

VERİ YÖNETİM PLANI ÖRNEĞİ

Bu Veri Yönetim Planı, TÜBİTAK'a Araştırma Destek Programları Başkanlığı (ARDEB) tarafından yürütülen destek programları kapsamında önerilen “**Yapay Zeka Destekli Görüntü İşleme ile Bitki Hastalıklarının Erken Teşhisi**” başlıklı proje kapsamında toplanacak verileri kapsar ve TÜBİTAK Açık Bilim Politikasıyla uyumlu olacak şekilde hazırlanmıştır. Bu plan, bu proje süresince üretilen verilerin kullanılabilirliği, erişimi ve korunmasını sağlamak için uygun şekilde yönetilmesini sağlamayı amaçlamaktadır.

1. Araştırmanız sırasında hangi tür veriler kullanılacak ve/veya elde edilecektir?

Açıklama: Araştırma sonuçlarını elde etmek için biriktirilen, gözlemlenen veya yaratılan verinin tamamı araştırma verisidir (sayısal veri, dokümanlar, laboratuvar notları, anketler, cevaplar, kopyalar, ses ve video kayıtları, fotoğraflar, filmler, tarihi eserler, numuneler, örnekler, modeller, algoritmalar, senaryolar, simülasyon sonuçları, metotlar ve iş akışları vb.). Proje süresince üretilen verinin tür ve formatlarının, verinin tanımlanmasında kullanılan izin ve dosya adlandırma kurallarının ile verinin tanımlanmasında kullanılacak standartların bu bölümde açıklanması gerekmektedir. Proje sırasında üçüncü taraflardan veri transferi yapılacaksa bu durumun da belirtilmesi gerekmektedir.

Proje süresince aşağıda belirtilen türde veriler elde edilecek ve kullanılacaktır:

- **Görüntü Verileri:**
 - **Tür:** Farklı bitki türlerine ait sağlıklı ve hastalıklı yaprak, gövde, meyve görüntüleri.
 - **Format:** JPEG, PNG.
 - **Boyut:** Her bir görüntü 1-5 MB arasında değişecektir. Toplam veri boyutu tahmini 500 GB olacaktır.
 - **Adlandırma Kuralı:** BitkiAdi_HastalikAdi_Tarih_SiraNo.jpg (örn: Domates_Mildiyo_20250718_001.jpg).
- **Meta Veriler:**
 - **Tür:** Görüntülerin çekildiği tarih, saat, konum (GPS), bitki çeşidi, hastalık şiddeti, çevresel koşullar (nem, sıcaklık).
 - **Format:** CSV, JSON.
 - **Adlandırma Kuralı:** MetaVeri_BitkiAdi_Tarih.csv.
- **Model ve Algoritma Verileri:**
 - **Tür:** Makine öğrenimi model ağırlıkları, eğitim ve test setleri, Python kodları (Jupyter Notebook, .py dosyaları).
 - **Format:** HDF5 (.h5), Pickle (.pkl), CSV, .py.
 - **Adlandırma Kuralı:** ModelAdi_Versiyon_Tarih.h5, AlgoritmaAdi_Modul.py.
- **Literatür ve Doküman Verileri:**
 - **Tür:** Bilimsel makaleler, raporlar, proje dokümantasyonları, laboratuvar notları.
 - **Format:** PDF, DOCX.

Üçüncü taraflardan herhangi bir doğrudan veri transferi yapılmayacaktır. Ancak, projenin başlangıç aşamasında açık kaynaklı genel bitki hastalıkları görüntü veri setleri (örn. PlantVillage veri seti) literatür taraması ve model geliştirme için referans amaçlı kullanılabilir. Bu veriler yalnızca literatür taraması ve kıyaslama amacıyla kullanılacak olup, proje kapsamında doğrudan yeni veri üretimi ve analizi ile ilişkilendirilmeyecektir.

