



BİR YAPAY ZEKÂ
YOLCULUĞU

HABER DEDEKTİFİ **ROBOKAN**

Tasarım ve Grafik Uygulama
TERMİNAL MEDYA LTD. ŞTİ.
Maslak Mah. Bilim Sokak No:5 SUN Plaza
Kat:13 Sarıyer/İSTANBUL
0 (212) 367 4988 ve 0 (532) 643 6959

Editör
ÖZLEM ÖZKAN

Grafik Uygulama
GÜLİSTAN ŞENOL

Baskı
SET POZİTİF MATBAA
Maslak Mahallesi Ahi Evran Caddesi
Rentaş İş Merkezi A Blok No: 62 Sarıyer/İSTANBUL
0 (212) 286 4933

Yayın Tarihi: Nisan 2020

SORUMSUZLUK BEYANI

İşbu raporda, son dönemin önemli teknolojilerinden yapay zekâ ve makine öğrenmesi hakkında BKM ve Crede tarafından hayata geçirilen kavram ispatı çalışmasına ait sonuçların özetlenmesi, alınan derslerin paylaşılması ve proje hakkında teknik bilgi verilmesi amaçlanmıştır. Rapor içeriği sadece bilgilendirme amaçlıdır, kişi ve kurumları bağlayıcı tavsiye veya görüş niteliği taşımaz. İşbu raporun içerdiği söz konusu bilgilerin güncel ve eksiksiz olduğu taahhüt edilmemektedir. Bu bağlamda işbu raporun içeriğini okuyan kişilere veya herhangi bir üçüncü kişiye karşı sorumluluğu ve yükümlülüğü bulunmamaktadır.



İÇİNDEKİLER

Sunuş	4
1. Yönetici Özeti	8
2. Robokan'ın Bileşenleri	12
2.1 Crawler	14
2.2 Öğrenme Kümesi	16
2.3 Öznitelik Adresleme	20
2.4 Algoritmalar	22
3. Performans	23
4. Sonuç	28



Dr. Soner Canko
BKM
Genel Müdürü

İçerisinde bulunduğumuz dijital dönüşüm sürecinde en önemli rollerden birini, aslında çok yeni bir teknoloji olmayan ancak kullanım alanları son yıllarda artan yapay zeka oynuyor. Yapay zekayı benimsediğimiz günlük yaşantımızda artık sesli asistanlarla konuşuyor, akıllı hoparlörlerle ev otomasyonu ve online alışveriş süreçlerimizi gerçekleştirebiliyoruz. Kurumların da biyometrik doğrulama, kredi riski değerlendirme, işlem güvenliği skorlama, müşteri hizmetleri gibi alanlarda yapay zekadan faydalandığını görüyoruz. Hızla gelişen teknolojinin sunduğu engin imkanlar, kurumların bu dönüşüme ayak uydurup uyduramayacağı sorusunu akıllara getiriyor. Hayal edebilmenin en önemli varlıklardan biri olduğu bir çağda yaşıyoruz. Bu çağda BKM olarak biz deneyerek öğrenmeye devam ediyoruz.

Yapay zekâ, derin öğrenme ve robotik otomasyon teknolojileri; görev tanımlarının, çalışma şekillerinin ve başarıya ulaşmak için gerekli olan yetkinliklerin değişmesine yol açıyor. Bunun bilincinde bir kurum olarak bugün uzman bir insanın yaptığı işlerin yapay zeka tarafından yapıp yapılamayacağını sınıadığımız Robokan'ı geliştirdik. Süreç boyunca yaptığımız denemeleri ve ulaştığımız sonuçları ise bu raporda siz yapay zeka meraklıları için derledik. Çalışmanın hayata geçmesinde emeği olan tüm arkadaşlarıma teşekkür eder, faydalı olmasını dilerim.

SUNUŞ



Celal Cündoğlu
BKM
Genel Müdür
Yardımcısı

Dijital dönüşümün en önemli unsurlarından biri olan yapay zeka teknolojisinin kullanım alanı birçok sektörde genişlerken finans dünyası, yapay zekanın potansiyelinden en iyi faydalanan sektörlerden biri olarak konumlanmayı başardı. Güvenlik çözümlerinden robo-danışmanlara, risk yönetiminden müşteri hizmetlerine kadar birçok alanda bugün yapay zekadan faydalanılıyor.

Ülkemizde yeni teknolojilerle ilgili bilincin oluşması için öncü çalışmalar yapan bir kurum olarak biz de bugüne kadar çeşitli alanlarda yapay zeka ile ilgili deneyimler yapıyoruz. Güvenlik çözümleri, chatbot, işlem adetleri tahmini, duygu analizi, bugüne kadar yapay zekayı denediğimiz konu başlıkları oldu. Yapay zekanın biraz daha derinliklerine inebilmek için kısa süre önce yeni bir çalışmaya daha imza attık. BKM'nin finans dünyasını bilinçlendirmek ve güncel konuları takip etmesine yardımcı olmak amacıyla gerçekleştirdiği haftalık içerik tarama, uygun haber seçimi ve rapor üretimi sürecini, yapay zeka tabanlı bir çözümle otomatikleştiren yeni çalışmamız "Haber Dedektifi: Robokan"ı Crede ekibi ile birlikte geliştirdik.

Okumakta olduğunuz raporda, Robokan projemizde izlediğimiz yolu, öğrendiklerimizi ve projenin çıktılarını paylaşmaktan mutluluk duyuyoruz. Dijital dönüşümün önemli bir bileşeni olarak gördüğümüz yapay zeka teknolojisinin sunduğu fırsatları ve öğrendiklerimizi ürün ve hizmetlerimize yansıtmaya devam edeceğiz. Türkiye'yi geleceğe taşımak için tüm paydaşlarımızı gelişen teknolojilerle yeni kullanım alanları bulmaya, denemeye ve paylaşmaya davet eder, raporumuzun yapay zeka ekosistemine küçük de olsa bir katkı sağlamasını dilerim.



Ahmet Kocamaz
Crede
Kurucu

Yapay zekâ teknolojileri, Türkiye’de üzerine çok tartışılan fakat yeni kullanım alanları bulmakta zorlanılan bir alan olarak karşımıza çıkıyor. BKM yöneticileri, bir yayın editörünün yaklaşımını yapay zekâ algoritmalarına anlatmak istediklerini söylediklerinde çok heyecanlandık ve hemen çözüm önerimizi sunduk.

Crede olarak 9 yıldır makine öğrenmesinin farklı endüstriyel uygulamaları üzerinde çalışıyoruz. Sadece 2 ay süren bu projede de, hem farklı bir endüstriyel probleme çözüm üretmiş, hem de daha derin projeler için bir adım atmış olduğumuza inanıyoruz. Oluşan analitik model artık her hafta çalışıyor ve editör kararları veriyor. Biz de sonuçlarını büyük bir zevkle takip ediyoruz.

Proje boyunca öğrenme verisi hazırlama, test ve uygulamaya alma aşamalarında bizlere çok hızlı ve özverili destek veren BKM İş Geliştirme Ekibi’ne; modelleme ve yazılım geliştirme adımlarında görev yapan Crede mühendislerine çok teşekkür ederim.

Projede makine öğrenmesinin bilinen yöntemlerini uygulamanın yanı sıra, bir çok yeni yaklaşım da kullandık ve bunları bu dokümana yansıtmaya gayret ettik. Umarım okuyucular da bu bilgilerden faydalanabilir ve proje sırasında hissettiğimiz heyecanı onlar da hisseder.

SUNUŞ



Prof. Dr. Cem Say
Boğaziçi Üniversitesi

Yapay zekâ, son yılların gözde konusu. Mesleğe başladığım yıllarda bilimkurgu değil de bilim olduğuna en yakınlarımızı bile ikna etmekte zorlandığımız bu alan, son on beş yılda yaşanan zincirleme ilerlemeler sayesinde potansiyelini ispatladı, kimi uygulamaları şimdiden insan performansını geride bıraktı. Sırada yaygınlaşması, hayatın her yönüne genel bir “akıllılık” artışı olarak yansımaları var. Bu yeni teknolojinin nasıl hayata geçirilebileceği ayrı ayrı her iş kolunun gündeminde. Yapay zekâyı kendi iş süreçlerimize nasıl entegre edebilir, ondan nasıl yarar sağlayabiliriz? Bu soruyu sormak önemli.

Finans sektöründe böyle yenilikçi yaklaşımların başını çeken BKM’yle bir süredir yapay zekâ konularında sıkı bir işbirliği içindeyiz. Çok şık bir kavram ispatı olduğunu düşündüğüm “Robokan” projesi için de fikir alışverişlerinde bulunduk. (Hatta kurulacak sistem sözü geçen işi çok iyi yapan Okan Yıldız arkadaşımızın becerilerini esas alacağı için adına “Robo-Okan” denilmesini ben önermiş olabilirim!) Sonuçta ortaya yerleşik yapay zekâ teknikleriyle mevcut bir sürecin nasıl yarı-otomatikleştirilebileceğine gayet güzel bir örnek çıktı. BKM’nin her övgüyü hak eden değerli ve çalışkan kadrosunu ve özelde proje ekibini kutlar, başarılarının devamını dilerim.

