

# Türkiye’de Makine Öğrenmesi ve Karar Ağaçları Alanında Yayınlanmış Tezlerin Bibliyometrik Analizi

Meryem PULAT\* İpek DEVECİ KOCAKOÇ\*\*

## ÖZ

*Bu çalışmada makine öğrenmesi ve karar ağaçları alanında yazılan tezlerin çeşitli parametreler kapsamında değerlendirilerek bibliyometrik analiz yöntemiyle incelenmesi amaçlanmıştır. Bu amaçla Yüksek Öğretim Kurumu'nun Ulusal Tez Merkezi veri tabanında "makine öğrenmesi", "machine learning", "karar ağaçları", "decision tree" anahtar kelimeler kullanılarak 2020 yılı mayıs ayında yapılan tarama sonucu 368 teze ulaşılmıştır. Ulaşılan tezler; yıl, tez türü, dili, üniversite, enstitü, anabilim dalı, sayfa sayısı, kullanılan program, kullanılan yöntem, anahtar kelimeler gibi değişkenler dikkate alınarak bibliyografik açıdan detaylı olarak incelenmiştir. Tezlerle ilişkin belirlenen araştırma soruları dikkate alınarak veriler elde edilerek Excel programına aktarılmıştır ve tüm hesaplamalar, grafikler ve tablolar bu program aracılığıyla yapılmıştır. Yapılan incelemeler sonucunda, denetimli öğrenme stratejisine dayanan makine öğrenmesi ve karar ağaçları alanında yazılmış tezlerle ilişkin önemli bulgular elde edilmiştir. Bunlar; yıllara göre tezlerin dağılımı, alana en çok katkı sağlayan üniversite, enstitü, anabilim dalı, en çok kullanılan makine öğrenmesi yöntemi, kullanılan öznelik seçim yöntemleri, kullanılan parametre optimizasyon yöntemleri, geliştirilen melez yaklaşımlar, geliştirilen topluluk öğrenme yöntemleri, en yaygın programlama dili/yazılım vb. şeklinde önemli bulgulardan oluşmaktadır.*

**Anahtar Kelimeler:** Bibliyometrik Analiz, Tez, Makine Öğrenmesi, Denetimli Öğrenme, Karar Ağaçları

**JEL Sınıflandırması:** C38

## Bibliometric Analysis of Theses Published on Machine Learning and Decision Trees in Turkey

### ABSTRACT

*In this study, it is aimed to examine the theses written in the field of machine learning and decision trees with the bibliometric analysis method by evaluating them within the scope of various parameters. For this purpose, 368 theses were reached as a result of the search performed in May 2020 by using the keywords "machine learning", "decision trees" in the database of the Council of Higher Education's National Thesis Center. Theses reached; It was examined in detail in terms of bibliography taking into account the year, thesis type, language, university, institute, department, the number of pages, the software used, the methods used, the keywords. In view of the research questions defined for the theses, the data were obtained and transferred to the Excel program. In this software, all calculations, graphs and tables were made. As a result of the examinations, important findings were obtained regarding the theses written in the fields of machine learning and decision tree based on the supervised learning strategy. These include important findings such as the distribution of theses by years, the university, institute, department that contributed the most to the field, the most commonly used machine learning method, the feature selection methods used, the parameter optimization*

\* Arş. Gör., Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, meryem.pulat@deu.edu.tr, ORCID Bilgisi: 0000-0003-0642-5619

\*\* Prof. Dr., Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, ipek.deveci@deu.edu.tr, ORCID Bilgisi: 0000-0001-9155-8269

(Makale Gönderim Tarihi: 29.01.2021 / Yayına Kabul Tarihi: 12.05.2021)

Doi Number: 10.18657/yonveek.870190

Makale Türü: Araştırma Makalesi

*methods used, the hybrid approaches developed, the ensemble learning methods developed, the most popular programming language / software, etc.*

**Key Words:** *Bibliometric Analysis, Thesis, Machine Learning, Supervised Learning, Decision Trees*

**JEL Classification:** C38

## GİRİŞ

Bibliyometri terimi, 1969 yılında Alan Pritchard tarafından ortaya atılmıştır (Lawani, 1981: 294). Pritchard’a göre bibliyometri, matematiğin ve istatistiksel yöntemlerin kitaplara ve diğer iletişim araçlarına uygulanması olarak tanımlanmaktadır (Pritchard, 1969: 348). Bibliyometri, filoloji, bilgi bilimi, matematik ve istatistik gibi birçok disiplinin kesişimi ve birleşimi ile nicel analize dayalı bir disiplindir (Li vd., 2020: 2). Bibliyometrik analiz; bilimsel araştırmaları ilerletmek için geniş bir perspektiften bakmaya olanak sağlayan, disiplinler arası bir öncüle sahiptir (Samiee ve Chabowski, 2012: 369).

Pritchard bibliyometri’nin yazılı iletişim süreçlerini nicelleştirmeyi amaçlayan tüm çalışmalarda kullanılabileceğini ve birçok bilimsel alanda hızla kabul göreceğini belirtmiştir ve literatür incelendiğinde Pritchard’ın bu tahmininde haklı çıktığı görülmektedir (Pritchard, 1969: 348). Bibliyometri, ilerlemeyi hedefleyen çoğu bilimsel alanda (tıp, matematik, ekonomi, bilgisayar bilimi, fizik, sosyoloji, psikoloji vb.) önemli bir araç haline gelmiştir (Martinez vd., 2015: 257).

Literatürdeki bibliyometrik çalışmalar incelendiğinde, belirli bir alan veya konudaki gelişimin araştırıldığı çalışmaların yanı sıra, dergilerde yayımlanan makaleleri, lisansüstü tezleri, kongre ve sempozyumlarda sunulan bildirimleri inceleyen ulusal ve uluslararası çalışmalara da sıklıkla rastlanmaktadır (Güzeller ve Çeliker, 2017: 2). Çalışmamızda denetimli öğrenme stratejisine dayanan makine öğrenmesi ve karar ağaçları alanında yazılmış tezlerin bibliyometrik analizi yapılmıştır.

Literatürde problemleri çözebilmek için çok sayıda makine öğrenmesi algoritması geliştirilmiştir. Çalışmalar incelendiğinde algoritmaların performansını arttırmak için farklı yaklaşımlar öne sürülmüştür. Bu yaklaşımlar genel olarak; öznelik seçimi yaptıktan sonra makine öğrenmesi algoritmalarını uygulamak, topluluk öğrenme metotları geliştirmek, makine öğrenmesi algoritmalarının parametrelerini belirlemede farklı optimizasyon algoritmaları uygulamak, makine öğrenmesi algoritmalarını sezgisel ya da farklı yöntemler kullanarak iyileştirmek şeklindedir. Çalışmanın odak noktası tezlerde kullanılan denetimli öğrenme stratejisine dayanan makine öğrenmesi, karar ağacı algoritmalarını ve algoritmaların performansını iyileştirmek için geliştirilen farklı yaklaşımları inceleyerek ileride yapılacak çalışmalarda tekrardan kaçınabilmek ve araştırmacılara bir bakış açısı sağlayabilmektir. Bu amaçla Ulusal Tez Merkezinde erişilebilen 368 tez bibliyometrik yöntemle incelenerek, araştırmacılara bu alanda yayınlanan tez çalışmaları hakkında bilgi vermek ve bu konuda çalışacak olan araştırmacılara yol haritası oluşturması hedeflenmektedir. Çalışmanın kapsamını oluşturan ve erişilebilen tezlere ait bilgiler Excel tablosuna aktarılmış,

hesaplamalar, tablolar ve grafikler ile ilgili işlemler belirtilen program üzerinden gerçekleştirilmiştir.

## I. MAKİNE ÖĞRENMESİ

Makine öğrenmesini anlamak için öncelikle öğrenme kavramının tanımlanması gerekmektedir. Öğrenme kavramı farklı şekillerde ifade edilmekle birlikte genellikle Simon tarafından öne sürülen tanımdan yola çıkarak farklı şekillerde ifade edilebilmektedir. Simon öğrenmeyi, "zaman içinde yeni bilgilerin keşfedilmesi yoluyla davranışların iyileştirilmesi süreci" olarak tanımlamıştır. Makine öğrenmesi ise bu öğrenme işinin bilgisayarlar tarafından gerçekleştirilmesidir. Makine öğrenmesi, bilgisayarın geçmişteki deneyimlerinden elde edilen bilgi ve tecrübeleri kullanıp bir model oluşturarak gelecekte oluşacak olan benzer olaylar hakkında karar verebilmesini ve problemlere çözüm bulabilmesini sağlayan bir yapay zekâ alanıdır. Başka bir deyişle makine öğrenmesi " bilgisayarın bir olay ile ilgili bilgileri ve tecrübeleri öğrenerek gelecekte oluşacak benzeri olaylar hakkında kararlar verebilmesi ve problemlere çözümler üretebilmesidir" denilebilir. Bilgisayar oyunları ve yapay zekâ alanında öncü kişilerden sayılan Arthur Lee Samuel 1959 yılında makine öğrenmesini, bilgisayarların yeniden programlamaya ihtiyaç duymadan görev yapmasını sağlayan bilim olarak tanımlamıştır. Mitchell (1997) makine öğrenmesini "Bir bilgisayar yazılımı, deneyim D'yi kullanarak istenilen bir görev G'yi gerçekleştirmek için performans P ölçüsü ile eğitilmesidir." şeklinde tanımlamıştır. Burada performans P ölçütü ile oluşturulan görev G, deneyim D'yi geliştirmektedir (Bilgin, 2018: 13; Öztemel, 2003: 21).

