



1001 – BİLİMSEL VE TEKNOLOJİK ARAŞTIRMA PROJELERİNİ DESTEKLEME PROGRAMI PROJE BAŞVURU FORMU

Proje Başlığı: Evrak Tanıma ve Kategorilendirme Sistemi

Proje Yürütücüsü: Zeynep Kıcıkoğlu, Mehmet Eren Ekiz, Abdullah Karataş

Projenin Yürütüleceği Kurum/Kuruluş: İmü

ÖZET

Türkçe ve İngilizce özetlerin projenin (a) özgün değeri, (b) yöntemi, (c) yönetimi ve (d) yaygın etkisinin ana hatlarını kapsaması beklenir. Her bir özet 600 kelime ile sınırlandırılmalıdır. Bu bölümün en son yazılması önerilir.

Proje Özeti

Bu proje, dijital belgelerin içeriğine göre otomatik olarak tanınması ve sınıflandırılması amacıyla tasarlanmış bir sistem geliştirmeyi hedeflemektedir. PDF, Word gibi dijital formatlarda bulunan belgeler, doğal dil işleme (NLP) ve makine öğrenmesi algoritmaları kullanılarak işlenip fatura, sözleşme, rapor, teklif mektubu gibi kategorilere ayrılacaktır. Sistem, belgelerdeki anahtar kelimeler ve yapısal özellikleri analiz ederek türlerini tanıyacak ve bu belgeleri ilgili kategoriye otomatik olarak yerleştirecektir.

Projenin özgün değeri, belge yönetimini hızlandırarak, hatasız ve otomatik bir sınıflandırma süreci sunmasında yatmaktadır. Özellikle işletmelerde evrak yönetiminin verimliliğini artıracak olan bu sistem, kullanıcıların belge işleme sürecindeki iş yükünü azaltacak, arşivleme ve erişim süreçlerini kolaylaştıracaktır.

Proje kapsamında, doğal dil işleme ve makine öğrenmesi tekniklerinin potansiyelini artırmayı hedefliyoruz. Proje yönetiminde, yapay zeka, veri bilimi ve belge yönetimi alanlarında uzmanlaşmış araştırmacılardan oluşan disiplinler arası bir ekip kurulacaktır. Geliştirilecek bu sistem, belge yönetimi süreçlerini optimize ederek küçük ve orta ölçekli işletmelere karar destek sistemi olarak hizmet edebilecektir. Ayrıca, farklı türde belgelerin tanınması ve sınıflandırılması sürecinde genişletilebilir bir altyapı sunarak yaygın etkisini artırabilecektir.

Anahtar Kelimeler: Evrak yönetimi, doğal dil işleme, makine öğrenmesi, belge sınıflandırma, veri bilimi, otomatik tanıma, belge işleme, yapay zeka, belge kategorilendirme

Title : Document Recognition and Categorization System

Summary

This project aims to develop a system designed to automatically recognize and classify digital documents based on their content. Documents in digital formats such as PDF, Word, etc. will be processed using natural language processing (NLP) and machine learning algorithms and categorized into categories such as invoices, contracts, reports, and proposal letters. By analyzing keywords and structural features in the documents, the system will recognize their types and automatically place them in the relevant category.

The unique value of the project lies in accelerating document management and providing an error-free and automatic classification process. This system, which will increase the efficiency of document management especially in enterprises, will reduce the workload of users in the document processing process and facilitate archiving and access processes.

Within the scope of the project, we aim to increase the potential of natural language processing and machine learning techniques. An interdisciplinary team of researchers specialized in artificial intelligence, data science and document management will be established to manage the project. This system will be able to serve as a decision support system for small and medium-sized enterprises by optimizing document management processes. It will also be able to increase its widespread impact by providing an extensible infrastructure in the process of recognizing and classifying different types of documents.

Keywords: Document management, natural language processing, machine learning, document classification, data science, automatic recognition, document processing, artificial intelligence, document categorization

1.ÖZGÜN DEĞER

1.1. Konunun Önemi, Projenin Özgün Değeri ve Araştırma Sorusu veya Hipotezi

Bu proje, dijital belgelerin içeriklerine göre otomatik olarak tanınması ve sınıflandırılması sürecine yeni bir perspektif sunarak, özellikle belge yönetiminde verimliliği artırmayı ve manuel işlemlerden kaynaklanan hataları en aza indirmeyi amaçlamaktadır. Literatürde mevcut belge sınıflandırma sistemleri genellikle anahtar kelime eşleştirme veya temel makine öğrenmesi algoritmalarına dayanırken, bu projede doğal dil işleme (NLP) ve makine öğrenmesi yöntemleri ile belgelerin içerik ve yapısal özellikleri dikkate alınarak daha karmaşık ve doğru bir sınıflandırma süreci gerçekleştirilecektir. Böylece, işletmelerin belge işleme süreçlerindeki iş yükünü azaltacak, belge türlerinin hatasız bir şekilde tanınması sağlanacaktır.

Özellikle küçük ve orta ölçekli işletmelerin belge yönetimi ve arşivleme süreçlerinde yaşadığı zaman kaybı, veri doğruluğu ve iş yükü gibi sorunlara çözüm olarak tasarlanan bu sistem, işletmelere önemli bir karar destek mekanizması sunmaktadır. Geliştirilen bu çözüm, belgelerin kategorize edilme sürecinde manuel müdahaleyi en aza indirgeyerek, arşivleme ve erişim süreçlerinde hız kazandıracaktır. Ayrıca, sistemin genişletilebilir ve esnek yapısı sayesinde farklı belge türleri ve diller için kolaylıkla uyarlanabilir olması, proje için yüksek bir yenilik değeri oluşturmaktadır. Bu özellikler, sistemin yalnızca mevcut ihtiyacı karşılamakla kalmayıp gelecekteki belge türleri ve formatları için de sürdürülebilir bir çözüm olmasını sağlayacaktır.

Projenin bilimsel katkısı, doğal dil işleme ve makine öğrenmesi yöntemlerini kullanarak belgelerin içerik odaklı sınıflandırma doğruluğunu artırmasıdır. Bu bağlamda proje, mevcut sistemlerin sınırlı doğrulukta kalmasını aşarak, yüksek doğruluk oranıyla sınıflandırma yapan ve yeni belge türlerini destekleyebilecek bir yapıyı hayata geçirecektir. Literatürde bu düzeyde genişletilebilirlik ve çeşitliliğe sahip sınıflandırma sistemleri nadir bulunmakta olup, bu projenin belge yönetimi alanına kavramsal, kuramsal ve metodolojik açıdan özgün katkılar sağlaması beklenmektedir.

Araştırma Sorusu: Belge türlerinin NLP ve makine öğrenmesi teknikleri kullanılarak içerik ve yapısal özellikleri göz önünde bulundurularak sınıflandırılması, dijital belge yönetiminde verimlilik, doğruluk ve iş yükü açısından nasıl bir etki yaratır?

Bu soruya yanıt ararken, proje ile geliştirilmesi hedeflenen çözüm, işletmelerin dijital belgelerle ilgili süreçlerinde hız ve doğruluk sağlayarak iş süreçlerini optimize edecek, böylece belge yönetimi alanında bilimsel yenilikler ve uygulamalı faydalar sunacaktır.