1. YÖNETİCİ ÖZETİ

Yapay zekâ teknolojisi tüm hızıyla sektörleri dönüştürmeye devam ederken finans sektöründe de iş zekası, risk yönetimi, skorlama, sahtekarlıkla mücadele, robotik süreç otomasyonu ve ileri analiz teknikleri gibi birçok kullanım alanıyla kendine yer ediniyor. Hem kurumlar hem de kullanıcılar için finansal süreçleri daha kolay ve verimli hale getirme yolunda atılan adımlar kuşkusuz, yapay zekâ teknolojisini geleceğin finans dünyasının merkezine koyuyor. Bankalararası Kart Merkezi (BKM) ve Crede ortaklığında hayata geçirilen Robokan uygulaması ile BKM İş Geliştirme Ekibi'nin haftalık olarak hazırladığı "Bankacılık ve Ödeme Teknolojileri Gelişmeleri" raporuna konu olan içerikler artık yapay

zekâ desteğiyle seçiliyor. 20'den fazla haber kaynağı üzerinde yapılan haftalık taramalar sonucunda Türkiye bankacılık ve finans ekosisteminin inovasyonlar ve fintech dünyasından son haberler hakkında güncel bilgiye sahip olmasını sağlayan içerikler Türkçe özetlenerek eposta, Medium, Twitter, LinkedIn gibi sosyal mecralarda paylaşıyor.

Robokan'nın çalışma prensibini üç adıma bölebiliriz:

1. Crawling: İnternet sayfaları üzerindeki içerikleri indirmek ve kategorize etmek için kullanılan ve crawler olarak adlandırılan web uygulaması, kaynak olarak adresle-

nen web sitelerine ziyaretler gerçekleştirerek ilgili hafta içerisinde yayına alınmış tüm haberleri tarar ve indeksler.

2. Öznitelik Adresleme: Crawler tarafından elde edilen içerikler, öznitelik adresleme (feature hashing) ile algoritmaları besleyecek şekilde sayısal olarak ifade edilir.

3. Haber Seçimi: Bugüne kadar üretilmiş içeriklerden oluşturulan öğrenme kümesi, algoritmanın vereceği kararlar için bir referans değeri taşımaktadır. Uygulamanın geliştirilmesinde her bir haber kaynağına özel crawler yazılımları geliştirilmiştir. Öznitelik adresleme işleminin tamamlanması-

nın ardından, haftalık içeriğimize eklemeyi uygun gördüğümüz haberlerin seçilmesi için altı farklı algoritma test edilmiştir.

Yapılan denemelerde kullanılan algoritmaların parametrelerinde yapılan optimizasyon çalışmaları sonucunda elde edilen çıktılarda %75'lik bir doğruluk payı ile Binary Locally Deep Support Vector Machine algoritması tercih edilmiştir.

BKM İş Geliştirme Ekibi, Robokan'ı kendi iş süreçlerini destekleyici bir yapay zeka aracı olarak kullanmaya başladı. Geçmişe bakarak doğru haber önerileri yapan Robokan'ın gelişimini sürdürebilmek ve geleceğe hazırlamak için elimizdeki modeli nasıl eğitmemiz gerektiğini ise bundan sonra çözmemiz gereken bir konu olarak belirledik.



BKM VE YAPAY ZEKÂ

İnsanlık için doğal bir yetenek olan “zekâ”, du-yularımızın ve yeteneklerimizin çevremizde olan bitenlerle baş edebilmek için beraber hareket et-mesi, ilişki kurabilme ve akıl yürütebilme yete-neğidir. İçinde yaşadığımız dönemde kimsenin dilinden düşürmediği ve hepimizi heyecanlandı-ran yeni bir zekâ türü var: “yapay zekâ.”

Peki, nedir bu yapay zekâ?

Yapay zekâ, karşılaştığı durumlar karşısında zeki ya da akıllı olarak adlandırılabilir bir şekilde davranabilme ve topladığı veri üzerine eyleme geçerek çevresiyle etkileşime geçebilme kapasite-sine sahip makineler ve bilgisayar programlarını ifade eden bir kavramdır. Bir başka deyişle, Prof. Dr. Cem Say'ın 50 Soruda Yapay Zekâ kitabında da tarif ettiği gibi doğal sistemlerin (zekice olsun veya olmasın) yapabildiği her bilişsel etkinliği (gerekirse bedenleri olan) yapay sistemlere, daha da yüksek başarımlı düzeylerinde nasıl yaptırabi-leceğimizi inceleyen bilim dalıdır.

BKM olarak yapay zekâ, nesnelerin interneti, artırılmış gerçeklik ve blokzincir gibi yeni teknolo-jilerin herkes tarafından anlaşılması ve Türkiye gündeminin üst sıralarında kalması için projeler hayata geçiriyoruz.

Birçok kurumsal şirket ve girişimin odaklandığı bir kavram olan yapay zekâyı, ödeme sistemle-rinde hizmet veren bir şirket olarak biz de ya-kından takip ediyor ve bu teknolojiyi kullanarak yeni çözümler üretmek için kapsamlı çalışmalar yapıyoruz.

Gerçekleştirmeye çalıştığımız projeler ile bildiği-miz tek zekâ türü olan insan zekâsını taklit eden makineler üretmeye çalışıyoruz. Hayal gücünün sınırının olmaması gerektiğine inandığımız bu konularda, çok önemli bir mücadele başlamış du-rumda. Bu mücadele, bilgisayarların ve insanla-rın, insan hayatını zenginleştirmek üzere nasıl bir araya geleceklerini ve ortaya çıkacak iş birliğinin hayatımızda neleri değiştireceğini belirleyecek.

Yapay zekânın finans ile ilişkisi

Nesnelerin interneti, blokzincir, büyük veri ve yapay zekâ gibi düşünce tarzını değiştiren tekno-lojilerin birçok sektörde olduğu gibi finans sektö-rünü de dönüştürmesi bekleniyor. Çok yönlü bir teknoloji olan yapay zekâ ve makine öğrenmesi beraberinde getirdiği iş zekası, chatbot, robotik süreç otomasyonu, skorlama ve ileri analiz tek-nikleri gibi kullanım senaryolarıyla kendisini fi-

Şekil 1: BKM'nin Yapay Zekadan Faydalandığı Alanlar



nans dünyasında da ispat etmeye çalışıyor. Yaptığımız araştırmalarda büyük veri ve kullanıcı deneyimi üzerine çalışmaların hızla devam ettiğini ve öğrenen makinelerin henüz insanların yerini alamadığını görüyoruz. Bununla birlikte günümüzde makine öğrenmesinden faydalanan çalışmaların sayısında hızlı bir artışın söz konusu olduğunu söylemek yanlış olmaz.

Yapay zekâ ile makinelerin insan davranışlarını ve zekasını taklit ederek çok daha fazla bilgi işlemesi mümkün. Yapılan araştırmalarda, yapay zekanın sağladığı imkanlarla bankacılık sektörünün 447 milyar dolar tasarruf edeceği öngörülmüyor. Bankalarda yapay zekâ uygulamalarının tasarruf sağlayacağı akışlar arasında ise ön ofis ve operasyonel süreçler olarak öne çıkıyor.

2. ROBOKAN'IN BİLEŞENLERİ

Robokan nedir?

BKM İş Geliştirme Ekibi olarak 2016 tarihinden bu yana her hafta bankacılık ve ödeme teknolojileri dünyasında yaşanan gelişmeleri Türkçe olarak BKM üyesi kurumlar ve ekosistemimizle paylaşıyoruz. Türkiye finans ekosisteminin faydalanabileceği bu içerikleri üretirken sayıları her gün artan web kaynaklarımızda yayınlanan

haberleri hafta boyunca tarıyor ve dikkatimizi çeken gelişmeleri Türkçe olarak özetliyoruz. İçeriklerimizi, e-posta, Medium, Twitter, LinkedIn gibi sosyal mecralarda paylaşarak ekosistemimizin ödeme sistemlerindeki inovasyon, düzenlemeler ve fintech haberleri hakkında güncel bilgiye ulaşması sağlıyoruz. Hazırladığımız içerikler kurum içerisinde ise iletişim sistemi Confluence'un yanı



sıra tuvaletlerde bulunan dijital ekranlarda da çalışma arkadaşlarımızın karşısına çıkıyor.

İçeriğimizi değerli kılan en temel faktörleri ise bilgi deryası web kaynakları içerisinden özenle yapılan tercihler ve Türkçe özetleme kabiliyetimiz olarak gösterebiliriz. Yayınladığımız her bir içerik için ortalama 2 iş günü efor sarf ettiğimizi, düşünürsek ortaya çıkan ürün için ne denli büyük bir emek harcadığımızı anlayabilirsiniz.

BKM olarak blokzincir, nesnelerin interneti, ma-

kine öğrenmesi ve yapay zekâ gibi günümüzde kritik öneme sahip teknolojilerle ilgili kavram ispatı çalışmaları yapmaya hep heves duyduk. İşte bu motivasyon sayesinde Haber Dedektifi Robokan isimli kavram ispatı projesi doğmuş oldu. Hem yapay zekâ teknolojisi keşfimizi genişletmek hem de teknolojinin iş süreçlerimize ne kadar yardımcı olacağını ölçümlemek adına nokta atışı bir kullanım senaryosu yakaladığımızıza inanıyoruz.