Örneklerle öğrenen sistemlerde farklı öğrenme yaklaşımları kullanılmaktadır. Öğrenme işlemini gerçekleştirecek olan sistem ve kullanılacak öğrenme algoritması bu yaklaşımlara bağlı olarak farklılaşmaktadır (Öztemel, 2003: 24). Genel olarak öğrenme yaklaşımları; denetimli, denetimsiz ve pekiştirmeli (takviyeli) olmak üzere üç grupta incelenmektedir.

(i) *Denetimli Öğrenme (Supervised Learning)*: Denetimli öğrenmede amaç, bir dizi girdi değişkenine dayanarak çıktı değerini tahmin etmektir (Friedman vd., 2001: 28). Bu tür yaklaşımda öğrenen sisteme öğrenmesi istenen olay ile ilgili örnekler girdi/çıktı seti olarak verilir. Sistemin görevi girdileri belirlediği çıktılara haritalamaktır. Böylece olayın girdileri ile çıktılarının arasındaki ilişkiler öğrenilmektedir (Öztemel, 2003: 25). Denetimli makine öğrenme yaklaşımı tahmin ve kestirim problemlerine odaklanmaktadır. Denetimli makine öğrenmesi regresyon ve sınıflandırma olarak iki alt grupta ele alınmaktadır (Akay, 2018:46).

(ii) *Denetimsiz Öğrenme (Unsupervised Learning)*: Denetimsiz öğrenmede; çıktı değeri yoktur ve amaç bir dizi girdi değişkeni arasındaki ilişkileri ve örüntüleri tanımlamaktır (Friedman vd., 2001: 21). Bu tür stratejide sisteme sadece girdi değerleri gösterilir ve sadece girdiler üzerinde öğrenme işlemini gerçekleştirmeye çalışır. Örneklerdeki parametreler arasındaki ilişkileri sistemin kendi kendisine öğrenerek ilişki ağını ortaya koyması beklenir (Öztemel, 2003: 25). Denetimsiz öğrenme stratejisi kümeleme, olasılık yoğunluk tahmini, öznitelikler

arasındaki ilişkilerin keşfedilmesi ve boyut indirgeme gibi amaçlar için kullanılmaktadır (Kutlugün, 2017: 31-32).

(iii) *Takviyeli Öğrenme (Reinforcement Learning)*: Bu yaklaşımda öğrenen sisteme her girdi seti için üretilmesi gereken çıktı setini sisteme vermek yerine sistemin kedisine gösterilen girdilere karşılık çıktıyı üretmesi beklenir ve ürettiği çıktının doğru veya yanlış olduğuna dair bir sinyal üretir. Sistem üretilen sinyali göz önüne alarak öğrenme sürecini devam ettirir (Öztemel, 2003: 25).

### **A. Topluluk Öğrenmesi**

Sınıflandırma, kümeleme ve regresyon problemlerinde, tek bir öğrenme algoritmasının kararlılığını ve tahmin doğruluğunu arttırabilmek için, topluluk öğrenmesi (ensemble learning) kavramı ortaya atılmıştır (Kapucu ve Çubukçu, 2019: 5). Topluluk öğrenmesi modeli, bireysel olarak eğitilmiş bir dizi temel modeli belirli bir toplam kuralı kullanarak birleştirerek tek bir nihai tahmin üreten güncel bir makine öğrenmesi araştırma alanıdır. Topluluk öğrenmesi yöntemleri, temel öğrenme algoritmalarının genelleştirme hatasını düşürerek ve doğru sınıflandırma oranını artırarak daha yüksek performans ile tahmin yapabilen modellerin oluşturulması amaçlanır. Topluluk tahmin modeli, geleneksel tek tahmin modelinden daha kararlı ve doğru tahminler sağlayabilir (Verma ve Mehta, 2017: 155; Onan, 2018: 60-61).

Topluluk öğrenmesi yöntemi, temel (baz-base) model oluşturma stratejisine göre heterojen topluluk ve homojen topluluk olmak üzere iki kategoriye ayrılabilir. Heterojen topluluk modelinde; temel modelleri, aynı eğitim verilerini farklı öğrenme algoritmalarına veya aynı algoritmalara farklı parametre ayarlarıyla uygulanarak oluşturulur. Buna örnek olarak bagging, boosting, random forest (RF), random subspace verilebilir. Homojen topluluk modelinde ise, orijinal verilerden yeniden örneklenen farklı eğitim verilerini aynı parametre ayarları ile aynı öğrenme algoritmasına uygulayarak temel modellerini oluşturur. Buna örnek olarak stacking verilebilir (Wang vd., 2018: 110).

### **B. Öznitelik Seçimi**

Veri setlerindeki öznitelikler sınıflama performansını etkileyen en önemli unsurlardan biridir. Bir veri setinde toplanan özniteliklerin sayısı genellikle nispeten büyüktür (yani boyutluluk laneti) ve bu özelliklerin tümü bilgilendirici değildir veya yüksek ayırt etme gücü sağlayamaz (Tsai vd., 2013: 241). Gerçek hayat problemlerini çözebilmek için kullanılan veri setlerinde gereksiz, ilgisiz, gürültülü, yanıltıcı vs. öznitelikler yer almakta ve problemleri çözmek için kurulacak modeller hakkında genellikle önsel bilgi bulunmamaktadır. Bazı makine öğrenme algoritmaları, gereksiz, ilgisiz, gürültülü, yanıltıcı gibi özniteliklere karşı duyarlıdır ve bu tür algoritmalarda performans düşüşleri ortaya çıkabilir. Bu nedenle özniteliklerin yeterli sayıda ve doğru olarak belirlenmesi gerekmektedir. Öznitelik seçimi (attribute selection), bir veri kümesindeki modeli oluşturmak için verilen orijinal öznitelik kümesinden alakasız ve gereksiz öznitelikleri ortadan kaldırarak orijinal veri setini en iyi temsil edebilecek altkümenin seçilmesi olarak tanımlanmaktadır. Öznitelik seçimi (diğer adıyla nitelik seçimi veya değişken seçimi), kullandığı algoritmaya göre öznitelikleri değerlendirip veri setinde

bulunan  $n$  adet öznelik arasından en iyi  $k$  adet özneliğin seçilmesi işlemidir (Budak, 2018: 21-22; Khalid vd., 2014: 373). Günlük hayatta karşılaştığımız gerçek problemlerin sonucu yanıltıcı gürültü adı verilen bileşenlere sahip olduğu düşünüldüğünde yapılan bu işlem aslında bir gerekliliktir (Bilgin, 2018: 24).

Öznelik sayısının azaltılması analiz aşamasında araştırmacıya birçok avantaj sağlamaktadır. Öznelik seçme işleminin avantajları (Khalid vd.,2014: 373):

- ✓ Öznelik kümesinin boyutunu düşürür ve algoritma hızını artırır,
- ✓ Gereksiz, ilgisiz veya gürültülü veriyi ortadan kaldırır,
- ✓ Veri kalitesini iyileştirir,
- ✓ Veri kümesi daha basit bir şekilde tanımlanabilir, görselleştirilebilir ve anlaşılabilir hale getirir, yani modelin karmaşıklığını en basit düzeye indirgemiş olur,
- ✓ Veri depolamak için gereken hafıza miktarını azaltır,
- ✓ Elde edilen modelin başarısını artırır,
- ✓ Ezberleme riskini de azaltır. Böylece; aşırı uyum (overfitting) problemi aşılır.

Öznelik seçiminin dezavantajı ise en iyi alt kümenin seçilmesi zaman yönünden ve ek işlem maliyeti açısından bir negatif oluşturulabilmesidir (Bilgin, 2018: 26).

Öznelik seçim yöntemleri; filtreleme (filter), sarmal (wrapper) ve gömülü (embedded) yöntem olmak üzere üç kategoriye ayrılabilir.

(i) *Filtreleme (Filter) Yöntemi*: Filtreleme yöntemleri makine öğrenmesinde kullanılan en eski öznelik seçim yöntemlerindedir. Filtreleme yöntemleri, özneliklerin teker teker modelin açıklama/doğrulama gücü üzerindeki etkisini incelemeye dayanmaktadır. Bu yöntemlerde herhangi bir sınıflandırma algoritması kullanmadan uzaklık, bilgi, bağımlılık ve tutarlılık ölçütleri gibi istatistiksel ölçütlere dayanan fonksiyonlar yardımıyla öznelik seçim işlemi yapılmaktadır (Budak, 2018: 23). Öznelik seçme ve sınıflandırma işlemleri birbirinden ayrı ve bağımsız olarak gerçekleştirilir. Filtreleme yöntemlerine örnek olarak Fisher Skor, ki-kare, Information Gain (IG), Gain Ratio (GR), Gini, F-score, Symmetrical Uncertainty (SU), Correlation Feature Selection (CFS), RELIEFF, ONER verilebilir.

(ii) *Sarmal (Wrapper) Yöntem*: Sarmal yöntemlerde, öznelik seçimi için doğru sınıflandırma oranına bağlı olarak performansının ölçüldüğü farklı öğrenme algoritmaları kullanarak en iyi performansı veren öznelikler seçilmektedir (Budak, 2018: 25). Bir başka deyişle, sarmal yöntemlerde makine öğrenmesi algoritmaları kullanılmaktadır ve özneliği seçmenin ölçütü kullanılan sınıflandırma oranının doğruluk oranıdır. Her bir iterasyonda belirli öznelik altkümesi için sınıflandırma sonucu elde edilmektedir. Sarmal yöntemlerde, öznelik alt küme uzayı büyüdükçe sezgisel yöntemler tercih edilmektedir (Yıldız vd., 2012: 2). Sarmal yöntemlere örnek olarak Ardışık ileri yönde seçim (sequential forward selection- SFS) L ekle- R çıkar (plus l – minus r), ardışık geri yönde seçim (sequential backward selection- SBS), ardışık ileri yönde kayan seçim (sequential

forward floating selection- SFFS), ardışık geri yönde kayan seçim (sequential backward floating selection- SBFS), özyinelemeli özellik eleme (recursive feature elimination-RFE), genetik algoritma (genetic algorithm-GA), parçacık sürü optimizasyonu (particle swarm optimization-PSO) verilebilir.