2. Proje kapsamında elde edilecek veya üretilen veri, paylaşım ve tekrar kullanım için uygun olacaktır mıdır?

Açıklama: Verinin keşfedilebilirliğini en üst düzeye çıkarmak için alınacak önlemler, veri paylaşımı ile ilgili koşul ya da kısıtlamalar, bunların bir veri paylaşım sözleşmesi ile düzenlenip düzenlenmeyeceği ve veri yayınlamanın zamanlaması gibi düzenlemelerin bu bölümde açıklanması gerekmektedir.

Akademik yayınlardaki şekillerin ve sonuçların elde edilmesini sağlayan sayıları ve/veya yazılım/kodları içeren veri setleri alandaki araştırmacıların tekrar kullanımına uygun olmalıdır.

Evet, proje kapsamında elde edilecek veya üretilecek veri, paylaşım ve tekrar kullanım için uygun olacaktır.

3. 2. soruya yanıtınız evet ise, lütfen 3, 4, 5 ve 6. Soruları da cevaplayınız.

Proje kapsamında elde edilecek görüntü verileri ve ilişkili meta veriler, makine öğrenimi modelleri ile kodlar, araştırmanın tamamlanmasının ardından açık erişimli veri havuzları aracılığıyla paylaşılacaktır. Verinin keşfedilebilirliğini artırmak amacıyla aşağıdaki önlemler alınacaktır:

- **Metadata (Üstveri) Kullanımı:** Her bir veri seti için Dublin Core veya DDI (Data Documentation Initiative) gibi standartlara uygun zengin metadata bilgileri hazırlanacaktır. Bu metadata, veri seti içeriğini, toplama yöntemlerini, formatlarını, anahtar kelimeleri ve ilgili yayınları içerecektir.
- **Veri Deposu Seçimi:** Veriler, uluslararası kabul görmüş ve kalıcı tanımlayıcı (DOI) sağlayan bir veri deposunda (örn. Zenodo, Dryad veya proje tamamlandıktan sonra kurulacak ULAKBİM açık veri deposu) yayınlanacaktır. Bu, verilere kalıcı ve kolay erişim sağlayacaktır.
- **Lisanslama:** Paylaşılan tüm veri setleri ve kodlar, Creative Commons Atıf 4.0 Uluslararası (CC BY 4.0) lisansı veya MIT Lisansı gibi açık lisanslar altında yayınlanacaktır. Bu, verinin tekrar kullanımını teşvik edecektir.
- **Veri Paylaşım Zamanlaması:** Proje kapsamında üretilen ve yayınlanan makalelerin kabul edilmesinden itibaren en geç 6 ay içinde ilgili veri setleri açık erişime sunulacaktır. Geliştirilen yazılım/kodlar GitHub gibi platformlar üzerinden sürekli güncel tutularak paylaşılacaktır.
- **Veri Bütünlüğü ve Kalite Kontrolü:** Paylaşımdan önce verinin tutarlılığı, doğruluğu ve eksiksizliği kontrol edilecektir. Anomaliler ve hatalar düzeltililecektir.
- **Veri Gizliliği ve Anonimleştirme:** Bitki hastalıkları verisinde kişisel veri bulunmamaktadır. Dolayısıyla anonimleştirme veya gizlilik endişesi taşınmamaktadır.

4. Veriler araştırma devam ederken nerede saklanacaktır?

Açıklama: Verilerin ne kadar süreyle saklanması gerektiği, verilerin planlanan yedekleme sıklığı ve proje süresince üretilen verilerden sorumlu kişi-bu kişinin sorumluluk alanlarının bu bölümde açıklanması gerekmektedir. Yedeklenen kopyaların hangi formatlarda saklanacağını ve bu formatlar için eğer gerekli ise yazılımların listelenmesi gerekmektedir.