Haftanın Önemli Gelişmeleri

28 Şubat 2020

B-Social
WeChat Pay-Network International
Yodlee-FinBit.io
Pix-Anlık Ödeme-Brezilya
Visa Delegated Authentication-Garmin Pay
Intuit-Credit Karma
BharatPe Mastercard-Viber
Amazon-Go Grocery
Youtility-Barclays
Monzo-RBS-Kredi Skoru
Tesco

Bu dosya, BKM tarafından hazırlanmıştır.

B K M
BANKALARARASI
KART MERKEZİ

Projemize verdiğimiz Robokan ismini öneren Prof. Dr. Cem Say'a teşekkür ederiz.

Şekil 2: Yöntemler

Haftalık ürettiğimiz içerikler için takip ettiğimiz iş adımları aşağıdaki gibi oluyor.



Takip ettiğimiz iş adımlarını yukarıdaki gibi parçalara bölmek Robokan'ın bize hangi adımlarda destek olabileceğini saptamamızda büyük kolaylık sağladı. Web sitelerindeki içeriklere erişmek ve bu verileri işleyebilmek için “Crawler” adı verilen içerik indeksleme yöntemiyle tanıştık.

2.1. Crawler nedir?

Web örümceği, web botu, web böceği olarak da adlandırılan web crawlerı, internet sayfaları üzerindeki içerikleri indirmek ve kategorize etmek için kullanılan bir web uygulamasıdır. Crawler, hedef gösterilen web sayfaları içerisindeki bilgileri otomatik olarak tarayıp indeksleyerek istenilen bilgiye daha kolay erişme imkânı sunar. Google, Yandex, Bing gibi popüler arama motorları, crawlerlar tarafından toplanan verilere bir arama algoritması uygulayarak kullanıcılarının sorgularına karşılık gelen sonuçları listeler.

Bir web crawlerını dağınık bir kütüphanedeki tüm kitapları düzene sokmak için her kitabın başlığını ve içeriğini okuyarak tüm kitapları buna

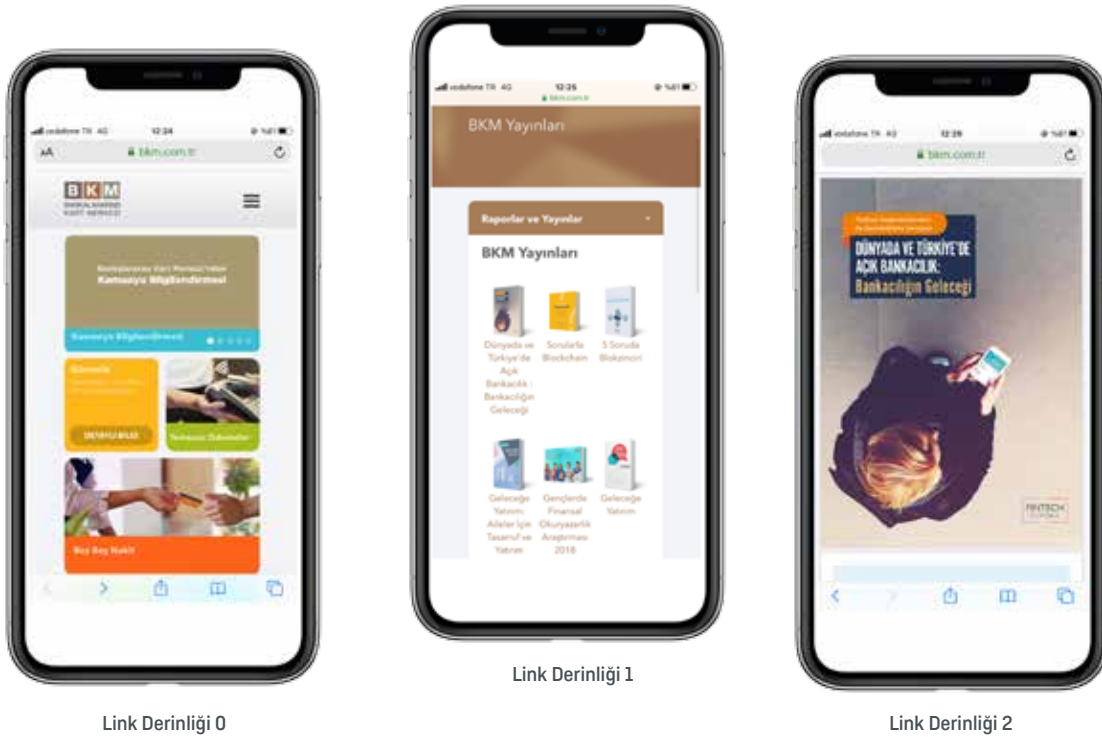
göre yeniden düzenleyen ve kütüphaneye gelen ziyaretçilerin istedikleri bilgiye kolayca erişmelerini sağlamak için katalog halinde sunan bir kişiye benzetebiliriz.

İş sürecimizin ilk adımı olan haber sitelerini tarama görevi bir web crawlerı tarafından zahmetsizce gerçekleştirilebilir. Web crawlerımızı bugüne kadar ürettiğimiz içeriklerin beslendiği tüm web sitelerine yönlendirerek güncel gelişmelere erişebilmesi için ilgili haber sayfalarının tarihi, başlığı ve içeriğine erişecek şekilde geliştirdik. Burada kritik olan nokta ise crawlerımızın istediğimiz gibi çalışması için hedef gösterdiğimiz web sitelerinin tasarımları arasındaki farklılıklarını tespit etmek ve ayırtmak oldu. Web siteleri dönemsel olarak gerek estetik gerek güvenlik sebebiyle tasarım de-

ğişikliklerine giderler. Erişmeye çalıştığımız web sitelerindeki bu farklılıklar, crawlerımıza her bir web sitesine özel geliştirme yapmamızı gerektirdi.

Crawlerımızın bize getireceği içerikleri sınırlandırmak için bazı filtreleme işlemleri yaptık. Bunlardan ilki tarih filtresiydi. Crawl edilen içeriklerin tarih bilgileri üzerinden en güncel haberlerin listelenmesini sağladık. Buna ek olarak crawlerın bizim için daha önemli olan kaynaklardan daha çok içerik bulabilmesi için link (bağlantı) derinliği kavramını hesaba kattık. Link derinliği, bir web sitesinin ana sayfası

Şekil 3: Link Derinliği Örneği



ile erişmeye çalıştığınız sayfanın arasındaki adım veya tıklama sayısı olarak tanımlanabilir.

Daha fazla faydalandığımız kaynakların derinliğini artırarak o sayfalardan daha çok içerik elde etmeyi hedefledik. İlk denemelerimizde her bir kaynak için link derinliğini 2 olarak varsaydık. Bunun sonucunda crawlerın bizlere 5.634 adet haber getirdiğini gördük. Yaptığımız kalibrasyonlara örnek vermek gerekirse bankacılık ve ödeme teknoloji odaklı web sitelerinden birinin derinliğini 2'den 3'e çıkarmak crawl edilen haber sayısında çok fazla bir değişikliğe sebep olmadı çünkü bu kaynak halihazırda spesifik bir konu üzerine içerik üreten bir web sitesiydi. Ancak aynı değişikliği, içerisinde bir çok kategori bu-

lunan başka bir kaynak üzerinde yaptığımızda 2'den 3'e çıkarılan link derinliği bize fazladan 4.876 tane haber getirdi.

Yaptığımız optimizasyon sonunda kaynaklarımızın link derinliğini 2 ile 4 arasında olacak şekilde belirledik. Crawler her haftanın içeriği için tekrar çalıştırılacağından bu rakamları isteğimize göre değiştirerek odaklanmak istediğimiz kaynaktan daha fazla içerik elde edebiliyoruz.

Artık crawlerımız hedeflediğimiz sitelerde istediğimiz tarih aralığı ve derinlikte çalışabilir hale geldi. Crawl edilen güncel içerikler içerisinden yapay zekâ algoritmalarının faydalanacağı bir öğrenme kümesini oluşturmak için çalışmalara başladık.

2.2. Öğrenme kümesi

Öğrenme kümesi oluşturmaktaki amacımız, algoritmanın bizim gibi düşünmesini sağlamaktır. Haftalık dosyamıza eklediğimiz her içeriği, değerlendirmeye almadığımız haberleri ise 0 olarak düşünürsek, elimizde bugüne kadar seçtiğimiz tüm içerikler, yani sadece 1'ler bulunuyordu. Amacımız güncel içeriklerden 0 ve 1'lerin bulunduğu bir öğrenme kümesi oluşturarak yapay zekânın bize en yakın kararları almasını sağlamak oldu. İlk crawl işleminin sonucunda 5.634 içerikten oluşan bir liste elde ettik. Her bir haberi tek tek değerlendirerek yanlarına 1'ler ve 0'lar koyduk. Tablo 1'de de görüleceği gibi, burada dikkatimizi dikkatimizi çeken detay, bazı kaynakların haber doğruluk oranının diğerlerine göre çok daha yüksek olduğunu gördük.