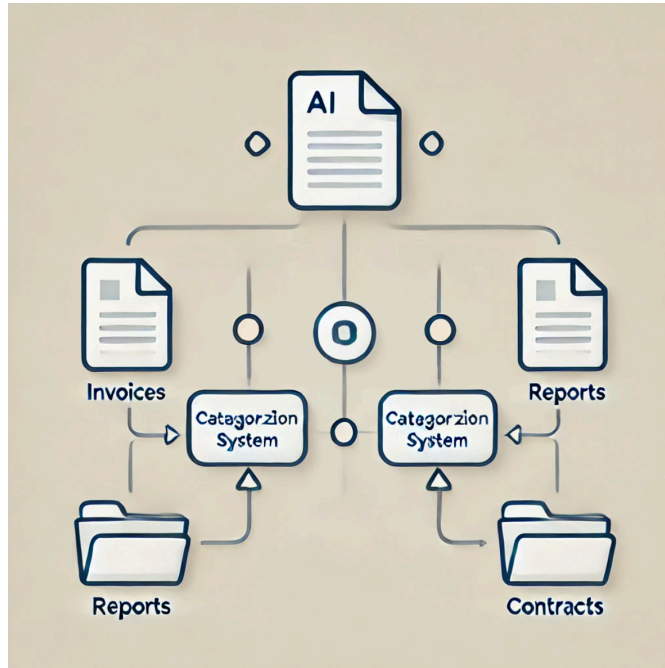
1.2. Amaç ve Hedefler

Bu projenin temel amacı, dijital belgelerin içeriklerine göre otomatik olarak tanınması ve sınıflandırılması için yapay zeka destekli bir sistem geliştirmektir. Bu sistem, işletmelerin belge yönetimi süreçlerindeki verimliliği artırarak, hataları en aza indirmek ve manuel işlemlerden kaynaklanan iş yükünü azaltmak amacıyla tasarlanmıştır. Proje kapsamında, doğal dil işleme (NLP) ve makine öğrenmesi tekniklerinin etkin kullanımıyla belgelerin türlerine göre sınıflandırılması hedeflenmektedir.

Hedefler:

1. Yüksek Doğruluk Oranı Sağlamak: Geliştirilecek sistemin, dijital belgeleri yüksek doğruluk oranıyla sınıflandırması ve türlerini otomatik olarak tanıması sağlanacaktır. Bu hedef, belgelerin içerik ve yapısal özelliklerinin derinlemesine analiz edilmesiyle ulaşılabilir niteliktedir.
2. Manuel İş Yükünü Azaltmak: İşletmelerde belgelerin manuel olarak sınıflandırılmasına yönelik zaman ve emek kaybını azaltmak hedeflenmektedir. Sistem, belgeleri otomatik olarak kategorilere ayırarak bu süreçteki insan müdahalesini minimuma indirecektir.
3. Belge Yönetimi Sürecinde Verimliliği Artırmak: Sistem sayesinde işletmelerin belge arşivleme, depolama ve erişim süreçleri hızlanacak; böylece belge yönetim süreçlerinde verimlilik artışı sağlanacaktır.
4. Farklı Belge Türleri ve Dilleri Desteklemek: Sistem, genişletilebilir bir altyapıya sahip olacak şekilde tasarlanacaktır. Bu sayede farklı belge türleri ve diller desteklenerek sistemin kullanım alanı genişletilecektir.
5. Karar Destek Mekanizması Sağlamak: İşletmelere karar destek mekanizması sunmak amacıyla sistemin, belgelerin doğru kategorilerde sınıflandırılması ve arşivlenmesini sağlaması hedeflenmektedir. Böylece, işletmelerin karar alma süreçlerinde belgeye erişim hızı artırılabilecek ve karar verme süreçleri hızlanacaktır.
6. Güvenilir ve Hatasız İşlem Sağlamak: Sistem, manuel süreçlerde oluşabilecek hataları en aza indirecek, güvenilir bir belge yönetim platformu sunacaktır. Bu sayede, iş süreçlerindeki doğruluk ve güvenilirlik artırılabilecektir.

Bu amaç ve hedefler, proje süresince ulaşılabilir, ölçülebilir ve işletmelerin günlük belge yönetim ihtiyaçlarına yönelik gerçekçi bir çözüm sunmayı hedeflemektedir.



Şekil 1. Yapay Zeka Destekli Evrak Kategorilendirme Sistemi Süreci

2.YÖNTEM

Bu projede dijital belgelerin sınıflandırılması ve türlerinin otomatik olarak tanınması için doğal dil işleme (NLP) ve makine öğrenmesi teknikleri kullanılacaktır. Literatürde benzer çalışmalarda başarıyla uygulanan teknikler dikkate alınarak, içerik ve yapısal özelliklere dayalı sınıflandırma yapılacaktır. Bu kapsamda veri toplama, ön işleme, model geliştirme, model doğrulama ve değerlendirme adımlarını içeren bir yöntem izlenecektir. Aşağıda bu adımlar ve her birine bağlı iş paketleri ayrıntılı olarak açıklanmıştır:

1- Veri Toplama ve Veri Hazırlama: Proje kapsamında kullanılacak veriler, dijital belge türlerine (fatura, sözleşme, rapor vb.) göre gruplandırılmış olan açık kaynak veri setlerinden toplanacaktır. Verilerin ön işlenmesi, NLP modellerine uygun hale getirilmesi için metin temizleme, normalizasyon, stop kelime çıkarma ve sözcük köklerine indirgeme gibi işlemler yapılacaktır. Bu süreç, belgedeki metinsel ve yapısal verinin daha anlamlı hale getirilmesini sağlayacaktır.

2- Özellik Çıkarma ve Temsil: Belgelerde sınıflandırmayı etkileyen özelliklerin belirlenmesi için TF-IDF (Term Frequency-Inverse Document Frequency), kelime vektörleri (Word2Vec, GloVe) ve benzeri NLP teknikleri kullanılacaktır. Bu aşamada, belgelerin sınıflandırılması için anahtar kelimeler, belge yapısı ve içerik özellikleri çıkarılarak modelde kullanılacak nitelikler belirlenecektir. Özellik çıkarma işlemi, bağımsız değişkenleri oluştururken doğru sınıflandırmayı sağlayacak verileri sağlamayı amaçlamaktadır.

3- Model Geliştirme: Belgelerin sınıflandırılması için denetimli makine öğrenmesi algoritmaları (örneğin Naive Bayes, SVM, Random Forest, XGBoost) ve derin öğrenme teknikleri (örneğin BERT, Transformer tabanlı modeller) kullanılacaktır. İlk aşamada farklı modeller denenerek, her bir modelin sınıflandırma başarısı ölçülecek ve en yüksek doğruluğu sağlayan model seçilecektir. Bağımlı değişkenimiz, her belgenin türü olup sınıflandırma hedefimizi temsil etmektedir.