Araştırma devam ederken veriler aşağıdaki şekillerde saklanacaktır:

- **Birincil Saklama:** Proje ekibinin çalıştığı bilgisayarların yerel disklerinde (SSD) ve proje ekibine tahsis edilmiş güvenli ağ depolama (NAS - Network Attached Storage) sunucusunda saklanacaktır. Bu sunucuya sadece proje ekibi ve yetkilendirilmiş IT personeli erişebilecektir.
- **Yedekleme Sıklığı:** Tüm veri setleri (ham veriler, işlenmiş veriler, modeller ve kodlar) haftalık olarak otomatik yedekleme yazılımları ile NAS sunucusundan ayrı bir yedekleme birimine (harici sabit disk veya bulut depolama) yedeklenecektir. Kritik model ve kod geliştirme süreçlerinde günlük manuel yedeklemeler yapılacaktır.
- **Yedekleme Formatları:** Yedeklenen veriler orijinal formatlarında (JPEG, PNG, CSV, JSON, HDF5, .py) saklanacaktır. Sıkıştırma için ZIP veya RAR formatları kullanılabilir.
- **Veri Sorumlusu:** Proje süresince üretilen verilerden Proje Yürütücüsü (Prof. Dr. [Ad Soyad]) sorumlu olacaktır. Proje Yürütücüsü, veri bütünlüğünün korunması, yedekleme prosedürlerinin takibi, erişim yetkilendirmeleri ve güvenlik önlemlerinin uygulanmasından sorumludur. Proje ekibindeki araştırma görevlileri, kendi ürettikleri verilerin düzenli olarak ana depolama alanına aktarılmasından ve meta verilerin doğru bir şekilde girilmesinden sorumlu olacaktır.

5. Proje tamamlandıktan sonra veriler uzun süreli olarak nasıl saklanacaktır ve üçüncü taraflarla nasıl paylaşılacaktır? Verilere kimler erişebilecektir?

TÜBİTAK'ın sunmuş olduğu veri koruma ve paylaşım planından farklı bir planınız var mıdır?

Açıklama: Veri için belirlediğiniz, yedekleme, saklama ve sürdürülebilirlik unsurlarını içeren uzun süreli koruma planının bu bölümde açıklanması gerekmektedir.

Proje tamamlandıktan sonra verilerin uzun süreli korunması ve paylaşımı aşağıdaki plana göre gerçekleştirilecektir:

- **Uzun Süreli Saklama:** Proje çıktısı olan nihai veri setleri (işlenmiş görüntüler, ilişkili meta veriler, eğitilmiş modeller ve nihai kodlar) proje bitiminden sonra en az 10 yıl süreyle kurumumuzun (Üniversite/Araştırma Merkezi Adı) merkezi güvenli veri sunucularında arşivlenecektir. Bu sunucular, RAID yapılandırması ve düzenli bant yedeklemeleri ile veri kaybına karşı korunmaktadır.
- **Kalıcı Veri Deposu:** Tüm paylaşılabılır veri setleri (görüntüler, meta veriler, modeller) kalıcı tanımlayıcılar (DOI) atanarak uluslararası kabul görmüş bir açık bilim veri deposunda (örn. Zenodo veya TÜBİTAK ULAKBİM'in uygun göreceği bir platform) yayınlanacaktır. Bu, verilere sürekli ve sürdürülebilir erişim sağlayacaktır.
- **Erişim ve Paylaşım:** Yayınlanan verilere herkes açık erişimle erişebilecektir. Kısıtlama olmaksızın, atıfta bulunularak yeniden kullanım serbest olacaktır (CC BY 4.0 lisansı). Geliştirilen yazılımlar GitHub üzerinden erişilebilir olacak ve sürekli güncel tutulacaktır.
- **TÜBİTAK Planından Farklılık:** TÜBİTAK'ın Açık Bilim Politikası'na tamamen uygun hareket edilecektir. Sunulan bu plan, TÜBİTAK'ın genel prensipleri doğrultusunda projemize özgü detayları içermektedir ve TÜBİTAK'ın veri koruma ve paylaşım planından farklı bir yaklaşım içermemektedir.

6. Araştırma sonrasında verileri saklamak için hangi altyapı olanakları (bilgi ve iletişim teknoloji altyapıları, güvenli arşiv, buzdolabı, yasal uzmanlık vb.) kullanılacaktır?

