

T.C.
İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİYOLOJİK VERİ BİLİMİ

İLAÇ-İLAÇ ETKİLEŞİMLERİNİ MAKİNE ÖĞRENMESİYLE
TESPİT EDEN TAHMİN MODELİ

YÜKSEK LİSANS TEZİ

AHMET ŞENER

OCAK 2022

T.C.
İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİYOLOJİK VERİ BİLİMİ

İLAÇ-İLAÇ ETKİLEŞİMLERİNİ MAKİNE ÖĞRENMESİYLE
TESPİT EDEN TAHMİN MODELİ

YÜKSEK LİSANS TEZİ

AHMET ŞENER

DANIŞMAN

DR. ÖĞR. ÜYESİ ARAFAT SALİH AYDINER

OCAK 2022

BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu, akademik ve etik kuralları gözeterek çalıştığımı ve her alıntıya kaynak gösterdiğimi taahhüt ederim

Ahmet ŞENER

Danışmanlığını yaptığım işbu tezin tamamen öğrencinin çalışması olduğunu, akademik ve etik kuralları gözeterek çalıştığını taahhüt ederim.

Dr. Öğr. Üyesi Arafat Salih Aydın

ONAY

Ahmet ŞENER tarafından hazırlanan İlaç-İlaç Etkileşimlerini Makine Öğrenmesi İle Tespit Eden Tahmin Modeli başlıklı bu yüksek lisans tezi, Medeniyet Üniversitesi Yüksek Lisans Entitüsü Biyolojik Veri Bilimi programında hazırlanmış ve jürimiz tarafından kabul edilmiştir.

JÜRİ ÜYELERİ

İMZA

Tez Danışmanı:

Dr. Öğretim Üyesi Arafat Salih Aydın

Kurumu: İstanbul Medeniyet Üniversitesi

Üyeler:

Dr. Öğr. Üyesi Muhammed Erkan KARABEKMEZ

Kurumu: İstanbul Medeniyet Üniversitesi

Prof. Dr. Kazım Yalçın Arğa

Kurumu: Marmara Üniversitesi

Tez Savunma Tarihi: 27.01.2021

TEŞEKKÜR

Tez çalışmasını yakından takip ederek akademik ve deneyimlerini benimle paylaşan, tezin her aşamasında bilimsel katkı, ilgi ve desteğini hiçbir zaman esirgemeyen, sabırla yönlendiren, kendisinden çok şey öğrendiğim değerli danışman hocam, Sayın Dr. Arafat Salih Aydınır’a çok teşekkür ederim.

Fedakârlıkları ve bana olan güveni ile tez çalışmam boyunca her zaman yanımda olan aileme, arkadaşlarıma ve özellikle eşim Selda Şener ve oğlum Erdem Yuşa Şener’e bana gösterdikleri sabır ve tahammülden ötürü tüm kalbimle teşekkür ederim.

İÇİNDEKİLER

BİLDİRİM	i
ONAY	ii
TEŞEKKÜR	iii
KISALTMALAR	vi
TABLO LİSTESİ	viii
ŞEKİL LİSTESİ.....	ix
ÖZET.....	x
ABSTRACT	xii
BÖLÜM I - GİRİŞ	1
BÖLÜM II – LİTERATÜR ARAŞTIRMASI	5
2. LİTERATÜR ÖZETİ.....	5
2.1 İLAÇ-İLAÇ ETKİLEŞİMİ.....	5
2.2 MAKİNE ÖĞRENMESİNE DAYALI TAHMİN MODELLERİ	6
2.2.1 DESTEK VEKTÖR MAKİNELERİ	7
2.2.2 K-EN YAKIN KOMŞULUK ALGORİTMASI	8
2.2.3 LOJİSTİK REGRESYON	9
2.2.4 KARAR AĞAÇLARI	10
2.2.5 ETİKET YAYIMI.....	11
2.2.6 NAİVE BAYES ALGORİTMASI	12
2.2.7 DERİN ÖĞRENME AĞLARI.....	13
2.3 İLAÇ –İLAÇ ETKİLEŞİMİNDE MAKİNE ÖĞRENMESİNİN KULLANILMASI.....	14
BÖLÜM III – METOD.....	19
3. METOD	19
3.1 VERİNİN ALINMASI ve YAPISAL HALE GETİRİLMESİ	19
3.2 MODEL İÇİN VERİNİN SEÇİLMESİ.....	22

3.3 VERİLERİN ÖNİŞLEME SOKULMASI	23
3.3.1 BMYGS VERİSİNİN DÖNÜŞTÜRÜLMESİ.....	23
3.3.2 ATK, ENZİM ve HEDEF VERİLERİNİN DÖNÜŞTÜRÜLMESİ.....	26
3.4 MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN UYGULANMASI	28
BÖLÜM IV – SONUÇLAR.....	32
4. SONUÇLAR	32
BÖLÜM V - TARTIŞMA ve SONUÇ	37
5. TARTIŞMA ve SONUÇ	38
KAYNAKÇA	42

KISALTMALAR

AGİY	: Amerikan Gıda ve İlaç Yönetimi (İng. U.S. Food and Drug Administration)
ATK	: Anatomik Terapötik Kimyasal (İng. Anatomical Therapeutic Chemical)
BMYGS	: Basitleştirilmiş Moleküler Yapı Giriş Sistemi (İng. Simplified Molecular-Input Line-Entry System)
ÇEA	: Çizge Evrişim Ağı (İng. Graph Convolution Network)
DMÖ	: Denetimli Makine Öğrenimi (İng. Supervised Machine Learning)
DSA	: Derin Sinir Ağları (İng. Deep Neural Networks)
DVM	: Destek Vektör Makineleri (İng. Support Vector Machine)
GB	: Ciga Bayt (İng. Giga Byte)
GİB	: Grafik İşleme Birimi (İng. Graphical Processing Unit)
GO	: Gen Ontolojisi
İHO	: İlaç Hedef Ontolojisi (İng. Drug Target Ontology)
İİE	: İlaç-İlaç Etkileşimi (İng. Drug-Drug Interaction)
K-EYK	: K-En Yakın Komşuluk (İng. K-Nearest Neighbourhood)
LR	: Lojistik Regresyon
MİB	: Merkezi İşleme Birimi (Central Processing Unit)
NB	: Naive Bayes
PPE	: Protein-Protein Etkileşimi (İng. Protein-Protein Interaction)
SA	: Sinir Ağları (İng. Neural Networks)
SEKK	: Sıradan En Küçük Kareler (İng. Ordinary Least Square)
SİP	: Sitokrom P450 enzimleri (İng. Cytochrome P450)
TBA	: Temel Bileşenler Analizi (İng. Principal Component Analysis)
TİR	: Ters İlaç Reaksiyonu (İng. Adverse Drug Reaction)

TK	: Tanimoto Katsayısı (İng. Tanimoto Coefficient)
XML	: Geniřletilebilir İřaretleme Dili (İng. eXtensible Markup Language)
YBP	: Yapısal Benzerlik Profili (İng. Structural Similarity Profile)

TABLO LİSTESİ

Tablo 1. İlaç Çiftlerinin Etkileşim Sayıları	23
Tablo 2. ATK Kodu Sınıflandırması Örneği	27
Tablo 3. Modellerde Test Edilecek Olan Veri Setlerinin İçerikleri	29
Tablo 4. Veri Setlerinin Etkileşimleri Tahmin Etmede Elde Ettikleri F1-Skorları.....	30
Tablo 5. DSA Çıktı Etiketlerinin Oluşturulması	31
Tablo 6. Derin Öğrenme Ağı Çapraz Doğrulama Sonuçları	34
Tablo 7. DSA Çıktı Etiketlerinin Oluşturulması – 2	35
Tablo 8. Derin Öğrenme Ağı Çapraz Doğrulama Sonuçları – 2	36
Tablo 9. Bazı Ciddi İlaç-İlaç Etkileşimleri	38
Tablo 10. Modelimizde Tespit Ettiğimiz Bazı İlaç-İlaç Etkileşimleri	39

ŞEKİL LİSTESİ

Şekil 1. İlaç-İlaç Etkileşimlerinin Sınıflandırılması	2
Şekil 2. Destek Vektör Makineleri	8
Şekil 3. Karar Ağacı (Desicion Tree)	11
Şekil 4. Basit Sinir Ağı yapısı ve Derin Öğrenme Sinir Ağı Yapısı	13
Şekil 5. XML İlaç Verilerinin Veritabanına Kaydedilmesi Akış Diyagramı	20
Şekil 6. Oluşturulan DRUG Veritabanı Diyagramı	21
Şekil 7. Firapiravir İlacının Moleküler Yapısı ve Üretilen BMYGS Kodu	23
Şekil 8. Moleküller İçin İkili Vektör Oluşturma Yöntemi	24
Şekil 9. Tanimoto Katsayısının Hesaplanması	25
Şekil 10. İlaçların TK Benzerlik Matrisi	25
Şekil 11. ATK Kodu Vektörü Oluşturulması	27
Şekil 12. Veri Eğitimi Başarı Skorları	32
Şekil 13. 10 Kat Çapraz Doğrulama İçin Eğitim ve Test Verisi Seçimi	33
Şekil 14. Farklı İterasyonlarda Majör Etkileşimler İçin Karışıklık Matrisleri	35
Şekil 15. İlaç-İlaç Etkileşimi Tahmin Modeli	37

ÖZET

İLAÇ-İLAÇ ETKİLEŞİMLERİNİ MAKİNE ÖĞRENMESİYLE TESPİT EDEN TAHMİN MODELİ

Şener, Ahmet

Yüksek Lisans Tezi, Lisansüstü Eğitim Enstitüsü; Biyolojik Veri Bilimi Programı

Danışman: Dr. Öğr. Üy. Arafat Salih Aydın

Ocak, 2022, 60 sayfa

İlaç-ilaç etkileşimleri birden fazla ilaç kullanan insanlarda görülen ve insanların kullandığı ilaçların etkileşimde bulunarak birbirlerinin etkisini olumlu veya olumsuz yönde etkilemesi durumudur. Bu görülen etkiler genellikle bir ilacın diğer ilaç etkisini artırması, diğer ilaç etkisini azaltması veya diğer ilacın etki ediş biçimini değiştirmesi şeklinde olabilmektedir. Bu durum birden fazla ilaç kullanan insanların sağlığında iyileşme yerine daha kötüye giden bir seyre neden olmakta, hatta ölümle bile sonuçlanabilen istenmeyen etkilerin oluşmasına sebebiyet vermektedir. Bu hem insan sağlığı açısından hem de hastanelerde oluşan ekstra sağlık masrafları açısından büyük maliyetlere neden olabilmektedir. İlaçların birbirleriyle olan etkileşimlerinin bir kısmı tespit edilmiştir. Ancak henüz test edilmemiş ve tespit edilememiş olan bir çok etkileşim bulunmaktadır. İlaç-ilaç etkileşimleri laboratuvar ortamlarında bilimsel yöntemlerle test edilerek bulunabilmektedir. Bu yöntemler hem zaman ve hem de iş gücü açısından çok fazla maliyete yol açabilmektedir. Bu bilimsel yöntemlere başlamadan önce etkileşmesi muhtemel olan ilaçların bilinmesi hem zaman, hem iş gücü hem de maliyet açısından büyük tasarruflar sağlayacaktır.

Bu tezde bilinmeyen ilaç-ilaç etkileşimlerinin tespit edilebilmesi için çeşitli yapay zeka ve makine öğrenmesi algoritmaları kullanılacaktır. Etkileşim tipi ve etkileşimin

ciddiyeti yönünden bir tahmin modelinin daha önceden hiç yapılmadığı tespit edilmiştir. Ayrıca etkileşimleri bulmaya yönelik böyle bir çalışma ülkemizde ilk defa yapılmaktadır. Bunu gerçekleştirmek için çeşitli veri girdilerini ve lojistik regresyon, derin sinir ağları, karar ağaçları gibi yöntemleri deneyerek en verimli modeli belirlemeye çalıştık. İlaçların molekül yapı bilgileri, sınıflandırma kodları, hedef bilgileri ve enzim bilgilerini çeşitli dönüşümler uygulayarak birbirinden farklı algoritmalar üzerinde denemeler yaparak en verimli yöntemi bulmaya çalıştık.

Bu çalışmanın çıktısı birden fazla ilaç kullanmanın gerektiği durumlarda hem ilacı veren hekimi hem de ilacı kullanan kişiyi uyarmak için kullanılabilir. Böylece istenmeyen ilaç reaksiyonlarının önüne geçilebilecektir. Yeni bir ilaç keşif aşamasında yine bu yöntemler kullanılarak mevcut hangi ilaçlar ile etkileşme olasılığının olduğu tahmin edilebilecektir.

Yaptığımız çalışma sırasında çeşitli dönüşümler ile ilaçların moleküler yapısını, hedeflerini sınıflandırma kodlarını ve enzim etkilerini girdi olarak kullandık. Tahmin modelini oluşturmak için çeşitli algoritmaları denedik ve en iyi başarımla derin sinir ağlarının diğer yöntemlerden daha iyi bir sonuç verdiğini gördük. Bu yöntemin kullanılması hem istenmeyen etkilerin ortaya çıkmasını engellemek açısından hem de ekstra maliyetlerin önüne geçmek açısından fayda sağlayacaktır.

Anahtar Kelimeler: İlaç-ilaç etkileşimi, Derin Öğrenme, Yapay Zeka, Ters İlaç Reaksiyonları

ABSTRACT

FORECAST MODEL DETERMINING DRUG-DRUG INTERACTIONS BY MACHINE LEARNING

Sener, Ahmet

Master Thesis, Graduate School of Education; Biological Data Science Program

Supervisor: Asst. Prof. Dr. Arafat Salih Aydiner

January, 2022, 60 pages

Drug-drug interactions are seen in people who use more than one drug, and when the drugs used by people interact and affect each other positively or negatively. These effects can usually be in the form of a drug increasing the effect of another drug, reducing the effect of another drug, or changing the way the other drug works. This situation causes a worsening rather than improvement in the health of people who use more than one drug, and even causes undesirable effects that can result in death. This can cause great costs both in terms of human health and in terms of extra health costs in hospitals. Some of the drug-drug interactions have been detected. However, there are many interactions that have not yet been tested and detected. Drug-drug interactions can be found in laboratory environments by testing them with scientific methods. These methods can be very costly in terms of both time and labor. Knowing the drugs that are likely to interact before starting these scientific methods will provide great savings in terms of time, labor and cost.

In this thesis, various artificial intelligence and machine learning algorithms will be used to detect unknown drug-drug interactions. It has been determined that a prediction model has never been made before in terms of the interaction type and the severity of the interaction. In addition, such a study to find interactions is carried out for the first time in our country. To achieve this, we tried to determine the most efficient model by testing various data inputs and methods such as logistic regression, deep neural

networks, and decision trees. We tried to find the most efficient method by trying different algorithms with various transformations on the molecular structure information, classification codes, target information and enzyme information of drugs.

The output of this study can be used to warn both the prescriber and the person using the drug in cases where more than one drug is required. Thus, unwanted drug reactions can be prevented. In a new drug discovery phase, it will be possible to predict which drugs are likely to interact with existing drugs by using these methods.

During our study, we used the molecular structure of drugs, their targets, classification codes and enzyme effects as inputs with various transformations. We tried various algorithms to build the prediction model and found that with the best performance, deep neural networks outperform other methods. The use of this method will be beneficial both in terms of preventing the emergence of undesirable effects and in terms of preventing extra costs.

