T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

METİN MADENCİLİĞİ YÖNTEMLERİ İLE 1991-2021 YILLARI ARASINDA ZOOTEKNİ ALANINDA YAZILAN TEZLERİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Fatih CAMCI
DANIŞMAN : Prof. Dr. Abdullah YEŞİLOVA

VAN-2022

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

METİN MADENCİLİĞİ YÖNTEMLERİ İLE 1991-2021 YILLARI ARASINDA ZOOTEKNİ ALANINDA YAZILAN TEZLERİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Fatih CAMCI

VAN-2022

KABUL VE ONAY SAYFASI

Zootekni Anabilim Dalı'nda Prof. Dr. Abdullah YEŞİLOVA danışmanlığında, Fatih CAMCI tarafından sunulan "Metin Madenciliği Yöntemleri ile 1991-2021 Yılları Arasında Zootekni Alanında Yazılan Tezlerin İncelenmesi" isimli bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince 30/06/2022 tarihinde aşağıdaki jüri tarafından oy birliği ile başarılı bulunmuş ve yüksek lisans tezi olarak kabul edilmiştir.

Başkan: Prof. Dr. AbdullahYEŞİLOVA	İmza:
Üye: Doç. Dr. Yılmaz KAYA	İmza:
Üye: Dr. Öğr. Ü. Hasan ÇELİKYÜREK	İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun …./..../….. tarih ve ..... sayılı kararı ile onaylanmıştır.

İmza
Prof. Dr. Harun AKKUŞ
Enstitü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

Fatih CAMCI

ÖZET

METİN MADENCİLİĞİ YÖNTEMLERİ İLE 1991-2021 YILLARI ARASINDA ZOOTEKNİ ALANINDA YAZILAN TEZLERİN İNCELENMESİ

CAMCI, Fatih
Yüksek Lisans Tezi, Zootekni Anabilim Dalı
Tez Danışmanı: Prof. Dr. Abdullah YEŞİLOVA
Temmuz 2022, 63 sayfa

Bu tez çalışmasında, zootekni alanında 1991-2021 yılları arasında yazılmış tezlerin metin madenciliği yöntemleri ile analiz edilmesi amaçlanmıştır. Çalışma kapsamında toplam 1665 tezin İngilizce özet kısımları incelenmiştir. Çalışmanın veri kaynağını ise yıllara göre ayrılmış olan bu tezler oluşturmaktadır. Veri kaynağımız metinsel yapıda olduğu için analize uygun hale getirmek adına öncelikle gerekli gürültülerden (noktalama işaretleri, sayılar, stopwordsler vb.) temizlenmiştir. Python dili ve kütüphaneleri ve web tabanlı bir yazılım olan Voyant kullanılarak bu tezlerin metin madenciliği yöntemleri ile analizleri yapıldığında her sene alanda yazılan tezlerin farklı konular üzerinde yoğunlaştığı tespit edilmiştir. Yıllara göre tek tek analizleri yapılan tez çalışmalarının son olarak hepsini tek bir metinsel veri kaynağı haline getirerek son 30 yılda en çok üzerinde durulan tez çalışma konuları ortaya konmuştur.

Anahtar kelimeler: Metin madenciliği, Python, Veri görselleştirme.

ABSTRACT

INVESTIGATION OF THESIS WRITTEN IN THE FIELD OF ZOOTECHNICS BETWEEN TEXT MINING METHODS AND THE YEARS OF 1991-2021

CAMCI, Fatih
M.Sc. Thesis, Department of Animal Science
Supervisor: Prof. Dr. Abdullah YEŞİLOVA
July 2022, 63 pages

In this study, it is aimed to analyze the text mining methods of theses written between 1991-2021 in the field of zootechnics. Within the scope of the study, the abstracts of a total of 1665 theses were examined. The data source of the study is these theses, which are separated by years. Since our data source is in a textual structure, first of all, necessary noises (punctuation marks, numbers, stopwords, etc.) were cleaned in order to make it suitable for analysis. When these theses are analyzed with text mining methods using Python language and libraries, it has been determined that the theses written in the field every year focus on different topics. Finally, the thesis studies, which were analyzed one by one according to the years, were brought together as a single textual data source and the most emphasized thesis studies in the last 30 years were revealed.

Keywords: Natural language processing, Python, Text mining.

ÖN SÖZ

Tez çalışma sürecimin her aşamasında bana yol gösteren, kıymetli vaktini ayıran her zaman çekinmeden yanına gidip yardım aldığım danışman hocam sayın Prof. Dr. Abdullah YEŞİLOVA ’ya teşekkürlerimi sunarım. Tüm hayatım boyunca bana madddi ve manevi destek olan kıymetli annem Sultan CAMCI ve değerli büyüğüm babam Salahattin CAMCI’ya, ayrıca verdiği desteklerle beni hiçbir zaman yalnız bırakmayan değerli eşim Şeyma CAMCI’ya sonsuz teşekkürlerimi sunarım. Oğlum Hamza CAMCI’nın daha güzel çalışmalar yapması temennisiyle…

2022

Fatih CAMCI

İÇİNDEKİLER

Sayfa

ÇİZELGELER LİSTESİ

Çizelge **Sayfa**

Çizelge **Sayfa**

ŞEKİLLER LİSTESİ

Şekil **Sayfa**

Şekil **Sayfa**

Şekil **Sayfa**

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	**Açıklama**
%	Yüzde
°C	Santigrat derece

Kısaltmalar	**Açıklama**
NLP	Doğal Dil İşleme

1. GİRİŞ

Metin madenciliği (text mining) en kısa tarifiyle yapılandırılmamış (unstructured) veri kaynağı olan metni veri kaynağı olarak alan ve bu kaynaktan önceden bilinmeyen, anlamlı ve yapılandırılmış (structured) bilgiler ortaya çıkarmaya yarayan bir veri madenciliği (data mining) alt dalıdır. Burada yapılandırılmamış terimi ile kastedilen şey diğer veri setleri gibi elimizde hazır satır ve sütunlara sahip, klasik bir veri tabanında ya da tabloda tutulan verinin olmamasıdır. Bu anlamlı bilgilerin elde edilme süreci diğer veri kaynakları ile karşılaştırıldığında uzun ve zorlu bir süreç olarak karşımıza çıkmaktadır.

Metin madenciliği, özel amaçlar için metinden değerli bilgileri çıkarmak adına, metnin analiz edilmesi işlemidir (Visa, 2001). Dünyada üretilen veri miktarı önceki yıllar ile kıyas ediliğinde hem çok hızlı bir şekilde artmakta hem de çeşitli hale gelmektedir. Bugün hepimiz adeta veri üreten ve paylaşan makineler haline geldik. Her gün atılan milyonlarca tweet, milyarlarca whatsapp mesajı, web sitelerinde paylaşılan blog yazıları ve e-posta gibi çeşitli kitle iletişim araçları ile zettabayt dediğimiz boyutlara varan veriler üretmekteyiz. Bu verilerin büyük bir kısmı da metinsel olarak üretilen sosyal medya paylaşımları, e-postalar, arama motorları kayıtları vb. gibi verilerden oluşmaktadır. Bu verilerin işlenmesi, faydalı bilgiler elde edip ona göre reaksiyon almak ise tüm dünya devi şirketlerinin ve devletlerinin en önemli amacı haline gelmiştir.

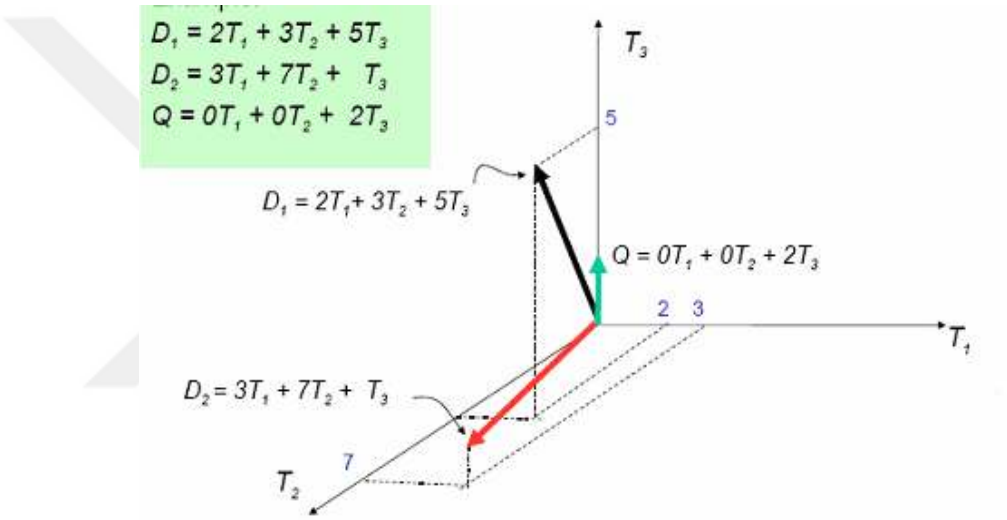
Metin madenciliği teknikleri ile işlenen metinden istenilen amaçlara göre çeşitli bilgiler elde etmek mümkündür. Bunlar arasında metinlerden özet çıkarma, metin yazarının tespit edilmesi, metin içerisindeki isim, sıfat, zamir gibi veya telefon numarası adres gibi verilerin elde edilmesi, kullanılan kelimelerin sıklıklarının tespiti gibi birçok faydalı ve anlamlı bilgi elde etmek mümkündür. Elde edilen bu bilgiler yapısal olarak depolanıp çeşitli veri görselleştirme araçları ile kullanıcıların bilgisine sunulmaktadır.

Bu çalışmada metin madenciliği teknikleri kullanılarak son 30 yılda (1991-2021) zootekni alanında yazılan tezlerde hangi konuların daha çok işlendiğini bulmak amaçlanmıştır. Veri kaynağı olarak ise bu tezlerin ingilizce özet kısımları kullanılmıştır. Çalışmada kullanılan bu metinsel veri kaynakları önce metin ön işleme sürecinden geçirilmiş daha sonra da üzerinde işlem yapılacak şekilde çeşitli tablolarda depolanmıştır. Metin ön işleme, tablolaştırma, görselleşrime gibi işlemler ise Python programlama dili

ve gerekli kütüphanlerin eklenmesi ile yapılmıştır. Çalışma sonucunda en çok tekrar eden kelimeler ilgili yıllarda yazılan tezlerde o konunun daha çok işlendiğini göstermektedir. Veri bilimine önem veren tüm dünya ülkelerinde metin madenciliği alanında önemli çalışmalar yapılmaktadır. Yapılan bu çalışmanın zootekni alanında yapılacak çalışmalara katkı sağlayacağı düşünülmektedir.

2. KAYNAK BİLDİRİŞLERİ

Metinsel haldeki Anadolu Ajansı haberlerinden derlenen yaklaşık 1370 adet Türkçe haber metni veri kaynağı olarak kullanılmıştır. Bu metinsel haldeki veriler analiz edilmeden önce çeşitli ön işleme basamaklarından geçirilmiştir. Bu işlemler metinsel haldeki verileri matematiksel olarak ifade edebilmek için kullanılan işlemlerdir. Dökümanları vektörel olarak ifade etmek için sınıflandırmak istenen doküman ve eğitim dökümanları vektörel uzayda ifade edilmiştir.