Açıklama: Bu altyapılara erişmeye yönelik proje ekibinin ve/veya yürütücü kuruluşun olanakları hakkında bilgi verilmelidir.

Araştırma sonrasında verileri saklamak için aşağıdaki altyapı olanakları kullanılacaktır:

- **Bilgi ve İletişim Teknolojileri (BİT) Altyapısı:**
 - **Üniversite Merkezi Sunucuları:** Üniversitemiz Bilgi İşlem Daire Başkanlığı tarafından yönetilen, yüksek kapasiteli (Petabyte seviyesinde) ve RAID 6 yapılandırmasına sahip sunucular kullanılacaktır. Bu sunucular, veri güvenliği için fiziksel güvenlik önlemleri, güvenlik duvarları ve düzenli sızma testleri ile korunmaktadır.
 - **Ağ Depolama (NAS):** Proje süresince kullanılacak NAS sunucusu, proje tamamlandıktan sonra da veri arşivleme ve kısa süreli erişim için kullanılmaya devam edecektir.
 - **Bant Yedekleme Sistemi:** Uzun süreli arşivleme için, üniversitemizin merkezi bant yedekleme sistemleri (LTO teyp kütüphanesi) kullanılacaktır.
 - **Yüksek Performanslı Hesaplama (HPC) Altyapısı:** Özellikle model eğitimleri için üniversitemizin HPC merkezindeki GPU destekli sunucular kullanılmaktadır. Bu altyapı, büyük veri setlerinin işlenmesi ve saklanması için uygun koşulları sağlamaktadır.
- **Güvenli Arşiv:** Dijital veriler için fiziksel bir arşiv ihtiyacı bulunmamaktadır. Ancak, projenin dokümantasyonları ve raporları, üniversitenin merkezi arşiv biriminde güvenli bir şekilde saklanacaktır.

- **Yasal Uzmanlık:** Proje ekibi, veri lisanslama ve açık erişim konularında üniversitemiz Hukuk Fakültesi'nden veya ilgili birimlerden (Teknoloji Transfer Ofisi gibi) gerektiğinde yasal danışmanlık alabilecektir.

Proje ekibinin ve yürütücü kuruluşun bu altyapılara erişim olanakları oldukça güçlüdür. Üniversitemiz, araştırmacılara bu tür altyapıları aktif olarak kullanma imkanı sunmaktadır. Proje ekibi, Bilgi İşlem Daire Başkanlığı ile yakın işbirliği içinde çalışacak ve veri depolama, yedekleme ve güvenlik konularında gerekli tüm destekleri alabilecektir.

7. Diğer Önemli Hususlar:

Açıklama: Verinin korunması ve paylaşılması için onay bilgileri, gerekliyse katılımcı kimliğinin saklı tutulması için alınacak önlemler, gerekliyse hassas verilerin güvenli şekilde saklanıp aktarılmasını sağlamak için alınacak önlemler, veri sahibinin adı, yeniden kullanım lisansı, üçüncü kişilerin kullanımına yönelik kısıtlamalar, patent başvurusu ya da bir derginin ambargo süresi gibi veri paylaşımında söz konusu olabilecek herhangi bir gecikme gibi konular için gerekli açıklamaları bu bölümde tanımlayınız.