(iii) *Gömülü (Embedded) Yöntem*: Gömülü yöntemlerde ise, makine öğrenmesi algoritması ve öznitelik seçim algoritması eş zamanlı olarak çalışmaktadır. Gömülü yöntemlerin yapısında hem sınıflandırma algoritması hem de öznitelik seçim algoritması içerdiğinden, sınıflandırma ve öznitelik seçme işlemleri eşzamanlı olarak gerçekleştirmektedir (Budak, 2018: 26-27). En uygun öznitelik alt kümesi arayışı, sınıflandırıcı yapısının içine yerleştirilmiştir ve öznitelik alt kümesi ve hipotezlerin birleşik alanında bir arama olarak görülebilir. Gömülü yöntemlere örnek olarak karar ağaçları, LASSO verilebilir (Subanya ve Rajalaxmi, 2014: 2).

### C. Parametre Optimizasyonu

Makine öğrenmesi modelleri tasarlanırken dışardan girilmesi gereken parametreler olan hiper parametrelerin optimizasyonu, çalışma maliyetini düşürdüğünden makine öğrenmesi modellerinin performansına olumlu etkide bulunmaktadır aynı zaman da algoritmanın performansı üzerinde önemli etkiye sahiptir. Makine öğrenmesi modelleri tasarlanırken kullanılan parametreler, eğitim sürecindeki verilerden elde edilebilen ve tasarımcı tarafından önceden tanımlanabilen olmak üzere iki gruba ayrılır. Bunlar model parametresi ve hiper parametrelerdir. Model parametreleri genellikle verilerden tahmin edilen veya öğrenilen parametrelerdir. Bu parametrelerde tasarımcının ayarlama yapması beklenmez. Öğrenilen modelin bir parçası olarak kaydedilir. Yapay sinir ağındaki ağırlıklar, bir destek vektör makinesinde (Support Vector Machine-SVM) destek vektörleri, doğrusal regresyonda veya lojistik regresyonda katsayılar, model parametrelerine örnek olarak verilebilir. Model parametrelerinden farklı olarak hiper parametreler, verilerden tahmin edilmez ve tasarımcı tarafından ayarlanması gerekmektedir. Destek vektör makinelerinde kullanılan kernel parametresi( $\gamma$ ), epsilon değeri ( $\epsilon$ ); K-En yakın komşu (KNN) algoritmasındaki komşuluk değeri( $k$ ); derin öğrenme algoritmalarında kullanılan filtre boyutu, filtre sayısı, nöron sayısı, katman sayısı, aktivasyon fonksiyonu vb., hiper parametreye örnek olarak verilebilir (Tanyıldızı ve Demirtaş, 2019: 1). Model oluşturulurken başarı oranı yüksek sonuçlar alabilmek için hiper parametrelerin optimize edilmesi gerekmektedir (Pınar vd., 2017: 262). Çalışmalarda Grid Search veya Random Search gibi en basit yöntemlerden meta-sezgisel gibi daha karmaşık olanlara kadar farklı algoritmalar hiper parametre değerini belirlemek için kullanılmaktadır.

## II. YÖNTEM

Bu çalışmada nicel araştırma yöntemlerinden bibliyometrik analiz tekniği kullanılmıştır. Bibliyometrik analiz, belirli bir alandaki literatürün nicel hale getirerek ve ortaya çıkan kavramları değerlendirerek alandaki eğilimlerin tespit edilmesini sağlamaktadır. Böylelikle bilimsel yayın politikalarına yön vermenin yanı sıra araştırmacılar için yol haritası oluşturmaktadır (Güzeller ve Çeliker, 2017: 89). Bu çalışmada, makine öğrenmesi ve karar ağaçları alanında yayımlanan

tezlerin belirli parametreler kapsamında değerlendirilerek bibliyometrik analiz yöntemiyle incelenmesi amaçlanmıştır. Bu amaç doğrultusunda çalışmada aşağıdaki sorulara cevap aranmıştır;

- Tezlerin yıllara göre dağılımı nedir?
- Tezlerin derecesine göre dağılımı nedir?
- Tezlerde kullanılan dil nedir?
- Tez çalışmalarının yapıldığı üniversitelerin dağılımı nasıldır?
- Tezlerin enstitülere göre dağılımı nasıldır?
- Tezlerin ana bilim dallarına göre dağılımları nasıldır?
- Tezlerin sayfa sayılarının dağılımı hangi aralıkta yoğunlaşmaktadır?
- Kullanılan programlar nelerdir?
- Tezlerin kullanılan yöntemlerin dağılımı nasıldır?
- Makine öğrenmesi algoritmalarının performansını iyileştirmek için geliştirilen farklı yaklaşımlar nelerdir?
- Tezlerde kullanılan anahtar kelimelerin analizi nasıldır?

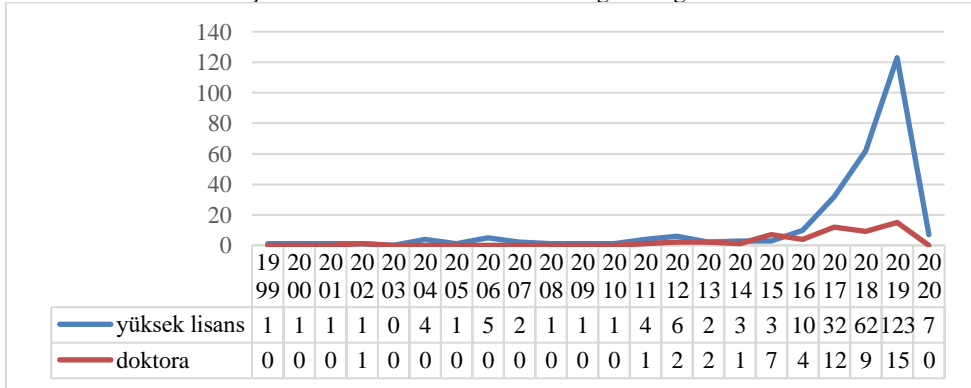
Tezlere ilişkin belirlenen araştırma soruları dikkate alınarak veriler elde edilerek Excel programına aktarılmıştır ve tüm hesaplamalar, grafikler ve tablolar bu program aracılığıyla yapılmıştır.

### III. BULGULAR

#### A. Makine Öğrenmesi Alanındaki Tezlere İlişkin Bulgular

Yüksek Öğretim Kurumu'nun Ulusal Tez Merkezi internet sitesinde “makine öğrenmesi”, “machine learning” anahtar kelimeler kullanılarak 2020 yılı mayıs ayında yapılan tarama sonucu 325 teze ulaşılmıştır. 2020 yılının verileri mayıs ayına kadar olduğundan 2020 yılının tamamını yansıtmamaktadır. Bu tezlerin 25 tanesi erişimi sınırlandırılmıştır. İncelenen tezlerin adlarına <https://tinyurl.com/2mhsaeky> linkinden ulaşabilirsiniz. Ulaşılan tezler; yıl, tez türü, dili, üniversite, enstitü, anabilim dalı, sayfa sayısı, anahtar kelime, kullanılan program, kullanılan yöntem dikkate alınarak bibliyografik açıdan detaylı incelenmiş ve Excel programına aktarılarak grafik ve tablolar oluşturulmuştur.

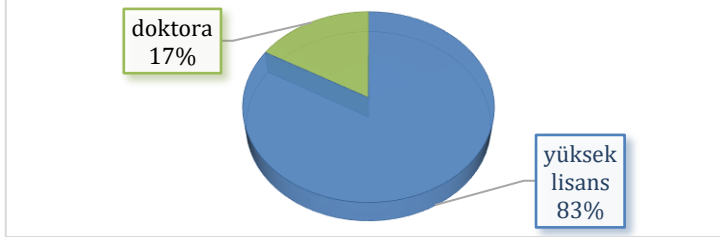
Şekil 1. Lisansüstü Tezlerin Yıllara göre Dağılımı



Lisansüstü tezlerin yıllara göre dağılımı Şekil 1’de verilmiştir. Tezlerin yıllara göre dağılımı incelendiğinde, 2016 yılına kadar büyük sıçramalar olmayan

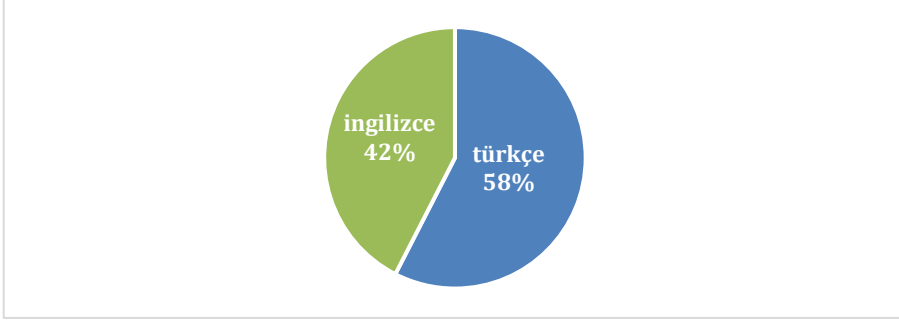
tez sayılarında 2016 yılından itibaren artışlar gözlenmektedir. Yazılan tezlerin çoğunluğunu yüksek lisans tezleri oluşturmaktadır.