Keywords: Drug-Drug Interaction, Deep Learning, Artificial Intelligence, Adverse Drug Reactions

BÖLÜM I - GİRİŞ

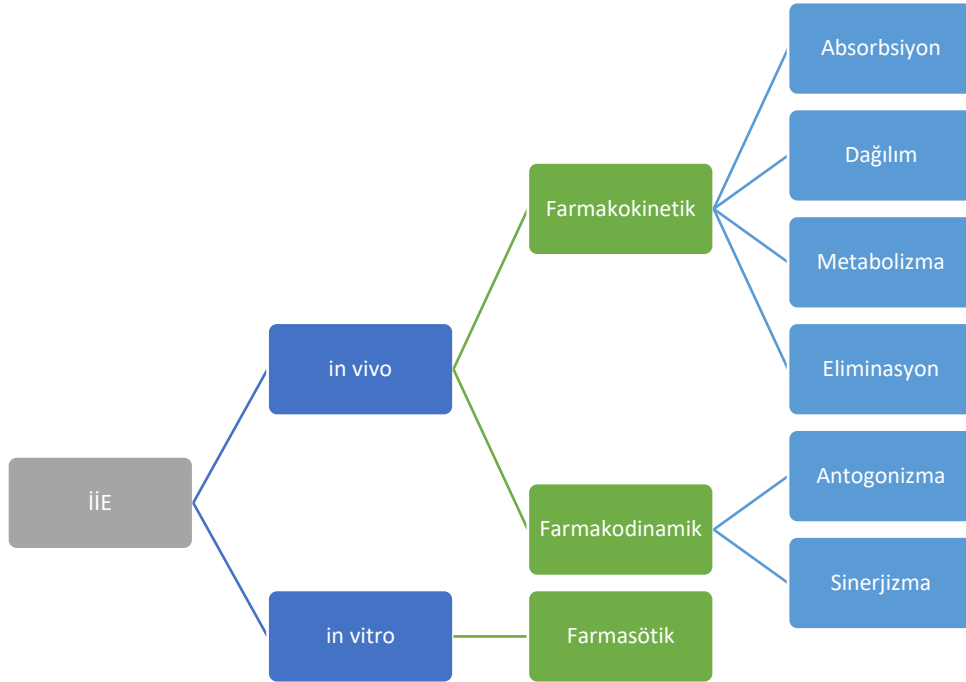
Günümüzde hastalıkların tedavisinde birden fazla ilacın bir arada kullanılması oldukça yaygındır. Özellikle yaşın artmasıyla birlikte kişideki hastalık sayısının artması birden fazla ilaç kullanılmasını kaçınılmaz hale getirmektedir. İlaç-İlaç etkileşimleri (İİE) birden fazla ilacın birlikte kullanılmasıyla ortaya çıkan etkilere dir. İstenmeyen ilaç etkileşimleri yan etkilerin ortaya çıkmasına, ters ilaç reaksiyonlarına ve bazı hastalarda ölüme kadar sebebiyet veren zehirlenmelere yol açabilir (Tekada ve diğerleri, 2017).

İlaçların neden olduğu yaralanmalar veya kötü etkiler ters ilaç reaksiyonu (TİR) olarak adlandırılır. İİE'ler ters ilaç reaksiyonlarının en yaygın nedenlerinden bir tanesidir. Bu sağlık giderlerinde önemli bir artışa sebep olmaktadır (Juurlink ve diğerleri, 2003). Bir çok İİE, klinik test aşamasında tanımlanmaz. İlaç klinik kullanım için onay aldıktan sonra bu İİE'ler rapor edilir. Bu bilinmeyen İİE'ler yatan hastalar üzerinde yapılmış olan ilaç hatalarının %3-%5'lik bir kısmını kapsar ve bu hatalar farklı hastalıkların baş göstermesiyle veya ölümle sonuçlanabilmektedir (Leape ve diğerleri, 1995). Klinik İİE'ler ayrıca çok ciddi sosyal ve ekonomik problemlere neden olabilmektedir.

İlaç-ilaç etkileşimleri hastanın iki ilacı birden aldığı herhangi bir zaman diliminde ortaya çıkabilir. İİE'ler ilaç etkisinde beklenmeyen artış, ilaç etkisinde azalma veya toksisite gibi olumsuz etkilere sebep olabilir. Bunun yanında ilacın yan etkilerinde azalma, ilaç toksisitesinin azalması veya yok olması, ilacın terapötik etkinliğinde bir artış gibi olumlu etkilere de sebep olabilir (Göksel, t.y).

İİE'ler *in vivo* etkileşimler ve *in vitro* etkileşimler olarak ikiye ayrılır. *in vitro* etkileşimler laboratuvar ortamında ya da yapay koşullarda meydana gelen etkileşim türleridir. *in vivo* ise canlının içinde meydana gelen etkileşimlerdir. Canlı organizma içerisinde meydana gelen etkileşimler farmakokinetik veya farmakodinamik biçimde olabilir. Etkileşme ilaçların etki yerinden uzakta oluyor ise farmakokinetik etkileşimdir. Farmakokinetik ilaç-ilaç etkileşimleri, bir ilacın, başka bir ilacın sistemik

konsantrasyonunu deęiřtirerek etki yerinde “ne kadar miktarda” ve “ne kadar süre ile” bulunacaęını etkilemesi sonucu meydana gelmektedir. Etkileřme ilaların etki yerinde veya yakınında oluyor ise farmakodinamik etkileřimdir. Farmakodinamik ila-ila etkileřimi; etkileřen ilaların, ya aditif etki ile toplam etkiyi artırması ya da ters etki ile toplam etkiyi azaltması veya tamamen ortadan kaldırması durumlarında meydana gelmektedir. Farmasötik ila etkileřimleri ise uygulama öncesinde bir ilacın formölasyonunun bařka bir ila tarafından deęiřikliğe uğratılması ile meydana gelmektedir (Elder, t.y).



řekil 1. İla-İla Etkileřimlerinin Sınıflandırılması (WEB 2)

İlaların geliřtirilmesi ařamasında etkileřimlerin arařtırılmasına raęmen bazı İİE’ler ilacın piyasaya sürölmesinden sonra da belirlenebilmektedir. Ciddi ila etkileřimleri yüzünden piyasadan çekilen ilalar da olabilmektedir. İİE’ler geleneksel olarak klinik ortamda ekonomik aıdan maliyetli ve ok zaman alan bir süreç sonunda tespit edilebilmektedir.

Birden fazla ila ile tedavinin ok yaygın olduęu günümüzde hızlı ve güvenilir bir řekilde İİE’lerin tespit edilmesi oldukça önem kazanmıřtır. Birbirleriyle etkileřebilecek ila çiftlerinin tahmin edilmesi için bir takım bilgi teknolojileri destekli hesaplama yöntemleri geliřtirilmiřtir.

İlaç-ilaç etkileşiminin tahmin edilmesinde temel olarak üç yöntem türünden bahsetmek mümkündür. Bunlar, ilaçların benzerliklerini esas alarak geliştirilen yöntemler, network ağları kullanılarak geliştirilen yöntemler ve makine öğrenmesi algoritmaları kullanan yöntemlerdir. Bunlardan birincisinde birbirleriyle etkileşen iki ilaçtan birincisine benzerlik gösteren bir ilacın ikinci ilaçla etkileşime girmesi olası olduğu varsayımı vardır. Örnek vermek gerekirse A ilacı ile B ilacı etkileşime giren iki ilaç olsun. C ilacı eğer yapısal yönden ve diğer yönlerden A ilacına bir benzerliği oranında C ilacının B ilacıyla etkileşime girebileceğini söyleyebiliriz.

Network tabanlı yöntemlerde ise sadece etkileşimi olan ilaçların veya bütün ilaçların birbirine bağlı olduğu bir ağ yapısı tasarlanır. Sadece bu ağda bulunan bağlantıları kullanan veya ilaçların birtakım özelliklerini de dahil ederek topolojik bir takım çıkarımlar yapan bir model oluşturulur. Yapılan çıkarımlar ile olası İİE'lerin tahmin edilmesi sağlanır.

Derin makine öğrenmesi tabanlı öğrenmelerde ise ilaçların özellikleri ve ilaç çiftlerinin birbirleriyle etkileşip etkileşmediği bilgisi derin öğrenme algoritmalarıyla eğitilir. Bu eğitim sonunda özellikleri girdi olarak verilen ilaç çiftlerinin etkileşip etkileşmeyeceğine dair bir kural seti oluşturulur. Bu kural setiyle olası İİE'ler tahmin edilir.

Bu tezi hazırlamaktaki amacımız ilaç-ilaç etkileşimlerini tahmin etmek üzere var olan makine öğrenmesi yöntemlerini kıyaslayarak en verimli modeli ortaya çıkarmaktır. İİE'lerin hesaplanması için geçmişte ve özellikle son yıllarda birbirinden farklı yöntemler ortaya koyulmuştur. Ortaya koyulan yöntemler yukarıda bahsettiğimiz tekniklerden sadece bir tanesini kullanmış veya bu teknikleri bir arada kullanarak daha kompleks modeller oluşturulmuştur. Ayrıca yapılan çalışmalarda benzerlikler hesaplanırken kimisinde ilaçların kimyasal molekül özellikleri dikkate alınırken kimisinde etki ettiği enzimlerin veya hedef proteinlerinin benzerlikleri dikkate alınmıştır (Lin ve diğerleri, 2010; Ryu ve diğerleri, 2018; Park ve diğerleri, 2015; Tatonetti ve diğerleri, 2012; Vilar ve diğerleri, 2014; Zhang ve diğerleri, 2015). Hem kullanılacak yöntemlerin çeşitli olması hem de kullanılacak özelliklerin çeşitli olması İİE'lerin tahmini için farklı modellerin ortaya çıkmasına neden olmuştur. Biz de çalışmamızda farklı teknik ve farklı ilaç özelliklerini bir arada deneyerek ilaç-ilaç

etkileşimlerini tahmin etmek üzere en etkili yöntemin ne olduğunu sunmaya çalışacağız.

Bu çalışma ile İİE'lerin tahmini ile ilgili derin öğrenme de dahil olmak üzere farklı makine öğrenmesi algoritmaları farklı parametre girişleri ile uygulanacaktır. Bu uygulama genel olarak hangi yöntemin hangi parametre ile gerçekleştirilmesinin daha etkin veya az etkin olduğu konusunda bilgiler ortaya koyacak ve en etkin model tespit edilmiş olacaktır.

Çalışmada ilk olarak veriler alınarak yapısal ve ilişkisel bir biçimde saklanması sağlanacaktır. Daha sonra bu verilerin makine öğrenmesinde kullanılabilmesi için gerekli temizlikler yapılacaktır ve verilerde dönüşüm işlemleri uygulanacaktır. Veriler işlemler için kullanıma uygun hale getirildikten sonra makine öğrenimi algoritmaları İİE'leri tahmin etmek üzere kullanılacaktır. Tahmin modelleri test edilerek çıkan sonuçlar doğrultusunda en uygun yöntem belirlenecektir.

Bölüm 2'de genel olarak bir literatür incelemesi yapılarak İİE'ler, makine öğrenmesi ve mevcutta yapılmış olan İİE tahminleme çalışmalarıyla alakalı bilgi verilmiştir. Bölüm 3'te modeller için uygun verinin oluşturulması için yapılan işlemler ve uygulanan yöntemler ayrıntılı olarak açıklanmıştır. Bölüm 4'te uygulanan yöntemlere ait sonuçlar bulunmaktadır. Bölüm 5'de ise çalışmanın sonuçlarının değerlendirilmesi yapılmıştır.

BÖLÜM II – LİTERATÜR ARAŞTIRMASI

İlaç-ilaç etkileşimleri ve İlaç-ilaç etkileşimlerini bulma çalışmaları ile ilgili kavramsal bilgilere bu bölümde yer verilmektedir.

2. LİTERATÜR ÖZETİ

Bu bölümde literatürde bulunan ilaç-ilaç etkileşimi, tahmine dayalı makine öğrenmesi metodları ve ilaç-ilaç etkileşimlerinin tespit edilmesi için yapılan makine öğrenmesi çalışmaları detaylı olarak sunulmuştur.

2.1 İLAÇ-İLAÇ ETKİLEŞİMİ

İlaç-ilaç etkileşimleri (İİE'ler) TİR'lerin en yaygın nedenlerinden biridir ve bu belirtilerin çoklu ilaç kullanan yaşlılarda yaygın olduğu bildirilmiştir (Gallelli ve diğerleri, 2002, 2010). TİR'ler yaygın bir klinik sorunu temsil eder ve hastanede artan yatış sayısından ve/veya yatış süresinden sorumlu olabilir (Lee ve diğerleri, 2000; Bordet ve diğerleri, 2001).

Çoklu ilaç kullanımı bir TİR'e sebep olduğunda ve bu yanlış olarak değerlendirildiğinde hastanın reçetesine yeni ilaçlar eklenmesine ve İİE'lerden kaynaklanan yeni TİR'lerin tetiklenmesine neden olabilir (Rochon ve Gurwitz, 1997).

İlaçların kombinasyonu, ilaç tedavisinin etkinliğini arttırmak için yaygın bir uygulamadır, ancak optimal kombinasyon ve optimal dozların seçimi, bir deneme yanılma meselesi olmaya devam etmektedir. Bu nedenle, *in vivo* olarak ilaç kombinasyonlarına karşı sinerjistik, katkı maddesi ve antagonistik tepkilerin tahmini oldukça ilgi çekici bir konu olmaya devam etmektedir (Jonker ve diğerleri, 2005).

İlaç keşif araştırma ekiplerine, ilaç keşif sürecinde başarılı ilaç adaylarını daha önce belirleme baskısı, kimyagerlere ve biyologlara karar verme süreçlerinde rehberlik etmeyi amaçlayan çeşitli *in vitro* tahlillerinin geliştirilmesine yol açmıştır. Bu erken testlerin rolü, yeni kimyasal varlıkların absorpsiyon, dağılım, metabolizma ve Sitokrom P450 (SİP) inhibisyon potansiyeli ile ilgili yapı iskelelerindeki yükümlülükleri belirtmektir. İİE taramasında biyoanalitik yöntem geliştirme, enzim kinetiği ile ilgili stratejiler, SİP varyantlarının tanınması ve SİP etkileşimlerini tahmin etmek için hesaplama modellerinin oluşturulması olarak dört temel kategoriye ayrılır (Hutzler ve diğerleri, 2005). Bu kategorilerin sonuncusunu tezimiz kapsamında ele alacağız.

2.2 MAKİNE ÖĞRENMESİNE DAYALI TAHMİN MODELLERİ

Tahmin, diğer ölçülen özniteliklerin değerleri göz önüne alındığında, incelenen bir sistemin bir özneliliğinin bilinmeyen değerinin tahmin edilmesini içerir. Tahmin makine öğrenmesinde, tahmin kuralı önceden çözülmüş durumlardan oluşan verilerden türetilir. Tahmine dayalı öğrenme yöntemlerinin çoğu, yıllar önce bilgisayar çağının şafağında ortaya çıkmıştır (Friedman, 2006).

Makine Öğrenmesi kavramı ilk olarak 1956'da Ray Solomonoff'un yazdığı makale ile ortaya çıkmıştır (Solomonoff, 1957). 1959 yılında ise Arthur Samuel'in yaptığı çalışmalar sonucu popülerliği artmış bir bilim dalıdır (Samuel ve diğerleri, 1959).

Denetimli Makine Öğrenimi (DMÖ), genel hipotezler üretmek için harici olarak sağlanan örneklerden akıl yürüten ve daha sonra gelecekteki örnekler hakkında tahminlerde bulunan algoritmaların aranmasıdır. Denetimli sınıflandırma, akıllı sistemler tarafından en sık gerçekleştirilen görevlerden biridir (Osisanwo ve diğerleri, 2017).

Sınıflandırma ile daha fazla ilgilenen denetimli temel makine öğrenimi algoritmaları olarak şunları sıralayabiliriz: Lineer Regresyon, Lojistik Regresyon, Naive Bayes Sınıflandırıcısı, Destek Vektör Makineleri, Karar Ağaçları, Rasgele Orman Sınıflandırıcısı, Yapay Sinir Ağları Bayes Ağları vb. (Ayodele, 2010).

Genomik ve görüntüleme teknolojilerindeki gelişmeler, çok sayıda örnekten elde edilen moleküler ve hücresel profil verilerinin patlamasına yol açmıştır. Biyolojik veri boyutundaki ve edinme hızındaki bu hızlı artış, geleneksel analiz stratejilerine meydan okumaktadır. Derin öğrenme gibi modern makine öğrenimi yöntemleri, içlerindeki gizli yapıyı bulmak ve doğru tahminler yapmak için çok büyük veri kümelerinden yararlanmayı vaat etmektedir (Angermueller ve diğerleri, 2016).

İlaç-ilaç etkileşimlerinin tahmin edilebilmesi için bugüne kadarki çalışmalar tezimiz yapılmadan önce incelenmiş ve ilaç-ilaç etkileşimlerinin tespitinde kullanılan bilgi tabanlı yöntemler olarak topolojik analiz, matris faktörizasyonu, lojistik regresyon, Naïve Bayes ağları, karar ağaçları, K-en yakın komşuluk, destek vektör makinaları, etiket propagasyonu, Bayes olasılık modeli, en kısa yol uzunluk ortalaması, derin öğrenme ağları, derin ileri beslemeli ağlar ve çizge evrimsel ağlarının kullanıldığı bilgisi elde edilmiştir (Angermueller ve diğerleri, 2016; Ayodele, 2010; Friedman, 2006; Osisanwo ve diğerleri, 2017; Samuel ve diğerleri, 1959; Vilar ve diğerleri, 2012).