Tablo 1: Öğrenme Kümesi Oluşturma Adımı - 1

Kaynaklar	Crawl Edilen Haber Adedi	Hedef (1) Olarak İşaretlenen Haber Adedi	İşaretlenen Tüm Haberler İçerisindeki Payı
Kaynak 1	9	2	1%
Kaynak 2	300	2	1%
Kaynak 3	49	1	0%
Kaynak 4	56	3	1%
Kaynak 5	62	29	12%
Kaynak 6	50	2	1%
Kaynak 7	80	1	0%
Kaynak 8	94	1	0%
Kaynak 9	1.196	10	4%
Kaynak 10	43	3	1%
Kaynak 11	169	81	33%
Kaynak 12	407	3	1%
Kaynak 13	107	0	0%
Kaynak 14	72	3	1%
Kaynak 15	245	2	1%
Kaynak 16	125	49	20%
Kaynak 17	31	13	5%
Kaynak 18	88	0	0%
Kaynak 19	814	13	5%
Kaynak 20	31	5	2%
Kaynak 21	20	7	3%
Kaynak 22	101	6	2%
Kaynak 23	1.202	1	0%
Kaynak 24	283	5	2%
	5.634	242	100%

Elde edeceğimiz öğrenme kümesinde haber kaynaklarının ağırlığını daha dağınık bir hale getirebilmek için bir hafta sonra Tablo-2'deki gibi sonuçlanan bir crawl işlemi daha crawl işlemi daha gerçekleştirdik. Bu sefer crawler'ın getirdiği 6.140 içerikten oluşan taramayı da puanlayarak yapay zekâ algoritmasının referans alacağı öğrenme kümemize ekledik.

Tablo 2: Öğrenme Kümesi Oluşturma Adımı - 2

Kaynaklar	Crawl Edilen Haber Adedi	Hedef (1) Olarak İşaretlenen Haber Adedi	İşaretlenen Tüm Haberler İçerisindeki Payı
Kaynak 1	10	5	0%
Kaynak 2	302	14	1%
Kaynak 3	50	6	0%
Kaynak 4	56	34	2%
Kaynak 5	94	47	3%
Kaynak 6	50	27	2%
Kaynak 7	80	35	2%
Kaynak 8	94	21	1%
Kaynak 9	1.205	54	3%
Kaynak 10	44	6	0%
Kaynak 11	337	257	15%
Kaynak 12	409	149	9%
Kaynak 13	107	41	2%
Kaynak 14	72	4	0%
Kaynak 15	246	136	8%
Kaynak 16	367	320	19%
Kaynak 17	54	50	3%
Kaynak 18	88	41	2%
Kaynak 19	824	282	17%
Kaynak 20	34	10	1%
Kaynak 21	24	12	1%
Kaynak 22	104	59	4%
Kaynak 23	1.203	22	1%
Kaynak 24	286	49	3%
	6.140	1.681	100%

Sonuç olarak toplam öğrenme kümemiz aşağıdaki gibi oluştu.

- 5 aylık içeriklerimizden toplam 301 adet link
- Bunların 242'si belirlediğimiz 22 kaynaktan
- Manuel olarak işaretlenen linklerden 1.439 adet link

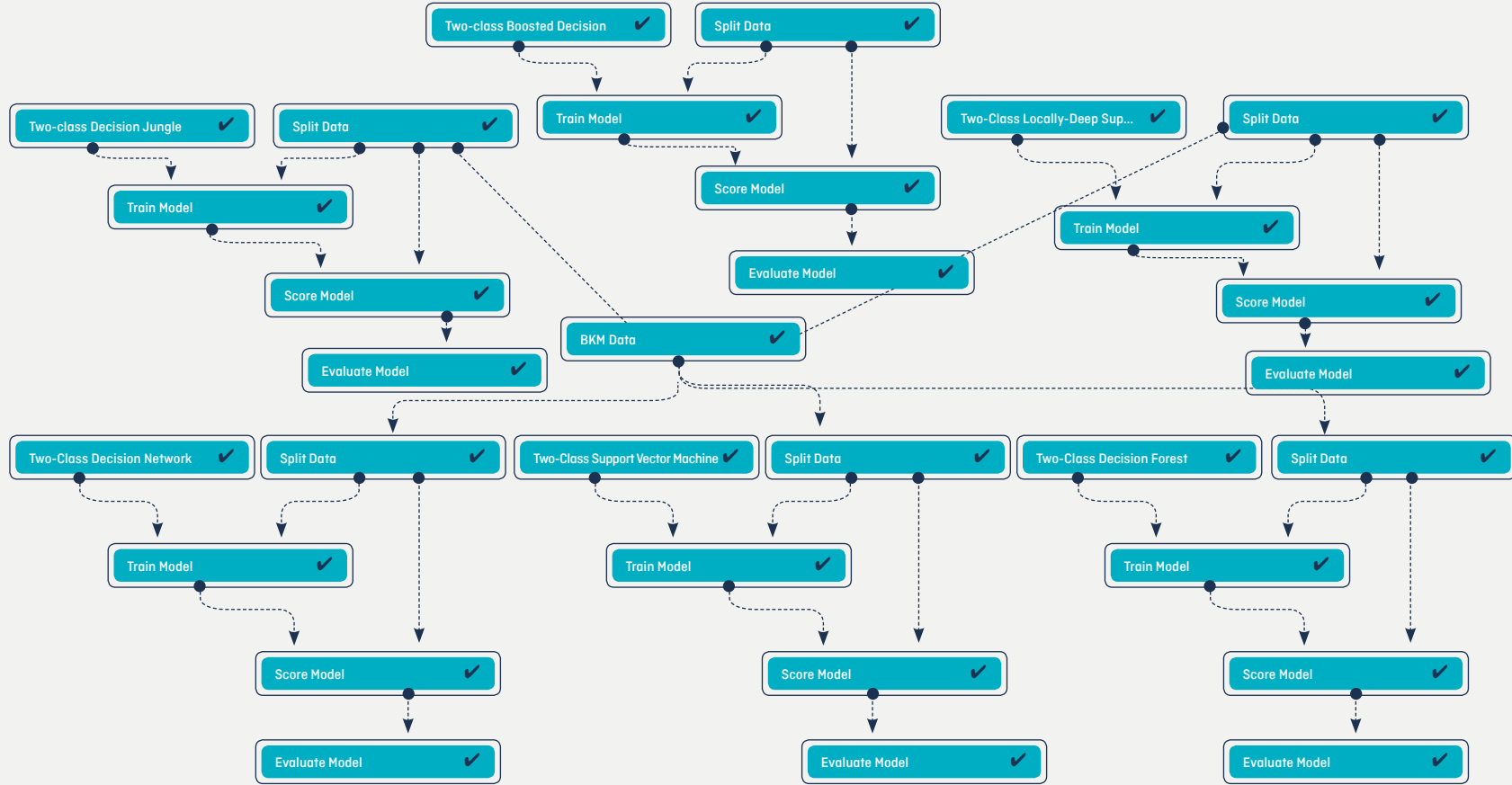
Şekil 4: Öğrenme Kümesinin Detaylı Görünümü



Azure ML

Microsoft'un bulut platformu Azure'un makine öğrenmesi ve yapay zekâ özelinde çalışan Azure ML (Machine Learning), makine öğrenmesi modellerini hızlıca oluşturmak, eğitmek ve çalıştırmak için geliştirilmiştir. Robokan uygulamamızda tercih ettiğimiz Azure ML sayesinde içeriklerimizin sayısal özeti öznetelik adresleme (feature hashing) düğümüyle elde ederek makine öğrenimi modellerini hızlıca oluşturduk ve denemelerimizi gerçekleştirdik. Herhangi bir kodlama gerektirmeyen arayüzü sayesinde modellerimizi Şekil 5'deki gibi zahmetsizce kurduk ve bizim için asıl önemli olan parametre tercihlerine odaklandık.

Şekil 5: Robokan Azure ML Geliştirme Şablonu



2.3. Öznitelik Adresleme (Feature Hashing) nedir?

Yahoo! Research tarafından geliştirilen Vowpal Wabbit, çevrimiçi öğrenme, özetleme, arama öğrenmesi (learning2search), aktif ve etkileşimli öğrenme gibi tekniklerle makine öğreniminin sınırlarını zorlayan bir açık kaynak makine öğrenmesi kütüphanesidir. Azure ML içerisinde “Metin Analizi” araçları arasında bulunan Feature Hashing İngilizce metinleri tam sayı dizileriyle ifade ederek öğrenme algoritmalarını beslemeye hazır hale getirir.

Feature Hashing’i Türkçe olarak “öznitelik adresleme” şeklinde ifade edebiliriz. Metinleri vektör-uzay modeline taşımak için her bir kelimeyi sayısallaştırmak gerekir. Bu nedenle her bir kelime veya kelime grubunu bir rakama denk düşecek şekilde vektör oluşturularak, modeller beslenir.

Öznitelik adresleme, benzersiz belirteçleri tam sayılara dönüştürerek, girdi aldığı metinleri herhangi bir dil analizi yapmadan sayılar halinde ifade etmeye yarar. Öznitelik adreslemenin en çok kullanıldığı alanlardan biri olan duygu analizi örneğini inceleyelim.

Kullanıcı Metni	Duygusalılık Puanı
I loved this book	3
I hated this book	1
This book was great	3
I love books	2

Öznitelik adresleme işlemi sonucunda verilen metin içerisinde bulunan kelime ve kelimelerden n-gram’lık bir sözlük oluşturulur. N-gram biricik olarak kabul edilen ve karşılığında tam sayı üretilen n kelimeden oluşan bir dizidir. Aşağıdaki örnekte, n-gram boyutu 1 (unigram) seçildiğinde her bir kelimenin karşılığı bir sayı ile ifade edilecektir.

