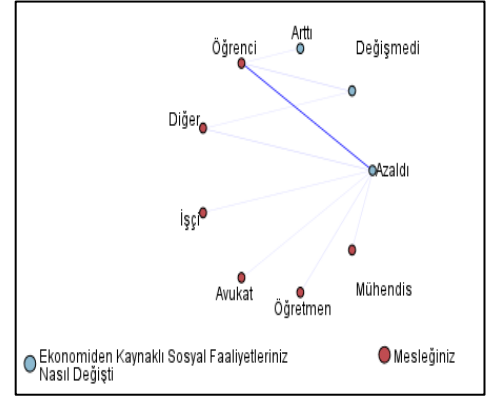
Aşağıdaki şekil 6, şekil 7, şekil 8’ de ekonomiden kaynaklı sosyal faaliyetlerinin cinsiyet, yaş ve meslek arasındaki ilişki matrisini görmekteyiz. İnsanların cevabının azaldı olarak verdiğini ve bu kişilerin 18 – 25 yaş aralığındaki öğrencilerden oluştuğunu söyleyebiliriz.



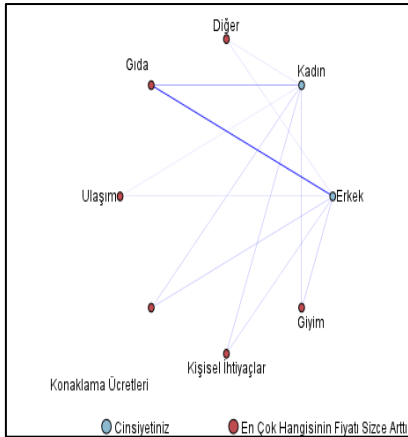
Şekil 11: İlişki Matrisi



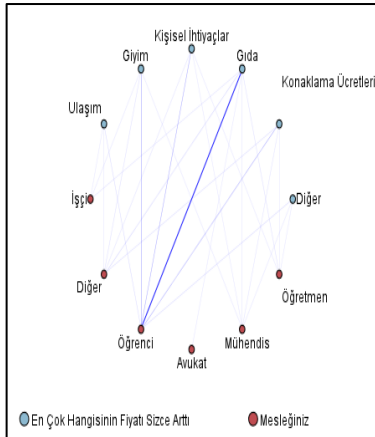
Şekil 11: İlişki Matrisi



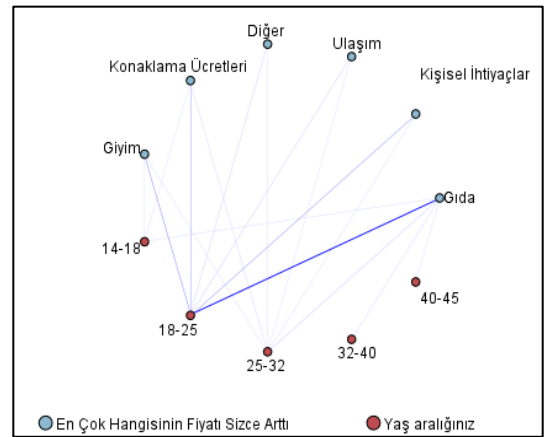
Şekil 11: İlişki Matrisi



Şekil 8: İlişki Matrisi



Şekil 8: İlişki Matrisi

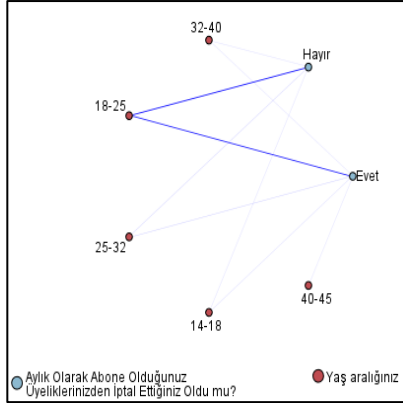


Şekil 8: İlişki Matrisi

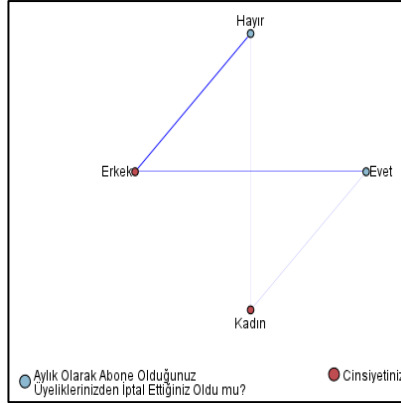
Aşağıdaki şekil 9, şekil 10, şekil 11’ de en çok fiyatı arttığı düşünülen gruplar ile cinsiyet, yaş ve meslek arasındaki ilişki matrisini görmekteyiz. En çok fiyatı arttığı düşünülen grup gıda grubudur. Öğrenciler genellikle gıda, kişisel ihtiyaç ve giyim arttığını düşünmektedir.

Aşağıdaki şekil 12, şekil 13, şekil 14’ te aylık olarak abone olunulan sürekli üyeliklerin iptali ile alakalı cinsiyet, yaş ve meslekler arasındaki ilişki matrisini gösterilmiştir. 18 – 25 yaş aralığındaki kişilerinin, öğrencilerin, erkeklerin cevaplarının birbirine yakın olduğunu cevapların herhangi birinin ağır basmadığını görmekteyiz.