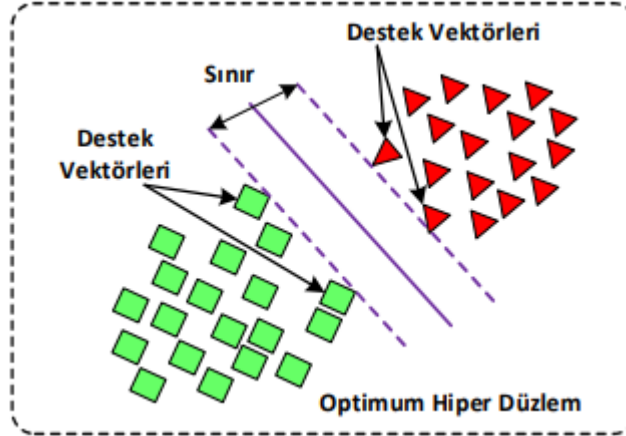
2.2.1 DESTEK VEKTÖR MAKİNELERİ

Destek vektör makineleri (DVM) girdi olarak verilen veri kümesini kullanarak öğrenmeyi gerçekleştiren, regresyon ve sınıflandırma için kullanıma elverişli olan denetimli bir makine öğrenme algoritmasıdır. DVM'nin amacı girdi olarak verilen eğitim verilerindeki iki sınıf arasında bir karar sınırı bulmaktır. Destek vektör makineleri, bir hiper düzlem kullanarak iki sınıfı en iyi şekilde ayıran bir sınırdır (Vapnik, 1999). Yöntem model karmaşıklığı üzerinde esnek bir kontrol sağladığı için doğrusal olmayan problemleri çözmede çok küçük eğitim setlerinde de başarılı sonuçlar vermektedir (Deng ve diğerleri, 2018). DVM eğitim hatasının yanında aynı zamanda bir güven aralığının toplamından oluşan genelleme hatasını da en aza indirmeye yönelik bir algoritmadır. Regresyon konusunda ise hiper düzlem sınırları içerisinde en fazla veri kalacak şekilde hatanın belli bir değer aralığında kalması sağlanır ve hata en aza indirgenir.

DVM'ler, doğrusal olmayan modelleri doğrusal bir probleme dönüştürmek için başlangıç verilerini daha yüksek boyutlu uzaya eşitlemek amacıyla çekirdek fonksiyonlarını kullanır (Qi ve diğerleri, 2019). Çekirdek fonksiyonunun seçimi destek vektör regresyonu performansı için büyük önem taşımaktadır (Zhong ve diğerleri, 2019). Literatürde en sık kullanılan çekirdek fonksiyonları Doğrusal, Polinomial, Sigmoid ve Gaussian Radyal tabanlı fonsiyondur (Yu ve diğerleri, 2006). Destek vektör regresyonunda amaçlanan eğitim verisi x_i değeri ile önceden belirlenen hata toleransından (ε) küçük olan bir uzaklıktaki y_i tahmin değerini hesaplayabilen $f(x)$ fonksiyonu oluşturmaktır (Li ve diğerleri, 2009). Bu fonsiyon Denklem 1'de gösterilmiştir. Şekil 2'de destek vektör makinelerinin sınıflar arasında nasıl ayırım yaptığı grafiksel olarak gösterilmektedir.

Denklem 1. Destek Vektör Makinesi

$$y = f(x) = W \cdot \phi(x) + b$$



Şekil 2. Destek Vektör Makineleri (WEB 3)

2.2.2 K-EN YAKIN KOMŞULUK ALGORİTMASI

K-en yakın komşuluk (K-EYK) makine öğrenmesi algoritması, veri kümesindeki K örneğini, uzaydaki benzerliği veya yakınlığı kullanarak tanımlamayı sağlar (Beretta ve Santaniello, 2016). Eksik değer ataması için kullanılan bir algoritma olan K-EYK

algoritması hem nicel hem de nitel özellikler için kullanılabilmekte ve bir dizi eksik değeri doğrudan işleyebilmektedir (Kim ve diğerleri, 2019).

Denklem 2'deki A_x , uzaklıkların artan sırasına göre yerleştirilen x girdi değişkeninin en yakın K komşusunu temsil etmektedir. K tane en yakın komşu seçildikten sonra, eksik veri A_x 'in j . özellik verilerinden elde edilen bir tahminle tamamlanmıştır. j . özellik sayısal bir değişken ise, tamamlanan veri, en yakın K komşusunun ortalama değeri kullanılarak elde edilir. K-EYK algoritması ile tamamlanan veri \tilde{x}_j , k 'ncı komşunun ağırlığı w_k ve x girdi değişkeninin en yakın komşusu v_{kj} ile hesaplanmaktadır (Denklem 3). K en yakın komşu değerlerine ait ağırlık w_k Denklem 4 ile elde edilir (Sanjar ve diğerleri, 2020).

Denklem 2. X değişkeninin en yakın K komşusu

$$A_x = \{v_k\}_{k=1}^K$$

Denklem 3. En yakın K komşunun ortalama değeri

$$\tilde{x}_j = \frac{1}{K} \sum_{k=1}^K w_k v_{kj} = 1$$

Denklem 4. En yakın komşu değerine ait ağırlık

$$w_k = \frac{1}{d(x, v_k)^2}$$

Her gözlemin x girdi değişkenine olan mesafesine göre, daha yakın komşulara daha büyük bir ağırlık atanır (Denklem 4).

2.2.3 LOJİSTİK REGRESYON

Lojistik regresyon ilk olarak 1940'larda, ikili sonuçları işleme konusunda sıradan en küçük kareler (SEKK) regresyonunun sınırlamalarının üstesinden gelmek için

alternatif bir teknik olarak önerilmiştir. 1980'lerin başlarında ise çeşitli istatistiksel paketlerde yerini almıştır. Bu tarihten sonra ise sosyal bilimlerde, yükseköğretimde eğitim araştırmalarında, epidemiyolojide, medikal uygulamalarda, tıp alanında, ekonomide, tarımda, veterinerlik, mühendislik ve taşıma sahalarında lojistik regresyon yaygın olarak kullanılmaktadır (Soyer, 2020).

Lojistik regresyon, bağımlı değişken üzerinde etki sahibi olan bağımsız değişkenlerin ne kadar etki ettiğini hesaplar. Lojistik regresyonda bağımsız değişkenlerin değerlerine göre bağımlı değişkenlerin beklenen değerini olasılık olarak elde ederiz. Lojistik regresyon, bağımlı değişkenin değerlerinin sınıflama ve atama işlemiyle en uygun gruba atanmasına yardımcı olmaktadır (Soyer, 2020).

Lojistik regresyon algoritması ikili, sıralı ve çok terimli olmak üzere üç farklı yöntemle ayrılır. Bağımlı olan değişkenin kategori sayısına ve ölçeğine göre bu yöntemlerden hangisinin kullanılacağı belirlenmektedir.

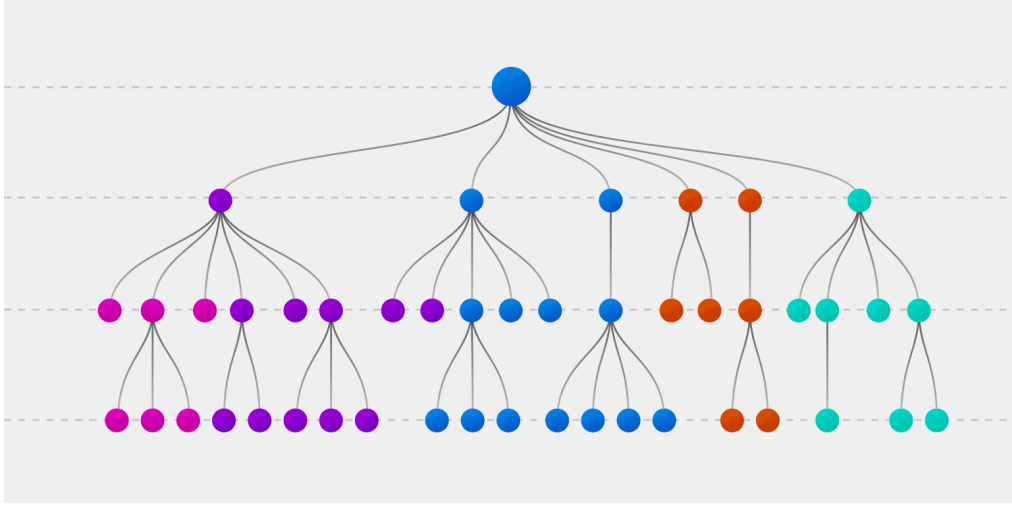
İkili lojistik regresyonda bağımlı değişken olarak 0/1, var/yok, evet/hayır, başarılı/başarısız gibi örnekler verilebilmektedir. Sıralı değişken olarak az/orta/çok, 0-3 yaş aralığı/4-6 yaş aralığı gibi örnekler, çok terimli değişken olarak ise çocuk-yetişkin-yaşlı, ortaokul-lise-üniversite gibi örnekler vermek mümkündür.

Lojistik regresyon algoritması uygulanabilmesi için bağımsız değişkenlerin herhangi bir varsayımı sağlama gerekliliği bulunmamaktadır. Bu yüzden lojistik regresyon bağımlı değişkenin yapısından kaynaklı olarak varsayımların sağlanamadığı durumlarda gözlemleri verilerin yapısında bulunan olası gruplara atamak için kullanılmaktadır. Diskriminant (ayırma) algoritması ve kümeleme algoritmalarında bağımsız değişkenlerin normal dağılım göstermesi gerekir ve çoklu doğrusal bağlantı içermemesi gerekmektedir. (Soyer, 2020).

2.2.4 KARAR AĞAÇLARI

Denetimli sınıflandırma modellerinden bir tanesi de karar ağaçlarıdır. Karar ağaçları bir kök, düğümler, dallar ve yapraklardan oluşan bir ağacın yapısından esinlenerek geliştirilmiştir. Şekil 3'te görüldüğü gibi karar ağaçları kökten başlayarak, aşağı doğru

hareket eder. Karar ağacının başlangıç düğümü kök düğüm olarak adlandırılır. Zincirin bittiği düğüm ise yaprak düğümdür. Karar ağacındaki her düğümün iki veya daha fazla dal çıkartılabilir. Her düğüm belirli bir özelliği temsil eder. Dallar ise bir dizi değer aralıklarını temsil eder. Bu değer aralıkları, verilen karakteristiğin değerler kümesi için bir bölme noktası görevini görmektedir (Ali ve diğerleri, 2012).



Şekil 3. Karar Ağacı (WEB 4)

2.2.5 ETİKET YAYIMI

Etiket Yayımları hem yarı denetimli hem de dönüştürücü bir öğrenme yöntemi olarak geçmektedir (Zhu ve Ghahramani, 2002). Etiket yayımı bir grafik yapısındaki bütün düğümlerin etiketlenmesi amacıyla kullanılır. Bu yöntemde öğrenme verilen veriler ile sınırlıdır ve örnek uzayın tamamını modellemeyi amaçlayan tümevarımsal algoritmalar kullanılmaktadır (Belkin ve diğerleri, 2004, 2006). Etiket yayımı algoritması iki ana varsayıma dayanmaktadır. İlk olarak, daha yakın düğümler, K-EYK'e benzer sınıf etiketlerine sahip olma eğilimindedir. İkincisi, aynı kümedeki noktalar aynı etikete sahiptir (Zhou ve diğerleri, 2004). Birincisi yerel, ikincisi ise küresel bir varsayımdır. Bu iki varsayım birlikte küme varsayımı olarak adlandırılır (Chapelle ve diğerleri, 2003; Wang ve diğerleri, 2008). Algoritma, ağırlıklı kenarlar ve yakınlık kullanarak bir grafikteki bir düğümün sınıf etiketini komşu düğümlere yayar. Benzersiz bir çözüme yakınsama elde edilene kadar, garanti edilen bu süreç

tekrarlanır (Zhu ve Ghahramani, 2002). Etiketlenmemiş verilere etiket atamak için etiketlenmiş veriler bir sınıflandırma modelinde kullanılır.

2.2.6 NAİVE BAYES ALGORİTMASI

Naive Bayes (NB) sınıflandırıcısının asıl amacı bir örneği eldeki veri setindeki sınıflardan bir tanesine olasılık hesabı ile yerleştirmektir (Friedman ve diğerleri, 2008). Bu yöntemde ilk olarak yeni gelen örneğin her sınıfa ait olma olasılığı hesaplanır. Daha sonra da hangi sınıfa ait olma olasılığı yüksek ise o sınıfa yerleştirilir.

NB yönteminde birbirlerinden bağımsız olan sürekli ve de kategorik değişkenler ele alınır. NB yönteminde verilen bir değişken kümesine $X = \{x_1, x_2, \dots, x_d\}$ karşılık C_j olaylarının gerçekleşme olasılığının hesaplanması söz konusudur. Bayes kuralı aşağıdaki gibi ifade edilir (Denklem 5).

Denklem 5. Bayes kuralı

$$p(C_j | x_1, x_2, \dots, x_d) = p(x_1, x_2, \dots, x_d | C_j) p(C_j)$$

Burada yer alan $p(C_j | x_1, x_2, \dots, x_d)$ X 'in C_j 'ye ait olma olasılığını ifade eder. NB yöntemi bağımsız değişkenlerin koşullu olasılıklarının istatistiksel olarak bağımsız olduğu varsayımı üzerine kurgulandığından dolayı yukarıdaki denklemde yer alan olasılık çarpımları Denklem 6'daki gibi ifade edilebilir.

Denklem 6. Bayes kuralı olasılık çarpımları

$$p(X | C_j) = \prod_{k=1}^d p(x_k | C_j)$$

Sonuç olarak aşağıdaki Denklem 7 elde edilir.

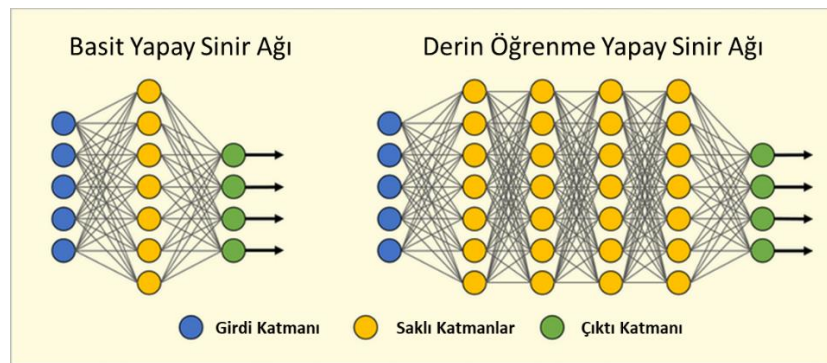
Denklem 7. Olasılık çarpımlarıyla birleştirilmiş Bayes kuralı

$$p(C_J | X) = p(C_J) \prod_{k=1} p(x_k | C_J)$$

$p(x_k | C_J)$ olasılıksal dağılım *Normal*, *Lognormal*, *Gamma* ve *Poisson* gibi farklı dağılım fonksiyonları seçilerek eldeki veri setinin doğasına göre NB farklı şekillerde modellenebilir (Çalışkan, 2017).

2.2.7 DERİN ÖĞRENME AĞLARI

Derin öğrenme (sinir) ağları (DSA) klasik sinir ağlarının (SA) mimarisinden esinlenerek tasarlanmış bir algoritmadır. SA’lardan farklı olarak DSA mimarisi ikiden fazla gizli katmandan oluşur. Son yıllarda özellikle grafik işleme birimlerinin (GİB) yaygınlaşması ile paralel işleme uygulamaları ön plana çıkmıştır. Bu durum DSA mimarisinin kullanımını artırmıştır. Derin modellerde milyonlarca hesaplamanın yapılarak ağırlıkların öğrenilmesi için paralel işlem kabiliyetleri sınırlı olan Merkezi işleme birimleri (MİB, İng. Central Processing Unit) yerine GİB’lar kullanılmaya başlanmıştır. DSA modelleri genel olarak aşırı öğrenmeye yatkındırlar. Başarılı bir genelleme öğrenabilmeleri için çok fazla verinin eğitim için kullanılması gerekmektedir. Verinin eskiye oranla çok daha ulaşılabilir olduğu günümüzde bu problem ortadan kalkmıştır.



Şekil 4. Basit Sinir Ağı Yapısı ve Derin Öğrenme Sinir Ağı Yapısı (WEB 5)

Derin öğrenme yöntemleri, her biri temsili bir düzeyde dönüştüren basit ancak doğrusal olmayan modüller oluşturularak elde edilen, birden çok temsil düzeyine sahip

temsil-öğrenme yöntemleri bütünüdür. Derin öğrenmenin kilit noktası, bu özellik katmanlarının mühendisler tarafından tasarlanmayıp, bu katmanların bir öğrenme prosedürü kullanılarak verilerden öğrenilmesidir (Lecun, Bengio ve Hinton, 2015).