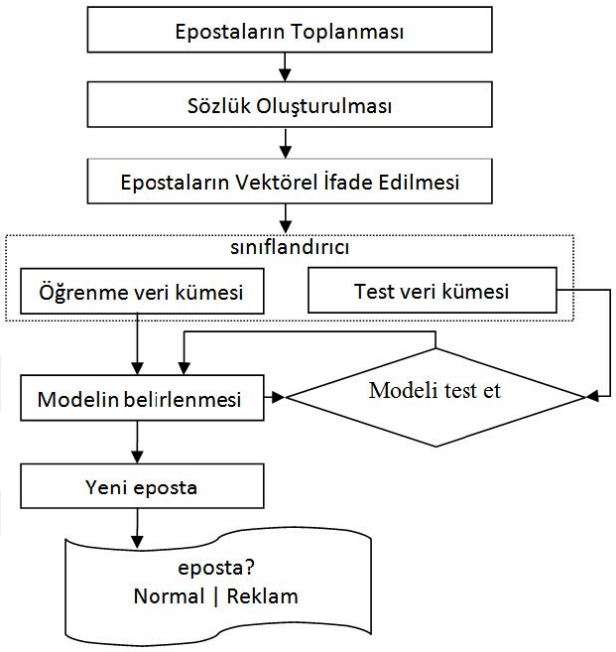


Şekil 2.1. Vektör uzay modeli.

Kaynak metindeki kategorileri otomatik olarak tahmin etmek için ise kategorisi önceden belirlenen makaleler ile veriler eğitilmiştir. Ön işleme adımlarından sonra ise metin sınıflamalarında kullanılan (Naive Bayes, k-NN) ve 3 farklı vektör tanımlama yöntemi (Binary vektör oluşturma, Frekans sıklığı ile vektör oluşturma, TF-IDF Ağırlıklandırma ile Vektör oluşturma) incelenmiş ve daha sonra VisualBasic.NET dili yazılan program vasıtasıyla bu metinsel veriler kategorilere ayrılmış ve yukarıda ismi geçen sınıflama tekniklerinin doğruluk olasıkları değenlendirilmiştir (Pilavcılar ve ark., 2007).

Kitle iletişim araçları içerisinde en çok kullanılan ve suistimal edilen araç hiç şüphesiz e-postalardır. Mail kutularına düşen Türkçe içerikli bu e-postaların spam yani

gereksiz mi yoksa faydalı mı olduğu metin madenciliği yöntemleri ile tespit edilmiştir. Bu amaçla 400'ü normal, 400'ü reklam içerikli olarak 800 eposta kullanılmış olup bu epostalardan rastgele seçilen 200 tanesi test, 600 tanesi ise eposta eğitim verisi olarak kullanılmıştır. Epostalar üzerinde matematiksel olarak çalışabilmek için bazı işlemlerden geçirilmiştir. Bu işlemlerin en başta geleni ise toplanan epostalardaki içerik kısımlarının Zemberek Kütüphanesi yardımıyla kelime köklerine ayrıştırılarak sözlük oluşturulmuştur.



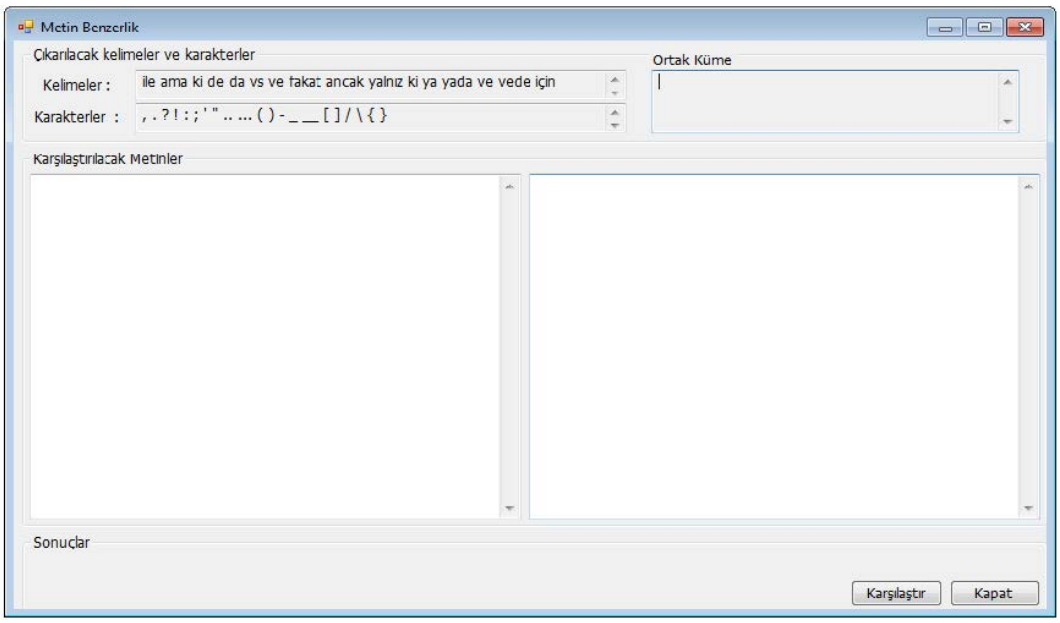
Şekil 2.2. Gerçekleştirilen çalışmanın yapısı.

Daha sonra ise kökleri bulunup sözlük haline gelen eposta verileri vektör uzayında tanımlanmıştır. Vektör oluşturmada öznitelik seçimi için "Bo W-Kelime Çantası" yöntemi kullanılmıştır. Bu seçimde 2 tür vardır. Bir tanesi tüm kelimeleri kapsayan 4263 kelime kökü, diğeri ise en çok tekrar eden yani frekans sıklığı fazla olan 144 kelimedir.

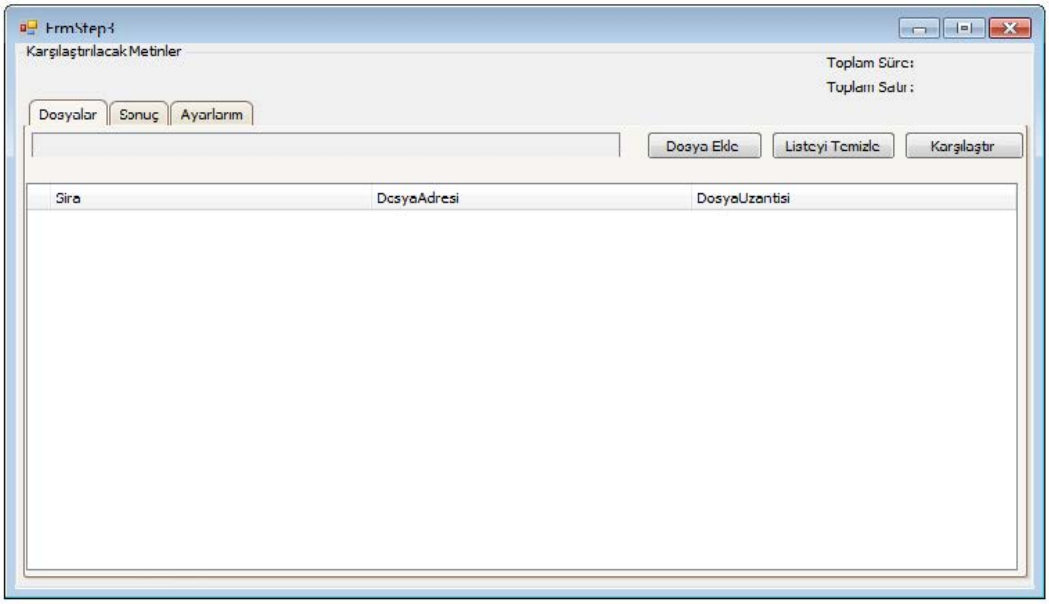
Bu metinsel reklam içeriklerinin sınıflandırılmasında ise Naive Bayes, k En Yakın Komşu ve Destek Vektör Makinesi algoritmaları tercih edilmiştir. Yapılan çalışma sonucunda ise en yüksek sınıflandırma başarısı Naive Bayes ve kNN ile elde edilirken binary vektörün diğer yöntemlere daha başarılı sınıflama yaptığı tespit edilmiştir. (Çalış ve ark., 2013).

Metin madenciliği yöntemleri ve. Net tabanlı C# programlama dili kullanılarak dökümanlar arasındaki benzerliklerin bulunması (örüntü tarama) amaçlamıştır. Sadece 2

doküman arasında değil istenilen n sayısaki doküman arasındaki benzerlikler de bulunmuştur. Uygulamaya yüklenen her dökümandaki cümlenin diğer tüm dökümanlardaki cümleler ile olan benzerlik ilişkisi hesaplanmıştır. Windows Form arayüzünde çalışan bu uygulama için bazı metin ön işleme hazırlıkları yapılmıştır. Bunlar arasında edat, bağlaç ve noktalama işaretlerinin silinmesi gibi aşamalar yer almaktadır. Daha sonra ise cümleler doküman içerisinde tespit edilmiştir. Uygulama için en çok kullanılan metin madenciliği benzerlik algoritmalarından olan kosinüs (cosine) ve jaccard kullanılmıştır. Uygulama 2 aşamada yapılmıştır. İlk uygulama ile basit düzeyde iki cümlenin birbirlerine olan benzerlikleri tespit etmektedir.



Şekil 2.3. Basit düzeyde iki cümlenin benzerlik tespiti.



Şekil 2.4 İleri düzey metinlerin benzerlik tespiti.

Uygulamanın 2.aşamasında ise yalnızca 2 cümlenin benzerliğini bulmaktan ziyade bir bütün olarak metin madenciliğinde kullanılabilecek düzeyde hazırlanmıştır (Döven ve ark., 2013).

Metin madenciliği yöntemleri ile Microsoft. Net programı üzerine inşa edilen ve Google arama motoru üzerinden yapılan anahtar kelime aramalarının sonucunda çıkan sitelerin gerçekten arama yapılan kelime ile uyumlu olup olmadığını ortaya çıkaran metin madenciliği tabanlı bir web sitesi sınıflandırma aracı tasarımı yapılmıştır. Arama sonucunda yer alan sitelerin e-ticaret ya da diğer kategorilere ait olma oranlarını bularak kullanıcıları gerçek sektör sitelerine yönlendirmektedir.

Naïve Bayes sınıflandırma algoritması kullanılarak, e-ticaret sitesi olan ve olmayan web sitelerine ait veriler 2 grup halinde toplanıp sql tablosuna kaydedilmiştir.

Column Name	Data Type	Allow Nulls
id_no	decimal(15, 0)	☐
kelime	varchar(50)	☑
ana_cumle	varchar(250)	☑
site_adres	varchar(500)	☐
kategori	varchar(1)	☑
olasilik	decimal(15, 13)	☑
		☐

Şekil 2.5. Verilerin sql tablosuna girilmesi

E-Ticaret sitesi olduğu bilinen ve bilinmeyen toplamda 20 ye yakın web sitesi ele alınmıştır. Microsoft. Net ile geliştirilen uygulama sayesinde sitelerin açılış sayfalarıjdaki html kodları kelimelerine ayrılarak veritabanındaki datamin isimli tabloya kaydedilerek analizleri yapılmıştır. Veri tabanındaki kelimelerin sitelerin açılış sayfalarında bulunma olasılığına göre bir formül kullanılmıştır.