4- Model Doğrulama ve Değerlendirme: Geliştirilen modelin doğruluğunu ölçmek için doğrulama veri seti üzerinde test edilecektir. Başarı ölçütleri arasında doğruluk, hassasiyet, özgüllük ve F1 skoru gibi sınıflandırma metrikleri kullanılacaktır. Modelin genel başarısını artırmak amacıyla k-fold çapraz doğrulama yöntemi uygulanarak, eğitim sürecindeki varyans azaltılacaktır. Ayrıca, modelin hata payını belirlemek ve iyileştirmeleri görmek için hiperparametre optimizasyonu yapılacaktır.

5- Uygulama ve Entegrasyon: Başarıyla eğitilen ve doğrulanan model, bir belge yönetim sistemine entegre edilerek, sınıflandırma sürecinin otomatik bir şekilde çalışmasını sağlayacak bir API olarak sunulacaktır. Bu API sayesinde işletmeler, dijital belgelerini kolaylıkla sisteme yükleyerek anında sınıflandırma sonuçlarını alabileceklerdir. Entegrasyon işlemleri, sistemin genişletilebilir ve kullanıcı dostu bir yapıya sahip olması amacıyla Docker gibi container teknolojileri ile desteklenecektir.

6- İstatistiksel Analiz ve Sonuçların Değerlendirilmesi: Çalışmanın sonuçları, elde edilen doğruluk ve verimlilik ölçütleri ile istatistiksel olarak değerlendirilecektir. Çeşitli bağımsız değişkenlerin sınıflandırma başarısına etkisi, varyans analizleri ve korelasyon analizleri ile ölçülecektir. Bulgular literatür ile karşılaştırılarak sistemin özgün katkıları belirgin hale getirilecektir.

Araştırmanın Tasarımı ve Değişkenler: Araştırmada bağımlı değişken olarak belgelerin sınıf etiketleri (fatura, sözleşme, rapor vb.) kullanılacaktır. Bağımsız değişkenler ise belgelerin anahtar kelime yoğunluğu, metin uzunluğu, belge yapısı gibi sınıflandırmaya yardımcı olan özelliklerden oluşacaktır. Çeşitli istatistiksel yöntemlerle bu bağımsız değişkenlerin sınıflandırma doğruluğuna etkisi değerlendirilecektir.

Ön Çalışma ve Fizibilite: Literatür çalışmaları, belge sınıflandırma sistemlerinin mevcut eksikliklerini ve iyileştirme alanlarını belirlememizde yardımcı olmuştur. Ayrıca, proje öncesinde sınıflandırma algoritmalarının belgelere uygulanabilirliği ve veri hazırlama süreçlerinin doğruluğu için küçük ölçekli testler yapılmıştır. Bu testler, yöntemin uygulanabilir olduğunu ve proje hedeflerine ulaşmaya elverişli olduğunu göstermiştir.

İş Paketleri ile İlişkilendirme: Yöntemin her aşaması belirli iş paketleri ile ilişkilendirilmiştir:

- İş Paketi 1: Veri toplama ve veri hazırlama
- İş Paketi 2: Özellik çıkarma ve temsil
- İş Paketi 3: Model geliştirme ve doğrulama
- İş Paketi 4: Entegrasyon ve uygulama
- İş Paketi 5: Sonuçların değerlendirilmesi ve raporlama

Bu yöntemler ve iş paketleri, proje süresince belirlenen amaç ve hedeflere ulaşmada etkin bir şekilde kullanılacaktır.

3.PROJE YÖNETİMİ

3.1. Yönetim Düzeni: İş Paketleri (İP), Görev Dağılımı ve Süreleri

3.1.1. İş-Zaman Çizelgesi

Projede yer alacak başlıca iş paketlerinin hangi sürede gerçekleştirileceği “İş-Zaman Çizelgesi” doldurularak verilir. Literatür taraması, gelişme ve sonuç raporu hazırlama aşamaları, proje sonuçlarının paylaşımı, makale yazımı ve malzeme alımı iş paketi olarak gösterilmemelidir.

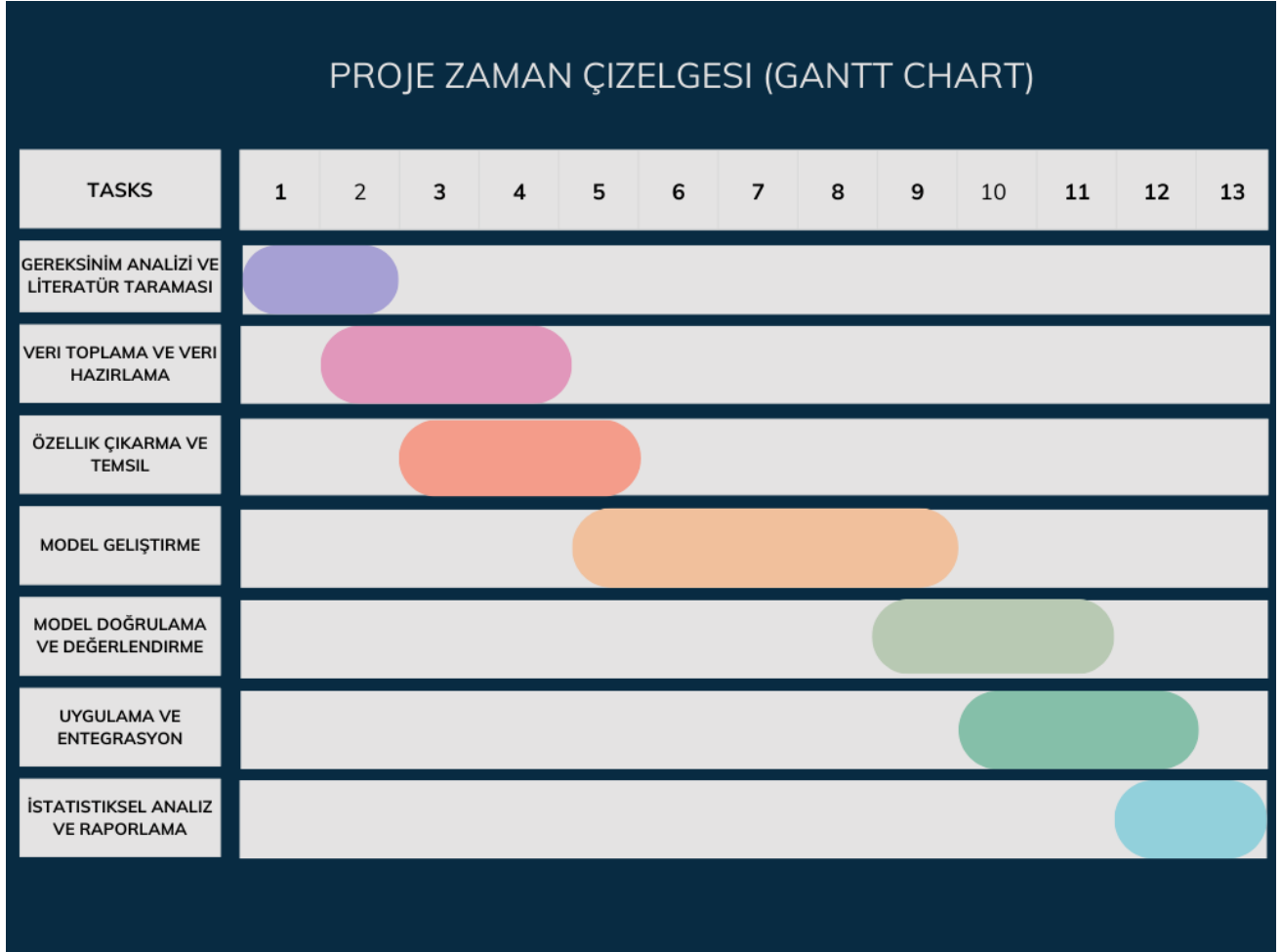
İŞ-ZAMAN ÇİZELGESİ (*)