2.3 İLAÇ –İLAÇ ETKİLEŞİMİNDE MAKİNE ÖĞRENMESİNİN KULLANILMASI

Yukarıda bahsi geçen metodlar ve başka teknikler İİE'lerin tespit edilmesi için çokça kullanılmıştır. Aşağıda veri işleme teknikleri kullanılarak gerçekleştirilmiş olan İİE tespit çalışmalarının bir özetini sunuyoruz:

Lin ve diğerleri yaptıkları çalışmada Amerikan Gıda ve İlaç Yönetimi (AGİY) tarafından bildirilen TİR vakalarını analiz eden bir sistem geliştirmişlerdir. Sistem dört aşamadan oluşmaktadır. Birinci aşamada TİR'e sebep olan ilaçlar ve ilaç hedef veritabanları birbirleriyle entegre edilmiştir. Daha sonra ilaçlar ve ilaç hedefleri anatomik terapötik kimyasal (ATK) sınıflandırmaya ve ilaç hedef ontolojisine (İHO) göre sınıflandırılır. Bir sonraki aşamada, ilaç ve ilaç hedeflerini içeren bir ilaç-hedef ağı oluşturulur. Son adımda ise AGİY tarafından rapor edilen her bir TİR vakası için ilaç etkileşim karmaşıklığını ortaya çıkarmak için topolojik analiz uygulanır (Lin ve diğerleri, 2009).

Tatonetti ve diğerleri klinik öneme sahip sekiz grup fenotipi manuel belirlemişlerdir. Daha sonra bu fenotiplere neden olan TİR'leri toplamışlardır. Her bir fenotip için bu TİR'lere sebep olan tedavilerdeki ilaçları ve ilaç çiftlerini bir matrise yerleştirerek çıktıyı Lojistic Regresyon yöntemiyle eğitmişlerdir. Bu eğitim sonucunda hangi ilaç çiftlerinin TİR'e sebep olabileceği tahmini yapılmıştır (Tatonetti ve diğerleri, 2012).

Vilar ve diğerleri çalışmalarında benzer moleküller benzer biyolojik etkilere sahiptir varsayımından yola çıkılarak yalnızca yapısal benzerlikleri dikkate almış ve ilaçların benzerlikleri için skor hesaplaması yapılır. Yapısal gösterim için bitlerden oluşan BIT_MACCS kodu vektör olarak tanımlanır. Burada her özelliğe belirli bir bit konumu atanır. Örneğin, C6H5-C(O)-NH2 molekülü için BIT_MACCS vektör kodundaki bazı yapısal özellikler şunlardır: bit 84 (NH₂, amin grubu), bit 154 (C=O, karbonil grubu),

bit 162 (aromatik, C₆H₅) ve bit 163 (altı üyeli halka, C₆H₅). Benzerliklerin hesaplanması için Tanimoto katsayısı (TK) yöntemi kullanılır. Elde edilen benzerlik skorları ve mevcut İİE bilgileri kullanılarak yeni İİE'lerin tahmin edilmesi sağlanır (Vilar ve diğerleri, 2012).

Gottlieb ve diğerleri yaptıkları çalışmada verilen bir ilaç çifti için var olan etkileşimlerdeki ilaç çiftleriyle olan benzerlikleri hesaplamışlardır. Yedi farklı benzerlik için yedi farklı skor hesaplaması yapılır. Daha sonra bu skorlar ilaçlar arasında bir etkileşim olup olmadığı tahmininde kullanılır. Bu benzerlik skorları kimyasal benzerlik, ligand bazlı kimyasal benzerlik, kayıtlı ve öngörülen yan etkilere dayalı benzerlikler, ATK sınıflandırma sistemi ve dizi benzerliği, protein-protein etkileşimi (PPE) ağı üzerindeki mesafe ve Gen Ontolojisi (GO) benzerliklerine göre hesaplanır. Eğer etkileşimde olup olmadığı tahmin edilecek olan ilaç çifti, mevcutta etkileşimde olan bir ilaç çiftine benziyor ise bu ilaç çiftinin etkileşime gireceği varsayılır (Gottlieb ve diğerleri, 2012).

Vilar ve diğerleri yine yaptıkları başka bir çalışmada belirli bir ilacın bilinen etkileşim ortaklarını 1 ve 0 lardan oluşan ikili bir vektör olarak kodlarlar. Daha sonra bu vektörün diğer ilaçların vektörleri ile ve mevcut İİE'ler ile benzerlik skoru çıkarılır. Bu benzerlik skoru ilaç etkileşimini tahminlemede kullanılır (Vilar ve diğerleri, 2013).

Cheng ve Zhao'nun geliştirdikleri sistemde ilaçların ATK sınıflandırma kodları, kimyasal yapı benzerlikleri, genomik yapı benzerlikleri ve TİR'e sebep olan fenotipik benzerlikleri hesaplanır. Her bir ilaç çifti için 4 tane benzerlik skoru elde edilir. Bu benzerlik skoru beş farklı tahmin modeline sokulur. Bu modeller Naive Bayes Ağları, Karar Ağaçları, K-En Yakın Komşuluk, Lojistik Regresyon ve Destek vektör makinalarıdır (Cheng ve Zhao, 2013).

Vilar ve diğerleri bir başka çalışmalarında ilaçların 2D ve 3D molkül yapılarının benzerlikleri ile etkileşim profili ve bilinen yan etkilerinin benzerlikleri hesaplayarak elde edilen skorlar ile ilaçların arasında bir etkileşim olup olmadığı tahmini yapmışlardır (Vilar ve diğerleri, 2014).

Zhang ve diğerleri yaptıkları çalışmada kimyasal yapılar ve yan etkiler birer vektör haline getirilmiş ve ilaç çiftlerinin birbirlerine benzerliği TK yöntemi kullanılarak

bulunmuştur. Daha sonra bütün ilaçların birbirleriyle bağlı olduğu bir ağ tasarlanmıştır. Bu ağda eğer iki ilacın bilinen bir etkileşimi var ise o kenar 1 değeri ile etiketlenmiştir. Etiket Yayımı (İng. Label Propagation) Algoritması kullanılarak etkiletlenmeyen kenarlarının ağırlıkları tahmin edilmiştir (Zhang ve diğerleri, 2015).

Huang ve arkadaşları farmakodinamik ilaç-ilac etkileşimlerini bulmak üzere bir ağ öğrenmesi çalışması yapmışlardır. Temel olarak iki ilacın aynı gen üzerine etki etmesi birbirlerinin etkinliğinde beklenmedik bozulmalara yolaçacağını varsaymışlardır. Çalışmalarında ilaç-hedef ilişkilerinden, protein-protein etkileşimi ağ topolojisinden, dokular arası gen ekspresyon korelasyonlarından gelen bilgileri entegre etmişler ve potansiyel farmakodinamik İİE'leri tahmin etmek için bir sistematik skor (S-skoru) hesaplamışlardır. Daha sonra ilaçların yan etkilerine dayanarak bir fenotipik benzerlik skoru (F-skoru) elde etmişlerdir. Elde edilen bu skorlara Bayes olasılık modelini entegre etmişlerdir (Huang ve diğerleri, 2013).

Park ve diğerleri iki ilacın aynı gen üzerine etki etmesinin ilaçların etkinliğinde değişiklik yapacağı varsayımından yola çıkarak farmakodinamik etkileşimleri bulmak üzerinde bir çalışma yapmışlardır. İlaç-hedef bilgileri, mevcut İİE bilgileri ve protein-protein etkileşim ağı bilgilerini kullanarak bir ağ yapısı tasarlanmış ve En Kısa Yol Ortalaması algoritması kullanılmıştır (Park ve diğerleri, 2015).

Ryu ve diğerleri sadece ilaç isimleri ve ilaçların yapısal bilgilerini girdi olarak alan bir makine öğrenmesi çalışması yapmışlardır. İlaç yapısı girdisi olarak basitleştirilmiş moleküler yapı kodu (BMYGS) kodu verilir. Her bir ilaç çifti için iki ayrı YBP oluşturulur. Daha sonra bu iki YBP birleştirilerek tek bir vektör haline getirilir. Bu vektörler DSA modeline sokulur. Her bir ilaç çiftinin etkileşim türü farklı olabileceğinden toplamda 86 çeşit etkileşim tipi için YBP ler kullanılarak öğrenme işlemi uygulanır. Bu işlem ayrıca ilaç-gıda etkileşimlerinin bulunması için de kullanılmıştır (Ryu ve diğerleri, 2018).

Qian ve diğerleri yaptıkları çalışma sonucu olumsuz etkileşen ilaçların hedeflerinin, etkileşmeyen ilaçların hedeflerinden daha fazla sinerjik genetik etkileşime sahip olma eğiliminde olduğunu göstermişlerdir. Bu bulgudan yola çıkarak ilaçların en çok kullanılan üç özelliği (yan etki, hedef, indükasyon) ve ilaç-ilac etkileşimi bilgilerinden

yararlanarak bir makine öğrenme çerçevesi sunmuşlar ve gradyan artırma tabanlı bir sınıflandırıcı geliştirmişlerdir (Qian ve diğerleri, 2019).

Lee ve diğerleri yapısal benzerlik, hedef gen benzerliği ve gen ontolojisi benzerliği şeklinde her bir ilaç çifti için üç adet benzerlik skoru hesaplarlar. İİE tahmini yapmak için ise üç adet otomatik kodlayıcı ve bir adet derin ileri beslemeli ağ kullanırlar. Otomatik kodlayıcılar birbirinden bağımsız olarak girdi ve çıktıların farkını en aza indirmek için eğitilir. Daha sonra ileri beslemeli ağ, girdi olarak indirgenmiş profil çiftleri ile eğitilir. Eğitim, üç otomatik kodlayıcıdan ve ileri beslemeli ağdan gelen maliyetlerin toplamını en aza indirmek için gerçekleştirilir. Bu nedenle, otomatik kodlayıcılar iki kez eğitilir ve İİE türünü daha doğru tahmin etmek için profilleri kodlar (Lee ve diğerleri, 2019).

Bir başka çalışmada Deng ve diğerleri öncelikle ilaçların özelliklerinden ilaç çiftlerinin benzerliğini hesaplar. Daha sonra topolojik yapı bilgisini ve ilaçların anlamsal ilişkilerini yakalamak üzere Grafik sinir ağlarından yararlanırlar. İİE'ler toplamda 65 kategoriye ayrılır. İlk adımda kimyasal yapı, hedefler, enzimler ve yolakların bilgisinden DSA kullanılarak dört farklı alt model oluşturulur. Daha sonra bu alt modeller çapraz etkileşim temsillerini öğrenmek ve İİE'leri tahmin etmek üzere yeni bir DSA çerçevesini oluştururlar (Deng ve diğerleri, 2020).

Mei ve diğerleri çalışmalarında bir ilaç, hedeflenen genler veya sinyal yolakları aracılığıyla diğer ilacın terapötik etkilerini değiştirdiğinde, iki ilacın potansiyel olarak etkileşime girdiğini varsayıldığından yola çıkılarak bir protein-protein etkileşim ağı ve protein yolaklarının ağını oluşturmuşlardır. Daha sonra iki ilaç arasındaki etkileşim yoğunluğunu, etkileşim etkinliğini ve etki aralığını ölçmek için elde ettiklerini toplamak suretiyle lojistik regresyon modeli uygulamışlardır (Mei ve Zhang, 2021).

Feng ve diğerleri ise çalışmalarında ilaçlarla alakalı hiç bir özellik girdi olarak alınmaz. Bu bir takım ilaç özelliklerinin eksik olduğu durumlarda da modelin kullanılmasını sağlamıştır. Yalnızca ilaçların etkileşim bilgileri ile bir çizge evrişim ağı (ÇEA) kurulur. DSA ise ağdaki ilaçların topolojik ilişkisini yakalayarak ilaçların düşük boyutlu özellik temsillerini öğrenir. Öğrenilen bu bilgiler sayesinde yeni tahminler yapılır (Feng ve diğerleri, 2020).

Bütün bu incelediğimiz çalışmalar sonunda ilaçların etkileşimini farmakokinetik veya farmakodinamik olarak tespit eden bir model olmadığı görülmüştür. Ayrıca ilaç etkileşimlerini ciddiyet bakımından da tahmin eden bir yöntem henüz denenmemiştir. Tezimizde bu ikisini tahmin etme kabiliyetine sahip bir model çıkarmak üzerine çalışacağız.

BÖLÜM III – METOD

Tezin bu bölümünde kullanılan ham veri setinin elde edilip yapısal olarak dönüştürülmesinden, yapısal haldeki verinin ön işleme sokularak makine öğrenmesi için girdi olarak sağlanacak şekilde çeşitli dönüşüm metodlarının nasıl uygulandığına dair bilgiler yer almaktadır. Ayrıca makine öğrenmesi algoritmalarının uygulanması ve çıkan sonuçlar hakkında da ayrıntılı bilgilere yer verilmektedir.

3. METOD

3.1 VERİNİN ALINMASI ve YAPISAL HALE GETİRİLMESİ

Çalışmamızda kullandığımız veriler Drugbank veritabanından alınmıştır. Drugbank'ın en son sürümü olan 5.1.8 sürümü kullanılmıştır. Ticari olarak kullanılması için lisans gereken Drugbank veritabanı akademik çalışmalarda ücretsiz olarak kullanılabilir. Drugbank veritabanında her ilaca ait iki yüzün üzerinde özellikle birlikte ilaçların birbirleriyle olan etkileşim bilgileri de bulunmaktadır (WEB1).

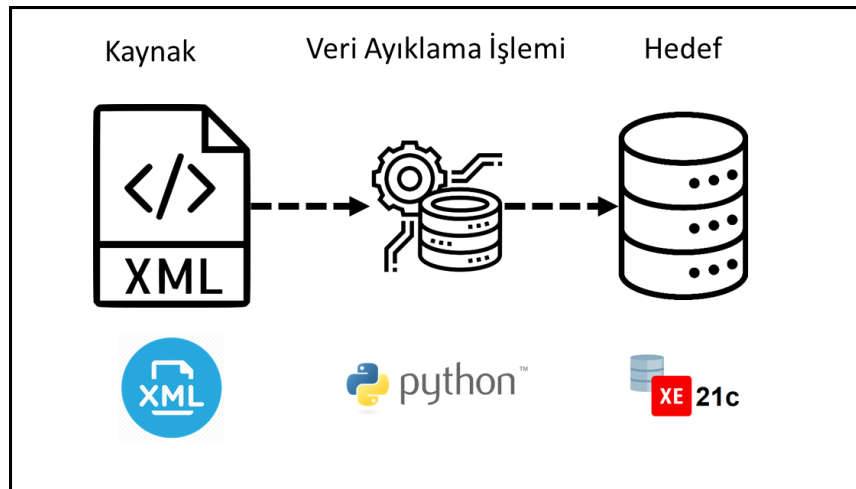
DrugBank, ilaçlar ve ilaç hedefleri hakkında bilgi içeren kapsamlı, erişimi ücretsiz bir veritabanıdır. Hem biyoinformatik hem de kimyasal bilişim kaynağı olarak ayrıntılı ilaç (yani kimyasal, farmakolojik ve farmasötik) verilerini kapsamlı ilaç hedefi (yani dizi, yapı ve yol) bilgileriyle birleştirir. DrugBank, ilaç endüstrisi, tıbbi kimyagerler, eczacılar, doktorlar, öğrenciler ve genel halk tarafından yaygın olarak kullanılmaktadır. Geniş kapsamı, kapsamlı referansları ve ayrıntılı veri açıklamaları nedeniyle DrugBank, veriye dayalı ilaç endüstrisinde büyük ilerlemeler sağlanmasına yardımcı olmuştur. DrugBank'ın en son sürümü (sürüm 5.1.8, 3 Ocak 2021'de yayınlandı), 2.706 onaylı küçük moleküllü ilaç, 1.505 onaylı biyolojik (proteinler, peptitler, aşılar ve alerjenikler), 132 nutrasötik ve 6.655'in üzerinde

deneysel (keşif aşaması ilaçlar) dahil olmak üzere 14.587 ilaç bilgisi içerir. Herbir ilaç için bilgilerin yarısı ilaç/kimyasal verilere ve diğer yarısı ilaç hedefi veya protein verilerine ayrılmış 200'den fazla veri alanı içerir (WEB1).

Drugbank veritabanını web üzerinden genişletilebilir işaretleme dili (XML) dosya formatında indirmek mümkün olmuştur. XML, yapısal verilerin düz yazı formatında tutulduğu basit bir dosya biçimidir. XML formatında yazılmış olan bilgiler hem insanlar tarafından hem de platform bağımsız olmak üzere makinalar tarafından okunabilmekte ve yorumlanabilmektedir.

Drugbank veritabanı dosyası yapısal olarak oldukça karmaşık bir yapıdadır. Çalışmada kullanılacak olan verileri her seferinde bu dosyadan ayıklayarak kullanmak hem yazılımsal olarak maliyet getireceği ve karmaşıklığı artıracığı için hem de çalışmaları yapacağımız makina sistemlerini fazladan kullanacağı için bu veriler bir ilişkisel veritabanına aktarılmıştır.