İfade (unigrams)	Kullanım Sıklığı
book	3
I	3
books	1
was	1

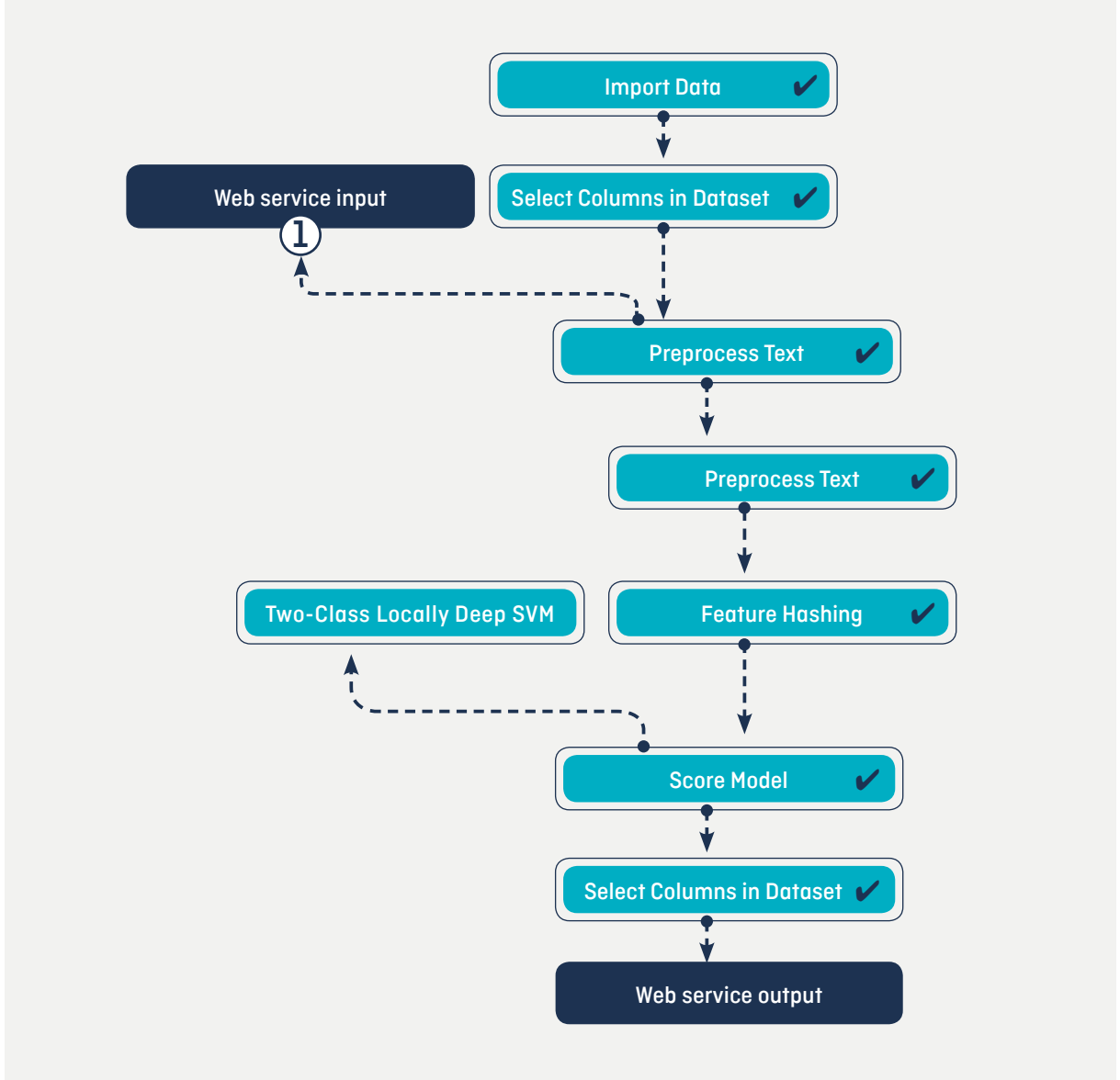
N-gram boyutunun 2(bigram) seçildiği durumlarda ise arka arkaya gelen her iki kelime için bir sayı karşılığı üretilecektir

İfadeler (bigrams)	Kullanım Sıklığı
This book	3
I loved	1
I hated	1
I love	1

Sözlük oluşturulduktan sonra, öznitelik adresleme modülü sözlük terimlerini tam sayı değerlerine dönüştürür ve ilgili kelime/kelime grubunun kullanılıp kullanılmadığını hesaplar. Oluşturulan tabloda metin girdileri satırlarda, hashlenen kelime/kelime grupları ise sütunlarda bulunur. Modül kelimelerin bulunmadığı durumlar için 0, bulunduğu durumlarda ise 1 olarak tabloyu doldurur.

Rating	Hashing feature 1	Hashing feature 2	Hashing feature 3
4	1	1	0
5	0	0	0

Şekil 6: Robokan Azure ML Uygulama Şablonu



Öznitelik adreslemeyi diğer modüllerden ayıran özellik ise içeriğinizin uzunlukları farklı olsa da, n-gram sayesinde boyutları sınırlandırılmış vektörler yani sayısal gösterimler elde edebilirsiniz. Çıktıların sayısal olması, sınıflandırma, kümeleme veya bilgiyi geri okuma gibi birçok

farklı makine öğrenme yönteminin kullanılmasını mümkün kılar. Arama işlemleri, metin karşılaştırmaları yerine tam sayı karmalarını kullanabildiğinden, özellik ağırlıklarının elde edilmesi de çok daha hızlıdır.

2.4. Algoritmalar

Proje’de binary sınıflandırma algoritmalarını kullandık. Bu alanda temel olarak kullanılan algoritmalar:

- Karar ağaçları
- Destek Vektör Makinesi
- Lojistik Regresyon
- Sinir Ağları’dır.

Çalışma sırasında lojistik regresyon kullanılmamıştır. Çünkü girdiler rakamsal değil, kategorik değişkenlerden oluşmaktadır. (Başlığın ve metnin içerisindeki kelimeler kategorik değişkenlerden oluşmaktadır.) Diğer algoritmalarında farklı varyasyonları bulunabilmektedir.

Proje kapsamında

- 3 farklı Karar ağacı
 - Binary Boosted Decision Tree
 - Binary Boosted Decision Forest
 - Binary Boosted Decision Jungle
- 2 farklı Destek Vektör Makinesi
 - Supported Vector Machine
 - Deep Supported Vector Machine
- 1 Nöral Sinir Ağı

algoritması iterasyonlarda kullanılmıştır.

Başlangıçta araç olarak KNIME uygulaması kullanılmış fakat buradaki her bir iterasyonun yaklaşık 48 saat sürdüğü ve 70 iterasyon hedeflendiği için bu şartlar altında projenin vaktinde bitmeyeceği görülmüştür. Çözüm olarak MS Azure’dan işlem süresi alınmış ve Azure ML ortamında yapılan iterasyonlarda her bir iterasyonun 5-7 dakika sürdüğü görülmüştür. Sonuç olarak, projeye Azure ML ile devam edilmesine karar verilmiştir.

Neler denedik?

Binary Boosted Decision Tree:

Artırılmış karar ağacı, ikinci ağacın ilk ağacın hatalarını düzelttiği, üçüncü ağacın birinci ve ikinci ağaçların hatalarını düzelttiği bir topluluk öğrenme yöntemidir. Tahminler, tahmini yapan tüm ağaç topluluğuna dayanır.

Binary Decision Forest:

Karar ormanları hızlı, denetimli topluluk modelleridir. Maksimum iki sonucu olan bir hedefi tahmin etmek için bu algoritma iyi bir seçimdir.

Binary Decision Jungle:

Decision Jungle, Decision Forest’in yeni bir uzantısıdır. Bir karar ormanı, karar yönelimli döngüsüz çizgilerden (DAG’ler) oluşur.

Karar ormanları aşağıdaki avantajlara sahiptir:

- Ağaç dallarının birleşmesine izin vererek, bir karar DAG tipik olarak, biraz daha uzun eğitim süresi pahasına da olsa, bir karar ağacından daha düşük bellek alanı ve daha iyi genelleme performansına sahiptir.
- Karar ormanları, doğrusal olmayan karar sınırlarını temsil edebilen parametrik olmayan modellerdir.
- Entegre özellik seçimi ve sınıflandırması yaparlar ve gürültülü özelliklerin varlığında dayanıklıdır.

Binary Locally Deep Support Vector Machine:

Destek vektör makineleri (SVM’ler), doğrusal ve doğrusal olmayan sınıflandırma görevlerinde kullanılabilen, son derece popüler ve iyi araştırılmış bir denetimli öğrenme modelleri sınıfıdır. Son araştırmalar, bu modelleri daha büyük eğitim setlerine verimli bir şekilde ölçeklendirmek için optimize etmenin yollarına odaklanmıştır. Microsoft Research’ün bu uygulamasında, veri noktalarını özellik alanına eşlemek için kullanılan çekirdek işlevi, sınıflandırma doğruluğu-

nun çoğunu korurken eğitim için gereken süreyi azaltmak için özel olarak tasarlanmıştır.

Binary Neural Network:

Bir sinir ağı birbirine bağlı katmanlar kümesidir. Girişler birinci katmandır ve ağırlıklı kenarlar ve düğümlerden oluşan döngüsüz çizge ile bir çıkış katmanına bağlanır.