İP N o	İş Paketi Adı	Projenin Başarısındaki Önemi (%)	Kim(ler) Tarafından Gerçekleştirileceği	AYLAR												
				1	2	3	4	5	6	7	8	9	10	11	12	13
1	Gereksinim Analizi ve Literatür Taraması	20	Herkes	X	X											
2	Veri Toplama ve Veri Hazırlama	20	Herkes		X	X	X									
3	Özellik Çıkarma ve Temsil	15	Herkes			X	X	X								
4	Model Geliştirme	30	Herkes					X	X	X	X	X				
5	Model Doğrulama ve Değerlendirme	20	Herkes									X	X	X		
6	Uygulama ve Entegrasyon	10	Herkes										X	X	X	
7	İstatistiksel Analiz ve Raporlama	5	Herkes												X	X

(*) Çizelgedeki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

(**) Sütun toplamı 100 olmalıdır.

(***) İP’de görev alacak kişilerin isimleri ve görevleri (araştırmacı, danışman, bursiyer ve yardımcı personel) yazılır. Bu aşamada bursiyer(ler)in isimlerinin belirtilmesi zorunlu değildir.



Şekil 2. Proje Zaman Çizelgesi için Gant Diyagramı

3.1.2. İş Paketleri

İŞ PAKETİ TABLOSU	
İP No: 1	İP Adı: Gereksinim Analizi ve Literatür Taraması
İP Hedefi: Projenin gereksinimlerini tanımlamak ve mevcut literatürü analiz ederek projenin bilimsel temellerini ve eksik noktalarını belirlemek.	
İP Kapsamında Yapılacak İşler/Görevler: 1.1. Proje için ihtiyaçların belirlenmesi ve dokümantasyonu. 1.2. Literatürde mevcut yöntemlerin ve yaklaşımların araştırılması. 1.3. Literatür analizi sonuçlarının raporlanması ve gereksinimlerle ilişkilendirilmesi.	Kim(ler) Tarafından Gerçekleştirileceği(*) 1.1. Mehmet Eren Ekiz 1.2. Abdullah Karataş 1.3. Zeynep Kıcıkoğlu
İP'nin Başarı Ölçütü: Proje gereksinimlerinin %100 doğrulukla belirlenmiş olması ve ilgili literatür incelemesinin tamamlanarak dokümanite edilmesi.	
Ara Çıktılar: İP için öngörülen ve başarı ölçütünün gerçekleşeceğini somut olarak gösteren (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, algoritma, yazılım, anket formu, verim, ham veri vb.) ara çıktılara ilişkin bilgi verilir. 1.1. Gereksinim analizi dokümanı. 1.2. Literatür tarama raporu.	

İŞ PAKETİ TABLOSU	
İP No: 2	İP Adı: Veri Toplama ve Veri Hazırlama
İP Hedefi: Projede kullanılacak dijital belgelerin toplanması ve veri temizleme, ön işleme aşamalarından geçirilerek analiz ve model eğitimi için uygun hale getirilmesi.	
İP Kapsamında Yapılacak İşler/Görevler: 2.1. İlgili veri kaynaklarının belirlenmesi ve veri setlerinin toplanması. 2.2. Veri temizleme işlemlerinin uygulanması (gürültü azaltma, eksik verilerin tamamlanması vb.). 2.3. Verilerin NLP ve makine öğrenmesi modellerine uygun hale getirilmesi (normalleştirme, stop kelimelerin kaldırılması vb.).	Kim(ler) Tarafından Gerçekleştirileceği(*) 2.1. Mehmet Eren Ekiz 2.2. Abdullah Karataş 2.3. Zeynep Kıcıkoğlu
İP'nin Başarı Ölçütü: Toplanan veri setinin %95 doğrulukla temizlenmiş olması ve analiz için uygun hale getirilmesi.	
Ara Çıktılar: İP için öngörülen ve başarı ölçütünün gerçekleşeceğini somut olarak gösteren (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, algoritma, yazılım, anket formu, verim, ham veri vb.) ara çıktılara ilişkin bilgi verilir. 2.1. Temizlenmiş ve hazır hale getirilmiş veri seti, ön işleme raporu.	

İŞ PAKETİ TABLOSU	
İP No: 3	İP Adı: Özellik Çıkarma ve Temsil
İP Hedefi: Sınıflandırma doğruluğunu artıracak özelliklerin belgelerden çıkarılması ve bu özelliklerin temsil edilmesi.	
İP Kapsamında Yapılacak İşler/Görevler: 3.1. Anahtar kelime, metin uzunluğu, belge yapısı gibi sınıflandırma için kullanılacak özelliklerin belirlenmesi. 3.2. Özellik çıkarma tekniklerinin uygulanması (TF-IDF, Word2Vec vb.). 3.3. Çıkarılan özelliklerin model eğitiminde kullanılabilir hale getirilmesi.	Kim(ler) Tarafından Gerçekleştirileceği(*) 3.1. Mehmet Eren Ekiz 3.2. Abdullah Karataş 3.3. Zeynep Kıcıkoğlu
İP'nin Başarı Ölçütü: Özelliklerin %90 doğrulukla çıkarılması ve modellerin bu özelliklerle eğitilmesi için uygun hale getirilmesi.	
Ara Çıktılar: İP için öngörülen ve başarı ölçütünün gerçekleşeceğini somut olarak gösteren (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, algoritma, yazılım, anket formu, verim, ham veri vb.) ara çıktılara ilişkin bilgi verilir. 3.1. Özellik çıkarma raporu, özellik seti dosyası.	