Şekil 2. Lisansüstü Tezlerin Türü



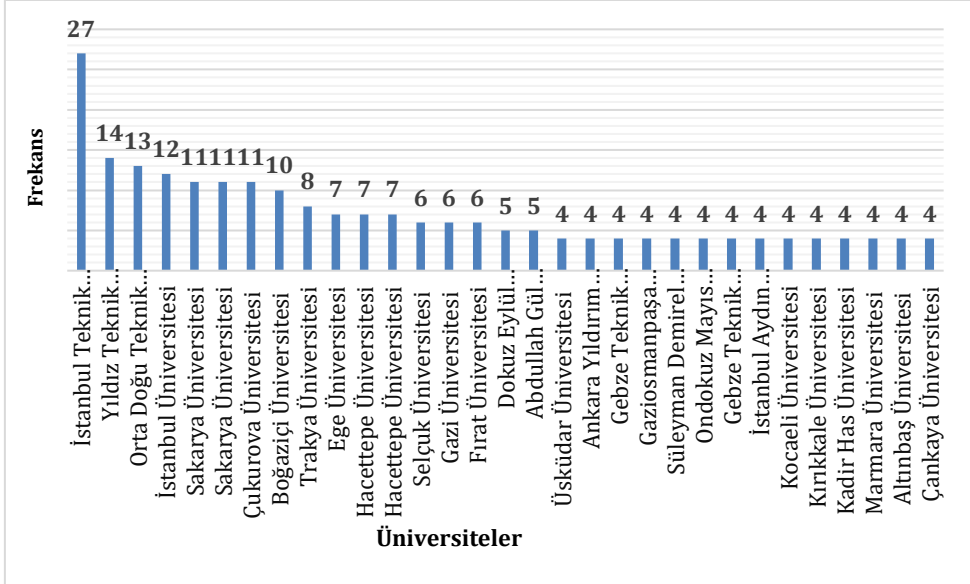
Şekil 2’de lisansüstü tezlerin türlerine göre dağılımı verilmiştir. Buna göre; çalışmada incelenen 325 lisansüstü tezin %17’si doktora, %83’ü ise yüksek lisans tezidir.

Şekil 3. Tezlerde Kullanılan Dil



Şekil 3 incelendiğinde; yayımlanan lisansüstü tezlerinin %58’i Türkçe, %42’si ise İngilizcedir.

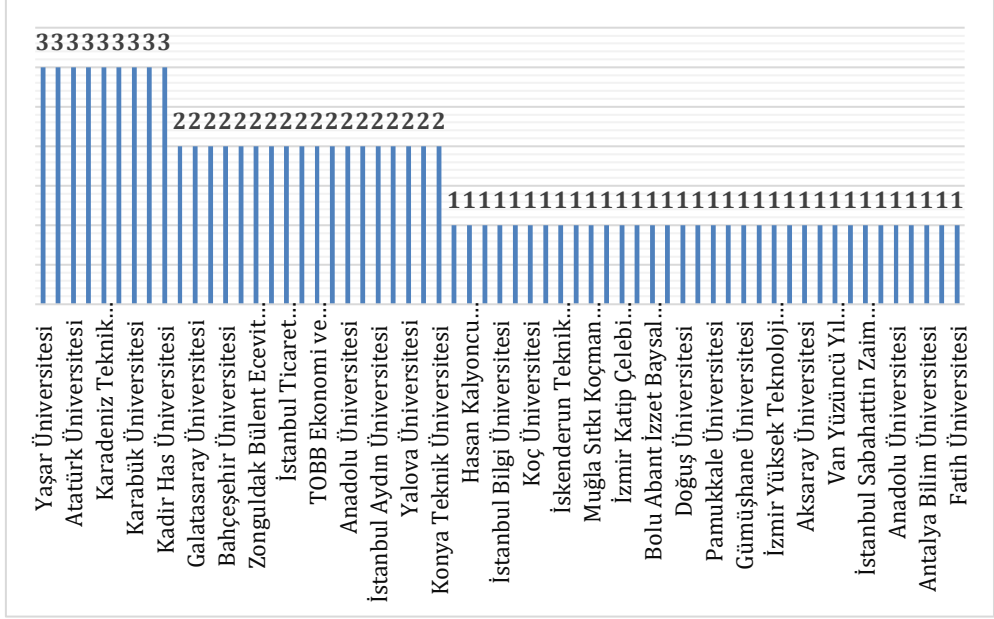
Şekil 4. 4 ve Üzeri Tez Yayınlayan Üniversitelerin Frekansı



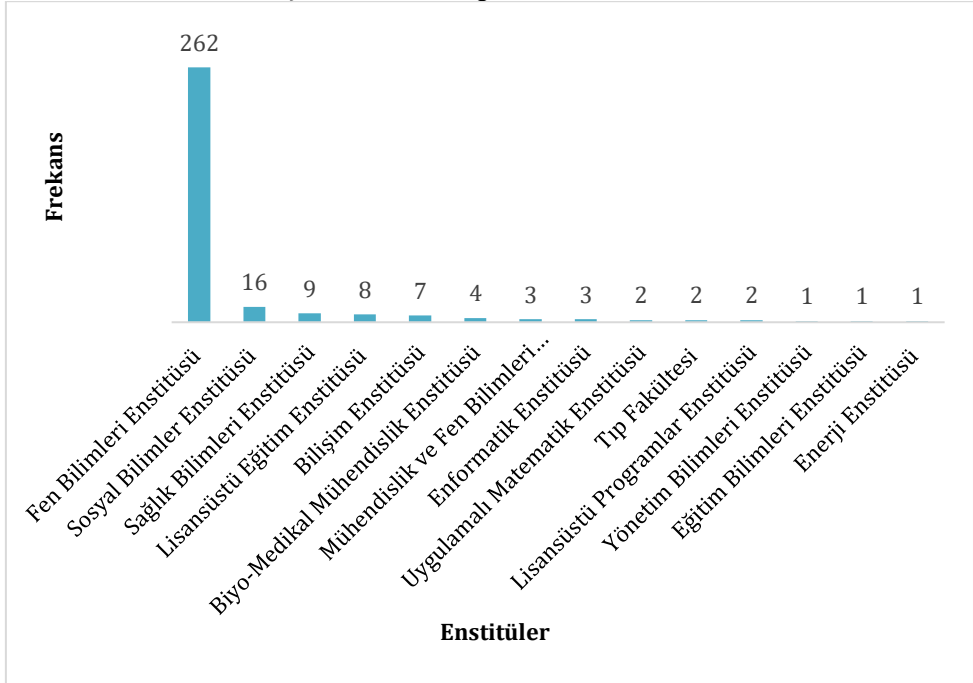


İncelenen tezlerin üniversitelere göre dağılımı Şekil 4 ve 5'te verilmiştir. Buna göre; İstanbul Teknik Üniversitesi, Yıldız Teknik Üniversitesi, Orta Doğu Teknik Üniversitesi, İstanbul Üniversitesi'nin alana en çok katkı sağlayan üniversiteler olduğunu söyleyebiliriz.

Şekil 5. 4' ten Az Tez Yayınlayan Üniversitelerin Frekansı

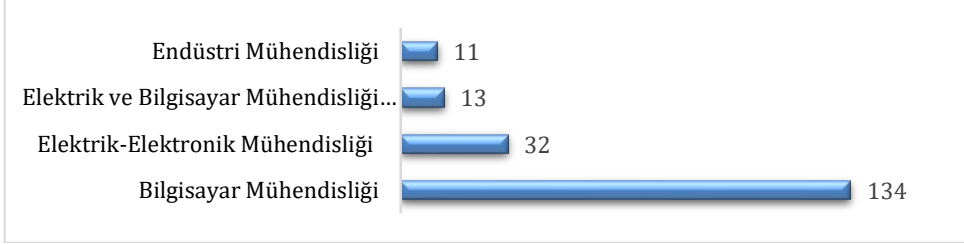


Şekil 6. Enstitülere göre Tezlerin Frekansı



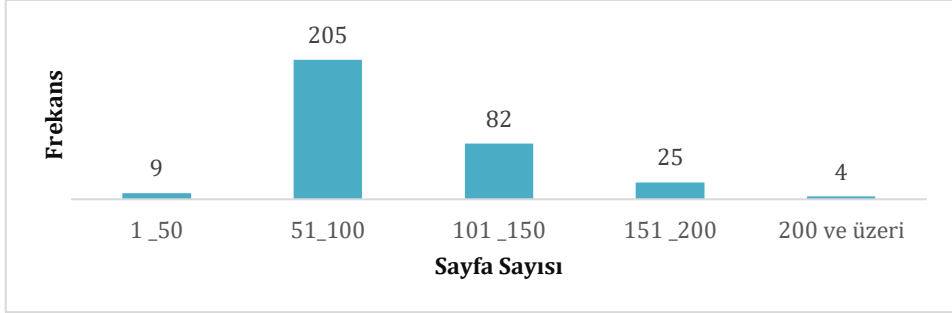
İncelenen tezlerin enstitülere göre dağılımı şekil 6'da verilmiştir. Grafik incelendiğinde; bu enstitüler içerisinde de en çok tezin fen bilimleri enstitüsü altında hazırlandığı görülmektedir.