Giriş ve çıkış katmanları arasına birden çok gizli katman eklenebilir. Öngörücü görevlerin çoğu yalnızca bir veya birkaç gizli katmanla kolayca gerçekleştirilebilir.

Sinir ağları kullanarak sınıflandırma denetimli bir öğrenme yöntemidir ve bu nedenle bir etiket sütunu içeren etiketli bir veri kümesi gerektirir

Binary Support Vector Machine:

Destek vektör makineleri (SVM'ler) iyi araştırılmış bir gözetimli öğrenme yöntemleri sınıfıdır. Bu özel uygulama, sürekli veya kategorik değişkenlere dayalı olarak iki olası sonucun tahmin edilmesine uygundur.

3. PERFORMANS

Performansı nasıl ölçtük?

Modelin ölçümü yapılırken Confusion Martix'ten (Karışıklık Matrisi) faydalanılmış, F-measure değerini maksimize etmek hedeflenmiştir. Doğruluk 0-1 dengesinin olmadığı öğrenme kümelerinde yanlış fikir verebileceği için ölçümlenmiş fakat dikkate alınmamıştır. Testler yapılırken yanlış pozitif değerleri azaltmak için Kesinlik değeri de göz önünde bulundurulmuştur.

Confusion Matrix

Accuracy (Doğruluk): $A+D/(A+B+C+D)$

Recall (Duyarlılık): $A/(A+C)$

Precision (Kesinlik): $A/(A+B)$

F1 Skoru: $2 \times \text{Recall} \times \text{Precision} / (\text{Recall} + \text{Precision})$

Tahminlenen	Gerçek	
	POSITIVE	NEGATIVE
	POSITIVE	NEGATIVE
POSITIVE	A	C
NEGATIVE	B	D

Öğrenme kümesi kitlenin %80'i, test kümesi %20'si olarak ayrılmıştır. Ayrıca işletim testi sırasında kesinlik tekrar ölçülmüş ve 0,83 olarak test kümesine paralel çıktığı gözlenmiştir. (Test kümesinin kesinlik değeri 0,73 olarak hesaplanmıştır.)

1.turda temel parametreler ile ulaşılan F1 değeri 0,70'in altında olan Binary Decision Forest, Binary Decision Jungle, Binary Neural Network algoritmaları elenmiştir

2. tur ileri iterasyonlara Binary Boosted Decision Tree, Binary Locally Deep Support Vector Machine, Binary Support Vector Machine algoritmaları ile ilerlenmeye karar verildi. Binary Boosted Decision Tree, Binary Support Vector Machine algoritmalarının daha fazla iyileşememesi nedeniyle Deep SVM algoritması seçilmiştir.

Tüm algoritmalar için toplam 70 iterasyon yapılmıştır.

İterasyon sonuçları ve kullanılan parametreler

Tablo 3: Binary Boosted Decision Tree İterasyon Sonuçları

İterasyon#	n-gram	Ağaç başına düşen maksimum yaprak sayısı	Yaprak düğümü başına minimum örnek sayısı	Öğrenme Oranı	Ağaç Sayısı	Cut-off değeri	F1 Skoru
1	1	20	10	0,15	100	0.5	0.725
2	1	20	8	0,15	100	0.5	0.726
3	1	20	5	0,15	100	0.5	0.710
4	1	20	12	0,15	100	0.5	0.707
5	1	50	10	0,15	100	0.5	0.706
6	1	50	8	0,15	100	0.5	0.709
7	1	50	5	0,15	100	0.5	0.716
8	1	50	12	0,15	100	0.5	0.711
9	2	20	10	0,15	100	0.5	0.692
10	2	20	8	0,15	100	0.5	0.696
11	2	20	5	0,15	100	0.5	0.667
12	2	20	12	0,15	100	0.5	0.688
13	2	50	10	0,15	100	0.5	0.663
14	2	50	8	0,15	100	0.5	0.686
15	2	50	5	0,15	100	0.5	0.690
16	2	50	12	0,15	100	0.5	0.661
17	1	20	8	0,1	100	0.5	0.702
18	1	20	8	0,15	200	0.5	0.735
19	1	20	8	0,15	500	0.5	0.733
20	1	20	9	0,15	200	0.5	0.733

Tablo 4: Binary Decision Forest İterasyon Sonuçları

İterasyon #	Yeniden örnekleme yöntemi	Karar ağacı sayısı	Karar ağaçlarının maksimum derinliği	Düğüm başına düşen rastgele bölünme sayısı	Yaprak düğümü başına düşen minimum örnek sayısı	F1 Skoru
21	Bagging	8	32	128	10	0.000
22	Bagging	80	32	128	10	0.000
23	Bagging	80	32	128	8	0.000
24	Replicate	8	32	128	10	0.413
25	Replicate	8	32	128	8	0.412
26	Replicate	80	32	128	10	0.399
27	Replicate	80	32	128	8	0.427
28	Replicate	80	32	512	8	0.580
29	Replicate	80	32	2048	8	0.639
30	Replicate	80	32	8192	8	0.679
31	Replicate	80	32	32768	8	0.666

Tablo 5: Binary Decision Jungle İterasyon Sonuçları

İterasyon #	Yeniden örnekleme yöntemi	Karar DAG'larının sayısı	Karar DAG'ların maksimum derinliği	Karar DAG'ların maksimum genişliği	Karar DAG'ı katmanı başına düşen optimizasyon adımı sayısı	F1 Skoru
32	Bagging	8	32	128	2048	0.311
33	Bagging	8	100	128	2048	0.385
34	Bagging	80	100	128	2048	0.325
35	Bagging	8	100	128	8192	0.522
36	Bagging	8	100	128	32768	0.592
37	Bagging	8	100	128	65536	0.641
38	Bagging	8	100	128	131072	0.640
39	Replicate	8	32	128	2048	0.299
40	Replicate	8	100	128	2048	0.343
41	Replicate	80	100	128	2048	0.348
42	Replicate	80	100	128	8192	0.527
43	Replicate	80	100	128	32768	0.603
44	Replicate	80	100	128	65536	0.628
45	Replicate	80	100	128	131072	0.640

Tablo 6: Binary Locally Deep Support Vector Machine İterasyon Sonuçları

İterasyon #	Ağaç derinliği	Lambda W	Lambda Theta	Lambda Theta Prime	Sigmoid keskinliği	İterasyon sayısı	Özel Normalleştirme	F1 Skoru
46	3	0.1	0.01	0.01	1	15000	Normalleştirme yok	0.703
47	4	0.1	0.01	0.01	1	15000	Normalleştirme yok	0.699
48	3	0.1	0.01	0.01	1	30000	Normalleştirme yok	0.714
49	3	0.2	0.01	0.01	1	30000	Normalleştirme yok	0.727
50	3	0.3	0.01	0.01	1	30000	Normalleştirme yok	0.690
51	3	0.4	0.01	0.01	1	30000	Normalleştirme yok	0.695
52	3	0.2	0.02	0.01	1	30000	Normalleştirme yok	0.710
53	3	0.2	0.01	0.02	1	30000	Normalleştirme yok	0.696
54	3	0.2	0.01	0.01	0.1	30000	Normalleştirme yok	0.724
55	3	0.2	0.01	0.01	0.01	30000	Normalleştirme yok	0.712
56	3	0.2	0.01	0.01	1	60000	Normalleştirme yok	0.684
57	3	0.2	0.01	0.01	0.1	60000	Normalleştirme yok	0.704
58	3	0.2	0.01	0.01	1	45000	Normalleştirme yok	0.745
59	3	0.2	0.01	0.01	1	50000	Normalleştirme yok	0.685

Tablo 7: Binary Neural Network İterasyon Sonuçları

İterasyon #	Saklı düğüm sayısı	Öğrenme oranı	Öğrenme iterasyonu sayısı	Normalleştirme tipi	F1 Skoru
60	200	0.1	200	Normalleştirme yok	0.643
61	400	0.1	200	Normalleştirme yok	0.700
62	800	0.1	200	Normalleştirme yok	0.622
63	400	0.1	400	Normalleştirme yok	0.607
64	400	0.05	200	Normalleştirme yok	0.620

Tablo 8: Binary Support Vector Machine İterasyon Sonuçları

İterasyon #	İterasyon sayısı	Normalleştirme özellikleri	Project to the unit -sphere	F1 Skoru
65	1	YOK	YOK	0.691
66	10	YOK	YOK	0.721
67	100	YOK	YOK	0.727
68	1000	YOK	YOK	0.725
69	1	YOK	VAR	0.691
70	200	YOK	YOK	0.726

Test Kümesi Sonuçları

Doğruluk: 0,857

Duyarlılık: 0,759

Kesinlik: 0,731

F1: 0,745

Tahminlenen	Gerçek	
	POSITIVE	NEGATIVE
	POSITIVE	NEGATIVE
POSITIVE	255	94
NEGATIVE	81	798

Tüm bu testler dışında model test amacıyla iki hafta üst üste işletilmiştir.

1. İşletim testi

Crawler 12.12.2019 Perşembe günü çalıştırılarak %50'nin üzerinde seçilme ihtimali görülen makaleler değerlendirilmiştir.