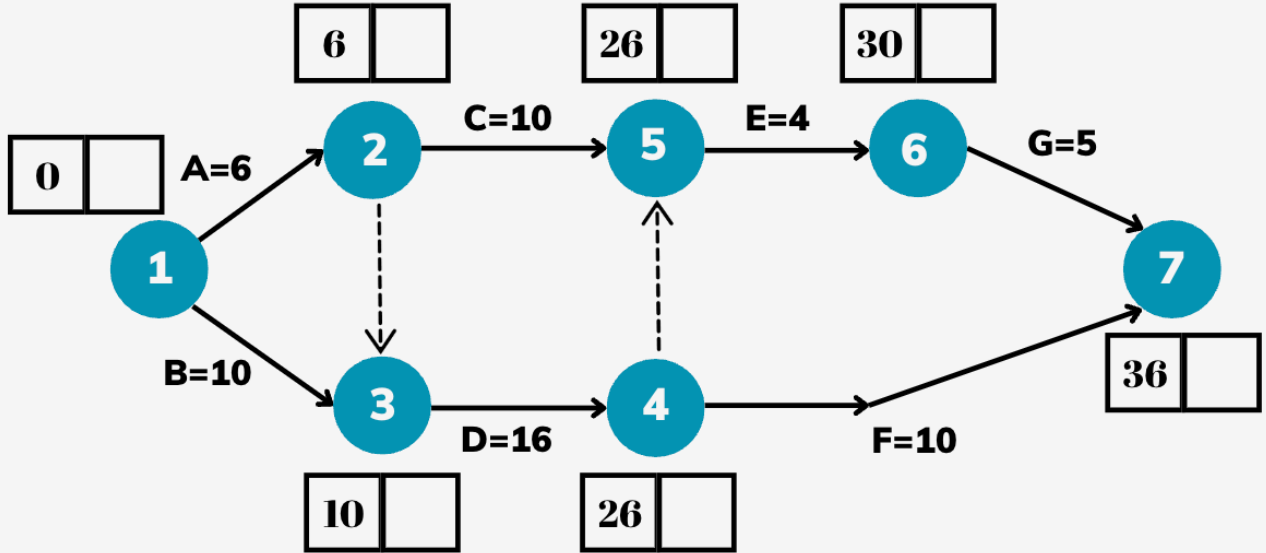
İŞ PAKETİ TABLOSU	
İP No: 4	İP Adı: Model Geliştirme
İP Hedefi: Dijital belgelerin sınıflandırılmasını sağlayacak makine öğrenmesi ve derin öğrenme modellerinin geliştirilmesi.	
İP Kapsamında Yapılacak İşler/Görevler: 4.1. Sınıflandırma modellerinin seçilmesi (Naive Bayes, SVM, BERT vb.). 4.2. Modellerin eğitim veri seti ile eğitilmesi. 4.3. Farklı modellerin doğruluk oranlarına göre karşılaştırılması ve en iyi modelin seçilmesi.	Kim(ler) Tarafından Gerçekleştirileceği(*) 4.1. Zeynep Kıcıkoğlu 4.2. Mehmet Eren Ekiz 4.3. Abdullah Karataş
İP'nin Başarı Ölçütü: Model doğruluğunun %90'ın üzerinde olması.	
Ara Çıktılar: İP için öngörülen ve başarı ölçütünün gerçekleşeceğini somut olarak gösteren (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, algoritma, yazılım, anket formu, verim, ham veri vb.) ara çıktılara ilişkin bilgi verilir. 4.1. Eğitilmiş model, model performans raporu.	

İŞ PAKETİ TABLOSU	
İP No: 5	İP Adı: Model Doğrulama ve Değerlendirme
İP Hedefi: Geliştirilen modellerin doğruluğunu ve etkinliğini ölçmek, en iyi modelin doğrulama verileri üzerinde test edilmesi.	
İP Kapsamında Yapılacak İşler/Görevler: 5.1. Doğrulama veri seti ile modelin performansının test edilmesi. 5.2. Modelin başarı metriklerinin (doğruluk, hassasiyet, özgüllük) hesaplanması. 5.3. Hiperparametre optimizasyonu yapılarak model doğruluğunun artırılması.	Kim(ler) Tarafından Gerçekleştirileceği(*) 5.1. Zeynep Kıcıkoğlu 5.2. Mehmet Eren Ekiz 5.3. Abdullah Karataş
İP'nin Başarı Ölçütü: Modelin doğrulama setindeki doğruluk oranının %90'a ulaşması.	
Ara Çıktılar: İP için öngörülen ve başarı ölçütünün gerçekleşeceğini somut olarak gösteren (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, algoritma, yazılım, anket formu, verim, ham veri vb.) ara çıktılarına ilişkin bilgi verilir. 5.1. Model doğrulama raporu, başarı metrikleri.	

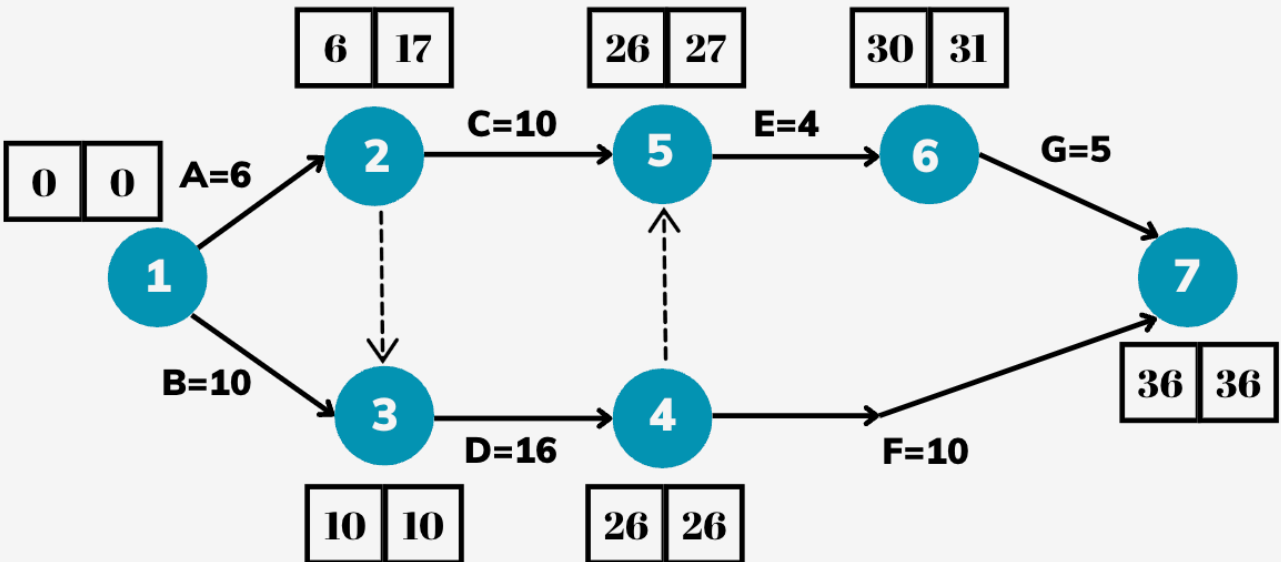
İŞ PAKETİ TABLOSU	
İP No: 6	İP Adı: Uygulama ve Entegrasyon
İP Hedefi: Geliştirilen modelin belge yönetim sistemine entegre edilmesi ve kullanıcıların sistem üzerinden belgelerini sınıflandırabilmesi.	
İP Kapsamında Yapılacak İşler/Görevler: 6.1. Modelin bir API olarak sunulması. 6.2. Belgelerin sınıflandırılması için kullanıcı dostu bir arayüz geliştirilmesi. 6.3. Sistem entegrasyonunun sağlanması ve son testlerin yapılması.	Kim(ler) Tarafından Gerçekleştirileceği(*) 6.1. Zeynep Kıcıkoğlu 6.2. Mehmet Eren Ekiz 6.3. Abdullah Karataş
İP'nin Başarı Ölçütü: Sistem entegrasyonunun başarılı şekilde tamamlanması ve sınıflandırma işleminin kullanıcılar tarafından kullanılabilir olması.	
Ara Çıktılar: İP için öngörülen ve başarı ölçütünün gerçekleşeceğini somut olarak gösteren (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, algoritma, yazılım, anket formu, verim, ham veri vb.) ara çıktılarına ilişkin bilgi verilir. 6.1. Entegre edilmiş belge yönetim sistemi, kullanıcı arayüzü.	