Şekil 7. Ana Bilim Dallarına göre Tezlerin Frekansı



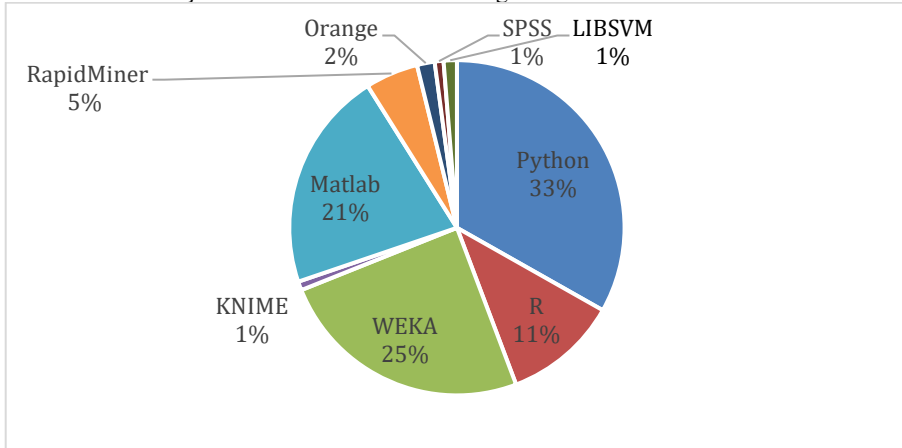
Şekil 7'de tezlerin en çok yayınladığı dört anabilim dalı verilmiştir. Tezlerin büyük bir kısmının Mühendislik alanında gerçekleştirildiği görülmektedir. Bulgular incelendiğinde, Bilgisayar Mühendisliği Anabilim Dalı ilk sırada yer almakta, Elektrik- Elektronik Mühendisliği Anabilim Dalı ikinci sırada yer almaktadır.

Şekil 8. Tezlerin Sayfa Sayısı Aralığı



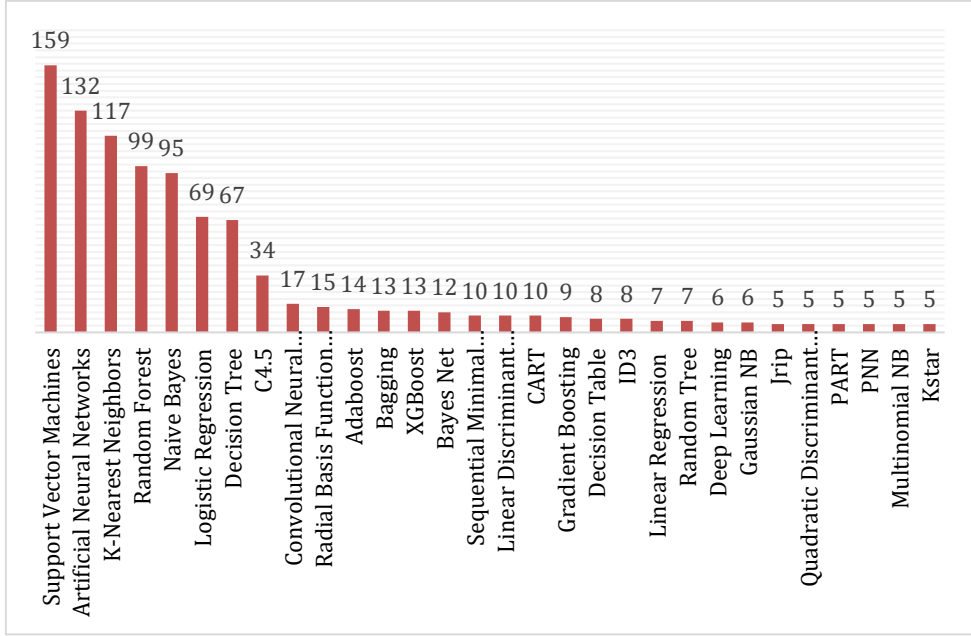
Lisansüstü tezlerin sayfa sayısı incelendiğinde ortalama 96 sayfa olduğu belirlenmiştir. En az sayfa sayısına sahip tez 35 sayfa iken en çok 498 sayfadan oluşmaktadır. Şekil 8'e göre, belirtilen sayfa aralıklarına göre tezlerin dağılımı incelendiğinde, sayfa sayılarının 51-100 arasında yoğunlaştığı görülmektedir.

Şekil 9. Tezlerde Kullanılan Programlama Dili / Yazılım

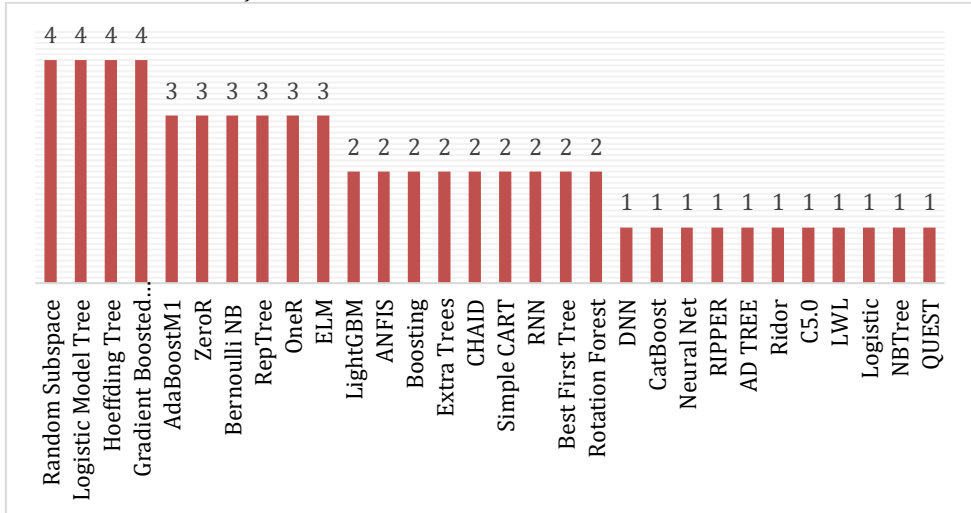


Kullanılan programları gösteren şekil 9 incelendiğinde en sık kullanılan program Python olduğu görülmektedir.

Şekil 10. 5 Defa ve Üzeri Kullanılan Yöntemlerin Frekansı

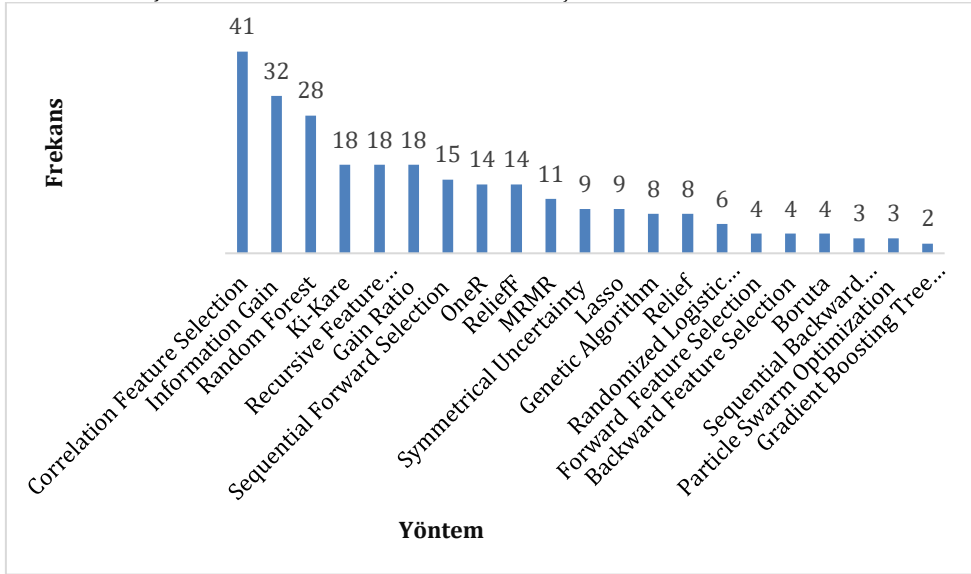


Şekil 11. 5'ten Az Kullanılan Yöntemlerin Frekansı



İncelenen tezlerde kullanılan makine öğrenmesi algoritmaları Şekil 10 ve Şekil 11'de verilmiştir. Şekiller incelendiğinde en sık kullanılan yedi yöntem: Support Vector Machines, Artificial Neural Networks (ANN), K-Nearest Neighbors, Random Forest, Naive Bayes (NB), Logistic Regression (LR), Decision Tree (DT) dir. Kullanılan yöntemler incelendiğinde topluluk öğrenme yöntemlerinin yeterince kullanılmadığı görülmektedir.

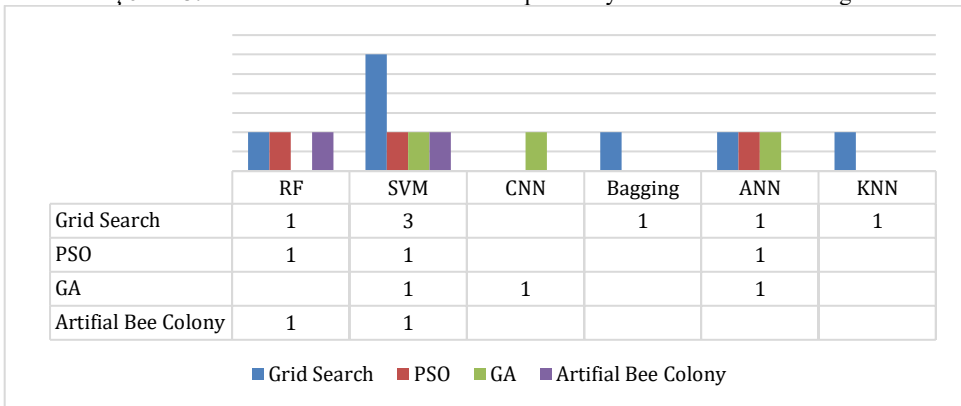
Şekil 12. Tezlerde Kullanılan Öznitelik Seçim Yöntemlerinin Frekansı



İncelenen tezlerde kullanılan öznitelik seçim yöntemleri Şekil 12’de görülmektedir. Grafikler incelendiğinde; Correlation Feature Selection, Information Gain, Random Forest, Ki-Kare, Recursive Feature Elimination, Gain Ratio en sık kullanılan öznitelik seçim yöntemleridir. Öznitelik seçim yöntemlerinde çoğunlukla klasik öznitelik seçim yöntemleri kullanılmaktadır. Öznitelik seçim yöntemlerinin hangi makine öğrenme algoritmalarıyla birlikte kullanıldığını gösteren grafik aşağıda verilmiştir.