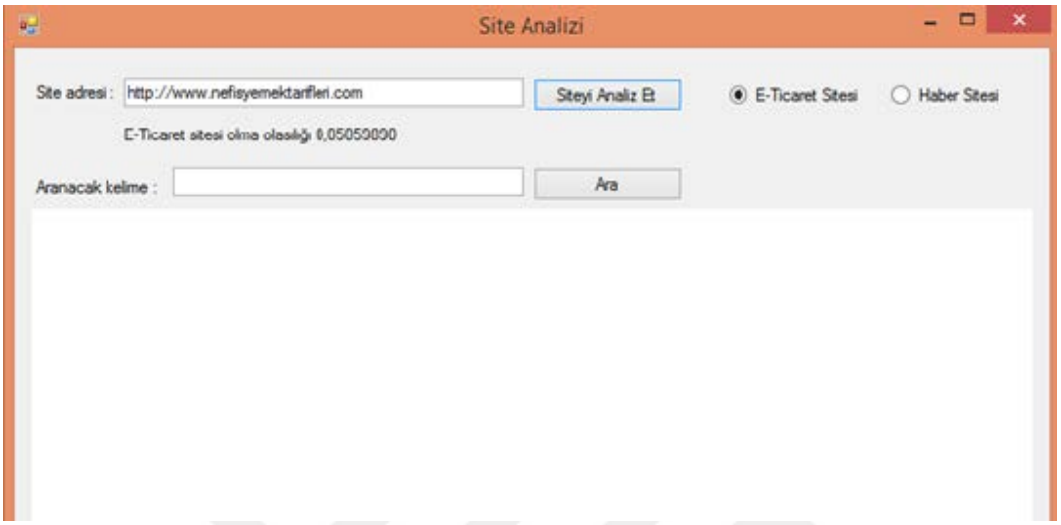
$$Oran = \frac{(kelimelerin\ kategorideki\ sayısı+1)}{(öğretilmiş\ kelime\ sayısı+kategorideki\ toplam\ kelime\ adedi)} \quad (2.1)$$

Örnek olarak “indirim” kelimesina bakacak olursak aşağıdaki gibi bir sonuçla karşılaşıyoruz.

	id_no	kelime	ana_cumle	site_adres	kategori	olasilik
1	34252	indirim	%50 indirim kampanyasını kaçırmayın	http://www.platinmarket.com/	H	0.8334401846000
2	9053	indirim	Baby Me Yenidoğan ıslak mendillerinde indirim	http://www.e-bebek.com/	E	3.9146690125000
3	9058	indirim	Baby Me Yenidoğan ıslak mendillerinde indirim	http://www.e-bebek.com/	E	3.9146690125000
4	33071	indirim	Bayiye ouml;zel indirim uygulama	http://www.noramedya.com/	H	0.8334401846000
5	33075	indirim	Bayiye ouml;zel indirim uygulama	http://www.noramedya.com/	H	0.8334401846000
6	594180	indirim	Çoklu İndirim Modülü Kullanımlarında Simge Eklenmiştir	http://www.tsoft.com.tr/	H	0.8334401846000
7	3885	indirim	Hem boyner com tr'de hem de Boyner Mağazaları'nd...	http://www.boyner.com.tr/	E	3.9146690125000
8	3907	indirim	Hem boyner com tr'de hem de Boyner Mağazaları'nd...	http://www.boyner.com.tr/	E	3.9146690125000
9	772	indirim	Hem ykm com tr'de hem de Ykm Mağazaları'nda size...	http://www.ykm.com.tr/	E	3.9146690125000
10	794	indirim	Hem ykm com tr'de hem de Ykm Mağazaları'nda size...	http://www.ykm.com.tr/	E	3.9146690125000
11	10270	indirim	indirim	http://www.sahibinden.com/	E	3.9146690125000

Şekil 2.6. İndirim kelimesinin kullanılma durumu.

Bu tablo bize “indirim” kelimesinin E-Ticaret sitelerinde daha yükek oranlarda karşımıza çıktığını göstermektedir. Bu şekilde tüm kelimeler için oran hesaplaması yapılır ve makine öğrenmesi için bir veri havuzu elde edilmiş olur.



Şekil 2.6. Uygulama arayüz ekranı.

Uygulama arayüzü yukarıdaki gibidir ve arama sonuçları Google üzerinde ilk 20 web sitesi için yapılmaktadır. Bu sitelerin her biri için ayrı ayrı Naïve Bayes sınıflandırma formülleri üzerinden oran hesaplaması yapılır (Erten ve ark., 2015).

Sosyal medyada müşteri etkileşimlerini inceleyen bir çalışmada kuyumculuk ve saat ticareti yapan bir firmanın instagram hesabından, paylaşım yaptığı içeriklerine göre belirlenen 6 aylık dönemdeki müşteri mesajları analiz edilmiştir. Bu 6 aylık dönemde (2019 Kasım-2020 Nisan) 1803 kişi tarafından gönderilen mesajlar ilk başta manuel olarak çekilmiş ve çeşitli ön işlemlerden geçirilmiştir. Daha sonrasında ise açık kaynak kodlu metin madenciliği hizmeti veren voyant sitesi ile bağımsız ve birbiri ardına en sık tekrar eden kelime ve cümle grupları tespit edilerek kelime ağaçları oluşturulmuştur. Kelime ağaçlarından elde edilen veriler kodlama, temaların bulunması, kodların ve temaların düzenlenmesi ve bulguların yorumlanması olarak 4 aşamada incelenmiştir.

Araştırmanın yapıldığı tarihte firmanın 18.500 adet takipçisi olduğu ve bu takipçilerin belirli bir yaş ve cinsiyet dağılımına sahip olduğu belirlenmiştir. Analiz sonuçlarına göre firmanın özel günlerde ve kampanyalı ürün sattığı dönemlerde müşteri etkileşimim çok fazla olduğu, kampanya, yarışma, çekiliş gibi içerik paylaşımlarının daha az yapıldığı dönemlerde ise etkileşim oranının düşük geldiği tespit edilmiştir.

Etkileşimlerin başarılı ya da başarısız olması ise aşağıdaki Akıncı (2019) formülüne göre hesaplanmıştır.

*(Toplam Beğeni + Toplam Yorum) / Takipçi Sayısı / Paylaşılan İçerik Sayısı*100 (2.2)*

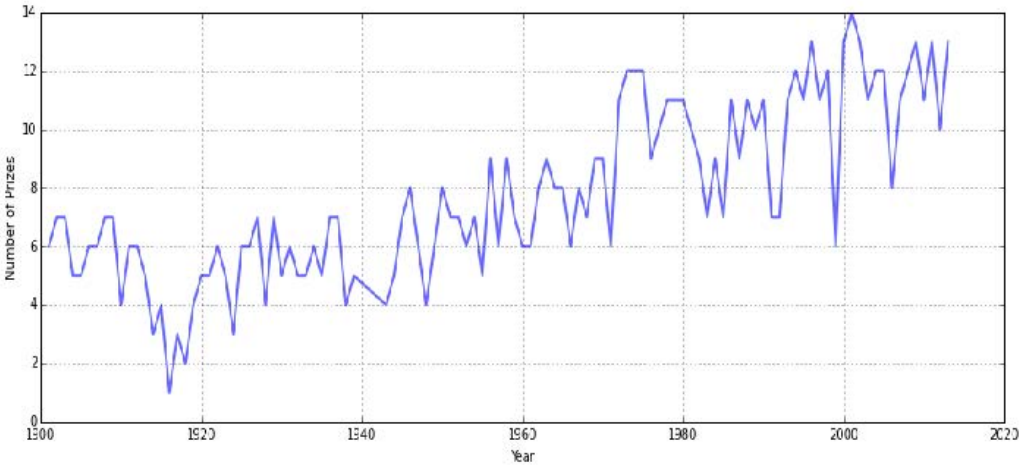
Bu formüle göre bir etkileşim tablosu ortaya konmuştur. Kampanya günleri daha çok etkileşim almış ve başarılı bulunmuştur.

Çizelge 2.1. Aylara göre paylaşılan içerik ve alınan tepkiler

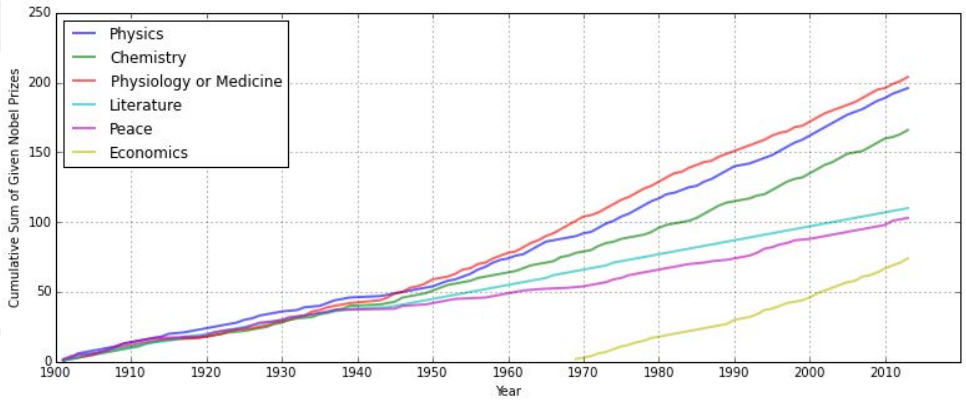
Ay	**Paylaşılan İçerik**	**Toplam Beğeni**	**Toplam Yorum**	**Etkileşim**
Kasım	9	490	675	1
Aralık	24	9329	26	2
Ocak	13	7817	15	3
Şubat	8	4824	112	3
Mart	8	3148	90	2
Nisan	6	2113	155	1

Fumegri (2020), formüller sonucunda çıkan oranın %1-2 aralığında ise başarılı olduğunu bildirmiştir. En yüksek etkileşim oranının ise %3 ile ocak ve şubat ayları olduğu tespit edilmiştir (İşler, 2021).

Metin madenciliği yöntemleri kullanılarak Wikipedia sitesi üzerinde bulunan nobel ödülü kazananlar tablosu analiz ve görselleştirmesi yapılmıştır. Veri kaynağı olarak Wikipedia, verileri analiz ve görselleştirme için ise Python, Anaconda Navigator, Jupyter Lab, Matplotlib araçları kullanılmıştır. Web Kazıma (Web Scraping) yöntemleri ile HTML formatına çevrilen bu metinsel veriler analize uygun hale getirilmeden önce Tokenizasyon, metindeki boşlukları ve kullanılmayan kelimeleri kaldırma, normalleştirme gibi işlemlerden geçirilip array içerisine atılmıştır. Ön işleme aşamasından sonra ise konular, url bilgileri ve yıllara göre dağılımları tablo halinde gösterilmiştir. Son olarak Matplotlib kütüphanesi ile yıllık verilen ödül sayısı ve verilen nobel ödüllerinin kümülatif toplamları görsel olarak tablolaştırılmıştır (Kâhya, 2021).



Şekil 2.7. Yıllara göre verilen ödül sayısı (Kâhya, 2021).