İŞ PAKETİ TABLOSU	
İP No: 7	İP Adı: İstatistiksel Analiz ve Raporlama
İP Hedefi: Model performansının değerlendirilmesi, sınıflandırma sürecinin etkisinin ölçülmesi ve proje sonuçlarının raporlanması.	
İP Kapsamında Yapılacak İşler/Görevler: <i>7.1. Modelin genel performansının istatistiksel analizinin yapılması.</i> <i>7.2. Başarı ölçütlerine göre elde edilen sonuçların karşılaştırılması.</i> <i>7.3. Sonuçların raporlanması ve projeye dair nihai raporun hazırlanması.</i>	Kim(ler) Tarafından Gerçekleştirileceği(*) <i>7.1. Mehmet Eren Ekiz</i> <i>7.2. Zeynep Kıcıkoğlu</i> <i>7.3. Abdullah Karataş</i>
İP'nin Başarı Ölçütü: Analiz raporlarının tamamlanması ve modelin beklenen doğruluk ve verimlilik hedeflerine ulaşması.	
Ara Çıktılar: İP için öngörülen ve başarı ölçütünün gerçekleşeceğini somut olarak gösteren (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, algoritma, yazılım, anket formu, verim, ham veri vb.) ara çıktılarına ilişkin bilgi verilir. <i>7.1. Proje sonuç raporu, istatistiksel analiz dokümanları.</i>	

PERT DİYAGRAMI



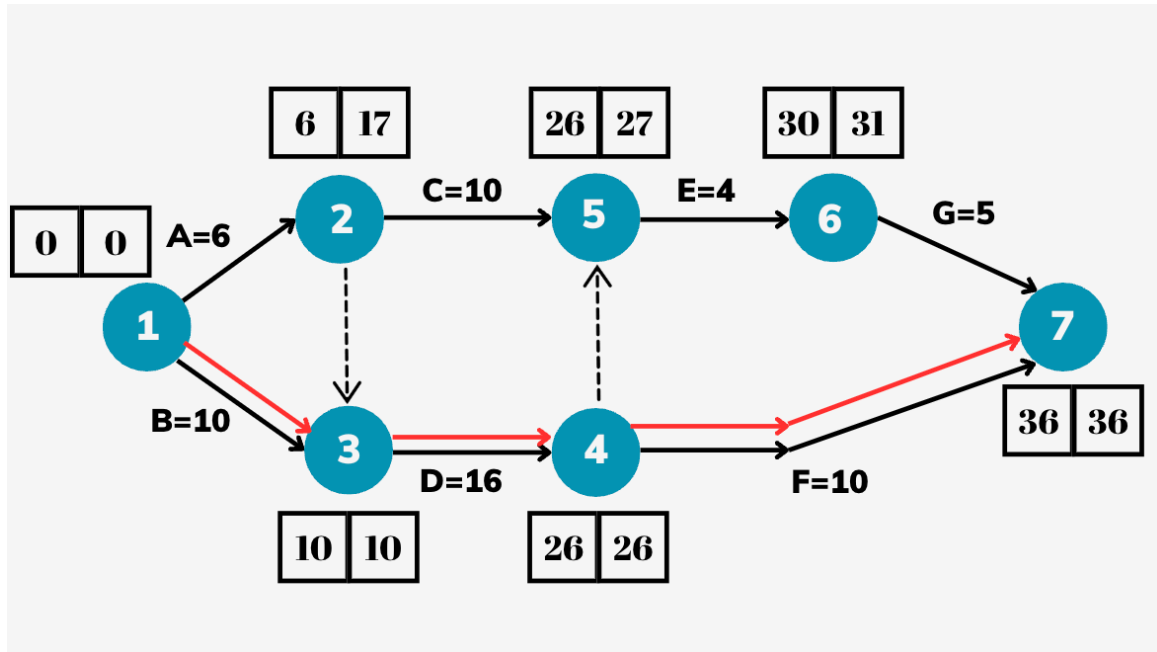
Şekil 3.



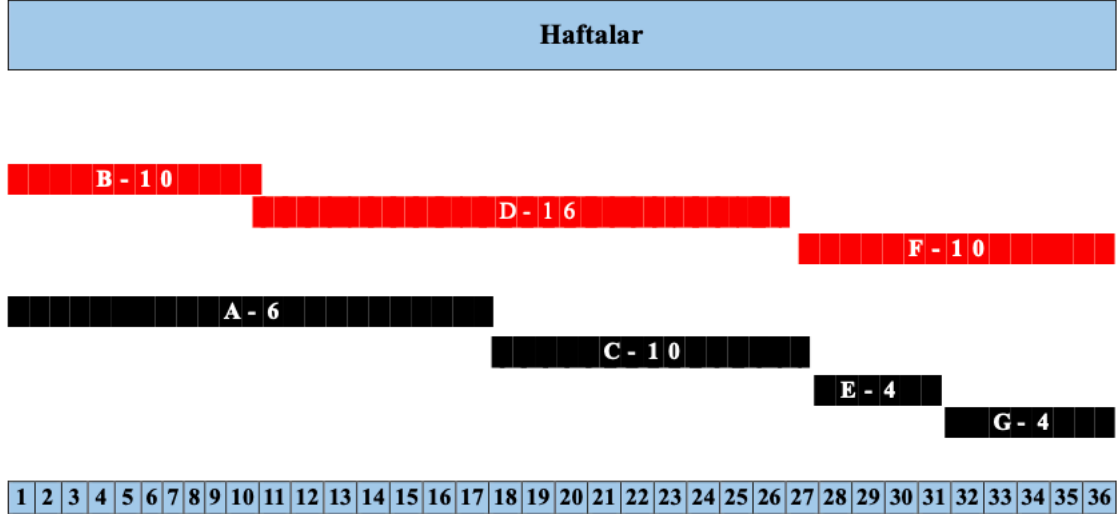
Şekil 4.

Faaliyetler	Önceki Faaliyetler	Tahmini Süreler (Hafta)			Ortalama	Varyans
		İyimser	Beklenen	Kötümser		
A	-	4	6	8	6	0.444444
B	-	8	10	12	10	0.444444
C	A	8	10	12	10	0.444444
D	A,B	12	16	20	16	1.777778
E	C,D	2	4	6	4	0.444444
F	D	8	10	12	10	0.444444
G	E	2	5	8	5	1

Şekil 5. PERT Diyagramı Görev ve Süre Tablosu



Şekil 6. Proje Yönetimi için PERT Diyagramı



Şekil 7. PERT Diyagramı Zaman Çizelgesi

Toplam bolluk baz alınarak grafik çizilmiştir. Kritik olan süreçlerin varyansları toplanır. Karakökü alınarak standart sapma tespit edilir.

$$\sqrt{2.666666} \approx 1.63299$$

RnD Costs (Araştırma ve Geliştirme Maliyetleri)

