**Makine Öğrenmesi Dönem Proje İsterleri**

**Proje Tanımı**

* Proje kapsamında, öğrenciler belirli bir konu içeriğinde veri seti seçerek/birleştirerek, temizleme ve ön işleme süreçlerini uygulayacak ve makine öğrenmesi algoritmalarını kullanarakanaliz yapacaklardır.
* Projede amaç makine öğrenmesi süreçlerini uygulamalı olarak gerçekleştirmektir.
* Proje tek kişilik olarak yapılabildiği gibi, **en fazla 2 kişi** olmak kaydı ile de gerçekleştirilebilir.
* Her grup **kendi seçtiği en az 3 algoritma ile** model performanslarını karşılaştıracaktır. (Derste anlatılanlar olmak zorunda değil.)
* **2 kişinin birlikte çalıştığı gruplarda daha fazla sayıda algoritma kıyaslaması gerekmektedir.**
* Sürecin sonunda gruplar, **sonuçlarını detaylı bir rapor ve sunum** ile açıklayacaktır.
* Projede yapılan işlemlerin ve kodlarının anlatıldığı max 10 dakikalık **proje sunumu Ara Sınav**, tüm sürecin ilgili literatür taraması ile **makale formatında hazırlandığı rapor ise Final sınavı** olarak değerlendirilecektir.
* ***Makale formatında hazırlanan raporda ilgili literatür taraması yapılmalı ve en az 10-15 kaynağa atıfta bulunulmalıdır.***
* Süreç ile ilgili son tarih ve duyurular belge sonunda bildirilmiştir.

**Proje Detayları**

**1.Veri Kaynağı- Çoklu Kaynaklardan Bütünleşik Veri Seti Oluşturma**

* **Öğrenciler en az iki** **açık veri kümesini anlamlı bir biçimde ilişkilendirerek tek ve tutarlı bir veri kümesi oluşturacaktır. Tek bir kaynaktan elde edilen veri setleri değerlendirilmeyecektir.**
* Veri kümeleri konu veya alan bakımından ilişkilendirilebilir olmalıdır.
* Nihai veri kümesi en az 1000 satır ve 9 özellik içermelidir.
* İlişkilendirme Kriteri:
  + Ortak bir anahtar (ID, tarih, coğrafi konum vb.) veya mantıksal eşleme tanımlanmalıdır.
  + Eşleştirme yöntemi (inner/outer join, yakın eşleme, fuzzy matching vb.) **raporda açıkça belirtilmelidir**.

**2. Veri Ön İşleme/Dönüştürme**

* Eksik verilerin doldurulma veya temizlenme, çıkarılma işlemleri,
* Veri aykırılıkları (outlier detection) belirlenmesi ve düzenlenmesi,
* Veri setinin genel dağılımı ve özet istatistikleri,
* Kategorik değişkenler için encoding işlemleri (One-hot encoding, Label encoding),
* Sayısal veriler için ölçeklendirme (Min-Max Scaling, Standardization),
* Gereksiz veya yüksek korelasyonlu değişkenler belirlenerek çıkarılması,
* Gerekirse veri transformasyonu işlemleri,
* Dağılım grafikleri, histogramlar, korelasyon/karmaşıklık matrisleri hazırlanması vs.

gibi işlemler yapılarak veri ön işleme aşamaları tamamlanacaktır.

Yukarıda bahsedilen bir veya birden çok veri ön işleme adımı yapılabilir. Tüm bu işlemler için her grup, kendi veri dönüşüm tekniklerini seçmekte özgürdür. Ancak açıklanmayan işlemler puan kaybına sebep olacaktır.