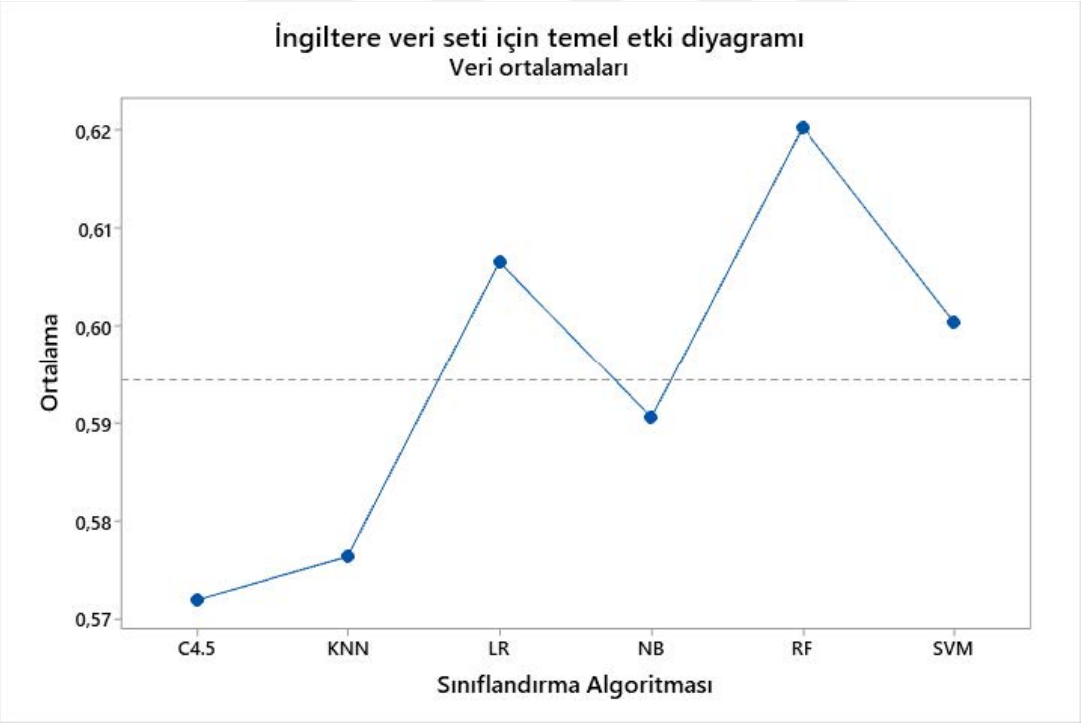
Şekil 2.8. Yıllara göre verilen nobel ödüllerinin kümülatif toplamı (Kahya, 2021).

Covid-19 salgını ilk görüldüğü günden bugüne kadar pek çok insanı etkilemiş ve büyük bir pandemiye dönüşmüştür. Dünyanın farklı bölgelerinde uygulanan sağlık politikaları ve tedavi yöntemlerine yönelik olarak birçok sosyal medya gönderisi ve haber makaleleri bulunmaktdır. Network ortamındaki gayri resmi kanallar aracılığı ile yapılan bu gönderi ve paylaşımların korona dönemindeki gönderilerin önemli bir bölümünü oluşturduğu, dünyadaki ilk ve zamanlı haber kaynakları olduğu tespit edilmiştir.

İngiltere ve İspanya'da Covid-19 sürecinde 2020 yılının mart, mayıs ve temmuz aylarında yayınlanan 299'ar tane haber makalesi metinsel veri kaynağı olarak kullanılmıştır. Bu veri seti zaman tahmini için zaman tahmini için farklı zamanlarda toplanan haber metinlerine yönelik anahtar sözcükleri öznitelik olark içermektedir. Bu metinsel verilerin temsilinde 3 temel n-gram modeli olan (1-gram, 2-gram, 3-gram) temsilleri ve cümlenin öğeleri 2-gram ve cümle öğeleri 3-gram öznitelikleri kelime/cümle

öğe çiftleri, karakter n-gram (n=2) ve karakter n-gram (n=3) öznitelikleri ve bu özniteliklerin elde edilen verinin öznitelik gruplarının etkinleri değerlendirilmiştir. Çalışmanın analiz işlemlerinden temel öğrenme algoritmaları ve metin temsil yöntemlerinde ise Python programlama dili Scikit-Learn Kütüphanesi kullanılmıştır. Oluşturulan bu öznitelik gruplarının başarım değerlendirmesinde ise yaygın olarak kullanılan makine öğrenmesi algoritmalarından olan Naïve Bayes, Lojistik Regresyon, Destek Vektör Makineleri, C4.5 karar ağacı, K-en yakın komşu algoritması ve Rastgele orman algoritmaları kullanılmıştır.

Deneysel analizlerdeki en yüksek başarı oranının, metin temsil yöntemi olarak 1-gram özniteliklerin karakter tabanlı 3-gram modeli ile kullanıldığı zaman elde edildiği görülmüştür. Temel sınıflandırma algoritmaları arasından ise en yüksek başarının Rastgele orman, ikinci başarının ise Lojistik Regresyon algoritmasından elde edildiği tespit edilmiştir (Onan, 2021).



Şekil 2.9. İngiltere için veri görseli.

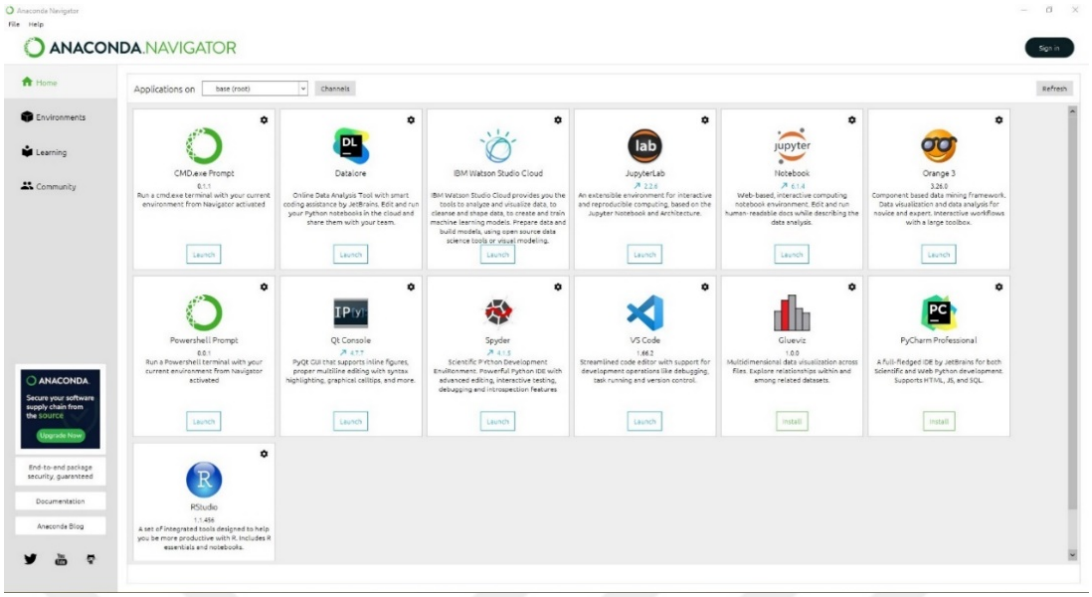
3. MATERYAL VE YÖNTEM

3.1. Materyal

Çalışmada kullanılan veri setini 1991-2021 (07.05.2021) yılları arasında Zootekni Ana Bilim Dalı alanında yazılmış 1665 tezin İngilizce özet kısımları oluşturmaktadır. Bu tezler YüksekÖğretim Kurulu'na ait Tez Merkezi sitesinden her bir yıl için ayrı ayrı filtreleme işlemleri yapılarak elde edilmiştir. Alanda yazılmış tüm yüksek lisans ve doktora tezlerini içeren son 30 seneye ait bu metinsel veri seti toplamda 1945 sayfadan oluşmaktadır.

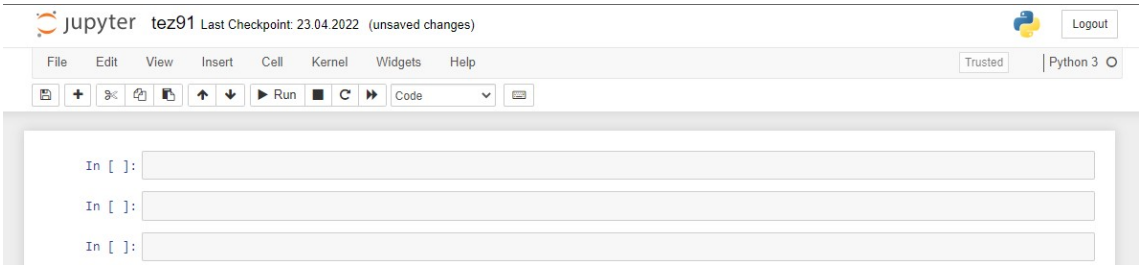
3.2. Yöntem

Çalışmada en iyi sonuçları almak için Veri Bilimi ve Yapay Zekâ çalışmalarında en çok tercih edilen platform olan açık kaynak kodlu Anaconda dağıtımı ve Python programlama dili ve web tabanlı Voyant yazılımı kullanılmıştır. Anaconda dağıtımı veri bilimi ile uğraşan kullanıcılar için yüzlerce python kütüphanesini içinde barındırır ve ihtiyaç olabilecek diğer paketleri kurulum için büyük kolaylıklar sağlamaktadır.



Şekil 3.1. Anaconda dağıtımının arayüzü.

Python programlama kısmı ise Anaconda dağıtımı içerisinde kurulu olarak gelen Jupyter Notebook üzerinden yapılmıştır. Jupyter notebook, python programlama dili için bize interaktif bir ortam sağlamaktadır ve aynı zamanda üzerinde istediğimiz sayıda irili ufaklı Markdownlar (notlar) ekleyebilmekteyiz. Web tabanlı Voyant yazılımı ile de metinsel veri kaynağımıza ait çeşitli kategorilerde analiz sonuçlarımız için görseller alabiliyoruz.



Şekil 3.2. Jupyter notebook arayüzü.

Yapılandırılmamış halde bulunan metinsel haldeki tez verilerimizi analiz etmek ve görselleştirmek için ise python kütüphanelerinden;

- Numpy-pandas (veri analizi),
- Nltk (metin madenciliği),
- Re (Regular Expressions),

- Matpolotlib, Image (görselleştirme),
- WordCloud (metin görselleştirme),

olmak üzere bir çok farklı tipte kütüphane kullanılmıştır.

3.2.1. Metin madenciliği avantajları ve kullanım alanları

Yazılı metinleri okuyup anlamak insanoğlu için her zaman en önemli amaçlardan birisi olmuştur. Hatta işi sadece metinsel verileri okumak olan bazı meslek grupları (editörler) bile vardır. Zaman ne kadar değişirse değişsin bu istek de hiçbir zaman değişmeyecek aksine hızla artacaktır. Peki, bu metinleri bizim yerimize okuyup özetleyecek, sınıflayacak, yazarlarını tanıyacak bir makine olsa ne kadar da güzel olurdu değil mi? Henüz yukarıda bazılarını saydığımız insan-makine etkileşimi ile alakalı konu başlıklarını tam manasıyla başarabildiğimiz söylenemez. Metin madenciliğinin çatısı konumunda bulunan Doğal Dil İşleme (Natural Language Processing) alanının nihai olarak varacağı nokta bir insanın bilgisayarlar ile tıpkı bir insanla konuşuyormuş gibi konuşabilecek olmasıdır.