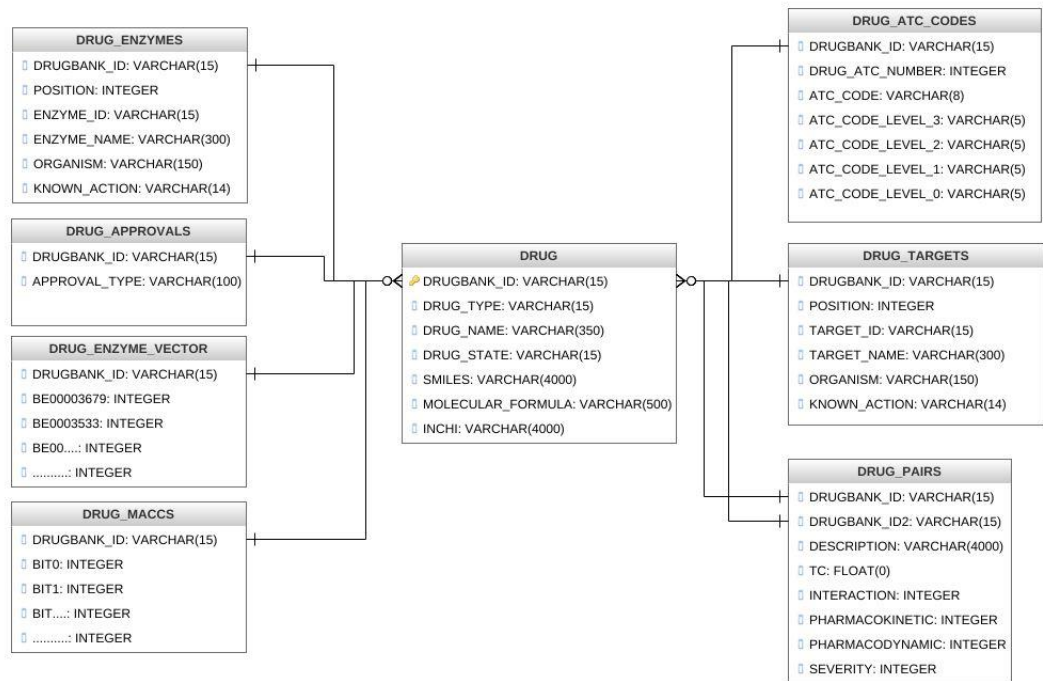
XML dosyasından verileri ayıklayıp ilişkisel veritabanı sistemine yüklemek için Python progama dili kullanılmıştır. Geliştirip deneyecek olduğumuz makine öğrenmesi modellerini de bu yazılım dili ile uygulayacağımızdan XML ayıklama işleminde de yine Python dili tercih edilmiştir. Python kullanım kolaylığıyla öne çıkan çok esnek bir programlama dilidir. Python dilinde erişilmesi mümkün olan bir çok açık kaynak kodlu makine öğrenmesi kütüphanesi bulunmaktadır. Bu iki özellik zamanı verimli kullanmak ve güvenilir modeller geliştirmek açısından Python dilini seçmemize neden olmuştur.



Şekil 5. XML İlaç Verilerinin Veritabanına Kaydedilmesi Akış Diyagramı

Verileri XML dosyasından ayrıştırıp koyduğumuz yapısal veritabanı ise Oracle XE 21c versiyonudur. Oracle XE belirli limitler dahilinde ücretsiz olarak indirilip kullanılabilen bir yapısal veritabanı programıdır. Oracle XE ile 12GB veri ile 2GB bellek kullanımı ücretsizdir. Yapacağımız çalışmada veritabanı olarak bu kaynakların kullanımı yeterli olduğu için ve kullanım ücretsiz olduğu için Oracle XE'nin kullanımı uygun görülmüştür.

Oluşturduğumuz ilaç veritabanı ise Şekil 6'da gösterildiği gibidir. DRUG tablosunda ilaçlara ait ilaç ismi, ilaç açıklaması, ilaç tipi, ilaç özellikleri gibi genel bilgiler bulunmaktadır. Bu tabloda bulunan DRUGBANK_ID alanı tablonun birincil anahtarıdır ve her ilaç için tekil bir kod numarasını tutar.



Şekil 6. DRUG Veritabanı Diyagramı

Bir ilaç birden fazla olabilecek bilgiler için yeni bir tablo oluşturulması uygun görülmüştür. Örneğin DRUG_ATC_CODES tablosunda ilaçları ATK kodları bulunur. Bir ilaç için yalnızca bir adet ATK kodu bilgisi bulunurken başka bir ilaç için otuzun üzerinde ATK kodu bilgisi bulunması mümkündür.

DRUG_PAIRS tablosu birbiriyle etkileşimde olan ilaç çiftlerinin bilgilerini ve bu etkileşimin nasıl bir etkileşim olduğu hakkındaki açıklamaları tutar. İlaçların etkileşim türleri (farmakokinetik, farmakodinamik) ve ilaçların etkileşimlerinin ciddiyet derecesi ve ayrıca ilaçların molekül yapılarının benzerlikleri hesaplanmış ve bu tabloya eklenmiştir. DRUG_TARGET tablosunda ilaçların hedef bilgileri, DRUG_ENZYMES tablosunda ilaçlara enzim bilgileri bulunmaktadır. DRUG_ATC_CODES tablosunda da ilaçların ATK sınıflandırma kodu bilgileri bulunmaktadır. DRUG_MACCS tablosuna ilaçların BMYGS kodları kullanılarak oluşturulmuş 167bit lik MACCS molekül vektörleri yerleştirilmiştir. DRUG_ENZYME_VECTOR tablosuna ilaçların etki ettikleri enzim bilgileri bir vektör haline getirilerek eklenmiştir. Toplamda 458 adet enzim bilgisi her ilaç için bu tabloda tutulmaktadır.

3.2 MODEL İÇİN VERİNİN SEÇİLMESİ

Drugbank veritabanında toplamda 14,315 adet ilacın bilgisi bulunmaktadır. Bu ilaçların büyük bir çoğunluğu uygulayacağımız modellerde kullanılmayacaktır. Modelin veri setine girecek ilaçlar bazı kriterler belirlenmiştir. İlk olarak seçtiğimiz ilaçların mutlaka onaylanmış olmasını beklemekteyiz. İkincisi seçtiğimiz ilaçların küçük molekül yapıda (moleküler ağırlığı 900 daltondan daha küçük) olması gerekmektedir. Bir başka kısıtımız ise ilaçların protein olmaması yani yaşayan organizmalardan elde edilmemiş olması gerekir. Ayrıca antikor olan ilaçlar da çalışmamıza dahil edilmemiştir. Bunun yanında tüm özellikleri sağlamasına rağmen ilaçların bazı bilgilerinin eksik olması hesaplamalarda ve tahminlerde hatalara yol açacağından eksik bilgiye sahip ilaçları da kasam dışında bıraktık. Neticede çalışmamızı yapmak için 650 adet ilaç kullanılmıştır.

Her bir ilaç diğer ilaçlar ile eşleştirildiğinde toplamda 210,925 adet ilaç çifti elde edilmiştir. Bu ilaç çiftlerinin etkileşimleri ve etkileşim tipleriyle alakalı rakamlar Tablo 1' de verilmiştir.

Tablo 1. İlaç Çiftlerinin Etkileşim Sayıları

	Açıklama	Adet
Etkileşen ve Etkileşmeyen İlaç Sayıları	Etkileşen İlaç Çifti Sayısı	79,498
	Etkileşmeyen İlaç Çifti Sayısı	131,427
	Toplam	210,925
Etkileşim Tipleri	Farmakokinetik Etkileşim	37,708
	Farmakodinamik Etkileşim	41,790
	Toplam	79,498
Etkileşim Ciddiyet Dereceleri	Minör Etkili Etkileşim	20,186
	Orta Etkili Etkileşim	58,176
	Majör Etkili Etkileşim	1,140
	Toplam	79,498

3.3 VERİLERİN ÖNİŞLEME SOKULMASI

Bir önceki bölümde verinin Drugbank kaynağından çekilerek yapısal bir biçimde nasıl tutulduğu hakkında bilgiler verilmiştir. Yapısal olarak tutulan verilerin makine öğrenmesi modellerinde kullanılabilmesi için bir takım önışlemden geçerek uygun formata getirilmesi gerekmektedir. Bunun için yapmış olduğumuz çalışmalar, kullandığımız teknikler ve kütüphaneler hakkında bilgiler bu kısımda sunulmuştur.

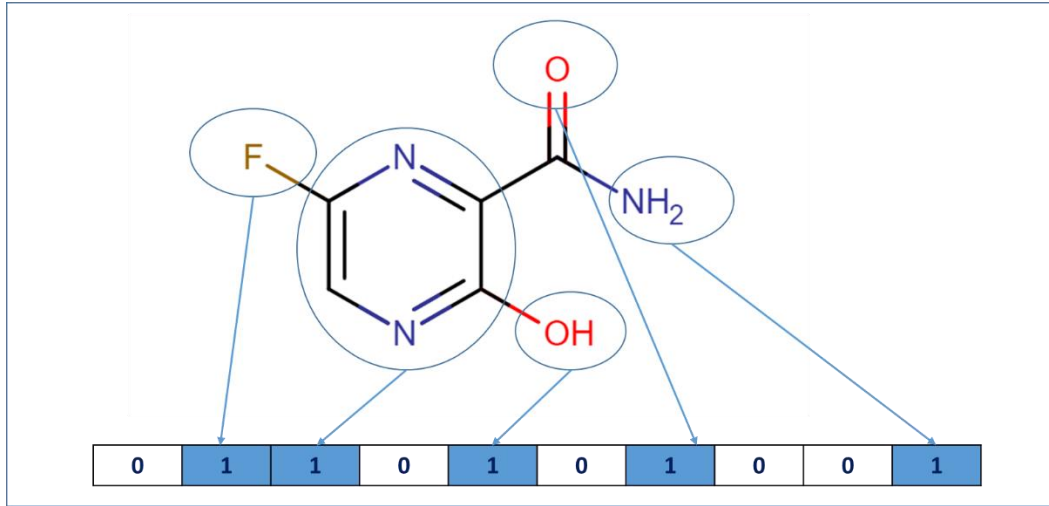
3.3.1 BMYGS VERİSİNİN DÖNÜŞTÜRÜLMESİ

BMYGS, molekül yapılarının alfabetik karakterler ve diğer noktalama işaretleri kullanılarak kesin bir şekilde tanımlanmalarına olanak veren bir sistemdir. Şekil 7’de Firapiravir in molekül yapısı ve buna karşılık üretilmiş olan BMYGS kodu bulunmaktadır.



Şekil 7. Firapiravir İlacının Moleküler Yapısı ve Üretilen BMYGS Kodu

BMVGS kodlarının bu haliyle herhangi bir modele sokulması veya herhangi bir istatistik hesabında kullanılabilmesi mümkün değildir. Bunun için BMVGS verisini ikili koda (İng. Binary Code) çeviren bir takım yöntemler geliştirilmiştir. Bu yöntemler her bir molekül için eşit uzunlukta olan, 0 ve 1’lerden meydana gelen bir vektör oluştururlar. Bunlardan en bilinenlerden bir tanesi MACCS parmak izidir (İng. MACCS Fingerprint). MACCS parmak izi her molekül için 167 bit uzunluğunda bir vektör olarak tanımlanır. Bu vektörler oluşturulurken moleküllerin sahip olduğu her özellik için bir bit atanır. Eğer molekülde belirli bir özellik varsa o özelliğe ait bit 1 olarak işaretlenir, yok ise 0 olarak işaretlenir. Şekil 8’de bunun örnek kavramsal bir açıklaması gösterilmeye çalışılmıştır.



Şekil 8. Moleküller İçin İkili Vektör Oluşturma Yöntemi

BMVGS kodlarından oluşturulan vektörler öğrenme modellerinde direk olarak kullanılabilir niteliktedir. Bunun yanında bu vektörler ilaçların molekül yapılarının benzerliklerini hesaplamak için de kullanılmaktadır. Molekül yapısı benzer olan ilaçların benzer etkiler göstereceği varsayımından yola çıkarak benzerliklerin modele sokulması çok yaygın bir yöntemdir. Molekül yapıların benzerliklerini hesaplamak için Tanimoto Katsayısı kullanılmaktadır. TK iki vektörün benzerliğini ölçmek için kullanılan bir skordur. TK 0 ile 1 arasında bir değer alır. Değerin 0’a yakın olması moleküllerin birbirine benzemediğini, 1’e yakın olması ise moleküllerin birbirine çok benzediğini ifade eder. TK’nin hesaplanma formülü Denklem 8’de gösterilmektedir. Şekil 9’da ise bu formülün nasıl kullanılacağı ayrıntılı olarak ifade edilmektedir.

Denklem 8. Tanimoto Katsayısı Formülü

$$TK = \frac{A \wedge B}{A \vee B}$$

A	1	0	1	1	0	0	0	1	0	1	A = 5
B	0	0	1	0	1	1	0	0	0	1	B = 4
A ∧ B	0	0	1	0	0	0	0	0	0	1	A ∧ B = 2
A ∨ B	1	0	1	1	1	1	0	1	0	1	A ∨ B = 7
$TC = \frac{2}{7}$											

Şekil 9. Tanimoto Katsayısının hesaplanması

TK değeri kapsam dahilindeki her bir ilaç çifti için hesaplanarak kaydedilmiştir. Daha sonra bu benzerlikler Şekil 10'da gösterildiği üzere bir matris haline getirilmiştir. Bu matrisin satırları ve sütun isimleri ilaçları temsil etmektedir. Değerler ise herbir ilaç çiftinin TK benzerlik değerleridir.

	ilaç 1	ilaç 2	ilaç 3	ilaç 4	ilaç 5	ilaç 6
ilaç 1	1	0.0254	0.0003	0.1532	0.7265	0.0055
ilaç 2	0.0254	1	0.0124	0.3565	0.9322	0.5560
ilaç 3	0.0003	0.0124	1	0	0.0623	0.0455
ilaç 4	0.1532	0.3565	0	1	0.1159	0.1418
ilaç 5	0.7265	0.9322	0.0623	0.1159	1	0.0074
ilaç 6	0.0055	0.5560	0.0455	0.1418	0.0074	1

Şekil 10. İlaçların TK Benzerlik Matrisi

Bu matristen yola çıkılarak her bir ilaç için toplam 650 (kapsamdaki ilaç sayısı kadar) değerli bir benzerlik vektörü elde edilmiştir. İlaçlar modele yalnızca moleküler benzerlik bakımından girdi sağlasa dahi bir ilaç çifti toplamda 1300 değerli bir vektör ile model girmiş bulunmaktadır. Makine öğrenmesi modellerinde bu kadar fazla özelliğin öğrenme işlemine tabi tutulması aşırı öğrenme problemine yol açmaktadır. Bundan kurtulmak amacıyla özellik azaltma yöntemleri geliştirilmiştir. Modellerde

kullanılacak olan veri sayısını azaltmak amacıyla Temel Birleşenler Analizi (TBA) kullanılmıştır. Bu yöntem kullanılarak her ilaç için 50 değerli bir vektör elde edilmiştir.

TBA’da temel amaç mümkün olan en yüksek varyans ile verinin boyutunun azaltılmasıdır. Boyut azaltma verideki bazı özelliklerin kaybolmasına sebebiyet vermesine rağmen kaybolan özellikler veri hakkında çok az bilgi içermektedir. Bu yöntemle yüksek korelasyonlu veriler bir araya getirilir. Verilerdeki en çok varyasyonu oluşturan ‘Temel Birleşenler’ seçilir. TBA sonucu oluşturulan her bir özellik birbirinden bağımsızdır.

Veriye TBA uygulanmadan önce kullanılan yaygın bir yöntem verilerin standartlaştırılmasıdır. Bu dağılımın standart sapması bire eşit olacak şekilde değerleri yeniden düzenlemek anlamına gelmektedir. Biz modellerimizi hem verileri standartlaştırarak hem de standartlaştırmadan test edip en uygun yöntemi bulmaya çalışacağız.