Veri Toplama ve Temizleme Araçları: \$3,000

Model Geliştirme ve Eğitim Maliyetleri (Bulut GPU): \$5,000

Yazılım Geliştirme ve Entegrasyon: \$4,000

Raporlama ve Analiz Araçları: \$2,000



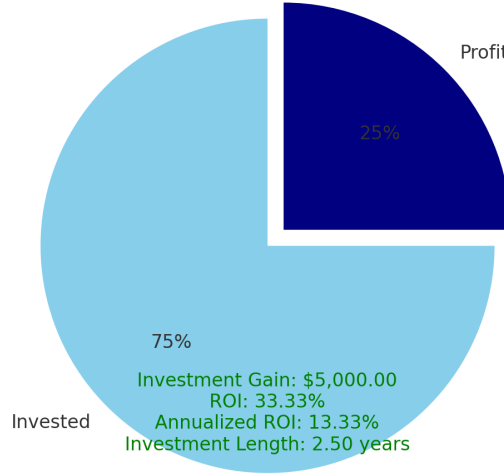
Proje Yönetimi ve Sertifikasyonlar: \$1,000
Toplam: \$15,000

Return on Investment (ROI - Yatırım Getirisi):

Başlangıç Tarihi: Ocak 2024
Tamamlanma Tarihi: Şubat 2025
Toplam Yatırım: \$15,000
Tahmini Getiri: \$20,000

Küçük ölçekli bir uygulama ile yatırımların ikiye katlanması mümkündür, dolayısıyla toplam gelirimiz \$20,000 olacaktır.

Investment & ROI Analysis for Document Classification Project



Şekil 6. Belge Sınıflandırma Projesi için Yatırım ve ROI Analizi

Result (Sonuç):

Investment Gain:	\$5,00
Annualized ROI:	33.33%
Investment Length:	1 years

Present Value Formulation (Şimdiki Değer Hesabı):

Formül: $PV = FV / (1 + i)^n$

FV (Future Value - Gelecek Değeri): \$20,000

Faiz Oranı: %5.25

Bugünkü Değer (PV): \$18,972

Tahmin edilen kısa vadeli kâr göz önüne alındığında, ürünün bugünkü değeri \$18,972 olarak hesaplanmıştır.

Belge Sınıflandırma ve Otomasyon Projesi için SMART Hedefler

Spesifik (basit, anlamlı, önemli):

Bu proje, OCR ve NLP tekniklerini kullanarak belgeleri otomatik olarak sınıflandıran, yüksek verimli bir yapay zeka destekli sistem geliştirmeyi amaçlamaktadır. Sistem, fatura, sözleşme, rapor ve niyet mektubu gibi belgeleri yüksek doğruluk ve hızla kategorize ederek manuel iş yükünü azaltacaktır.

Ölçülebilir (anlamlı, motive edici):

Projenin başarısı, %90'ın üzerinde sınıflandırma doğruluğu elde edilmesi ve her belgenin 5 saniyeden daha kısa sürede işlenmesi ile ölçülecektir. Bu, belge iş akışlarını optimize ederek işletmelerin otomasyon çözümlerine geçişini teşvik edecektir.

Ulaşılabilir (uygun, gerçekleştirilebilir):

Gelişmiş OCR ve NLP araçlarının, ayrıca eğitim ve doğrulama için mevcut veri setlerinin varlığı, sistemin uygulanabilirliğini desteklemektedir. Modüler tasarım sayesinde sistemin sürekli geliştirilmesi ve ölçeklendirilmesi de mümkündür.

İlgili (mantıklı, gerçekçi, kaynakları doğru kullanan, sonuç odaklı):

Proje, işletmelerde belge yönetimi ihtiyacını ele almaktadır. Sınıflandırma süreçlerini otomatikleştirerek hataları azaltır, zamanı etkin kullanır ve üretkenliği artırır, bu da projeyi günümüz ihtiyaçlarına son derece uygun hale getirir.

Zamana Bağlı (zaman odaklı, sınırlı, maliyet/zaman sınırlı, zaman duyarlı):

Proje, 36 ay içinde tamamlanacak şekilde tasarlanmıştır ve her aşamada ölçülebilir kilometre taşları ile işleyiş takip edilerek zamanında işleyen, optimize edilmiş bir sistem sağlanacaktır.

Satın Alma Kalemleri

Satın Alma Kalemleri ve Gerekçeleri

Kalem Adı	Maliyet (TL)	Gerekçe
GPU ve Bulut Hizmetleri	100,000 TL	Yapay zeka model eğitimi için yüksek hesaplama gücüne ihtiyaç duyulmaktadır.
OCR ve NLP Yazılımları	50,000 TL	Belgelerin metin işleme ve sınıflandırma süreçlerinde kullanılacak araçlar (ör. Tesseract, Google Vision API).
Veri Seti Alımı	30,000 TL	Belgelerin sınıflandırılması için etiketlenmiş veri seti ihtiyacı bulunmaktadır.
Veri Depolama Hizmetleri	20,000 TL	Büyük boyutlu belgelerin saklanması ve işlenmesi için bulut tabanlı depolama hizmetleri (örn. AWS S3).

Kalem Adı	Maliyet (TL)	Gerekçe
Sertifikasyon ve Testler	10,000 TL	Proje sonrası sistemin doğrulanması ve kullanıcı testleri için gerekli sertifikasyon maliyetleri.

Açıklama:

- GPU ve Bulut Hizmetleri: Model eğitimleri ve yüksek hesaplama gereksinimleri için AWS veya Google Cloud gibi hizmetlerden yararlanılacaktır.
- OCR ve NLP Yazılımları: Metin tabanlı belgeleri işlemek ve sınıflandırmak için açık kaynaklı veya profesyonel araçlar kullanılacaktır.
- Veri Seti Alımı: Etiketlenmiş belgeler (fatura, sözleşme vb.) proje doğruluğunu artırmak için satın alınacaktır.
- Veri Depolama Hizmetleri: Projede kullanılan büyük boyutlu veri setlerini depolamak ve erişilebilir kılmak için bulut tabanlı hizmetler kullanılacaktır.
- Sertifikasyon ve Testler: Proje sonunda kalite kontrol, kullanıcı testleri ve sertifikasyon işlemleri için kullanılacaktır.

2.Risk Yönetimi

RISK YÖNETİMİ TABLOSU (*)

İP No	Risk(ler)in Tanımı	Alınacak Tedbir (ler) (B Planı)
1	Problemin yanlış tanımlanması	Problem tanımlandıktan sonra sürekli olarak gözden geçirilmeli ve gerekirse güncellenmelidir.
2	Verimsiz veya yetersiz çözüm yöntemleriyle sonuca ulaşılması	Kararlı ve kanıtlanmış çözümlere odaklanarak bu çözümler iyileştirilmeli ve gerektiğinde alternatifler düşünülmelidir.
3	Modelin doğruluk oranının düşük olması veya hedeflenen başarıya ulaşamaması	Geri adım atarak model parametrelerini ve algoritma seçimini yeniden değerlendirme yapılmalı; ek araştırmalar yapılmalıdır.
4	Yanlış özellik çıkarma veya veri hazırlama tekniklerinin seçilmesi	Kullanılan teknikler doğrulanmalı ve alternatif özellik çıkarma yöntemleri test edilmelidir.
5	Prototip aşamasının beklenenden uzun sürmesi	Tasarım süreci optimize edilmeli, beklenmeyen zorluklara karşı önlemler alınarak sürecin daha hızlı tamamlanması sağlanmalıdır.
6	Model test aşamasında çok sayıda hata ile karşılaşılması	Test aşamasında karşılaşılabilecek hatalar önceden tahmin edilmeli, buna uygun olarak önceden karşı önlemler belirlenmelidir.
7	Entegrasyon sürecinde teknik uyumsuzlukların ortaya çıkması	Entegrasyon sürecinde uzman desteği alınmalı ve alternatif entegrasyon seçenekleri gözden geçirilmelidir.