Metin madenciliği ise özellikle son yıllarda cep telefonlarının ve buna paralel olarak internetin ve sosyal medya mecralarının hızla yayılmasıyla çok popüler bir dal haline gelmiştir. Milyarlarca kullanıcya sahip olan bu sosyal ağlarda konuşulan, tartışılan, beğenilen tüm olaylar dünya üzerindeki pek çok devlet ve şirket tarafından dikkatle takip edilmektedir. Özellikle internet ortamında depolanan bu metinsel haldeki yapılandırılmamış verileri analiz etmek ve onlardan faydalı bilgiler elde etmek için ise Metin Madenciliği (Text Mining) yöntemleri kullanılmaktadır. İnsan gücü ile asla erişilemeyecek olan bu faydalı bilgiler gelecek adına karar vermek için en önemli meta konumunda bulunmaktadır. Sadece dijital dünyadaki belgeler değil yıllar önce el yazısı ile yazılmış olan dökümanları da bazı uygulamalar vasıtasıyla taratıp metinsel veri haline getirebilmekteyiz. Metin madenciliğinin avantajlarından bazıları şunlardır;

- Zaman ve iş gücü tasarrufu sağlar,
- Günlük hatta saatlik dilimlerde ortaya çıkan yeni tartışma başlıklarını çok kolay bir şekilde analiz edilebilir,
- Metinsel haldeki verileri yapılandırıp tablolara aktararak özellikle veri bilimi ve yapay zekâ için veri setleri oluşturmaya yarar,

- Metinsel temeldeki roman, hikaye, makale vb.. dökümanlarını anazli ederek özetlemeye, yazar tanımaya, metin sınıflandırması yapmaya yardımcı olur.

Metin madenciliğinin pek çok farklı kullanım alanı da bulunmaktadır. Bunlardan birkaç tanesini şöyle sıralayabiliriz;

- Mail, mesaj gibi iletilerin sınıflandırılması,
- Metinlerde duygu analizi,
- Sosyal medya gönderilerinin analiz edilmesi,
- Web sitelerinden bilgi çıkarılması,
- Doğal dil işleme çalışmaları için kaynak veri elde etmek.

3.2.2. Metin madenciliği uygulama adımları

Veri analiz işlemlerinde standart olarak elde bir tablo bulunur ve bu tabloda satırlar ve sütunlar arasında çeşitli gözlem ve deneylerle elde edilmiş veriler vardır. Metin madenciliği işlemlerinde ise eldeki verimiz yapılandırılmamış (unstructured) yapıda olduğu için bazı ön işleme adımlarını takip etmek gerekmektedir. Bu adımları bitirdikten sonra ise verimize istediğimiz matematiksel müdaheleyi yapabilmekteyiz.

3.2.3. Gerekli kütüphanelerin kurulması

Metinsel haldeki verilerimizin analiz ve görselleştirme işlemlerine başlamadan önce gerekli kütüphaneleri kurmalıyız. Aşağıdaki şekildeki kütüphanelerin tamamı çalışmaya başlamadan önce kurulmuştur.

```
import numpy as np
import pandas as pd
import re
import nltk
import textblob
from os import path
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt
!pip install nltk
!pip install textblob
!pip install wordcloud
```

Şekil 3.3. Gerekli kütüphanlerin kurulması.

3.2.4. Metnin uygulama alanına aktarılması

Gerekli kütüphane kurulumlarını yaptıktan sonra verilerimizi uygulama arayüzüne aktarmamız (import) gerekmektedir. İndirdiğimiz tez özet verilerinin pdf, word ve txt uzantılı halleri bilgisayara kaydedilmiştir. Biz burada. txt uzantılı dosyaları kullanacağız.

```
tez=open("E:/Yüksek Lisans/TEKLİ ÖZETLER-TR ve ENG/English/2021 tekli özetler/2021 Toplu Özet.txt","r")
icerik=tez.read()
print(icerik)
```

Şekil 3.4. Metnin uygulama alanına aktarılması.

Yukarıdaki şekilde görüldüğü gibi tezlerin özet kısımları içerik adlı değişkene aktarılmıştır. Üzerinde işlem yapmak istediğimizde bu isim ile kullanacağız.

3.2.5. Metnin cümlelerine ayrılması

Metnimizi uygulama alanına aktardıktan sonra sırada cümlelelerine ayırma işlemi (tokenize) var. Analiz için en önemli adımlardan olan tokenleme işlemi için python nltk kütüphanesi içerisinden punkt paketini indirilmiştir ve içeriye aktarılan tezimizi nltk kütüphanesinin sent_tokenize metoduna parametre olarak verilmiştir.

```
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\camci\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

```
True
```

```
cumleler= nltk.tokenize.sent_tokenize(icerik)
print(cumleler)
```

Şekil 3.5. Metnin cümlelere ayrılması.

Bu işlemin ardından içeriye aktarılan metinsel verimizi tek tek cümlelerine ayırmış olduk.

3.2.6. Cümlelerin seriye dönüştürülmesi

Tokenlerine ayrılan cümlelerimizin üzerinde matematiksel olarak işlem yapabilmemiz için ilk adımlardan olan seriye dönüştürme işlemi yapılmıştır. Bunun için pandas kütüphanesinin Series metoduna bir önceki adımda cümlelerine ayırdığımız değişken ismi verilip seriye dönüştürme işlemi tamamlanmıştır.

```
vektor_cumleler=pd.Series(cumleler)
vektor_cumleler
```

Şekil 3.6. Cümlelerin seriye dönüştürülmesi.

3.2.7. Cümlelerin dataframe'e dönüştürülmesi

Seriye dönüştürme işlemini de bitirdikten sonra Excel tablosu gibi satır ve sütunlara sahip bir tablo elde edebilmek için verimizin pandas DataFrame yapısına çevrilmesi gerekmektedir. Önceki adımda seriye dönüştürülen verimiz pandas kütüphanesi DataFrame metoduna parametre olarak verilmiş ve sütun ismi olarak ise cümleler değişkeni seçilmiştir.

```
tez_df_cumleler = pd.DataFrame(tez_seri, columns = ["cumleler"])
tez_df_cumleler
```

Şekil 3.7. Cümlelerin dataframe dönüştürülmesi.

3.2.7. Cümlelerin kelimelerine ayrılması

Kelimeler üzerinde de analiz işlemi yapabilmek için tez verimizin kelime bazında parçalara ayrılması gerekmektedir. Bu işlem için ise python dilinde yerleşik olarak gelen string metotlarından split kullanılmıştır.

```
kelimeler=icerik.split()
kelimeler
```

Şekil 3.8. Cümlelerin kelimelerine ayrılması.

3.2.8. Kelimelerin seriye dönüştürülmesi

Kelimelerine ayırdığımız verimizde istenilen index değerindeki kelimeye ulaşabilmek ve istatistiksel işlem yapabilmek için kelimelerimiz seriye dönüştürülmüştür. Bunun için önceki adımda elde edilen kelimeler değişkeni pandas kütüphanesinden Series metoduna parametre olarak verilmiştir.

```
vektor_kelimeler=pd.Series(kelimeler)
vektor_kelimeler
```

Şekil 3.9. Kelimelerin seriye dönüştürülmesi.

3.2.9. Kelimelerin dataframe'e dönüştürülmesi

Cümleler üzerinde uyguladığımız DataFrame'e dönüştürme işlemini kelimeler üzerine de uygulamak zorundayız. Seri haline çevirdiğimiz değişkenimiz pandas kütüphanesinden DataFrame metoduna parametre olarak verilmiş ve sütun ismi de kelimeler olarak atanmıştır.

```
tez_df_kelimeler = pd.DataFrame(vektor_kelimeler, columns = ["kelimeler"])
tez_df_kelimeler
```

Şekil 3.10. Kelimelerin dataframe dönüştürülmesi.

3.2.10. Metnin tamamının küçük harfe dönüştürülmesi

Veri setimiz üzerinde sağlıklı analizler yapabilmek için gürültü diye tabir edilen bazı fazlalıkların atılması gerekmektedir. Bunlardan ilki tüm metnin tek tipte olması ve rahat okunması için küçük harfe dönüştürme işlemidir. Daha önce dataframe yapısına çevirdiğimiz cümlelerimizin tamamı python dilinde lambda metodu kullanılarak küçük harfe çevrilmiştir.

```
tez_df_cumleler_lower=tez_df_cumleler
tez_df_cumleler_lower

tez_df_cumleler_lower["cumleler"].apply(lambda x: " ".join(x.lower() for x in x.split()))
```

Şekil 3.11. Metnin tamamının küçük harfe dönüştürülmesi.

3.2.11. Metindeki noktalama işaretlerinin silinmesi

Tamamı küçük harfe çevrilen metnimizdeki diğer bir gürültü temizleme işlemi de noktalama işaretlerinin silinmesidir. Noktalam işaretleri analiz işlemleri sırasında bize faydalı bilgi sağlamadı için silinmelidir. Temizleme işlemi için python dilindeki str.replace metodu kullanılmıştır.

```
df_vektor=tez_df_cumleler_lower.copy()
df_vektor

df_vektor= df_vektor.str.replace("[^\w\s]","")
```

Şekil 3.12. Noktalama işaretlerinin silinmesi.

3.2.12. Metindeki sayıların silinmesi

Analiz işlemleri sırasında işimize yaramayan yani faydalı bilgi elde edemediğimiz diğer fazlalıklardan birisi de sayılardır. Bu sayılar da yine str.replace metodu yardımı ile silinmiştir.

```
df_vektor=df_vektor.str.replace("\d","")
df_vektor
```

Şekil 3.13. Sayıların silinmesi.

3.2.13. Metindeki stopwordslerin (durak kelimelerin) silinmesi

Her doğal dil kendi içerisinde bir takım durak kelimeleri barındırmaktadır. Bu kelimeler kendi başına bir anlam ifade etmeyen fakat kelimeleri ve cümleleri birbirine bağlayan yapılardır. Örnek verecek olursak “what, which, this, that” inglizcedeki bazı durak kelimelerindendir. Bu kelime setlerine ilave olarak tezler içerisinde sıklıkla geçen fakat analiz değeri olmayan (Prof, Dr, abstract) gibi bazı kelimeler de stopwords listesine ilave edilerek çalışmanın analize uygun hale getirilmesi sağlanmıştır.

```
sw = set(STOPWORDS)
sw.update(["Prof", "Dr", "THE", "OF", "ABSTRACT","±","%","The","+","Prof.","Dr.",
           "The","the","%","if","If","in","SUMMARY","abstract","THE","OF","+","±"])

tez_df_cumleler_sw=tez_df_cumleler["cumleler"].apply(lambda x: " ".join(x for x in x.split() if x not in sw))
tez_df_cumleler_sw
```

Şekil 3.14. Stop wordslerin temizlenmesi.

3.2.14. Tokenization (cümleleri liste halinde kelimelere ayırma)

Metinsel verimizdeki her bir cümleyi kendi içinde parçalara ayırıp liste halinde görüntülemek için TextBlob kütüphanesi kullanılmıştır.