3.3.2 ATK, ENZİM ve HEDEF VERİLERİNİN DÖNÜŞTÜRÜLMESİ

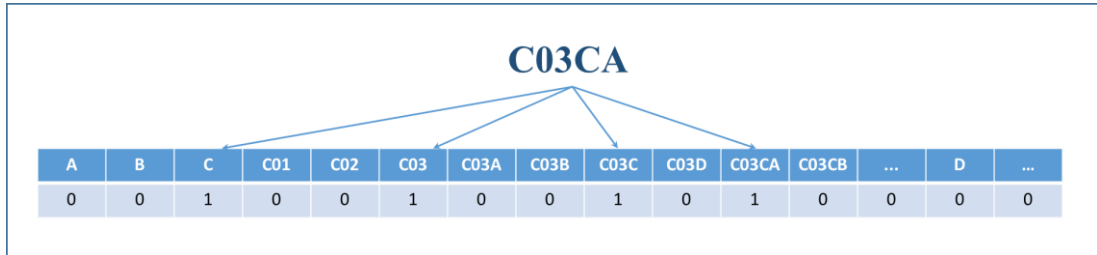
ATK, ilaç kullanımını izlemek ve kullanım kalitesini artırmak amacıyla araştırmalar yapmak için geliştirilmiş bir sınıflandırma sistemidir. İlaçları etken maddelerine göre sınıflandırmaya yarar. Bu sınıflandırma yapılırken ilacın etki ettiği yer ve ilaçların terapötik, kimyasal ve farmakolojik özellikleri dikkate alınır.

Bu sistemde ilaçlar beş seviyede gruplanırlar. Birinci seviye anatomik ana grubu belirtir. İkinci seviye terapötik alt grubu belirtir. Üçüncü seviye terapötik/farmakolojik alt grubu belirtir. Dördüncü seviye kimyasal/terapötik/farmakolojik alt grubu belirtir. Beşinci seviye kimyasal maddeyi belirtir. Tablo 2’de örnek bir ATK kodu sınıflandırması verilmiştir. Tabloda birinci seviyeden başlayarak beşince seviyeye kadar uzanan sınıflandırmanın aşamaları görülmektedir. Birinci seviye bir harften oluşur. İkinci seviye iki haneli bir rakamdan, üçüncü ve dördüncü seviye yine birer harften oluşmaktadır. Beşinci seviye ise iki haneli bir rakam olup kimyasal maddeyi temsil etmektedir. Beşinci seviyenin varyansı çok fazla olduğundan dolayı bu bilgiyi modellerimize girdi olarak vermeyeceğiz.

Tablo 2. ATK Kodu Sınıflandırması Örneği (WEB 6)

ATK Seviyesi	ATK Kodu	ATK Açıklaması
Anatomik Ana Grup	C	Kardiyovasküler Sistem
Terapötik Alt Grup	C03	Diüretikler
Farmakolojik Alt Grup	C03C	Yüksek Tavan Diüretikleri
Kimyasal Alt Grup	C03CA	Sülfonamidler
Kimyasal Madde	C03CA01	Furosemid

Her ilaç bir ATK kategorisine dahil olabileceği birden fazla ATK kategorisine giren ilaçlar da bulunabilmektedir. Bu durum ATK kodlarının her ilaç için bir özellik olarak model verisine girmesine imkan tanımamaktadır. Bu yüzden ATK kodlarının anlamlı bir şekilde modele girdi olarak verilebilmesi için bu veriyi her bir kolona bir sınıflandırma kodu gelecek şekilde bir vektör hazırladık. Örneğin yukarıda Tablo 2’deki örnekte görülen C, C03, C03C, C03CA kategorileri için birer kolon oluşturduk. Bunun gibi ilk dört seviyedeki bütün kategorileri birer kolon haline getirdik. Eğer ilaç ATK sınıflandırmasında herhangi bir sınıfa giriyor ise o sınıfı 1 olarak girmiyor ise 0 olarak kodladık. Şekil 11’de yaptığımız işlem sembolik olarak görünmektedir.



Şekil 11. ATK Kodu Vektörü Oluşturulması

Oluşan ATK vektörünün uzunluğu 712 kolon olmuştur. Bu kadar fazla özellik yine aşırı öğrenme problemine yol açacağından TBA kullanılarak bu sayı 50’ye düşürülmüştür.

İlaç enzimlerinin ve ilaç hedeflerinin makine öğrenmesi modeline girmesini sağlamak amacıyla veriler ATK kodlarına benzer şekilde bir ön işleme tabi tutulmuştur. Enzim ve hedef verileri için birer vektör oluşturularak bu vektör içinde ilacın bilgilerindeki enzim veya hedef var ise 0, yok ise 1 ile işaretlenmiştir. Hedef bilgilerini çeren vektörün boyutunu azaltmak için yine TBA kullanılmıştır. Enzimlerdeki vektör

büyükliğini azaltmak için yalnızca SİP enzimlerinin bilgisi dikkate alınmıştır. Bu durum, “*SİP enzimlerini inhibe eden ilaçlar, diğer ilaçların parçalanmasını engellediği için güvenlik problemleri var denir*” görüşünden kaynaklanmaktadır (Küçükçalık, 2019:4)

3.4 MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN UYGULANMASI

Makine öğrenmesi tekniklerinin uygulanması için ilk olarak modele sokulacak olan verinin uygun bir formatta hazırlanmış olması gerekmektedir. Bir önceki bölümde bahsedildiği üzere her bir ilacın toplamda dört özelliği için vektörler oluşturduk. En verimli yöntemi bulmak için bu vektörlere çeşitli dönüşümler uygulayarak farklı varyasyonlar elde ettik. Hangi vektör dönüşümlerinin birlikte daha iyi çalıştığını görmek için 8 farklı veri setini makine öğrenmesi algoritmalarına girdi olarak hazırladık. Bu veri setlerindeki vektörleri herhangi bir kurala bağlı olmaksızın rasgele seçme yöntemini kullanarak hazırladık. Bu veri setlerine ilaçların dört özelliğine ait vektörlerle birlikte ilaç-ilac etkileşimi bilgisini de ekledik.

En son düzenlemelerle modele girecek olan verinin formatı A ve B ilaç çifti için A ilacının benzerlik vektörü, B ilacının benzerlik vektörü, A ilacının ATK kodu vektörü, B ilacının ATK kodu vektörü, A ilacının enzim vektörü, B ilacının enzim vektörü, A ilacının hedef vektörü, B ilacının hedef vektörü, iki ilacın etkileşim bilgisi, farmakokinetik etkileşim bilgisi, farmakodinamik etkileşim bilgisi, etkileşimin ciddiyet bilgisini içerecek şekilde düzenlenmiştir. Tablo 3 hazırlanan 8 farklı veri setinin içeriğini göstermektedir. Her modelin bu sekiz farklı veri seti üzerinde çalıştırılması sağlanmıştır. Bu bizim hangi veri setinde daha başarılı sonuçları aldığımız konusunda bize bilgi sağlayacaktır. Ayrıca hangi modelde hangi veri setinin daha iyi çalıştığı bilgisini almamıza da yardımcı olacaktır.

Tablo 3. Modellerde Test Edilecek Olan Veri Setlerinin İçerikleri

	Veri 1	Veri 2	Veri 3	Veri 4	Veri 5	Veri 6	Veri 7	Veri 8
Benzerlik Vektörü 50	✓	✓						
Benzerlik Vektörü 50 Std					✓			
Benzerlik Vektörü 100 Std.				✓				
Benzerlik Vektörü 100			✓					
Benzerlik Vektörü – Tümü						✓		
Enzim Vektörü SİP	✓	✓	✓	✓	✓	✓		
Enzim Vektörü – Tümü							✓	✓
Hedef Vektörü – Tümü	✓							
Hedef Vektörü 50		✓	✓					✓
Hedef Vektörü 50 Std				✓	✓	✓	✓	
ATK Vektörü 50	✓	✓	✓					
ATK Vektörü 50 Std				✓	✓	✓	✓	✓
Etkileşim Bilgisi	✓	✓	✓	✓	✓	✓	✓	✓
Farmakokinetik Bilgisi	✓	✓	✓	✓	✓	✓	✓	✓
Farmakodinamik Bilgisi	✓	✓	✓	✓	✓	✓	✓	✓
Ciddiyet Bilgisi	✓	✓	✓	✓	✓	✓	✓	✓

Ryu ve arkadaşları 2018 yılında yaptıkları çalışmada ilaçların benzerlik vektörünü TBA yöntemi ile 10, 20, 30, 40 ve 50 kolonlu olarak denemiş ve en iyi sonucu 50 kolon için almışlardır. Biz de çalışmamızda vektörleri 50 kolon olarak ve 100 kolon olarak denedik. Bunun yanında tüm vektör setini de bazı modellerde başarıımı ölçmek maksadıyla denedik.

İlk olarak verilerimizi ilaçlar arasında etkileşim olup olmadığı tahmini yapma konusunda eğittik. Bu eğitim için verileri ayrı ayrı Lojistik Regresyon (LR), Naive Bayes, Karar Ağaçları, K-En Yakın Komşuluk ve Derin Öğrenme Ağları uygulanmıştır. Bu uygulamalar yapılırken toplam verinin %90'lık bir kısmı eğitime sokulurken %10'luk kısmı testler için ayrılmıştır. Test verileriyle uygulanan veriler

için performansı belirlemek adına F1-skorları hesaplanmış ve Tablo 4 içerisinde gösterilmiştir.

Tablo 4. Veri Setlerinin Etkileşimleri Tahmin Etmede Elde Ettikleri F1-Skorları

	LR	NB	DT	K-EYK	DSA
Veri Seti 1	0.78	0.43	0.85	0.82	0.92
Veri Seti 2	0.75	0.57	0.86	0.83	0.90
Veri Seti 3	0.91	0.56	0.84	0.82	0.90
Veri Seti 4	0.75	0.55	0.84	0.79	0.88
Veri Seti 5	0.74	0.54	0.85	0.79	0.85
Veri Seti 6	0.79	0.64	0.83	0.79	0.82
Veri Seti 7	0.73	0.51	0.85	0.79	0.83
Veri Seti 8	0.76	0.42	0.86	0.83	0.90
Ortalama	0.77625	0.5275	0.8475	0.80625	0.875

Yukarıdaki tabloda görüldüğü üzere DSA ilaç-ilaç etkileşimini tahmin etmede diğer metodlara göre daha üstün bir başarı sağladığı görülmektedir. Karar ağaçları yine ona çok yakın bir trend göstermektedir. Bir sonraki aşamada yine aynı veri setleriyle, bu sefer sadece ilaç-ilaç etkileşimlerini değil bunun yanında etkileşim türünü (farmakokinetik veya farmakodinamik) ve etkileşimin sağlık adına ciddiyetini (minör, orta veya majör) de tahmin edilmesi adına çalışmalar yapılmıştır. Bu çalışmalarda hem bir önceki aşamada yüksek performans sergilediği için hem de katman sayılarıyla ve katmandaki düğüm sayılarıyla daha çok optimizasyon yapılabileceği için DSA seçilmiştir.

Bu aşamada ilaç çiftlerinin etkileşimi, etkileşim tipleri, etkileşimin ciddiyeti olmak üzere 3 farklı tahminde bulunulmuştur. Bunu mümkün kılmak için yapay sinir ağının son katmanı 7 farklı değer ile etiketlenmiştir. Her bir etiket değeri hem etkileşim olup olmadığını hem de etkileşimin tipini ve ciddiyetini verecek şekilde Tablo 5 üzerinde gösterildiği gibi tasarlanmıştır.

Tablo 5. DSA Çıktı Etiketlerinin Oluşturulması

Sınıf No	Farmakokinetik E.	Farmakodinamik E.	Ciddiyet	Etiket
0	0	0	0	0
1	1	0	1	101
2	1	0	2	102
3	1	0	3	103
4	0	1	1	11
5	0	1	2	12
6	0	1	3	13

Tablo 5’de görüldüğü üzere her bir etiket ilaçların birlikte kullanıldığında birbirleri arasında oluşabilecek bir etkileşim durumunu temsil etmektedir. Örneğin 0 numara iki araç arasında hiç bir etkileşim olmadığını, 3 numara ise ilaçlar arasında bir etkileşim olacağını, bu etkileşim türünün farmakokinetik olacağını ve ciddiyet derecesinin ise 3 olduğunu yani majör bir etkileşim olacağını ifade etmektedir. Bu durumda 0 numaralı etiket etkileşim olmadığını, 101 numaralı etiket farmakokinetik minör etkileşim olduğunu, 102 numaralı etiket farmakokinetik orta etkileşim olduğunu, 103 farmakokinetik majör etkileşim olduğunu, 11 numaralı etiket farmakodinamik minör etkileşim olduğunu, 12 numaralı etiket farmakodinamik orta etkileşim olduğunu, 103 farmakodinamik majör etkileşim olduğunu göstermektedir.

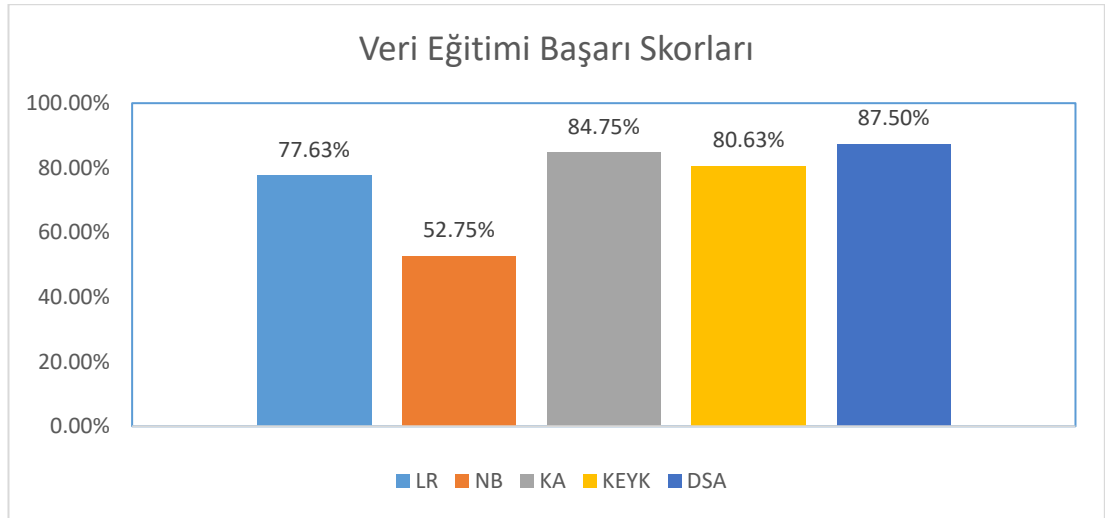
Derin sinir ağları verilen girdilerle eğitim yaptıktan sonra test verileriyle öğrenmenin performansı ölçülmektedir. Derin sinir ağlarıyla yapılan tahmin sonucunda her etiketin olma olasılığını veren bir çıktı alırız. Bütün etiketlerin olasılığını topladığımızda bu her zaman 1’e eşit çıkar. Yani bizim modelimizde bir tahmin yaptığımızda elimizde 7 etiket türü için toplamda 7 tane olasılık değeri olacaktır. Modelin performansı hesaplanırken ise en yüksek olasılığa sahip etiket çıktı olarak alınır. Daha sonra bu değer asıl olması gereken etiket değeri ile karşılaştırılarak performans skoru elde edilir.

BÖLÜM IV – SONUÇLAR

Bu bölümde uygulanan yöntemlerin sonuçları, sonuçlar sonrası yapılan düzenlemeler ve yorumlar yer almaktadır.

4. SONUÇLAR

İlk olarak yapılan makine öğrenmesi yöntemlerinin sadece İİE bulma konusundaki başarıları Şekil 12’de grafik ile gösterilmiştir. Bu grafikte görüldüğü üzere en başarılı yöntemin DSA olduğu görülmektedir. Bu sebeple asıl tahmin edeceğimiz ilaç etkileşim tipi ve ciddiyeti için eğiteceğimiz yöntem DSA olacaktır.



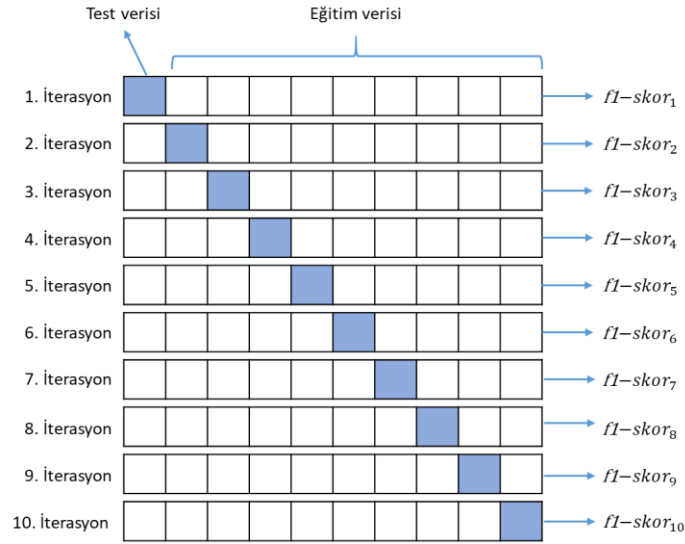
Şekil 12. Veri Eğitimi Başarı Skorları

Derin sinir ağlarıyla bu daha önceden belirlediğimiz 7 kategoriye test etmek için modelimiz 7 çıktı alacak şekilde düzenlendi. Daha sonra model çalıştırılarak performansı incelendi. Model performansını artırmak amacıyla modeldeki katman sayıları ve düğüm sayıları değiştirilerek yeni denemeler yapıldı. Aşırı öğrenme problemini çözmek için bazı düğümlerin etkinlikleri azaltıldı. Sonuç olarak en yüksek

performansa sahip, optimuma en yakın çözüm belirlenmiş oldu. Bu çözüm elimizdeki 8 veri setine uygulandı.