Crawl edilen sayfa adedi	10.501
Son 1 haftadaki paylaşılan içerikler	1.206
Son bir hafta için seçilen sayfa (Cut-off %50)	103
BKM tarafından onaylanan	79

Test kümesi sonuçları

Doğruluk: 0,857

Duyarlılık: 0,759

Kesinlik: 0,731

F1: 0,745

İşletim testi sonuçları

Doğruluk: NA

Duyarlılık: NA

Kesinlik: 0,767

F1: NA

2. İşletim testi

Crawler 20.12.2019 Cuma günü çalıştırılarak %50'nin üzerinde seçilme ihtimali görülen makaleler değerlendirilmiştir. Eşleşme algoritması da kullanılarak 7 makale eklenmiştir.

Crawl edilen sayfa adedi	9.745
Son 1 haftadaki paylaşılan içerikler	1.358
Son bir hafta için seçilen sayfa (Cut-off %50)	60
BKM tarafından onaylanan	50

Test kümesi sonuçları

Doğruluk: 0,857

Duyarlılık: 0,759

Kesinlik: 0,731

F1: 0,745

İşletim testi sonuçları

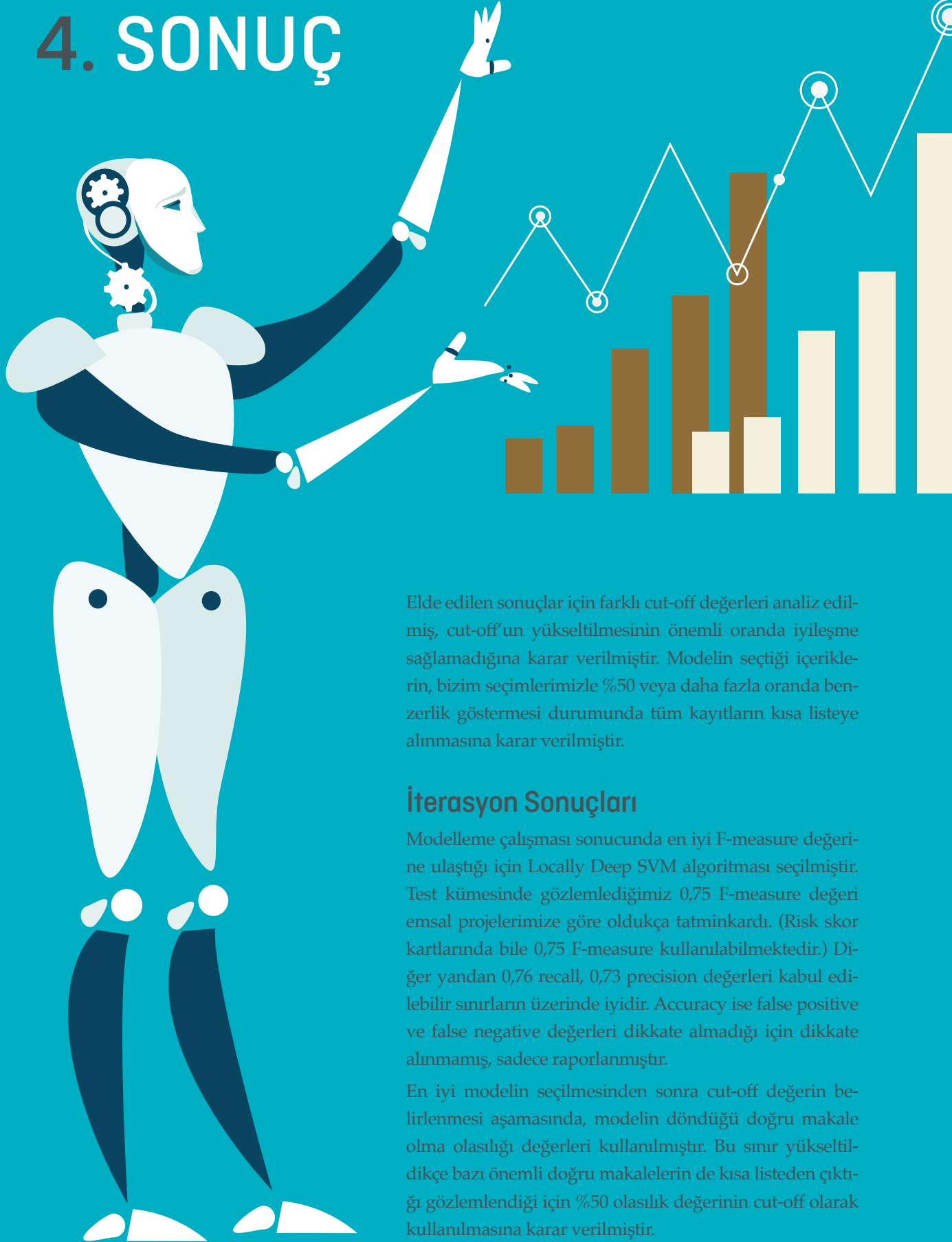
Doğruluk: NA

Duyarlılık: NA

Kesinlik: 0,833

F1: NA

4. SONUÇ



Elde edilen sonuçlar için farklı cut-off değerleri analiz edilmiş, cut-off'un yükseltilmesinin önemli oranda iyileşme sağlamadığına karar verilmiştir. Modelin seçtiği içeriklerin, bizim seçimlerimizle %50 veya daha fazla oranda benzerlik göstermesi durumunda tüm kayıtların kısa listeye alınmasına karar verilmiştir.

İterasyon Sonuçları

Modelleme çalışması sonucunda en iyi F-measure değerine ulaştığı için Locally Deep SVM algoritması seçilmiştir. Test kümesinde gözlemlediğimiz 0,75 F-measure değeri emsal projelerimize göre oldukça tatminkardı. (Risk skor kartlarında bile 0,75 F-measure kullanılabilir.) Diğer yandan 0,76 recall, 0,73 precision değerleri kabul edilebilir sınırların üzerinde iyidir. Accuracy ise false positive ve false negative değerleri dikkate almadığı için dikkate alınmamış, sadece raporlanmıştır.

En iyi modelin seçilmesinden sonra cut-off değerinin belirlenmesi aşamasında, modelin döndüğü doğru makale olma olasılığı değerleri kullanılmıştır. Bu sınır yükselttikçe bazı önemli doğru makalelerin de kısa listeden çıktığı gözlemlendiği için %50 olasılık değerinin cut-off olarak kullanılmasına karar verilmiştir.



Proje çıktısı program, haftada ortalama 120 adet makale seçmektedir ve bunların ortalama %70-%75'i BKM ekipleri tarafından da doğru seçim olarak işaretlenmektedir.

Gelecek Adımlar

Mevcut durumda %75'in üzerinde bir başarı oranı yakaladığımız Robokan uygulamamızın gelecek aşamalarında modelimizin başarı oranını yükseltmeyi hedefliyoruz. Kullandığımız algoritmanın parametreleri ve öznitelik adresleme alanında yapacağımız denemelerle skorumuzu geliştirmeyi amaçlıyoruz.

Bankacılık ve ödeme teknolojileri başlığı altında devam ettiğimiz çalışmamız, kurum içerisindeki diğer ekiplerimizin de ilgisini çekti. Gelecekte, benzer yapıyı temassız ödemeler, bilgi güvenliği, blokzincir ve yapay zekâ gibi başlıklar için de kullanmayı değerlendiriyoruz. Önümüzdeki dönemde mevcut eğitim kümемizin içeriğini güncel tutmak ve yeni teknolojilere dair haberleri yakalamak için bir arayüz üzerinden, içerik besleme kanalı oluşturmayı planlıyoruz. Bu geliştirmeye ek olarak, Twitter, LinkedIn ve Medium gibi sosyal mecralarda konularında uzman kişilerin içeriklerini tarayarak gündem oluşturan gelişmeleri eğitim kümемize dahil etmeyi hedefliyoruz.

Terimler Sözlüğü

> **Accuracy (Doğruluk):** Veriler üzerinde yapılan tahminin doğruluk oranını gösteren ölçüm.

> **Algoritma (Algorithm):** Bir görevi tamamlaması için bilgisayara verilen bir formül (yani bilgisayar için bir kurallar kümesi)

> **Bayes ağları (Bayesian networks):** Bayes ağı, Bayes modeli ve inanç ağı (belief network) olarak da bilinir, bir dizi değişken ve bağımlılıklarını temsil eden grafik tabanlı bir modeldir.

> **Bilgisayar görüşü (Computer vision):** Bir makine görüntü dosyalarından (JPEG'ler) veya kamera yayınlarından görsel girdiyi işlemesi.

> **Bilişsel hesaplama (Cognitive computing):** Veri madenciliği, NLP ve örüntü tanıma yoluyla insan düşünce süreçlerini taklit eden bilgisayarlı model.

> **Büyük veri (Big data):** Standart veri işleme yazılımı tarafından işlenemeyecek kadar karmaşık olan büyük miktarda yapılandırılmış veya yapılandırılmamış veri.

> **Çekişmeli üretici ağlar (GAN—Generative adversarial networks):** Yüzeysel bir anlatımla insan gözüne gerçek görünüşte fotoğraflar üretebilen bir tür sinir ağı. GAN tarafından oluşturulan görüntüler, fotoğraf verilerinin öğelerini alır ve bunları insanların, hayvanların ve yerlerin gerçekçi görüntülerine dönüştürür. Basit girdilerle anlamlı ve karmaşık çıktılar oluşturabiliyor.