3.Araştırma Olanakları

Projenin yürütüleceği kurum ve kuruluşlarda var olan ve projede kullanılacak olan altyapı/ekipman (laboratuvar, araç, makine-teçhizat, vb.) olanakları belirtilir.

ARAŞTIRMA OLANAKLARI TABLOSU (*)

Kuruluşta Bulunan Altyapı/Ekipman Türü, Modeli (Laboratuvar, Araç, Makine-Teçhizat, vb.)	Projede Kullanım Amacı

(*) Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

4.YAYGIN ETKİ

Proje başarıyla gerçekleştirildiği takdirde projeden elde edilmesi öngörülen çıktılar, bu çıktıların paylaşımı ve yayılımına yönelik faaliyet(ler)/ürün(ler)/hizmet(ler) ile projeden oluşması öngörülen etkiler kısa ve net cümlelerle ilgili bölümde belirtilmelidir.

4.1.Projeden Elde Edilmesi Öngörülen Çıktılara İlişkin Bilgiler

Bu bölümde, projeden elde edilmesi öngörülen çıktılara yer verilmelidir. Söz konusu çıktılar, amaçlarına göre belirlenen kategorilere ayrılarak belirtilmeli, nicel gösterge ve hedeflere dayandırılmalı ve varsa bu çıktıları kullanacak kurum/kuruluş(lar)a ilişkin bilgi verilmelidir. Her bir çıktının elde edilmesinin öngörüldüğü zaman aralığı belirtilmelidir.

Çıktı Türü	Öngörülen Çıktı (lar)	Çıktının Elde Edilmesi için Öngörülen Zaman Aralığı (*)
Bilimsel/Akademik Çıktılar (Bildiri, Makale, Kitap Bölümü, Kitap vb.):		
Ekonomik/Ticari/Sosyal Çıktılar (Ürün, Prototip, Patent, Faydalı Model, Üretim İzni, Tescil, Görsel/İşitsel Arşiv, Envanter/Veri Tabanı/Belgeleme Üretimi, Spin-off/Start-up Şirket vb.):		
Araştırmacı Yetiştirilmesi ve Yeni Proje(ler) Oluşturulmasına Yönelik Çıktılar (Yüksek Lisans/ Doktora/Tıpta Uzmanlık/Sanatta Yeterlik Tezleri ve Ulusal/Uluslararası Yeni Proje vb.):		

(*) Proje başlangıcından itibaren 6 aylık süreler halinde belirtilmelidir (Örn. 0-6 ay/6-12 ay/12-18 ay, Proje sonrası vb.).

4.2.Proje Çıktılarının Paylaşımı ve Yayılımı

Proje faaliyetleri boyunca elde edilecek çıktıların ve ulaşılabilecek sonuçların ilgili paydaşlar ve olası kullanıcılara ulaştırılması ve yayılmasına yönelik yapılacak olan toplantı, çalıştay, eğitim, web sitesi, medya, fuar, proje pazarı ve benzeri etkinlikler aşağıdaki tabloda verilmelidir.

PROJE ÇIKTILARININ PAYLAŞIMI VE YAYILIMI TABLOSU (*)

Etkinlik Türü (Toplantı, Çalıştay, Eğitim, Web Sayfası, Görsel/Yazılı/ Sosyal Medya, Fuar, Proje Pazarı vb.)	Paydaş / Olası Kullanıcılar	Etkinliğin Zamanı ve Süresi

(*) Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

4.3.Projeden Oluşması Öngörülen Etkilere İlişkin Bilgiler



- Proje başarıyla gerçekleştirildiği takdirde projeden oluşması öngörülen
- Toplumsal/kültürel etki,
 - Ekonomik etki,
 - Ulusal Güvenlik etkisi

Proje Başvuru Sisteminde (PBS) seçilen [12. Kalkınma Planı](#) hedefleri ve politikaları çerçevesinde hedef kitle/alan belirtilerek açıklanmalıdır. Beklenen etkiler doğrulanabilir ve ölçülebilir olmalıdır. Etkilerin oluşma zamanına ilişkin öngörüler belirtilmelidir. Önerilen projeden oluşması öngörülen etkiler ile bu etkilerin 12. Kalkınma Planı hedef ve politikalarına sağlayacağı katkılar bu bölümde yapılacak açıklamalar çerçevesinde değerlendirilecektir. Şayet projeden oluşması öngörülen herhangi bir etki ve bu etkilerin 12. Kalkınma Planı hedef ve politikaları ile ilişkisi mevcut değilse bu durum açıkça belirtilmelidir.

Etki Türü	Öngörülen Etki Türü ve Kalkınma Planıyla İlişkisi	Etkinin Oluşması Öngörülen Zaman (*)
Toplumsal/Kültürel Etki: <ul style="list-style-type: none">• Yaşam Kalitesine Katkı,• Sürdürülebilir Çevre ve Enerjiye Katkı,• Refah veya Eğitim Seviyesinin İyileştirilmesine Katkı,• Ülke ya da Dünya Düzeyinde Önemli Bir Sosyal Soruna Getirilecek Çözümler vb.• Proje Sonuçlarını Uygulayan Kurum/ Kuruluş		
Ekonomik Etki: <ul style="list-style-type: none">• Potansiyel Sektörel Uygulama Alanları,• Küresel Pazar Öngörülleri,• İstihdam Katkısı,• Rekabetçilik (İhracata Etkisi, İthal İkamesi, Yabancı Sermaye Yatırımının Tetiklenmesi vb.)		
Ulusal Güvenlik Etkisi: <ul style="list-style-type: none">• Siber güvenlik,• Enerji güvenliği,• Sınır güvenliği,• Gıda güvenliği,• Ekonomik güvenlik vb.		

(*) Proje başlangıcından itibaren 6 aylık süreler halinde belirtilmelidir (Örn. 0-6 ay/6-12 ay/12-18 ay, Proje sonrası vb.)

BELİRTMEK İSTEDİĞİNİZ DİĞER KONULAR

Sadece proje önerisinin değerlendirilmesine katkı sağlayabilecek bilgi veya veri (grafik, tablo, vb.) eklenebilir.

BAŞVURU FORMU EKLERİ

EK-1: KAYNAKLAR

EK-2: BÜTÇE VE GEREKÇESİ

EK-3: PROJE EKİBİNİN DİĞER PROJELERİ VE GÜNCEL YAYINLARI (Proje Başvuru Sistemi (PBS)'ne girilen bilgiler doğrultusunda Sistem tarafından otomatik olarak oluşturulmaktadır.)