**3.Makine Öğrenmesi Süreçleri**

* Veri önişleme adımları gerçekleştirildikten sonra gruplar kendi seçtikleri en az 3 makine öğrenmesi algoritmasını uygular.
* İlk model sonuçları elde edilir ve karşılaştırmalar başlar.
* Model değerlendirme metrikleri hesaplanır (Accuracy, F1-Score, Precision, Recall vb.)
* İlk sonuçların neden iyi veya kötü olduğu analiz edilir.
* Gruplar özgünlüklerini gösterebilmeleri için ekstra yöntemler kullanabilirler.
* **İlk yöntemler de sonrasında yapılan iyileştirmeler de aşama aşama raporda bulunmalıdır.**

**4. Model Optimizasyonu ve İyileştirme**

* Hyperparametre ayarlamaları (GridSearch, RandomizedSearch vb.) yapılabilir.
* K-fold çapraz doğrulama yapılabilir.
* Model performansını artırmak için özellik mühendisliği ve farklı dönüşümler denebilir.

**5. Sonuçların Kıyaslanması ve Raporlama**

* Seçilen modelin avantajları ve dezavantajları tartışılır.
* Modelin doğruluğu ve genelleme yeteneği değerlendirilir.
* Kıyaslama Kriterleri:
  + En yüksek doğruluk oranı
  + En iyi görselleştirme
  + Modelin açıklanabilirliği
  + Gerçek dünyadaki anlamlılığı (uygulanabilirliği)
* Tüm aşamalar detaylı bir şekilde rapor haline getirilir. Model sonuçları grafikler, tablolar ve istatistiklerle desteklenir.

**6. Sunum ve Değerlendirme**

* Her grup maximum 10 dakikalık bir sunum yaparak projelerini tanıtmalıdır. Bu tanıtım Ara sınav olarak değerlendirilecektir. Sunumu slayt şeklinde yapmak zorunlu değildir. Öğrencinin anlatımı kendine özgü olabilir.
* Tanıtımda, kullanılan veri seti, sebebi, içeriği ve yapılan işlemler ile kullanılan algoritmalardan bahsedilmeli, sonuçları gösterilmeli ve kıyaslanmalıdır.
* Tanıtımı grup üyelerinden birinin yapması yeterlidir.
* Yapılan tüm işlemlerin bildiri formatında rapor olarak sunumu da Final sınavı olarak değerlendirilecektir.
* Projenin teknik doğruluğunun yanı sıra yenilikçilik ve uygulama potansiyeli de değerlendirmeye dahil edilecektir.

**7. SABİS’ e Yüklenmesi Gereken Belgeler**

* **Ara Sınav**
  + **Kod Dosyası:** Tüm veri işleme ve analiz adımlarını içeren Jupyter Notebook veya .py dosyası.
  + **Nihai Veri Kümesi:** Birleştirilmiş veri kümeleri CSV, XLSX veya Parquet formatında.
* **Final**
  + **Makale formatında hazırlanmış rapor dosyası (Times New Roman,11 Punto,2 yana yasla,1 satır aralığı)**
* Sunum yapmayan grupların ara sınav ve final notu 0 olarak girilecektir. Final ödevini yapmalarının da SABİS’e dosya yüklemelerinin de bir anlamı yoktur.
* SABİS’E YÜKLENMEYEREK SON DAKİKA MAİLİME ATILAN HİÇBİR ÇALIŞMA DEĞERLENDİRİLMEYE ALINMAYACAKTIR.
* SABİS ödev teslim tarihini kontrol etmeniz önerilir.
* Ara Sınav Sunum tarihleri linki: Linkte size tahsis edilen zamanda sunumunuzu yapmanız gerekmektedir.

<https://docs.google.com/spreadsheets/d/1JmYy8uaol1moMa2KqwNSuO30n4Dyba_TUyNKsMxddJE/edit?usp=sharing>

* Sunum tarihleri ile ilgili benim tarafımdan bir değişiklik veya opsiyon söz konusu değildir. Ancak sizler kendi aranızda **ANLAŞARAK-**haftalık sunum yapacak öğrenci sayısı değişmeyecek şekilde- sunum zamanlarınızı değiştirebilir ve ilgili drive linki üzerinden isim değişikliği yapabilirsiniz.

**8. Proje Takvimi**

Ara sınav SABİSE son yükleme tarihi**: 24 Aralık 2025**

Final sınav son teslim tarihi**: 07 Ocak 2026**

**Başarılar Dilerim.**