> **Dar yapay zekâ (ANI—Artificial narrow intelligence):** Dar yapay zekâ belirli bir zamanda yalnızca bir göreve veya soruna odaklanabilen (örneğin bir insan rakibine karşı oyun oynayabilir) bir yapay zekâ türüdür. Bu şu anda mevcut yapay zekâ şeklidir. Zayıf yapay zekâ (weak AI) olarak da bilinir.

> **Denetimsiz öğrenme (Unsupervised learning):** bir algoritmanın sınıflandırılmayan veya etiketlenmeyen bilgilerle eğitildiği bir tür makine öğrenmesi, böylece algoritmanın rehberlik (veya denetim) olmadan hareket etmesine izin verme. Gözetimsiz veya kontrolsüz öğrenme de denir.

> **Derin öğrenme (deep learning):** Bilgisayarlara nasıl öğrenileceğini öğreten bir makine öğrenme tekniği (yani sınıflandırma tekniklerini kullanarak, bir insan zihni gibi öğrenmeyi taklit eden makineler)

> **Doğal dil işleme (NLP—Natural language processing):** Bilgisayarların doğal dil verilerini kullanarak insan dilini ve özelliklerini işlemesine, yorumlamasına ve analiz etmesine yardımcı olur.

> **Evrişimsel sinir ağı (CNN—Convolutional neural network):** Çok katmanlı algılayıcılar kullanarak görsel görüntüleri analiz etmek, sınıflandırmak ve kümelemek için özel olarak oluşturulan bir tür sinir ağı.

> **FN (False negative -Yanlış Negatif):** Yanlış bilgiye yanlış demek.

> **FP (False positive-Yanlış Pozitif):** Yanlış bilgiye doğru demek.

> **Genel yapay zekâ (AGI—Artificial general intelligence):** İnsanlarla benzer kabiliyete sahip yapay zekâ türüdür ayrıca güçlü yapay zekâ (strong AI) olarak da bilinir.

> **Genetik algoritma (Genetic algorithm):** Zor problemlerin çözümlerini etkili ve hızlı bir şekilde bulmak için kullanılan genetiğin prensiplerine dayanan bir algoritma.

> **Geri yayılım (Backpropagation):** İngilizce 'backward propagation of errors' kelimelerinden üretilen terim; kısaca, sistemin ilk çıktısının istenen çıktı ile karşılaştırıldığı, daha sonra (çıktılar arasındaki) arasındaki fark minimum seviyeye gelene kadar ayarlanan sinir ağlarının eğitimi yöntemidir.

> **Görüntü tanıma (Image recognition):** Bir görüntüdeki veya videodaki nesnenin özelliğini tanımlama veya algılama işlemi.

> **Gözetimli öğrenme (Supervised learning):** Çıktı veri setlerinin istenen çıktıları veya algoritmaları üretmeleri için makineleri öğrettiği bir tür makine öğrenmesi. (öğretmen-öğrenci ilişkisine benzer) Denetimli / kontrollü öğrenme olarak da bilinir.

> **Kümeleme (Clustering):** Makinelerin benzer verileri daha büyük veri kategorilerine ayırmasını sağlayan algoritma tekniği.

> **Makine çevirisi (Machine translation):** Metin veya konuşmaya dayalı çevirilerde (insandan insana) kullanılan dil işleme (NLP) uygulaması.

> **Makine öğrenimi (ML—Machine learning):** Bilgisayarların algılayıcı verisi ya da veritabanları gibi veri türlerine dayalı öğrenimini olanaklı kılan algoritmaların tasarım ve geliştirme süreçlerini konu edinen bir bilim dalıdır.

> **Optik Karakter Tanıma (OCR—Optical Character Recognition)**: Metin görüntülerinin (klavyeyle yazılan, elle yazılmış veya yazdırılan) elektronik veya mekanik olarak makine kodlu metne dönüştürülmesi.

> **Öğrenim aktarma (Transfer learning)**: Önceden öğrenilmiş verileri kullanan ve yeni bir görev grubuna uygulayan bir sistem. Örneğin, arabaları tanımayı öğrenirken edinilen bilgiler kamyonları tanımaya çalışırken geçerli olabilir.

> **Örüntü tanıma (Pattern recognition)**: Verilerde bulunan örüntülerin otomatik olarak tanınması.

> **Pekiştirmeli öğrenme (Reinforcement learning)**: Davranışlıktan esinlenen, öznelere bir ortamda en yüksek ödül miktarına ulaşabilmesi için hangi eylemleri yapması gerektiğiyle ilgilenen bir makine öğrenmesi yaklaşımıdır.

> **Precision (Kesinlik)**: Doğru olarak değerlendirilen verilerin gerçekten kaçının doğru olduğunu gösteren ölçümdür.

> **Reaktif makineler (Reactive machines)**: Deneyimler hakkında analiz yapabilir, algılayabilir ve tahminlerde bulunabilir, ancak veri depolamaz; durumlara tepki verirler ve verilen anı temel alarak hareket ederler.

> **Recall (Duyarlılık)**: İstenilen bilgiyi doğru olarak tespit etme oranını gösteren ölçümdür.

> **Robotik (Robotics)**: İnsan zekasını veya hareketlerini daha etkili hale getiren robotların üretim ve tasarımı.

> **Robotik işlem otomasyonu (RPA—Robotic process automation)**: İnsanlar tarafından tamamlandığında tekrarlanan görevleri gerçekleştirmek için yapay zekâ ve makine öğrenimi özelliklerine sahip yazılımlar.

> **Sezgisel (Heuristic)**: Hızlı, optimum, çözüme dayalı problem çözme için tasarlanmış bir bilgisayar bilimi tekniği. Diğer kullanımları arasında sezgisel algoritma da vardır.

> **Sınıflandırma (Classification)**: Makinelerin veri noktalarına kategori atamasını sağlayan algoritma tekniği.

> **Sınırlı hafıza (Limited memory)**: Belirli bir zaman dilimine sınırlı kısa süreli hafızaya sahip sistemler.

> **Sohbet botları (Chatbots)**: İnsan kullanıcıyla yazılı veya sesli komutlarla sohbet edebilen bir sohbet robotu. E-ticaret, eğitim, sağlık ve iş endüstrileri tarafından iletişim kolaylığı ve kullanıcı sorularını yanıtlamak için kullanılır.

> **Süper yapay zekâ (ASI—Artificial super intelligence)**: Bir yapay zekâ insan davranışı veya zekasını taklit etmekle kalmıyor, bunları aşıyorsa işte buna süper yapay zekâ diyoruz. Süper yapay zekâ şimdilik sadece üzerinde tahminler yapabileceğimiz bir şey. Sonuçlarının ne olacağı ise şimdiden büyük bir tartışma konusu. Neyse ki uzmanlar uzun bir süre daha hayata geçeceğini öngörmüyor.

> **Tekrarlayan sinir ağı (RNN—Recurrent neural network)**: Sıralı bilgilere ve örüntü tanıma temelinde mantıklı çıktı üreten bir tür sinir ağı.

> **TN (True negative-Doğru Negatif)**: Doğru bilgiye yanlış demek.

> **TP (True positive-Doğru Pozitif)**: Doğru bilgiye doğru demek.

> **Turing testi (Turing test)**: Bilgisayar bilimcisi Alan Turing (1950) tarafından oluşturulan test, basitçe ayrı ortamlardan soru cevap yoluyla karşdakinin makine olup olmadığını anlamak için kullanılır.

> **Veri madenciliği (data mining)**: Büyük ölçekli veriler arasından faydalı bilgiye ulaşma, bilgiyi madenleme işidir. Büyük veri yığınları içerisinde gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programı kullanılarak aranması olarak da tanımlanabilir.

> **Yapay zekâ**: Doğal sistemlerin (zekice olsun veya olmasın) yapabildiği her bilişsel etkinliği (gerekirse bedenleri olan) yapay sistemlere, daha da yüksek başarımlı düzeylerinde nasıl yaptırabileceğimizi inceleyen bilim dalıdır. [Referans: 50 Soruda Yapay Zekâ-Cem Say]

> **Yapay sinir ağı (ANN—Artificial neural network)**: İnsan beyninden sonra yapay bir sinir sistemi yaratarak, duyuşal verileri öğrenen, yorumlayan ve sınıflandıran örüntü tanıyan bir bilgisayar algoritması aracılığıyla modellenmiş bir ağ.

> **Yapılandırılmamış veriler (Unstructured data)**: Kolayca aranabilir düzenli olmayan veriler (örneğin ses, video, sosyal medya içeriği)

> **Yapısal veri (Structured data)**: Kolayca aranabilir düzende açıkça tanımlanmış veriler.

ROBOKAN

UYGULAMASINI GELİŞTİREN EKİPLER

BANKALARARASI KART MERKEZİ



Dr. Soner Canko
Genel Müdür



Celal Cündoğlu
Genel Müdür
Yardımcısı



Prof. Dr. Cem Say
Danışman,
Boğaziçi Üniversitesi



Özge Çelik
İş Geliştirme
Direktörü



Okan Yıldız
İş Geliştirme
Müdürü



Enes Türk
İş Geliştirme
Mühendisi

CREDE



Ahmet Kocamaz
Kurucu



Batuhan Demirci
Veri Analisti

BİR YAPAY ZEKÂ
YOLCULUĞU

HABER DEDEKTİFİ **ROBOKAN**