Performansı test etmek amacıyla 10 katlı çapraz doğrulama testi yapıldı. Çapraz doğrulama, elde ettiğimiz performansın rasgele mi elde ettik yoksa gerçekten bu performansı yakalayabiliyor muyuz anlamamıza yardımcı olmaya yarar.

Daha önceden yaptığımız doğrulama testinde verinin %90'ını eğitim için %10'unu test için kullanmıştık. Çapraz doğrulama işleminde yine aynı testi farklı eğitim ve test kümeleriyle 10 defa tekrarladık. Her defasında çıkan F1 skorlarının ortalamasını alarak modelin gerçek performansını hesapladık. Şekil 13'te eğitim ve test verilerinin ne şekilde seçildiği verilmiştir. Denklem 9 ise ortalama performansın nasıl hesaplandığını göstermektedir.



Şekil 13. 10 Kat Çapraz Doğrulama İçin Eğitim ve Test Verisi Seçimi

Denklem 9. 10 Kat Çapraz Doğrulama İçin Performans Hesaplaması

$$performans = \frac{1}{10} \sum_{i=1}^{10} F1_skor_i$$

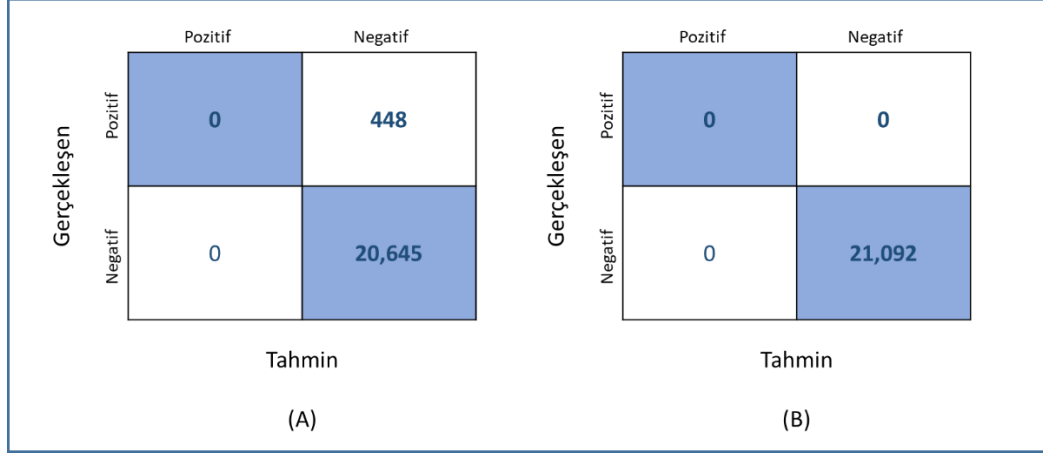
8 farklı veri seti için hazırladığımız DSA yapısı çalıştırıldığında ve 10 katlı çapraz doğrulama yapıldığında Tablo 6'da görülen sonuçlar elde edilmiş oldu.

Tablo 6. Derin Öğrenme Ağı Çapraz Doğrulama Sonuçları

İterasyon	Veri 1	Veri 2	Veri 3	Veri 4	Veri 5	Veri 6	Veri 7	Veri 8
1	0.93	0.92	0.92	0.89	0.89	0.90	0.87	0.92
2	0.92	0.90	0.91	0.88	0.87	0.88	0.85	0.90
3	0.92	0.91	0.92	0.89	0.89	0.89	0.87	0.91
4	0.78	0.73	0.74	0.66	0.66	0.67	0.60	0.74
5	0.72	0.64	0.65	0.56	0.55	0.59	0.48	0.66
6	0.72	0.64	0.66	0.54	0.50	0.53	0.46	0.65
7	0.72	0.64	0.66	0.60	0.56	0.59	0.56	0.65
8	0.98	0.98	0.98	0.98	0.97	0.97	0.96	0.98
9	0.98	0.98	0.98	0.97	0.97	0.96	0.96	0.97
10	0.94	0.93	0.93	0.92	0.91	0.91	0.89	0.93
Ortalama	0.861	0.827	0.835	0.789	0.777	0.789	0.75	0.831

Tabloda görüldüğü gibi kurulan derin öğrenme ağ modelinde en iyi sonucu birinci veri seti vermiştir ve 0.86 gibi bir F1 skoru üretmiştir. Çıkan sonuçları dikkatli olarak incelediğimizde 4, 5, 6 ve 7 numaralı iterasyonlarda genel olarak bir başarı düşüşü görülmektedir. Bunun nedeni bulmak için iterasyonlardaki dağılımlara daha yakından bakmamız gerekmektedir. Her bir iterasyon için test verisi dağılımlarını incelediğimizde Tablo 5’de görülen 3 numaralı ve 6 numaralı sınıfların dağılımlarınının bu iterasyonlarda test verisine daha çok dağıldığını görmekteyiz. Bu dağılımlarda dikkatimizi öncelikle çeken konu 3 ve 6 numaralı sınıfların ikisinin de etkileşim ciddiyetinin majör, yani çok önemli etkilere yol açan etkileşim olmalarıdır. Yani bu iterasyonlarda ciddiyet seviyesi majör olarak nitelendirilen ilaç ilişkileri bu derin öğrenme ağında başarılı bir şekilde yakalanamamıştır. Her ne kadar %86 lık bir başarı oranı başarılı sayılsa da, örneklem sayısı az olan fakat çok önemli olan bir özelliğin tahmin başarısının düşük olması derin öğrenme modelinin amacına çok hizmet edemediğini düşündürmektedir. Şekil 14’ te gösterilen karışıklık matrisi bu sonuca nasıl ulaştığımız hakkında bilgi vermektedir. Şekilde belirtilen şemalar üçüncü sınıfa dahil ilaçların, yani farmakokinetik etkileşime sahip majör etkileşimlerin farklı iterasyonlarda gösterdiği performansın değerleridir. F1- skoru değerleri A ve B için sırasıyla 0.72 ve 0.98 olarak çıkmıştır. (A) iterasyonunda 3. sınıfa ait 448 adet ilacın olduğunu fakat bunlardan hiç bir tanesinin dahi doğru olarak tahmin edilemediğini

görmekteyiz. (B) iterasyonunda ise test kümesinde bu sınıfa ait hiç ilaç bulunmamaktadır. Bu durum (B) iterasyonundaki başarı oranının yüksek çıkmasının sebebi olmuştur diyebiliriz.



Şekil 14. Farklı İterasyonlarda Majör Etkileşimler İçin Karışıklık Matrisleri

Derin öğrenme ağları, etkileşim ciddiyeti dışında kalan kısmı tahmin etmesi için tekrar eğitildi. Sadece etkileşim tiplerini içeren sınıflar Tablo 7’de görüldüğü gibi seçildi. Bu sınıflar ile iki ilaç arasında etkileşim olup olmadığını bulmak ve etkileşimin tipi tahmin etmek mümkün olmuştur.

Tablo 7. DSA Çıktı Etiketlerinin Oluşturulması - 2

Sınıf No	Farmakokinetik E.	Farmakodinamik E.	Etiket
0	0	0	0
1	1	0	10
2	0	1	1

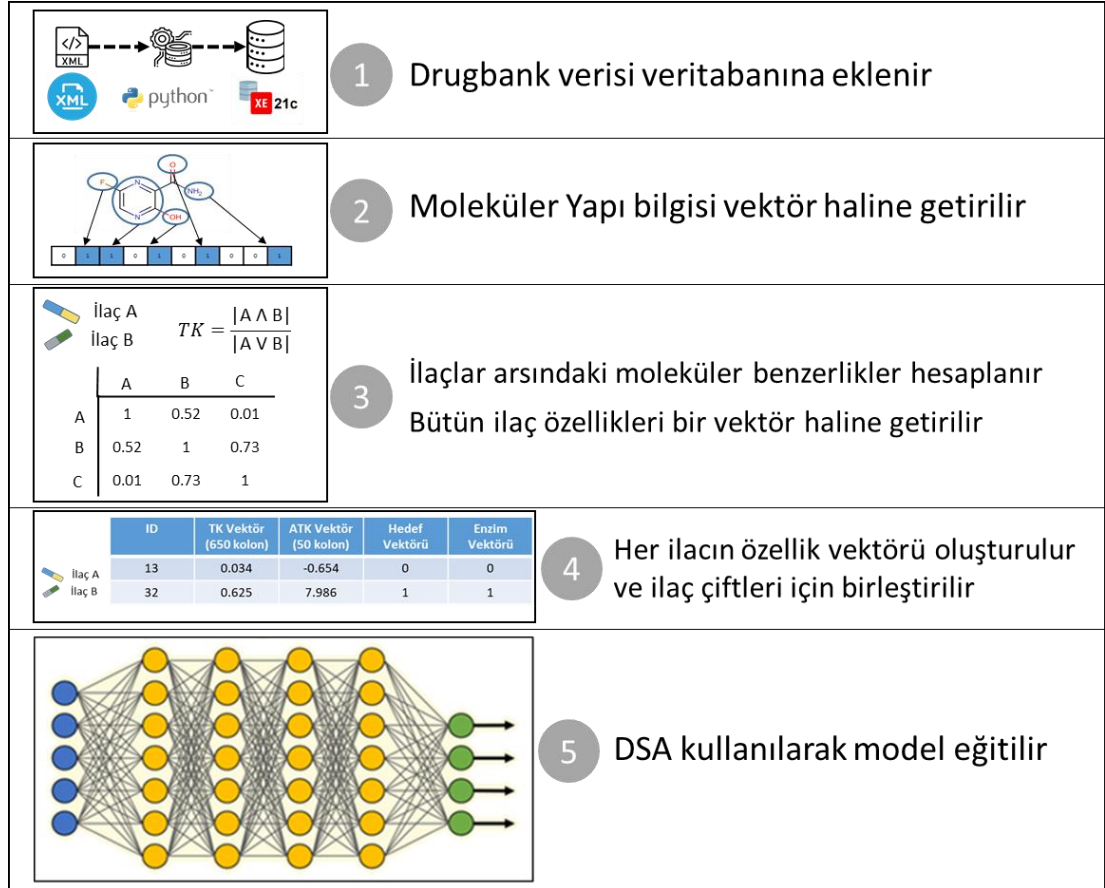
Yapılan eğitim ve arkasından yapılan testlerin sonucu Tablo 8’de gösterilmiştir. Bu modelde ortalama olarak her veri seti için %2 lik bir iyileşme sağlanmıştır.

Tablo 8. Derin Öğrenme Ağı Çapraz Doğrulama Sonuçları - 2

İterasyon	Veri 1	Veri 2	Veri 3	Veri 4	Veri 5	Veri 6	Veri 7	Veri 8
1	0.95	0.93	0.94	0.90	0.90	0.91	0.88	0.93
2	0.94	0.91	0.92	0.89	0.86	0.88	0.85	0.92
3	0.94	0.92	0.93	0.90	0.89	0.90	0.85	0.93
4	0.80	0.76	0.78	0.70	0.69	0.70	0.64	0.76
5	0.78	0.70	0.73	0.63	0.61	0.65	0.61	0.72
6	0.76	0.69	0.74	0.60	0.58	0.60	0.54	0.71
7	0.73	0.67	0.69	0.60	0.59	0.62	0.59	0.70
8	0.98	0.98	0.98	0.96	0.97	0.94	0.95	0.98
9	0.98	0.98	0.98	0.97	0.96	0.95	0.95	0.97
10	0.95	0.94	0.94	0.91	0.91	0.90	0.89	0.93
Ortalama	0.881	0.848	0.863	0.806	0.796	0.805	0.775	0.855

İlaçların ATK kodu, moleküler yapısı, hedef proteinleri ve enzim bilgilerini kullanarak kullanarak yaptığımız denemelerde en verimli yöntemi birinci veri seti üzerinde derin öğrenme sinir ağıları kullanarak ilaçların farmakokinetik ve farmakodinamik etkileşimlerini tahmin ederken tespit ettik. Bu süreci bir modele dönüştürdüğümüzde Şekil 15 üzerinde görüldüğü gibi bir akış elde ederiz. İlk adımda verileri Drugbank’tan XML formatında indiririz. Burdaki verileri yapısal hale getirmek için Python dilinden yararlanırız. Yapısal hale dönüştürülen verileri Oracle XE veritabanına yükleriz. İkinci adımda ilaçların molekül yapı bilgilerini 0 ve 1’lerden oluşan bir vektör haline getiririz. Bunu yaparken python da Rdkit kütüphanesinden yararlanırız. Rdkit kimya informatiği için geliştirilmiş içinde çeşitli araçları barındıran bir kütüphanedir. Üçüncü adımda her bir ilacın diğer ilaçlar ile olan kimyasal yapı benzerliklerini Tanimoto katsayısı kullanarak hesaplarız. Her ilaç için 650 kolonluk bir vektör elde ederiz. Bu adımda ATK kodu, hedef ve enzim bilgisi için de 0 ve 1’lerden oluşan birer vektör elde ederiz. ATK kodunu içeren vektörü TBA kullanılarak 50 kolona düşürülür. Dördüncü adımda her bir ilaç için vektörler birleştirilerek ilaca ait özellik vektörleri oluşturulur. İlaç çiftleri için bu özellik vektörleri ve ilaçların etkileşim bilgileri eklenerek DSA için girdi tablomuz ortaya çıkmış olur. Beşinci adımda girdiler kullanılarak oluşturulan DSA yapısı eğitilir. Bu proje için DSA yapısı python proglama dilindeki Keras kütüphanesi ile oluşturulmuştur. Keras python ile yazılmış

olan ve neredeyse her türlü derin öğrenme modelini oluşturabilen bir kütüphanedir. Keras DSA modelinin girdiler ile eğitilmesi sonucunda tahmin modelimiz tamamlanır. Tahmin modeli sonucunda iki çeşit çıktımız olabilmektedir. Birincisi ilacın etkileşim bilgisi ve etkileşim tipi (farmakokinetik veya farmakodinamik), ikincisi ise üç sonuç kategorisi (etkileşim yok, farmakokinetik etkileşim, farmakodinamik etkileşim) için gerçekleşme yüzdeleridir.



Şekil 15. İlaç-İlaç Etkileşimi Tahmin Modeli

Keras ile oluşturduğumuz DSA yapısı 1 giriş katmanı, 1 çıkış katmanı ve 4 ara katman olmak üzere toplam 6 katmanlı bir yapıya sahiptir. Giriş katmanı 1280 adet girdi bilgisini alarak orta katmanlara iletir. Orta katmanlar sırası ile 128, 64, 64, 64 adet düğümlere sahiptir. Son olarak çıktı katmanı ise 3 adet çıkış düğümüne sahiptir ve her düğüm bir olasılık sayısını çıktı olarak verir.

İlk olarak yalnızca ilaç etkileşimi tahmini yapan modelleri test ederek en iyi modelin DSA olduğunu olduğunu ve sonrasında en iyi sonuç veren veri setini tespit ettik.

BÖLÜM V – TARTIŞMA ve SONUÇ

5. TARTIŞMA ve SONUÇ

İİE önceden tespit edilemediğinde ve hasta kişiye etkileşen ilaç çiftleri verildiğinde önemli sonuçlara neden olabilmektedir. Tablo 9’da bazı ciddi etkileşen ilaç çiftleri ve bunların ne tür etkileşime sebep olduğu bilgisi bulunmaktadır.

Tablo 9. Bazı Ciddi İlaç-İlaç Etkileşimleri (Büyükkorkuoğlu ve diğerleri, 2019)

Etkilenen İlaç	Etkilenen İlaç	Potansiyel Etki
Varfarin	Asetil salisilik asit	Kanamada artış, INR’ de artış
Karbamazepin	Rifampin	Karbamazepin seviyelerinde azalma
Fenitoin	Simetidin, Eritromisin, Klaritromisin, Flukonazol	Fenitoin seviyelerinde artış
Fenobarbital	Simetidin Eritromisin, Klaritromisin, Flukonazol	Fenobarbital seviyelerinde artış
SSRI	Trisiklik antidepresanlar	Trisiklik antidepresan etkisinde artış
SSRI	Tramadol	Nöbet oluşum riskini artırır, Serotonin sendromu

Yaptığımız çalışma sırasında kullandığımız Drugbank veri tabanında bulunmayan fakat çalışmamızın tespit ettiği bazı ilaç ilaç etkileşimleri ise Tablo 10’da verilmiştir.