Ele alınan tezlerde Makine Öğrenmesi Algoritmalarının parametrelerini belirlemede kullanılan parametre optimizasyon yöntemlerine ilişkin grafik şu şekildedir:

Şekil 13. Tezlerde Kullanılan Parametre Optimizasyon Yöntemlerinin Sıklığı

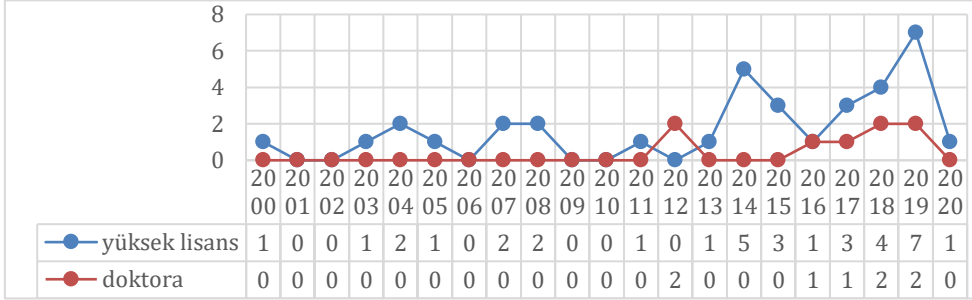


Makine öğrenmesi algoritmalarının parametrelerini belirlemede kullanılan parametre optimizasyon yöntemleri algoritmalarla birlikte Şekil 13’te verilmiştir.



linkinden ulaşabilirsiniz. Ulaşılan tezler; yıl, tez türü, dili, üniversite, enstitü, danışman akademik unvanı, anabilim dalı, sayfa sayısı, kullanılan program, kullanılan yöntem dikkate alınarak bibliyografik açıdan detaylı incelenmiş ve Excel programına aktarılarak grafik ve tablolar oluşturulmuştur.

Şekil 15. Lisansüstü Tezlerin Yıllara göre Dağılımı



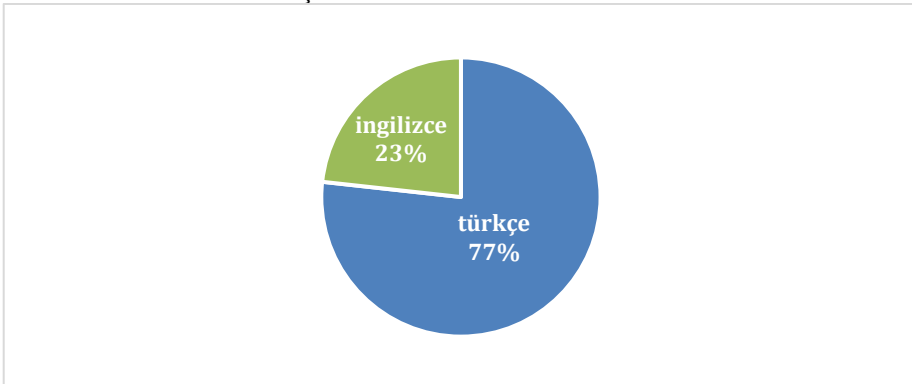
Lisansüstü tezlerin yıllara göre dağılımı Şekil 15’de verilmiştir. Tezlerin yıllara göre dağılımı incelendiğinde, ilk tezin 2000 yılında yazıldığı 2012 yılına kadar doktora tezinin yazılmadığı gözlenmektedir. Yazılan tezlerin çoğunluğunu yüksek lisans tezleri oluşturmaktadır.

Şekil 16. Lisansüstü Tezlerin Türü

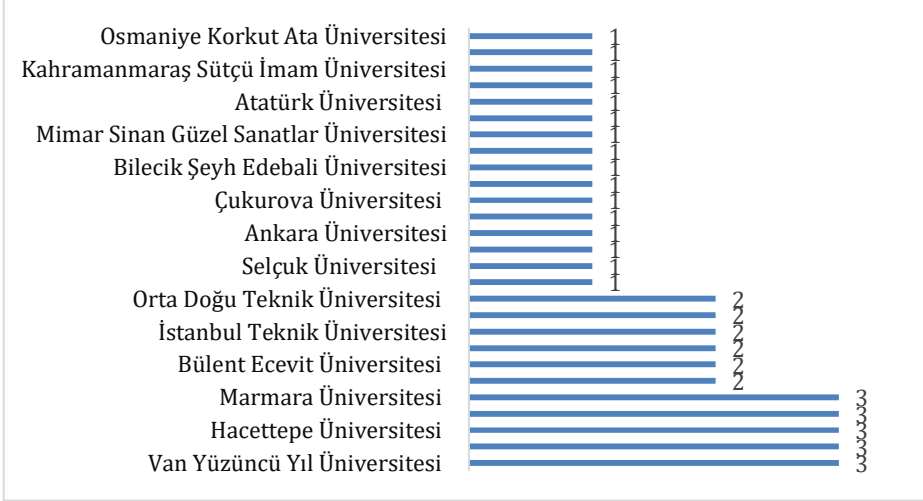


Şekil 16’da lisansüstü tezlerin türlerine göre dağılımı verilmiştir. Buna göre; çalışmada incelenen 43 lisansüstü tezin %19’u doktora, %81’i ise yüksek lisans tezidir.

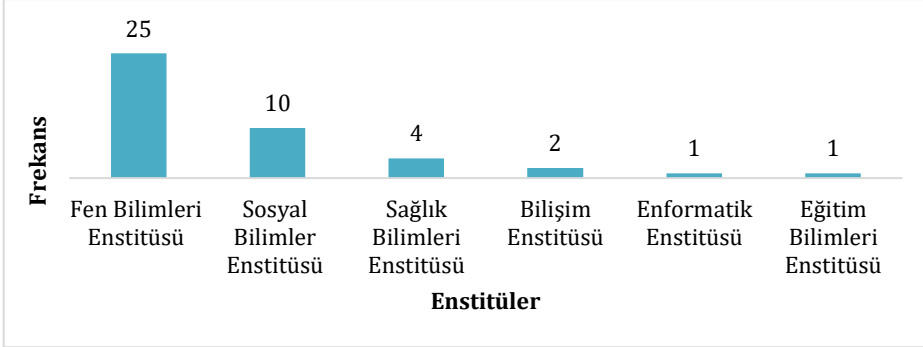
Şekil 17. Tezlerde Kullanılan Dil



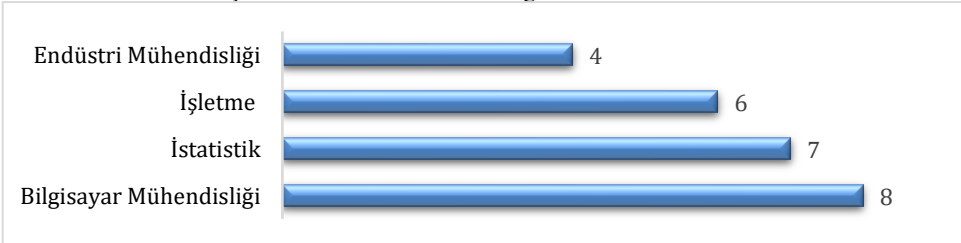
Şekil 17 incelendiğinde; yayınlanan lisansüstü tezlerinin %77’si Türkçe, %23’ü ise İngilizcedir.

**Şekil 18.** Tezlerin yayınlandığı Üniversitelere Göre Frekansı

İncelenen tezlerin üniversitelere göre dağılımı Şekil 18’de verilmiştir. Buna göre; Van Yüzüncü Yıl Üniversitesi, Gazi Üniversitesi, Hacettepe Üniversitesi, Boğaziçi Üniversitesi ve Marmara Üniversitesi’nin alana en çok katkı sağlayan üniversiteler olduğunu söyleyebiliriz.

**Şekil 19.** Enstitülere göre Tezlerin Frekansı

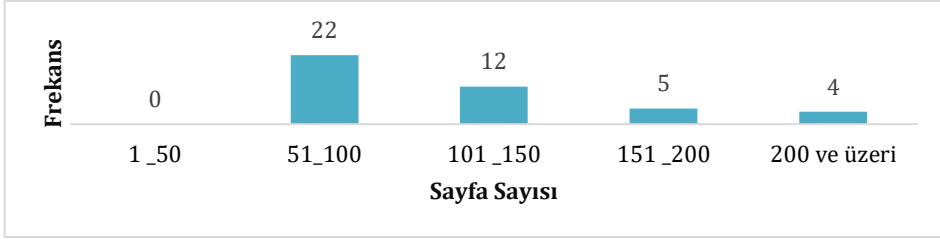
İncelenen tezlerin enstitülere göre dağılımı şekil 19’da verilmiştir. Grafik incelendiğinde; bu enstitüler içerisinde en çok tezin fen bilimleri enstitüsü altında hazırlandığı görülmektedir.