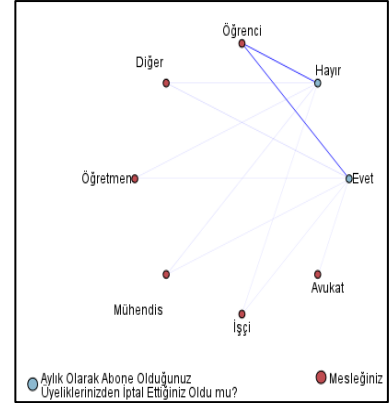
İnsanların Gelir Düzeyleri ile Enflasyon Artışından Kaynaklı Harcama Alışkanlıklarının Veri Madenciliği Algoritmaları ile Analizi



Şekil 14 İlişki Matrisi



Şekil 14 İlişki Matrisi



Şekil 14 İlişki Matrisi

Sonuç

Veri madenciliği elimizde bulunan verileri çeşitli yöntem ve tekniklerle analiz etmemizi sağlar. Bu çalışmada anket ile dijital ortamda alınan veriler SPSS modeler programında K – Means algoritması ile analiz edildi. Analizde veriler 5 farklı kümeye ayrılacak şekilde ayarlandı. En kalabalık kümenin hangisi olduğu tespit edildi. Bu kümedeki kişilerin cinsiyeti erkektir. Aylık ortalama gelirleri 0 – 2500 TL arasındadır ve meslekleri öğrencidir. Enflasyondan dolayı aylık sürekli aboneliklerini iptal etmişlerdir. Bu kişiler aktif olarak çalışmıyordur. 18 – 25 yaş aralığındaki bireylerden oluşmaktadır. Geçen yıla oranla gelirleri 0 - %15 arasında artış göstermiştir. Aylık gelirlerinden 0 - %10 aralığında biriktirebilmektedirler. Gelirlerine oranla giderlerindeki artışın daha yüksek olduğunu düşünmektedirler. Ekonomiden kaynaklı sosyal faaliyetlerini azaltan kişilerdir. Aldıkları ürünleri değerlendirirken ihtiyaçlarını karşılayan en uygun fiyatlı ürünü tercih etmektedirler. En çok gıda fiyatlarının arttığını düşünmektedirler. İnsanların verdiği cevaplardan önemli olanları cinsiyet, yaş aralığı ve meslek kriterlerine göre ikili matrisle incelenmiştir. Bu sayede incelenen verideki insanların harcama alışkanlıkları ile alakalı bilgi edinilmiştir.

Bu çalışmada kullanılan algoritma ile veri setindeki verilerin gruplandırılması yapılabilmeye aynı zamanda o verilerin birbiri ile ilişkileri ikili matrisler kullanılarak yapılabilir. Elde edilecek bilgilere göre bu veri setiyle alakalı aksiyonlar alınıp daha verimli ve nitelikli sonuçlar çıkarılabilir. Örnek verecek olursak öğrenciler gıda fiyatlarının çok arttığını düşünmektedir. Öğrencilerin mutluluğu için gıda fiyatlarını düşürmek için okulda veya okul çevresinde çeşitli çalışmalar yapılarak öğrencilerin refah seviyesi artırılabilir.

Kaynakça

- Akküçük, U. (2011), Sınıflama, K. V. Veri Madenciliği, Yalın Yayınları, 1-123.
- Arslantekin, S. (2003), Veri madenciliği ve bilgi merkezleri. *Türk Kütüphaneciliği*, 17(4), 369-380.
- Cengiz, Ö. Z. E. L., & Topsakal, A. (2014). Veri madenciliği kullanarak beton basınç dayanımının belirlenmesi. *Cumhuriyet Üniversitesi Fen Edebiyat Fakültesi Fen Bilimleri Dergisi*, 35(1), 1-11.
- Coşlu, E. (2013). Veri madenciliği. *Akademik bilişim*, 23-25.
- Han, J.-Kamber, M., Morgan Kaufmann Publishers, Data Mining Concepts and Techniques, 1st Ed., San Francisco, USA, 2000.
- Karypis G. - Han E.-Kumar V, IEEE Computer, Chameleon: Hierarchical Clustering Using Dynamic Modeling, 1999 : 68-75.
- Koyuncugil, A., & Özgülbaş, N. (2009). Veri madenciliği: Tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. *Bilişim Teknolojileri Dergisi*, 2(2).
- Osman, A. S. (2019). Data mining techniques.
- Özbay, Ö. (2015). Veri madenciliği kavrami ve eğitimde veri Madenciliği uygulamaları. *Uluslararası Eğitim Bilimleri Dergisi*, (5), 262-272.
- Özçınar, H. (2006). *KPSS sonuçlarının veri madenciliği yöntemleriyle tahmin edilmesi* (Master's thesis, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü).
- Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları.
- Özkan, M., & Boran, L. (2014). Veri Madenciliğinin Finansal Kararlarda Kullanımı. *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(1), 59-82.
- Rygielski, C., Wang, J. C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in society*, 24(4), 483-502.
- Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.
- Yurdakul, S. (2015). *Veri madenciliği ile lise öğrenci performanslarının değerlendirilmesi* (Master's thesis, Kırıkkale Üniversitesi).