Bu bilgiler başka bir çevrimiçi etkileşim sorgulama sayfası olan webmd.com/interaction-checker/default.htm adresinden elde edilmiştir. Bu durum bize Drugbank veri tabanının bütün etkileşimleri içermediğini göstermektedir.

Tablo 10. Modelimizde Tespit Ettiğimiz Bazı İlaç-İlaç Etkileşimleri (WEB 7)

Etkilenen İlaç	Etkilenen İlaç	Potansiyel Etki
Phenelzine	Armodafinil	Yüksek tansiyon riskinde artış (akut hipertansif dönem)
Tranylcypromine	Armodafinil	Yüksek tansiyon riskinde artış (akut hipertansif dönem)
Primidone	Selinexor	Belirtilmemiş etkileşim mekanizması
Primidone	Tinidazole	Tinidazol'un seviyesinde veya etkisinde azalış
Primidone	Dopamine	Primidone artar ve Dopamine sedasyonu azaltır

Bu tezde ilaç-ilaç etkileşimlerini tahmin etmek amacıyla makine öğrenmesi yöntemlerini uygulayarak bir tahmin modeli oluşturduk. En uygun sonuca ulaşabilmek için lojistik regresyon, naive bayes, K-en yakın komşuluk, karar ağaçları ve derin öğrenme yöntemlerini verilerimiz üzerinde test ettik. En verimli sonucun derin öğrenme ağları kullanıldığında alındığını görerek modelimizi bu yönde ilerlettik.

Bu çalışmamızda ilk defa olarak ilaçların etkileşimlerinin tipini yani farmakokinetik etkileşim mi yoksa farmakodinamik etkileşim mi var olduğunu ve etkileşimin ciddiyet derecesini tahmin eden bir model üzerinde çalıştık. Daha önce etkileşimin ciddiyeti üzerine herhangi bir çalışma yapıldığını görmemiştik. Bu da bizi motive eden unsurlardan bir tanesi olur. Elimizdeki verilerin ve uyguladığımız yöntemlerin ilişkilerin ciddiyeti konusunda tahmin yapma performansının çok iyi olmadığını gördük. Bunun veri kümesindeki yüksek ciddiyetli örneklerin az olmasından kaynaklanma olasılığı vardır. Girdi olarak verdiğimiz verilerin bu tür örnekler için bir korelasyon oluşturmuyor olma ihtimali de bulunmaktadır.

Modellerin eğitilmesi sonucu en verimli çalışan modelimizde giren örneklem kümesine göre %73 ile %98 arasında değişen ve ortalaması %88 olan başarı oranını elde ettik. Daha önceden farmakokinetik ve farmakodinamik etkileşim tiplerini tahmin eden bir model ile karşılaşmadık. Benzer girdi verileri ile ilaç etkileşimi tahmini yapan Cheng ve Zhao 2014 yılında ilaçların benzerlik profillerini hesaplayarak DVM kullanarak %67 oranında bir başarı elde etmişlerdir. Ryu ve arkadaşları 2018 yılında önerdikleri metodda yalnızca ilaçların benzerlik profillerini kullanarak DSA ile yapılan eğitim 86 farklı etkileşim türü için %92.4 lük bir başarı göstermiştir. Bu modeli bir basamak daha geliştiren Lee ve arkadaşları 2019 yılında ileri beslemeli DSA modeli ile ilaçların hedefleri ve gen ontolojisi verilerinin yapısal benzerlik profilini çıkararak modele dahil etmişler ve bu başarıyı %97 oranlarına taşımışlardır.

Benzerlik profillerinin yapılan tahminlerde yüksek başarı sağladığı literatürdeki çalışmalardan anlaşılmaktadır. Bir sonraki adım olarak çeşitli benzerlik profilleri üretilerek yeni modeller denenebilir.

Ayrıca ilaç etkileşimlerindeki ciddiyet derecelerini tespit edebilmek adına hangi özelliklerin bu ciddiyeti tetiklediği konusunda araştırmalar yapılarak bunu veri setlerimize eklememiz faydalı olabilecektir.

Biz her etkileşim tipini ve etkileşmenin olmadığı durumu etiketleyerek tek seferde tahmin yapan bir çalışma gerçekleştirdik. İlk önce etkileşim olup olmadığını bulan ve daha sonra etkileşimin çeşidini tahmin eden bir model başarısının artmasına sebep olabilecektir.

Oluşturduğumuz ATK sınıflandırma vektörlerinde sınıflandırma seviyeleri arasındaki ilişkilerin makine öğrenmesi algortimalarına önceden tanıtılması işlemi de başarı skorunu artırabileceği göz önünde bulundurularak daha sonraki çalışmalarda dikkat edilecek bir husus olarak değerlendirilebilir.

Vektör boyutlarını azaltmak için TBA yöntemini kullandık. Bazı parametrik olmayan özellik azaltma yöntemlerinin denenmesi daha yüksek başarı elde edilmesine neden olabilir. Bu durum da ileriki bir çalışma konusu olarak değerlendirilebilir.

Modelimizde her bir ilaç çiftinin hangi sonuç sınıfına ait olduğunu bulurken etiketlerden en yüksek olasılık değeri olanı seçtik. Bu yöntem yerine belirli bir olasılık

değerinin üzerindeki kayıtların sınıfa dahil edilmesi hususunda yapılacak bir çalışma modelin başarı derecesini yükseltebilecektir.

Modelde kullandığımız veri setleri bütün ilaçları kapsamak yerine sadece belirli bilgilere sahip olan ilaçları içermektedir. Bilgi eksikliği olan bir ilaç için tahminde bulunma durumu bizim modelimiz için geçerli olamamaktadır. Bir sonraki aşamada eksik bilgili ilaçların tahmini için de bir yöntem bulmak iyi bir araştırma konusu olacaktır.

KAYNAKÇA

- Ali, J., Khan, R., Ahmad, N., Maqsood, I. (2012). Random Forests and Decision Trees, *IJCSI International Journal of Computer Science Issues International Journal of Computer Science Issues*, 9(3).
- Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. *Molecular Systems Biology*, 12(7), 878.
- Ayodele, T.O. (2010). Types of Machine Learning Algorithms, *New Adv. Mach. Learn*, 3, 19–48.
- Belkin, M., Niyogi, P., & Sindhvani, V. (2006). Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. *J. Mach. Learn. Res.*, 7, 2399-2434.
- Belkin, M., Matveeva, I., & Niyogi, P. (2004). Regularization and Semi-supervised Learning on Large Graphs. *COLT*, 17.
- Beretta, L., & Santaniello, A. (2016). Nearest neighbor imputation algorithms: a critical evaluation. *BMC Medical Informatics and Decision Making*, 16(3), 197-208.
- Büyükkokuroğlu, M.E., Tanyeri, P., & Keleş, R. (2019). İlaç-İlaç Etkileşimleri Konusunda Farkındalık. *Online Türk Sağlık Bilimleri Dergisi.*, 4(3), 377-391.
- Chapelle, O., Weston, J., & Schölkopf, B. (2002). Cluster Kernels for Semi-Supervised Learning. *NIPS*, 15.
- Cheng, F., & Zhao, Z. (2014). Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *Journal of the American Medical Informatics Association : JAMIA*, 21(e2), e278–e286. <https://doi.org/10.1136/amiajnl-2013-002512>
- Çalışkan, A. (2017). Derin sinir ağının sınıflandırma başarımının artırılması için yeni bir eğitim stratejisi geliştirilmesi ve biyomedikal veri setleri üzerinde test edilmesi, Yayınlanmamış doktora tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü.
- Deng, H., Fannon, D.J., & Eckelman, M.J. (2018). Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy and Buildings*, 163, 34-43.

- Elder, J. (t.y). "Introduction to Machine Learning and Pattern Recognition," http://www.eecs.yorku.ca/course_archive/2011-12/F/4404-5327/lectures/01%20Introduction.pdf adresinden erişilmiştir.
- Friedman, J.H. (2006). Recent Advances in Predictive (Machine) Learning. *Journal of Classification*, 23, 175-197.
- Friedman, N., Geiger, D., & Goldszmidt, M. (2004). Bayesian Network Classifiers. *Machine Learning*, 29, 131-163.
- Gallelli, L., Colosimo, M., Tolotta, G.A., Falcone, D.M., Luberto, L., Curto, L.S., Rende, P., Mazzei, F., Marigliano, N.M., De Sarro, G., & Cucchiara, S. (2009). Prospective randomized double-blind trial of racecadotril compared with loperamide in elderly people with gastroenteritis living in nursing homes. *European Journal of Clinical Pharmacology*, 66, 137-144.
- Gallelli, L., Ferreri, G., Colosimo, M., Pirritano, D., Guadagnino, L., Pelaia, G., Maselli, R., & De Sarro, G. (2002). Adverse drug reactions to antibiotics observed in two pulmonology divisions of catanzaro, Italy: a six-year retrospective study. *Pharmacological research*, 46 5, 395-400.
- Gottlieb, A., Stein, G. Y., Oron, Y., Ruppin, E., & Sharan, R. (2012). INDI: a computational framework for inferring drug interactions and their associated recommendations. *Molecular systems biology*, 8, 592. <https://doi.org/10.1038/msb.2012.26>
- Göksel, Sibel. (t.y). "İlaç Etkileşimlerinin Farmakokinetik ve Farmakodinamik Temelleri" http://tfd.org.tr/sites/default/files/Klasor/Dosyalar/kongreler/2005/tfd2005_029_ulkergoksel.pdf [Erişim 27.12.2021]
- Gündüz, M., & Soyer, M. (2019). Küresel Rekabetçilik ile Ekonomik Özgürlük Arasındaki İlişkinin Binary Lojistik Regresyon Analizi ile İncelenmesi. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*.
- Huang, J., Niu, C., Green, C. D., Yang, L., Mei, H., & Han, J. D. (2013). Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. *PLoS computational biology*, 9(3), e1002998. <https://doi.org/10.1371/journal.pcbi.1002998>

- Hutzler, M.J., Messing, D.M., & Wienkers, L.C. (2005). Predicting drug-drug interactions in drug discovery: where are we now and where are we going? *Current opinion in drug discovery & development*, 8 1, 51-8.
- Jonker, D.M., Visser, S.A., van der Graaf, P.H., Voskuyl, R.A., & Danhof, M. (2005). Towards a mechanism-based analysis of pharmacodynamic drug-drug interactions in vivo. *Pharmacology & therapeutics*, 106 1, 1-18.
- Juurink, D.N., Mamdani, M.M., Kopp, A., Laupacis, A., & Redelmeier, D.A. (2003). Drug-drug interactions among elderly patients hospitalized for drug toxicity. *JAMA*, 289 13, 1652-8.
- Kim, T., Ko, W., & Kim, J. (2019). Analysis and Impact Evaluation of Missing Data Imputation in Day-ahead PV Generation Forecasting. *Applied Sciences*, 9(1), 204.
- Küçükçalık, Ahmet. (2019). “CYP-450 Enzimi Sınıflarını İnhibe Eden İlaçların, Diğer İlaçların Metabolizmine Etkisi,”
https://web.itu.edu.tr/kucukcalik/enzim_CYP-450.pdf [Erişim 07.01.2022]
- Leape, L.L., Bates, D.W., Cullen, D.J., Cooper, J., Demonaco, H.J., Gallivan, T., Hallisey, R., Ives, J., Laird, N., Laffel, G. (1995). Systems analysis of adverse drug events. *JAMA*, 274(1), 35-43.
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
<https://doi.org/10.1038/nature14539>
- Lee, C.E., Zembower, T.R., Fotis, M.A., Postelnick, M.J., Greenberger, P.A., Peterson, L.R., & Noskin, G. (2000). The incidence of antimicrobial allergies in hospitalized patients: implications regarding prescribing patterns and emerging bacterial resistance. *Archives of internal medicine*, 160 18, 2819-22.
- Lee, G., Park, C., & Ahn, J. (2019). Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC bioinformatics*, 20(1), 415.
<https://doi.org/10.1186/s12859-019-3013-0>
- Li, Q., Meng, Q., Cai, J., Yoshino, H., & Mochida, A. (2009). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, 86(10), 2249-2256.
- Lin, S. F., Xiao, K. T., Huang, Y. T., Chiu, C. C., & Soo, V. W. (2010). Analysis of adverse drug reactions using drug and drug target interactions and graph-based

- methods. *Artificial intelligence in medicine*, 48(2-3), 161–166.
<https://doi.org/10.1016/j.artmed.2009.11.002>
- Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138.
- Park, K., Kim, D., Ha, S., & Lee, D. (2015). Predicting Pharmacodynamic Drug-Drug Interactions through Signaling Propagation Interference on Protein-Protein Interaction Networks. *PloS one*, 10(10), e0140816.
<https://doi.org/10.1371/journal.pone.0140816>
- Qi, M., Luo, H., Wei, P., & Fu, Z. (2019). Estimation of low calorific value of blended coals based on support vector regression and sensitivity analysis in coal-fired power plants. *Fuel*, 236, 1400-1407.
- Rochon, P.A., & Gurwitz, J.H. (1997). Optimising drug treatment for elderly people: the prescribing cascade. *BMJ*, 315, 1096 - 1099.
- Ryu, J. Y., Kim, H. U., & Lee, S. Y. (2018). Deep learning improves prediction of drug-drug and drug-food interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 115(18), E4304–E4311.
<https://doi.org/10.1073/pnas.1803294115>
- Samuel, A.L. (1959). Some Studies in Machine Learning Using the Game of Checkers.
- Sanjar, K., Olimov, B., Kim, J., Paul, A., & Kim, J. (2020). Missing Data Imputation for Geolocation-based Price Prediction Using KNN-MCF Method. *ISPRS Int. J. Geo Inf.*, 9, 227.
- Snyder, B.D., Polasek, T.M., & Doogue, M.P. (2012). Drug interactions: principles and practice. *Australian Prescriber*, 35, 85-88.
- Solomonoff, R.J. (1957). An Inductive Inference Machine.
- Tatonetti, N. P., Fernald, G. H., & Altman, R. B. (2012). A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *Journal of the American Medical Informatics Association : JAMIA*, 19(1), 79–85. <https://doi.org/10.1136/amiajnl-2011-000214>
- Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

- Vilar, S., Uriarte, E., Santana, L., Lorberbaum, T., Hripcsak, G., Friedman, C., & Tatonetti, N. P. (2014). Similarity-based modeling in large-scale prediction of drug-drug interactions. *Nature protocols*, 9(9), 2147–2163. <https://doi.org/10.1038/nprot.2014.151>
- Vilar, S., Uriarte, E., Santana, L., Tatonetti, N. P., & Friedman, C. (2013). Detection of drug-drug interactions by modeling interaction profile fingerprints. *PloS one*, 8(3), e58321. <https://doi.org/10.1371/journal.pone.0058321>
- Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., & Friedman, C. (2012). Drug-drug interaction through molecular structure similarity analysis. *Journal of the American Medical Informatics Association : JAMIA*, 19(6), 1066–1074. <https://doi.org/10.1136/amiajnl-2012-000935>
- Wang, F., & Zhang, C. (2008). Label Propagation through Linear Neighborhoods. *IEEE Trans. Knowl. Data Eng.*, 20, 55-67.
- WEB 1, <https://go.drugbank.com/about> [Eriřim 01.12.2022].
- WEB 2, <http://howmed.net/pharmacology/drug-interactions/> [Eriřim 29.11.2021].
- WEB 3, <https://datascientistsolution.com/data-science/support-vector-machine-with-python/> [Eriřim 01.12.2021]
- WEB 4, <https://www.explorium.ai/blog/the-complete-guide-to-decision-trees/> [Eriřim 25.12.2021]
- WEB 5, <https://towardsdatascience.com/machine-learning-fundamentals-ii-neural-networks-f1e7b2cb3eef> [Eriřim 05.01.2022]
- WEB 6, https://tr.wikipedia.org/wiki/Anatomical_Therapeutic_Chemical_Classification_System [Eriřim 31.12.2021]
- WEB 7, webmd.com/interaction-checker/default.htm [Eriřim 12.01.2022]
- Yu, P.S., Chen, S., & Chang, I. (2006). Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, 328, 704-716.
- Zhang, P., Wang, F., Hu, J., & Sorrentino, R. (2015). Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects. *Scientific reports*, 5, 12339. <https://doi.org/10.1038/srep12339>
- Zhong, H., Wang, J., Jia, H., Mu, Y., & Lv, S. (2019). Vector field-based support vector regression for building energy consumption prediction. *Applied Energy*, 242, 403-414.

- Zhou, D., Bousquet, O., Lal, T.N., Weston, J., & Schölkopf, B. (2003). Learning with Local and Global Consistency. *NIPS, 16*.
- Zhu, X., & Ghahramani, Z. (2002). Learning from labeled and unlabeled data with label propagation, Technical report, CMU CALD tech report.