**Şekil 20.** Ana Bilim Dallarına göre Tezlerin Frekansı

Şekil 20’de tezlerin en çok yayınladığı dört ana bilim dalı verilmiştir. Bulgular incelendiğinde, bilgisayar mühendisliği ana bilim dalı ilk sırada yer

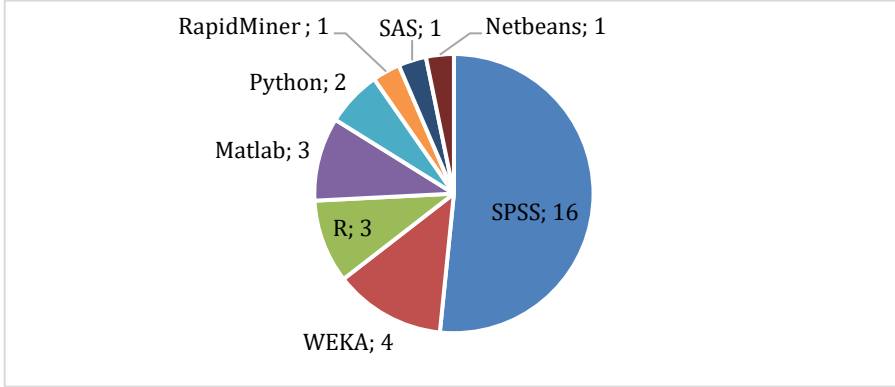
almakta, istatistik ana bilim dalı ikinci sırada yer almaktadır. Bunları işletme ve endüstri mühendisliği ana bilim dalları takip etmektedir.

Şekil 21. Tezlerin Sayfa Sayısı Aralığı



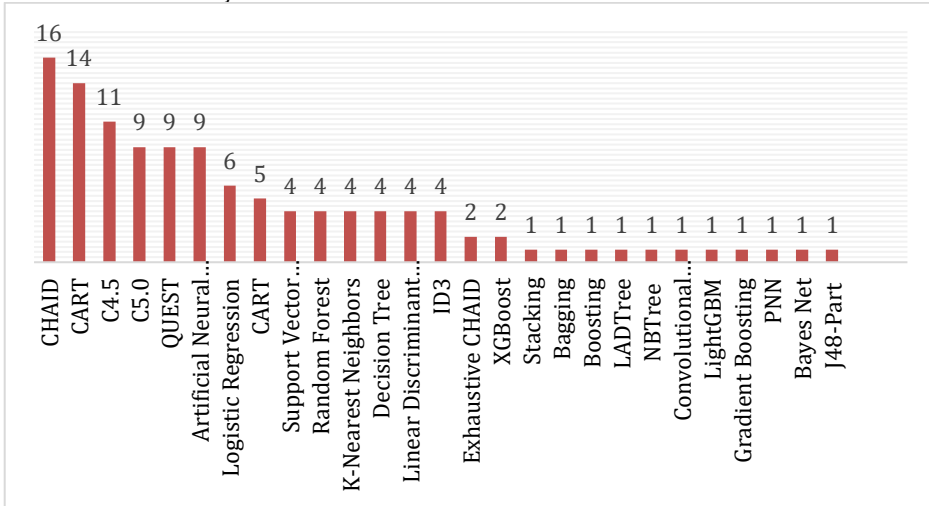
Lisansüstü tezlerin sayfa sayısı incelendiğinde ortalama 120 sayfa olduğu belirlenmiştir. En az sayfa sayısına sahip tez 54 sayfa iken en çok 341 sayfadan oluşmaktadır. Şekil 21'e göre, belirtilen sayfa aralıklarına göre tezlerin dağılımı incelendiğinde, sayfa sayılarının 51-100 arasında yoğunlaştığı görülmektedir.

Şekil 22. Tezlerde Kullanılan Programlama Dili / Yazılım



Kullanılan programı gösteren şekil 22 incelendiğinde en sık kullanılan programın SPSS olduğu görülmektedir.

Şekil 23. Tezlerde Kullanılan Yöntemlerin Frekansı







**Makine öğrenmesi alanında yazılan tezlere ilişkin elde edilen bulgular:**

– Tezlerin yıllar bazında dağılımı incelendiğinde, 2016 yılından sonra yayınlanan tez sayısında önemli bir artış olduğu görülmektedir. Aynı zamanda incelenen tezlerin çoğunluğunun yüksek lisans tezi olduğu görülmektedir.

– Alana en çok katkı sağlayan üniversitelerin İstanbul Teknik Üniversitesi, Yıldız Teknik Üniversitesi, Orta Doğu Teknik Üniversitesi, İstanbul Üniversitesinin olduğu; tezlerin büyük çoğunluğunun Fen Bilimler Enstitüsü altında yayımlandığı ve ana bilim dalı olarak da Bilgisayar mühendisliği alanının çoğunlukta olduğu görülmektedir.

– Makine öğrenmesi alanında çok sayıda algoritma mevcuttur. İncelenen tezlerde Support Vector Machines, Artificial Neural Networks, K-Nearest Neighbors, Random Forest, Naive Bayes, Logistic Regression, Decision Tree sırasıyla en sık kullanılan yöntemleridir. Aynı zamanda topluluk öğrenme yöntemlerinin yeterince kullanılmadığı bu alanda önemli bir boşluk olduğu görülmektedir.

– Tezlerde Correlation Feature Selection, Information Gain, Random Forest, Ki-Kare, Recursive Feature Elimination, Gain Ratio en sık kullanılan öznelik seçim yöntemleridir. İncelenen çalışmalarda çoğunlukla klasik öznelik seçim yöntemleri kullanıldığı sarmal ve gömülü yaklaşımlarının kullanımının çok yaygın olmadığı görülmektedir.

– Makine öğrenmesi algoritmalarının parametrelerini belirlemede kullanılan parametre optimizasyon yöntemleri algoritmalarla birlikte incelendiğinde; GA, PSO, Grid Search ve Artificial Bee Colony kullanılan parametre optimizasyon yöntemleridir. Hangi makine öğrenme algoritmasının parametrelerinin optimizasyonun yapıldığına bakıldığında en sık SVM yönteminin parametrelerini belirlemede farklı yaklaşımlar kullanılmış ancak tezlerde parametre optimizasyonu çok az sayıdadır ve sezgisel yaklaşımların kullanımına da çok az rastlanmıştır. Bu alanda önemli bir boşluk bulunmaktadır. Kullanılan yöntemlerden yola çıkarak yeni yöntemler denenebilir.

– Tezlerde geliştirilen topluluk yöntemlerine bakıldığında da bu alanda boşluk bulunmaktadır ve farklı metotlar bir arda kullanılarak yeni yaklaşımlar geliştirilebilir.

– Sadece bir tezde melez yaklaşım ortaya atılmıştır. Geliştirilen melez yaklaşıma ilişkin verilen tablodan yola çıkarak yeni yaklaşımlar geliştirilebilir.

– Python ve WEKA sırasıyla en yaygın programlama dili / yazılımdır.

**Karar ağaçları alanında yazılan tezlere ilişkin elde edilen bulgular:**

– Tezlerin yıllar bazında dağılımı incelendiğinde, ilk tezin 2000 yılında yazıldığı 2012 yılına kadar doktora tezinin yazılmadığı gözlenmektedir. Yazılan tezlerin çoğunluğunu yüksek lisans tezidir.

– Alana en çok katkı sağlayan Van Yüzüncü Yıl Üniversitesi, Gazi Üniversitesi, Hacettepe Üniversitesi, Boğaziçi Üniversitesi ve Marmara Üniversitesi; tezlerin büyük çoğunluğunun Fen Bilimler Enstitüsü ikinci olarak ise

Sosyal Bilimler Enstitüsü altında yayınlandığı ve ana bilim dalı olarak da Bilgisayar mühendisliği alanının çoğunlukta olduğu görülmektedir.

– Makine öğrenmesi alanında çok sayıda algoritma mevcuttur. İncelenen tezlerde CHAID, CART, C4.5, C5.0, QUEST ve ANN sırasıyla en sık kullanılan yöntemleridir. Aynı zamanda topluluk öğrenme yöntemlerinin yeterince kullanılmadığı bu alanda önemli bir boşluk olduğu görülmektedir.

– Tezlerde sadece bir tane çalışmada öznitelik seçimi kullanılmıştır. Karar ağacı algoritmaları ile birlikte öznitelik seçim yöntemlerinin kullanımı alanında önemli bir boşluk bulunmaktadır.

– Makine öğrenmesi algoritmalarının parametrelerini belirlemede kullanılan parametre optimizasyon yöntemleri karar ağacı alanında yazılan tezlerde kullanılmamıştır.

– Tezlerin çok az bir kısmında melez yaklaşım geliştirilmiştir. Geliştirilmiş melez yaklaşıma ilişkin verilen tablodan yola çıkarak yeni yaklaşımlar geliştirilebilir.

– SPSS ve WEKA sırasıyla en yaygın programlama dili / yazılımdır.

Erişilen tezlerin birçok parametre açısından detaylı analiz edildiği bu çalışma ile bu kapsamda ileride çalışacak araştırmacılara birçok yönden (kullanılabilecek yöntem, program, geliştirilecek melez yaklaşım vb.) yol göstereceği düşünülmektedir.

### **Araştırma ve Yayın Etiği Beyanı**

Makalenin tüm süreçlerinde Yönetim ve Ekonomi Dergisi'nin araştırma ve yayın etiği ilkelerine uygun olarak hareket edilmiştir.

### **Yazarların Makaleye Katkı Oranları**

Yazarlar çalışmaya eşit oranda katkı sağlamıştır.

### **Çıkar Beyanı**

Yazarın herhangi bir kişi ya da kuruluş ile çıkar çatışması yoktur.