```
from textblob import TextBlob

TextBlob(tez_df_cumleler_sw["cumleler"][1]).words
#1.Cümlemizdeki kelimeleri liste haline getirmiş olduk.
```

Şekil 3.15. Cümlelerin liste halinde kelimelere ayrılması.

3.2.15. Lemmatization (cümlelerdeki kelimeleri kök halinde gösterme)

Metinde geçen kelimelerin ek almadan yalın halde gösterimi için yine TextBlob kütüphanesi kullanılmıştır.

```
from textblob import Word
nltk.download("wordnet")
tez_df_cumleler_sw["cumleler"].apply(lambda x: " ".join([Word(word).lemmatize() for word in x.split()]))
```

Şekil 3.16. Kelimeleri kök halinde gösterme.

3.2.16. Metin görselleştirme (bar plot)

Metin görselleştirme işlemleri için ilk olarak Bar plot grafiği kullanılmıştır. Bunun için metinde geçen terimlerin frekans değerleri bulunmuştur ve tf1 değişkenine atanmıştır. Bu grafikte örnek olarak 15'den fazla frekans değerine sahip kelimeler grafiğe aktarılmıştır.

```
tf1 = (tez_df_cumleler_sw["cumleler"]).apply(lambda x:
                            pd.value_counts(x.split(" "))).sum(axis = 0).reset_index()
tf1.columns = ["kelimeler","terim_frekansı"]
a = tf1[tf1["terim_frekansı"] >15]
a.plot.bar(x = "kelimeler", y = "terim_frekansı");
```

Şekil 3.17. Metnin barplot olarak gösterilmesi.

3.2.17. Metin görselleştirme (word cloud)

Diğer görselleştirme aracımız ise word cloud'tur. Bu grafik frekans değeri yüksek olan kelimeyi diğer kelimelerden daha büyük halde göstermektedir.

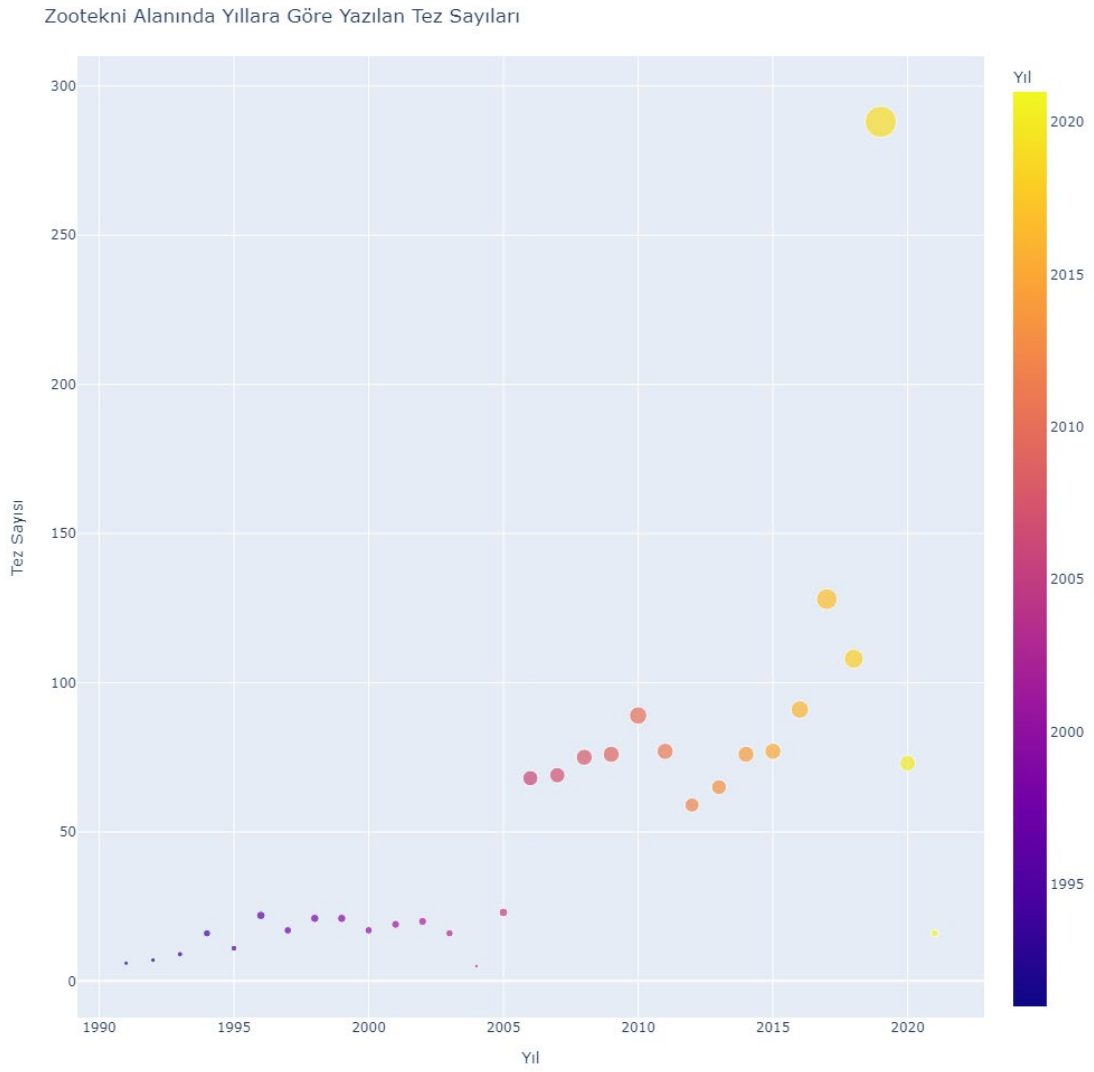
```
text = " ".join(i for i in tez_df_cumleler_sw.cumleler)
sw = set(STOPWORDS)
wordcloud = WordCloud(width=1080, height=1080,stopwords=sw, background_color="white").generate(text)
plt.figure( figsize=(15,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

Şekil 3.18. Metnin wordcloud olarak gösterilmesi.

4. BULGULAR

4.1. Tanıtıcı İstatistikler

Çalışmada kullanılan veri seti 1991-2021 yılları arasında Zootekni alanında yazılan 1665 adet tezin İngilizce özet kısımlarını ele almaktadır. Her seneye ait ayrı ayrı filtreleme işlemleri yapılarak elde edilen bu metinsel veri seti toplamda 1945 sayfadan oluşmaktadır. Bu tezlere ait adet bilgisi aşağıdaki şekilde gösterilmektedir.



Şekil 4.1. Yıllara göre yazılan tez sayıları.

4.2. Yıllara Göre Tez Analizlerinin Sonuçları

Çalışmada kullanılan metinsel veri kaynakları metin ön işleme işlemlerinden geçirildikten sonra Python ve Voyant yazılımı kullanılarak yıllara göre en çok işlenen konular ve bu konulara ait çeşitli istatistiksel çıkarımlar yapılmıştır.

4.2.1. 1991 yılı analiz sonuçları

1991 yılında yazılan tezler incelendiğinde zootekni alanında toplam 6 adet tez yazıldığı görülmektedir. Gerekli text preprocessing (metin ön işleme) aşamaları yapıldıktan sonra 1991 yılına ait yazılan tezler içerisinde en çok tekrar edilen kelime grupları aşağıdaki şekildeki gibi tespit edilmiştir.

Çizelge 4.1. 1991 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	feed	22
2	groups	22
3	kg	20
4	significant	17
5	differences	14
6	protein	14
7	meal	13
8	type	13
9	control	12
10	respectively	12

Yukarıdaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.2. 1991 yılı analizine göre wordcloud grafiği.

4.2.2. 1992 yılı analiz sonuçları

1992 yılında ise zootekni alanında toplam 7 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.2. 1992 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	types	30
2	fat	27
3	feed	25
4	tf	23
5	weight	21
6	protein	18
7	higher	15
8	kg	15
9	significantly	15
10	groups	13

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.3. 1992 yılı analizine göre wordcloud grafiği.

4.2.3. 1993 yılı analiz sonuçları

1993 yılına geldiğimizde ise zootekni alanında toplam 9 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en çok tekrar eden kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.3. 1993 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	protein	26
2	weight	20
3	diets	11
4	meal	11
5	affected	10
6	df	10
7	lambs	10
8	rumen	10
9	age	9
10	body	9

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.4. 1993 yılı analizine göre wordcloud grafiği.

4.2.4. 1994 yılı analiz sonuçları

1994 yılında ise zootekni alanında toplam 16 adet tezin yazıldığı tespit edilmiştir. Metin ön işleme teknikleri uygulandıktan sonra 1994 yılı içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.4. 1994 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	meal	38
2	weight	37
3	milk	36
4	tf	35
5	feed	33
6	fat	32
7	cattle	30
8	fattening	30
9	yield	30
10	diet	24

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.5. 1994 yılı analizine göre wordcloud grafiği.

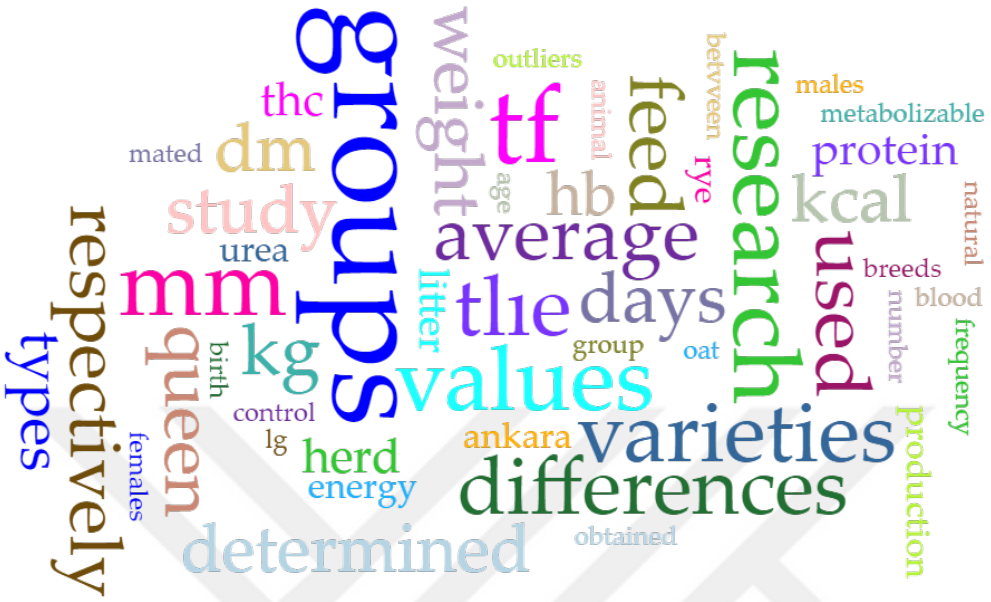
4.2.5. 1995 yılı analiz sonuçları

1995 yılında ise zootekni alanında toplam 11 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.5. 1995 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	groups	22
2	tf	22
3	research	15
4	tlıe	14
5	values	14
6	varieties	14
7	differences	13
8	average	12
9	feed	12
10	mm	12

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.



Şekil 4.6. 1995 yılı analizine göre wordcloud grafiği.