- **Etik Onaylar:** Bu projede doğrudan insan veya hayvan denekleri ile çalışılmadığı için etik kurul onayı gerekmemektedir. Ancak, bitki hastalıkları verilerinin toplanmasında çevresel düzenlemelere ve biyoçeşitlilik koruma prensiplerine uyulacaktır.
- **Veri Sahibi:** Oluşturulan tüm veri setleri, algoritmalar ve modellerin fikri mülkiyet hakları, TÜBİTAK ve projenin yürütücü kuruluşu (Üniversite/Araştırma Merkezi Adı) arasında imzalanacak sözleşmeler çerçevesinde belirlenecektir. Genel olarak, proje ekibi tarafından üretilen veriler kamuya açık hale getirildikten sonra "Creative Commons Atıf 4.0 Uluslararası (CC BY 4.0)" lisansı ile yeniden kullanıma açılacaktır.
- **Yeniden Kullanım Lisansı:** Tüm veri setleri için CC BY 4.0 lisansı kullanılacaktır. Geliştirilen yazılımlar için MIT Lisansı tercih edilecektir. Bu lisanslar, atıf yapılması koşuluyla ticari veya ticari olmayan amaçlarla verinin serbestçe kopyalanmasına, dağıtılmasına, gösterilmesine, değiştirilmesine ve yeniden kullanılmasına izin vermektedir.
- **Üçüncü Kişilerin Kullanımına Yönelik Kısıtlamalar:** Herhangi bir kısıtlama bulunmamaktadır. Açık erişim prensipleri doğrultusunda veriler herkese açık olacaktır.
- **Gecikmeler:** Proje sonuçlarının akademik dergilerde yayınlanması durumunda olası bir ambargo süresi (genellikle 6-12 ay) veri paylaşımını geciktirebilir. Bu süre zarfında veri, yayıncı kuruluşun politikalarına uygun olarak kapalı tutulabilir, ancak ambargo süresi sona erdiğinde hemen açık erişime sunulacaktır. Proje kapsamında olası bir patent başvurusu ihtimali olması durumunda, patent başvuru süreci tamamlanana kadar ilgili veri setleri gizli tutulabilir. Ancak bu durum, projenin başlangıcında belirlenecek ve TÜBİTAK ile paylaşılacaktır.

Nasıl Kabul Alınır – Örnek Başvuru Üzerinden İpuçları:

Bu örnek başvuru, kabul alma şansınızı artırmak için aşağıdaki noktalara dikkat edilerek hazırlanmıştır:

1. **Tam ve Detaylı Cevaplar:** Her bir soruya, "Açıklama" kısmında istenen tüm detaylar verilerek eksiksiz yanıtlar verilmiştir. Örneğin, veri türleri, formatları, adlandırma kuralları spesifik olarak belirtilmiştir.
2. **Açık Bilim ve FAIR İlkeleri Vurgusu:** TÜBİTAK Açık Bilim Politikası'na uyumluluk baştan belirtilmiş, verilerin keşfedilebilirliği (Findable), erişilebilirliği (Accessible), birlikte çalışabilirliği (Interoperable) ve tekrar kullanılabilirliği (Reusable) (FAIR ilkeleri) vurgulanmıştır. Özellikle metadata kullanımı, kalıcı tanımlayıcılar ve açık lisanslama bu prensipleri destekler.

3. **Proaktif Yaklaşım:** Veri yönetimi süreçleri (yedekleme sıklığı, sorumlu kişiler, uzun süreli saklama planları) detaylandırılarak olası sorunlara karşı hazırlıklı olduğu gösterilmiştir.
4. **Güçlü Altyapı Vurgusu:** Üniversitenin mevcut BİT altyapısı, güvenli depolama ve HPC olanakları belirtilerek projenin veri yönetimi için gerekli teknik kapasiteye sahip olduğu gösterilmiştir.
5. **Gerçekçi ve Uygulanabilir Plan:** Sunulan planın hem iddialı hem de projenin kapsamına uygun ve uygulanabilir olduğu gösterilmiştir.
6. **Olası Riskler ve Çözümler:** Ambargo süresi veya patent başvurusu gibi veri paylaşımını geciktirebilecek durumlar öngörülmüş ve bu durumlara yönelik yaklaşımlar belirtilmiştir.
7. **Sorumlulukların Net Tanımı:** Veri yönetiminden sorumlu kişiler (Proje Yürütücüsü, araştırma görevlileri) ve sorumluluk alanları net bir şekilde tanımlanmıştır.

Bu örnek başvuru, sizin projenizin spesifik detaylarına göre doldurulduğunda, değerlendiricilere projenizin veri yönetiminin titizlikle planlandığı ve TÜBİTAK beklentileriyle uyumlu olduğu izlenimini verecektir.