### **KAYNAKÇA**

- Akay, E. Ç. (2018). Ekonometride yeni bir ufuk: büyük veri ve makine öğrenmesi. *Sosyal Bilimler Araştırma Dergisi*, 7(2), 41-53.
- Altan, S.N. (2018). *Metin Sınıflandırma için Makine Öğrenmesi Tekniklerine Dayalı Bir Yöntem Geliştirilmesi*. (Yüksek lisans Tezi), Ege Üniversitesi, Fen Bilimleri Enstitüsü.
- Ateş, E. (2019). *Hisse Senetleri ve Sosyal Medya Arasındaki İlişkinin Makine Öğrenmesi Teknikleri ile Belirlenmesi*. (Yüksek lisans Tezi). Doğu Üniversitesi Fen Bilimleri Enstitüsü.
- Bilgin, M. (2018). Makine Öğrenmesi. İstanbul: Papatya Yayınları.
- Budak, H. (2018). Özellik seçim yöntemleri ve yeni bir yaklaşım. *Journal of Natural & Applied Sciences*.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). The Elements of Statistical Learning. Vol. 1, No. 10. New York: Springer Series in Statistics.
- Güldan, S. (2014). *Makine Öğrenmesi Yöntemleriyle Gerçek Olmayan Tüketici Yorumlarının Tespiti*. (Doktora Tezi), İstanbul Kültür Üniversitesi, Fen Bilimleri Enstitüsü.
- Gür, H. (2018). *Kredi Değerlendirmesi için Bulanık Karar Ağaçları Tabanlı Bir Karar Destek Sistemi*. (Yüksek lisans Tezi), Gazi Üniversitesi, Bilişim Enstitüsü.
- Güzeller, C. O. ve Çeliker, N. (2017). Geçmişten günümüze gastronomi bilimi: bibliyometrik analiz. *Journal of Tourism and Gastronomy Studies*, 5/Special Issue2, 88-102. Doi: 10.21325/jotags.2017.114.

- Hotamışlı, M., ve Erem, I. (2014). Muhasebe ve finansman dergisi'nde yayınlanan makalelerin bibliyometrik analizi. *Muhasebe ve Finansman Dergisi*, (63), 1-20. <https://doi.org/10.25095/mufad.396474>.
- İşkesen, S. E. (2014). *Development and Implementation of a Price Prediction System Using Machine Learning Techniques*. (Yüksek lisans Tezi), Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü.
- Kapucu, C., ve Çubukçu, M. (2019). Fotovoltaik sistemlerde topluluk öğrenmesi temelli hata tespiti. *Bilişim Teknolojileri Dergisi*, 12(2), 83-91. <https://doi.org/10.17671/gazibtd.508475>.
- Khalid, S., Khalil, T., & Nasreen, S. (2014, August). A survey of feature selection and feature extraction techniques in machine learning. In 2014 Science and Information Conference (pp. 372-378). IEEE. Doi: 10.1109/SAI.2014.6918213.
- Köksal, İ. H. (2018). *Fashion Trend Prediction Using Machine Learning Techniques*. (Yüksek lisans Tezi), Yaşar Üniversitesi, Fen Bilimleri Enstitüsü.
- Kutlugün, M. A. (2017). *Gözetimli Makine Öğrenmesi Yoluyla Türe Göre Metinden Ses Sentezleme*. (Yüksek lisans Tezi), İstanbul Sabahattin Zaim Üniversitesi, Fen Bilimleri Enstitüsü.
- Lawani, S. M. (1981). Bibliometrics: its theoretical foundations, methods and applications. *Libri*, 31, 294.
- Li, Y., Xu, Z., Wang, X., & Wang, X. (2020). A bibliometric analysis on deep learning during 2007–2019. *International Journal of Machine Learning and Cybernetics*, 1-20. <https://doi.org/10.1007/s13042-020-01152-0>.
- Martínez, M. A., Cobo, M. J., Herrera, M., & Herrera-Viedma, E. (2015). Analyzing the scientific evolution of social work using science mapping. *Research on Social Work Practice*, 25(2), 257-277. Doi: 10.1177/1049731514522101.
- Mitchell, T. M. (1997). *Machine Learning*. USA: McGraw Hill Companies, Inc.
- Onan, A. (2018). A clustering based classifier ensemble approach to corporate bankruptcy prediction. *Alphanumeric Journal*, 6(2), 365-376. <http://dx.doi.org/10.17093/alphanumeric.333785>.
- Onan, A. (2016). *Görüş Sınıflandırma için Makine Öğrenmesi Algoritmalarına Dayalı Bir Yöntem Tasarımı ve Gerçekleştirimi*. (Doktora Tezi), Ege Üniversitesi Fen Bilimleri Enstitüsü.
- Özdemir, S. (2014). *A Decision Tree Based Intrusion Detection System with Bootstrap Aggregating, Discretization, and Feature Selection*. (Yüksek lisans Tezi), Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü.
- Öztemel, E. (2003). *Yapay Sinir Ağları*. İstanbul: PapatyaYayıncılık.
- Öztürk, S. (2015). *An Experimental Evaluation for Fuzzy Decision Trees*. (Yüksek lisans Tezi), Çukurova Üniversitesi, Fen Bilimleri Enstitüsü.
- Pınar, M., Okumuş, O., Turgut, U. O., Kalıpsız, O., ve Aktaş, M. S. (2017). Büyük veri içeren öneri sistemleri için hiperparametre optimizasyonu. *Ulusal Yazılım Mühendisliği Sempozyumu*, 22-272.
- Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, 25(4), 348-349.
- Samiee, Saeed – Chabowski, Brian R (2012). Knowledge structure in international marketing: a multi-method bibliometric analysis. *Journal of the Academy of Marketing Science*, 40 (2), 364-386. Doi: 10.1007/s11747-011-0296-8.
- Sezer, Ü. (2008). *Karar Ağaçlarının Birliktelik Kuralları ile İyileştirilmesi*. (Yüksek lisans Tezi), Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü.
- Subanya, B., & Rajalaxmi, R. R. (2014, February). Feature Selection Using Artificial Bee Colony for Cardiovascular Disease Classification. In 2014 International Conference on Electronics and Communication Systems (ICECS) (pp. 1-6). IEEE.
- Tanyıldızı, E., ve Demirtaş, F. (2019). Hiper Parametre Optimizasyonu. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-5). IEEE.
- Tozlu, İ. (2019). *Simplifying Balance Sheet Adjustment Process in Commercial Loan Applications Using Machine Learning Methods*. (Yüksek lisans Tezi), İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü.
- Tsai, C. F., Eberle, W., & Chu, C. Y. (2013). Genetic algorithms in feature and instance selection. *Knowledge-Based Systems*, 39, 240-247. <https://doi.org/10.1016/j.knosys.2012.11.005>.

- Türker, S. (2019). *Zararlı Android Yazılımlarının Makine Öğrenmesi ile Ailelerine Göre Sınıflandırılması*. (Yüksek lisans Tezi), Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü.
- Verma, A., & Mehta, S. (2017, January). A Comparative Study of Ensemble Learning Methods for Classification in Bioinformatics. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence* (155-158). IEEE.
- Wang, Z., Wang, Y., & Srinivasan, R. S. (2018). A novel ensemble learning approach to support building energy use prediction. *Energy and Buildings*, 159, 109-122. <https://doi.org/10.1016/j.enbuild.2017.10.085>.
- Yıldız O., Tez M., Bilge H. Ş., Akcayol M. A., Güler İ. (2012). Meme Kanseri Sınıflandırması İçin Gen Seçimi. *IEEE 20nd Signal Processing and Communications Applications Conference; 18-20 April 2012; Muğla, Turkey*: IEEE, 988-991.

## SUMMARY

Bibliometry is an important tool for evaluating and analyzing academic research developed in countries, universities, research centers, research groups, journals, authors, and a particular field or discipline. It provides objective criteria for evaluating academic research and is therefore increasingly valued as a tool for evaluating academic quality and productivity. Bibliometry contributes to the advancement of science in many different ways: assessing progress, identifying the most reliable sources of scientific publications, establishing the academic basis for evaluating new developments, identifying major scientific actors, developing bibliometric indexes to evaluate academic output, current issues in the literature, methods, findings, and It creates a roadmap for future researchers by revealing the issues that are out of date, determining the current situation and trends, and determining the existing gaps in a particular research discipline.

Machine learning is a field of artificial intelligence that enables the system to create a model using the learning gained from its past experiences and to make a prediction for future situations. Numerous machine learning algorithms have been developed to solve problems. When the studies are examined, different approaches have been proposed to increase the performance of algorithms. These approaches generally; It is in the form of applying machine learning algorithms after choosing the feature, developing ensemble learning methods, applying different optimization algorithms in determining the parameters of machine learning algorithms, improving machine learning algorithms by using heuristic or different methods. The focus of the study is to examine machine learning and decision tree algorithms based on the supervised learning strategy, to avoid duplication in future studies by examining different approaches developed to improve performance, thus providing researchers with a perspective.

325 theses were reached as a result of the scanning performed in May 2020 by using the keyword "machine learning", in the database of the Council of Higher Education's National Thesis Center. 43 theses were reached as a result of the scanning performed in May 2020 by using the keywords "decision tree", in the database of the Council of Higher Education's National Thesis Center. Theses reached; It was examined in detail in terms of bibliography taking into account the year, thesis type, language, university, institute, department, the number of pages, the software used, the methods used, the keywords. the data were obtained and

transferred to the Excel program. In this software calculations, graphs and tables were made.

As a result of bibliometric analysis in the field of machine learning; There was a significant increase in the number of theses published after 2016, It has been observed that the majority of the theses are master's thesis, the university that contributes the most to the field is Istanbul Technical University, Support Vector Machines is the most frequently used algorithm, ensemble learning methods are not used enough, the use of embedded and wrapper approaches is not very common, and when we look at the developed community methods there is a gap in this area, Python is the most common programming language / software.

As a result of bibliometric analysis in the field of decision trees; the first thesis was written in 2000, It has been observed that the majority of the theses are master's thesis, the university that contributes the most to the field is Van Yüzüncü Yıl University,, CHAID is the most frequently used algorithm, ensemble learning methods are not used enough, this field where only one study uses attribute selection, parameter optimization methods are not used, SPSS is the most common programming language / software.

It is thought that this study, in which the accessed theses are analyzed in terms of many parameters, will guide researchers who will work in this context in many ways (method, program, hybrid approach to be developed, etc.).