4.2.6. 1996 yılı analiz sonuçları

1996 yılına gelindiğinde ise zootekni alanında toplam 22 adet tezin yazıldığı tespit edilmiştir. Metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en çok tekrar eden kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.6. 1996 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	milk	70
2	groups	48
3	lactation	39
4	values	39
5	respectively	35
6	kg	34
7	gene	32
8	majör	30
9	genes	26
10	day	25

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.7. 1997 yılı analizine göre wordcloud grafiği.

4.2.7. 1997 yılı analiz sonuçları

1997 yılında ise zootekni alanında toplam 17 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en çok araştırması yapılan kelime gruplarının listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.7. 1997 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	kg	43
2	weight	32
3	period	29
4	respectively	26
5	study	24
6	days	21
7	milk	20
8	dm	16
9	carcass	15
10	characteristics	15

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.8. 1997 yılı analizine göre wordcloud grafiği.

4.2.8. 1998 yılı analiz sonuçları

1998 yılında ise zootekni alanında toplam 21 adet tezin yazıldığı tespit edilmiştir. Analiz için gerekli metin ön işleme teknikleri uygulandıktan sonra 1998 yılı içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.8. 1998 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	55
2	groups	36
3	feed	31
4	significant	29
5	effect	27
6	lambs	26
7	live	24
8	used	22
9	different	21
10	experiment	21

Yukardaki sık kullanılan kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.9. 1981 yılı analizine göre wordcloud grafiği.

4.2.9. 1999 yılı analiz sonuçları

1999 yılında ise zootekni alanında toplam 21 adet tezin yazıldığı tespit edilmiştir. Metin ön işleme teknikleri uygulandıktan sonra 1999 yılında en çok kullanılan kelime gruplarının listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.9. 1999 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	42
2	lambs	35
3	levels	25
4	growth	22
5	performance	22
6	cows	21
7	effects	20
8	research	20
9	broiler	19
10	feed	19

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.10. 1999 yılı analizine göre wordcloud grafiği.

4.2.10. 2000 yılı analiz sonuçları

2000 yılına geldiğimizde ise zootekni alanında toplam 17 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra 2000 yılında en çok tekrar eden kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.10. 2000 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	52
2	groups	38
3	days	35
4	respectively	30
5	experiment	28
6	feed	28
7	kg	28
8	effect	27
9	group	27
10	significant	27

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.11. 2000 yılı analizine göre wordcloud grafiği.

4.2.11. 2001 yılı analiz sonuçları

2001 yılında ise zootekni alanında toplam 19 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en çok tekrar eden kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.11. 2001 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	significant	37
2	age	35
3	effect	27
4	length	26
5	levels	26
6	milk	25
7	yield	25
8	diets	22
9	traits	21
10	measurements	19

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.12. 2001 yılı analizine göre wordcloud grafiği.

4.2.12. 2002 yılı analiz sonuçları

2002 yılında ise zootekni alanında toplam 20 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.12. 2002 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	40
2	groups	28
3	egg	25
4	animal	22
5	fattening	22
6	using	22
7	feed	20
8	methods	19
9	science	19
10	group	18

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.13. 2002 yılı analizine göre wordcloud grafiği.

4.2.13. 2003 yılı analiz sonuçları

2003 yılında ise zootekni alanında toplam 16 adet tez yazılmıştır. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.13. 2003 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	29
2	body	28
3	study	24
4	egg	22
5	yield	22
6	acid	21
7	model	20
8	animal	18
9	experiment	18
10	production	18

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.14. 2003 yılı analizine göre wordcloud grafiği.

4.2.14. 2004 yılı analiz sonuçları

2004 yılında ise zootekni alanında toplam 5 adet tezin yazıldığı tespit edilmiştir. Metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.14. 2004 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	milk	27
2	kg	22
3	period	21
4	length	17
5	weight	17
6	respectively	16
7	yield	16
8	lactation	15
9	traits	15
10	groups	13

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.15. 2004 yılı analizine göre wordcloud grafiği.

4.2.15. 2005 yılı analiz sonuçları

2005 yılında ise zootekni alanında toplam 23 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.15. 2005 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	51
2	feed	50
3	lactation	44
4	yield	41
5	study	34
6	groups	33
7	live	33
8	fed	32
9	feeding	31
10	group	29

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.16. 2005 yılı analizine göre wordcloud grafiği.

4.2.16. 2006 yılı analiz sonuçları

2006 yılında ise zootekni alanında toplam 68 adet tezin yazıldığı tespit edilmiştir. Metin ön işleme teknikleri uygulandıktan sonra 2006 yılı içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.16. 2006 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	176
2	feed	143
3	group	127
4	groups	113
5	study	94
6	day	88
7	significant	85
8	animal	82
9	body	81
10	respectively	81

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.17. 2006 yılı analizine göre wordcloud grafiği.

4.2.17. 2007 yılı analiz sonuçları

2007 yılına gelindiğinde ise zootekni alanında toplam 69 adet tezin yazıldığı tespit edilmiştir. Metin ön işleme teknikleri uygulandıktan sonra 2007 yılı içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.17. 2007 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	134
2	feed	133
3	egg	127
4	kg	105
5	study	94
6	age	80
7	animal	80
8	respectively	79
9	significant	75
10	groups	73

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.18. 2007 yılı analizine göre wordcloud grafiği.

4.2.18. 2008 yılı analiz sonuçları

2008 yılına geldiğimizde ise zootekni alanında toplam 75 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.18. 2008 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	130
2	feed	122
3	groups	114
4	significant	100
5	egg	90
6	group	89
7	study	89
8	control	88
9	kg	86
10	animal	85

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.19. 2008 yılı analizine göre wordcloud grafiği.

4.2.19. 2009 yılı analiz sonuçları

2009 yılında ise zootekni alanında toplam 76 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra 2009 yılında yazılan tezlere ait en çok araştırılan kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.19. 2009 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	144
2	feed	132
3	egg	105
4	kg	100
5	groups	97
6	significant	94
7	study	93
8	animal	91
9	milk	89
10	group	87

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.20. 2009 yılı analizine göre wordcloud grafiği.

4.2.20. 2010 yılı analiz sonuçları

2010 yılında ise zootekni alanında toplam 89 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.20. 2010 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	feed	115
2	study	115
3	animal	106
4	milk	106
5	science	104
6	kg	95
7	group	88
8	groups	87
9	egg	78
10	weight	77

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.21. 2010 yılı analizine göre wordcloud grafiği.

4.2.21. 2011 yılı analiz sonuçları

2011 yılında ise zootekni alanında toplam 77 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.21. 2011 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	milk	134
2	weight	114
3	study	113
4	groups	105
5	science	99
6	group	95
7	animal	93
8	yield	90
9	significant	85
10	period	79

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.22. 2011 yılı analizine göre wordcloud grafiği.

4.2.22. 2012 yılı analiz sonuçları

2012 yılında ise zootekni alanında toplam 59 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.22. 2012 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	study	93
2	milk	85
3	animal	79
4	feed	78
5	groups	69
6	kg	69
7	weight	68
8	group	67
9	science	64
10	production	59

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.23. 2012 yılı analizine göre wordcloud grafiği.

4.2.23. 2013 yılı analiz sonuçları

2013 yılında ise zootekni alanında toplam 65 adet tezin yazıldığı tespit edilmiştir. Metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.23. 2013 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	egg	143
2	weight	121
3	study	110
4	milk	108
5	feed	99
6	significant	96
7	kg	94
8	groups	93
9	effect	83
10	animal	78

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.24. 2013 yılı analizine göre wordcloud grafiği.

4.2.24. 2014 yılı analiz sonuçları

2014 yılında ise zootekni alanında toplam 76 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.24. 2014 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	119
2	groups	115
3	egg	114
4	animal	108
5	feed	107
6	kg	105
7	study	103
8	group	97
9	milk	93
10	different	87

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.25. 2014 yılı analizine göre wordcloud grafiği.

4.2.25. 2015 yılı analiz sonuçları

2015 yılında ise zootekni alanında toplam 77 adet tezin yazıldığı tespit edilmiştir. Metin ön işleme teknikleri uygulandıktan sonra 2015 yılında en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.25. 2015 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	milk	147
2	study	130
3	weight	120
4	group	106
5	feed	92
6	animal	84
7	kg	78
8	groups	76
9	performance	73
10	egg	67

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.26. 2015 yılı analizine göre wordcloud grafiği.

4.2.26. 2016 yılı analiz sonuçları

2016 yılında ise zootekni alanında toplam 91 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.26. 2016 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	study	153
2	weight	134
3	feed	129
4	animal	118
5	group	117
6	kg	114
7	milk	107
8	groups	100
9	performance	95
10	science	91

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.27. 2016 yılı analizine göre wordcloud grafiği.

4.2.27. 2017 yılı analiz sonuçları

2017 yılına gelindiğinde ise zootekni alanında toplam 128 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.27. 2017 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	study	203
2	group	190
3	feed	175
4	weight	172
5	kg	164
6	groups	157
7	milk	156
8	egg	143
9	animal	141
10	performance	140

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.28. 2017 yılı analizine göre wordcloud grafiği.

4.2.28. 2018 yılı analiz sonuçları

2018 yılında ise zootekni alanında toplam 108 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.28. 2018 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	feed	201
2	study	186
3	weight	179
4	egg	164
5	group	147
6	determined	141
7	animal	137
8	groups	115
9	Science	115
10	production	112

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.29. 2018 yılı analizine göre wordcloud grafiği.

4.2.29. 2019 yılı analiz sonuçları

2019 yılına geldiğimizde ise zootekni alanında toplam 288 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.29. 2019 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	study	531
2	feed	441
3	weight	415
4	milk	408
5	animal	398
6	determined	339
7	egg	325
8	kg	308
9	science	307
10	group	305

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.30. 2018 yılı analizine göre wordcloud grafiği.

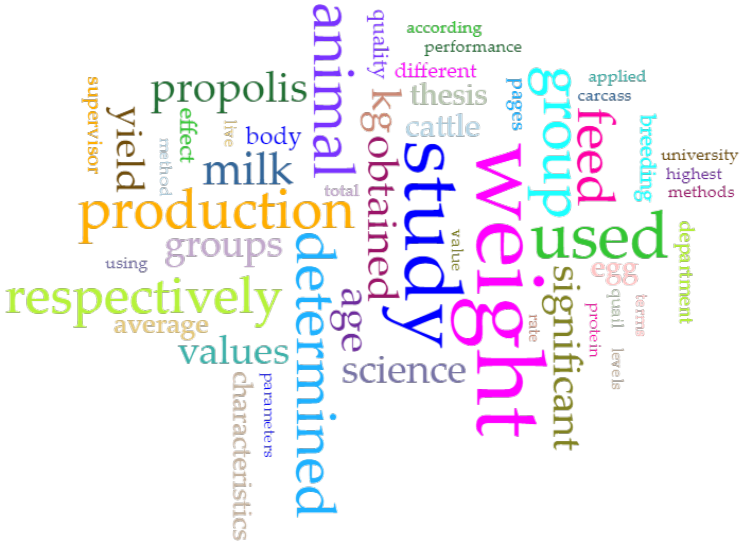
4.2.30. 2020 yılı analiz sonuçları

2020 yılında ise zootekni alanında toplam 73 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.30. 2020 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	174
2	study	144
3	used	104
4	animal	95
5	production	95
6	group	94
7	determined	93
8	respectively	90
9	feed	89
10	milk	78

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.



Şekil 4.31. 2020 yılı analizine göre wordcloud grafiği.

4.2.31. 2021 yılı analiz sonuçları

2021 yılında ise 07.05.2021 itibariyle zootekni alanında toplam 16 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.31. 2021 yılı kelime frekans çizelgesi

Sıra No	Kelime	Frekans
1	weight	65
2	study	33
3	egg	31
4	feed	30
5	group	28
6	groups	26
7	production	26
8	yield	26
9	animal	23
10	performance	23

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

Şekil 4.32. 2021 yılı analizine göre wordcloud grafiği.

4.2.32. 1991-2021 yılı arası toplu analiz analiz sonuçları

1991-2021 yılları arası yazılan toplam 1665 adet tezin yazıldığı tespit edilmiştir. Gerekli metin ön işleme teknikleri uygulandıktan sonra ilgili yıl içerisinde en sık frekans dağılımına sahip kelime listesi aşağıdaki gibi karşımıza çıkmaktadır.

Çizelge 4.32. 1991-2021 yılları arası kelime frekans çizelgesi -1

Sıra No	Kelime	Frekans
1	weight	2802
2	study	2522
3	feed	2469
4	milk	2044
5	groups	1976
6	group	1973
7	animal	1970
8	kg	1926
9	egg	1810
10	significant	1675

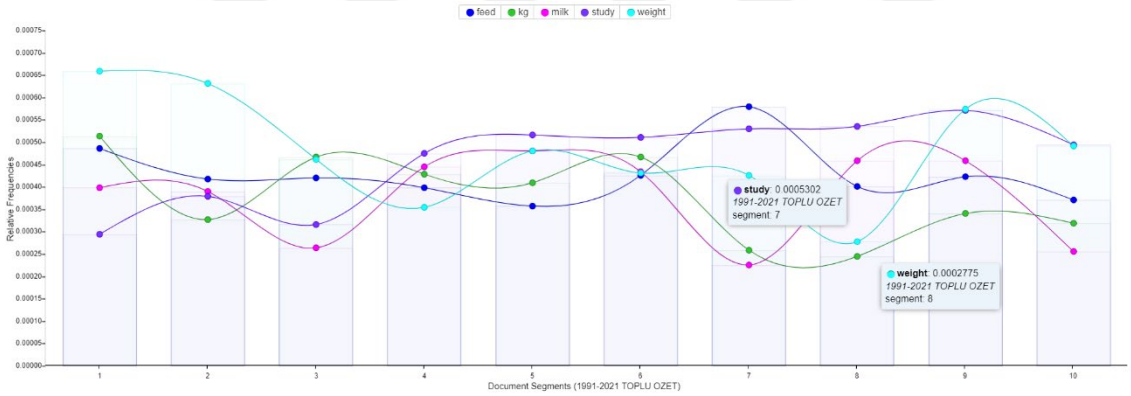
Çizelge 4.33. 1991-2021 yılları arası kelime frekans çizelgesi -2

Sıra No	Kelime	Frekans
11	science	1590
12	performance	1472
13	respectively	1467
14	yield	1460
15	determined	1453
16	used	1451
17	effect	1401
18	production	1384
19	different	1364
20	thesis	1307

Yukardaki işlenen kelime gruplarının wordcloud grafiği de aşağıdaki şekilde gösterilmiştir.

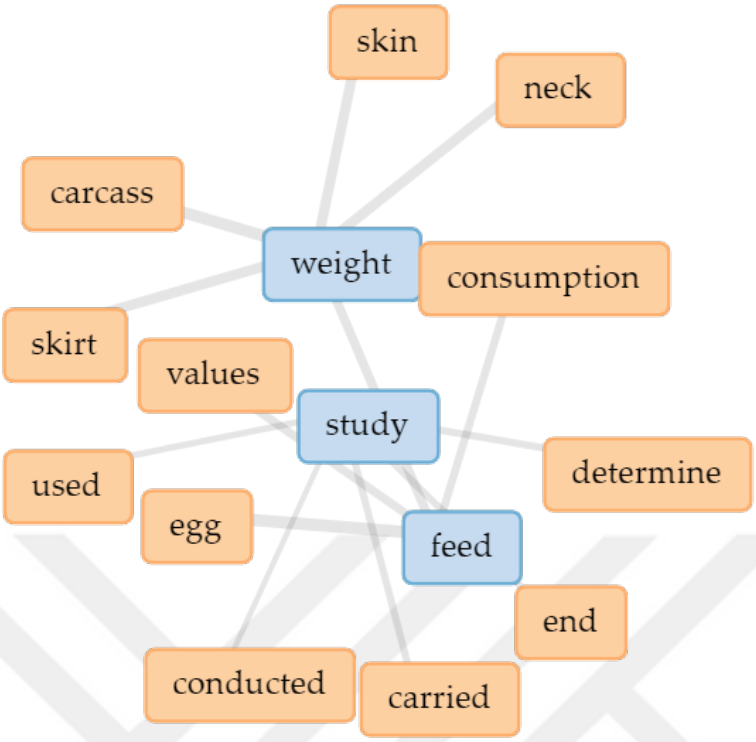
Şekil 4.33. 1991-2021 yılları arası analizine göre wordcloud grafiği.

Bu terimleri toplu olarak birleştirilen tez kaynağımızda kullanım sıklıklarını gösteren diğer bir grafiğimiz ise aşağıda gösterilmiştir.



Şekil 4.34. En sık kullanılan terimlerin kullanım eğrisi.

Bu terimlerin birbirleriyle olan kullanım sıklıklarını gösteren bir diğer grafiğimiz ise aşağıda gösterilmektedir. Terimlerin arasındaki çizgi kalınlıkları terimlerin birlikte kullanım sıklığını belirtmektedir.



Şekil 4.35. En sık kullanılan terimlerin birlikte kullanım grafiği.

5. TARTIŞMA VE SONUÇ

Bu çalışmada 1991-2021 yılları arasında zootekni alanında yazılan tezler yöktez sitesi üzerinden derlenerek elde edilmiştir. Her sene için ayrı ayrı filtreleme işlemleri yapılarak derlenen bu tezler son olarak toplu bir doküman haline getirilmiştir. Bu doküman metinsel veri kaynağı olarak kullanılmış olup bu veri setinden Python programlama dili ve web tabanlı bir yazılım olan voyant ile çeşitli çıkarımlar yapılmıştır.

Yapılan incelemeler sonucunda her sene zootekni alanında yazılan tezlerin inceleme konuları ve tez sayıları farklılık göstermekle birlikte bazı konuların neredeyse her sene yoğun bir şekilde araştırma konusu yapıldığı tespit edilmiştir. Bu konulardan bazıları "weight, feed, milk yield, egg, kg" gibi terimlerdir.

Çalışmada kullanılan veri seti ve analiz sonuçları ile Türkiye'de zootekni alanında son 30 yılda yapılan çalışmaların odak noktaları tespit edilebilmektedir. Konu hakkında literatür çalışması yapılmış olup ülkemizde Zootekni alanında bu tarz bir çalışmanın olmadığı gözlenmiştir. Çalışma bu yönüyle de özgünlük kazanmaktadır.

KAYNAKLAR

Çalış, K., Gazdağı, O., Yıldız, O., 2013. Reklam içerikli epostaların metin madenciliği yöntemleri ile otomatik tespiti. ***Bilişim Teknolojileri Dergisi,*** **6** (1): 1-7.

Döven, S., 2013. ***Metin Madenciliği ile Dokümanlar Arasındaki Benzerliklerin Bulunması*** (yüksek lisans tezi, basılmamış). BÜ, Fen Bilimleri Enstitüsü, İstanbul.

Erten, F., 2015. ***Metin Madenciliği Tabanlı Bir Web Sitesi Sınıflandırma Aracı Tasarımı*** (yüksek lisans tezi, basılmamış). MÜ, Fen Bilimleri Enstitüsü, İstanbul.

İşler, F., 2021. Sosyal Medyada Etkileşimi Etkileyen Faktörlerin İncelenmesi: Kuyumculuk Sektöründe Bir Örnek Olay İncelemesi. ***İşletme Bilimi Dergisi,*** **9** (1): 193-215.

Güler, N., 2007. ***Metin Madenciliği ile Metin Sınıflandırma*** (yüksek lisans tezi, basılmamış). YTÜ, Fen Bilimleri Enstitüsü, İstanbul.

Kahya, A., 2021. Wikipedia'daki Verilere Metin Madenciliği Yöntemlerinin Uygulanması. ***ESTUDAM Bilişim Dergisi,*** **11** (5): 11-14.

Onan, A., 2021. COVID-19 ile ilgili sosyal medya gönderilerinin metin madenciliği yöntemlerine dayalı olarak zaman-mekansal analizi. ***Avrupa Bilim ve Teknoloji Dergisi,*** **11** (5): 138-143.

Visa, A.2001. Technology of text mining. ***In International Workshop on Machine Learning and Data Mining in Pattern Recognition***.25-27 Temmuz 2001, London. 1–11.

ÖZ GEÇMİŞ

İlkokul öğrenimini Rauf Orbay ilkokunda, ortaokul öğrenimini Başöğretmen Atatürk İlköğretim Okulu'nda, lise öğrenimini Sivas Teknik Lise ve Endüstri Meslek Lisesi'nde (2000-2004) Sivas merkezde tamamladı. Lisans eğitimini Kocaeli Üniversitesinde Teknik Eğitim Fakültesi Bilgisayar Öğretmenliğinde 2006-2012 yılları arasında tamamladı. 2019 yılında Van Yüzüncü Yıl Üniversitesi Ziraat Fakültesinde Zootekni Anabilim Dalı'nda yüksek lisans öğrenimine başladı. 2014 yılından itibaren Van'ın İpekyolu ilçesinde öğretmen olarak görev yapmaktadır. Evli ve bir çocuk babasıdır.

VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
LİSANSÜSTÜ TEZ ORİJİNALLİK RAPORU

Tarih: 27/06/2022

Tez Başlığı / Konusu: Metin madenciliği yöntemleri ile 1991-2021 yılları arasında zootekni alanında yazılan tezlerin incelenmesi.

Yukarıda başlığı/konusu belirlenen tez çalışmamın Kapak sayfası, Giriş, Ana bölümler ve Sonuç bölümlerinden oluşan toplam 63 sayfalık kısmına ilişkin, 2706/2022 tarihinde şahsım/tez danışmanım tarafından Turnitin intihal tespit programından aşağıda belirtilen filtreleme uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %6 (yüzde 6) dır.
Uygulanan filtreler aşağıda verilmiştir:
- Materyal ve yöntem hariç,
- Kaynaklar hariç,
- Tezden çıkan yayınlar hariç,
- 7 kelimeden daha az örtüşme içeren metin kısımları hariç (Limit inatch size to 7 words)

Van Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinallik Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim.

Tarih ve İmza

Adı Soyadı: Fatih CAMCI

Öğrenci No: 18910001226

Anabilim Dalı: Zootekni Anabilim Dalı

Programı: Biyometri ve Genetik

Statüsü: □ Yüksek Lisans □ Doktora

DANIŞMAN ONAYI
UYGUNDUR
Prof. Dr. Abdullah YEŞİLOVA

ENSTİTÜ ONAY
UYGUNDUR
Prof. Dr. Harun AKKUŞ