**2024-2025 BAHAR Dönemi**

ISE- 216 Veri Bilimi İçin İstatistik Dersi Proje Dokümantasyonu

**Projenin Amacı**

* Bu proje; ilgili derste hem teorik hem de pratik olarak bahsedilen bilgilerin gerçek bir dünya problemi içeren bir veri seti kullanılarak uygulanmasını amaçlamaktadır.
* Proje tamamlanmış hali ile ders içeriğinde öğrenilen bilgilerin bir yansıması olarak değerlendirilecektir.
* Veri Bilimi için İstatistik dersini alan öğrenci sayısı 91 olarak görülmektedir. Projelerinizi 5’er kişilik gruplar halinde yapmanız gerekmektedir. **ÖZEL ve GEÇERLİ bir mazeretiniz yoksa gruplardaki kişi sayısı değiştirilmeyecektir.**

**Fonksiyonel Gereksinimler**

1. Ücretsiz ve halka açık veri kümeleri içeren UCI (https://archive.ics.uci.edu/datasets) ya da Kaggle (https://www.kaggle.com) gibi web sitelerinden elde edilen veriler, gerçek dünya problemleri kapsamında (Pazar-sepet analizi, borsa tahmini, siber saldırı tahmini, hastalık tahmini, anomali tespiti…vs.) seçilmelidir.
2. Hali hazırda kullandığınız başka bir veri seti var ise ilgili uygulamayı üzerinde yapabilirsiniz ancak öncesinde bu veri seti içeriği ile ilgili benimle temasa geçmelisiniz.
3. Veri kümeleri seçimi yapılırken verilerin hem kategorik hem de sayısal veriler içermesi gerekmektedir. Seçilen veriye göre çeşitli işlemlerin yapılması proje isterleri içerisinde değerlendirilecektir.
4. Seçilen veri seti **en az 500+ örnekten oluşmalı ve en az 6 özelliğe sahip olmalıdır.**
5. Seçilen veri seti üzerinde **veri temizleme, dönüştürme ve ön işleme** süreçleri uygulanmalıdır. Veri seti temiz bir veri seti ise *kirletilerek* işlemler gerçeklenmelidir.
6. Veri seti üzerinde yapılacak işlemler tamamlandıktan sonra tarafınızdan seçilecek ilgili istatistiksel yöntemler ve gerekiyorsa tekrar veri önişleme adımları derste bahsedilen şekilde veya farklı yöntemler kullanılarak titizlikle yapılmalıdır.
7. Veriye ait istatistiksel analizlerde mutlaka görselleştirme yöntemleri de kullanılmalıdır.
8. Veriniz üzerinde mutlaka derste anlatılan tüm bilgi ve testler uygulanmalı, hipotez kurgusu yapılmalı ve hipotez değerlendirilmelidir.
9. Yapılacak projede kullanılacak programlama dili bağımsızdır. Puanlamaya etki etmeyecektir.
10. Aynı veri kümesinin birden çok proje ekibi tarafından **kullanılmaması** gerekmektedir. Her grubun çalıştığı veriler birbirinden ayrı olmak zorundadır. Farklı platformdan bulunan aynı verilerin de kabulü söz konusu değildir. Verilerin içeriklerini bilmek ve değerlendirmek öğrencilerin sorumluluğundadır.
11. **Veri setinizi seçmeden önce proje gruplarının oluşturulduğu ve seçilen veri setinin özelliklerinin yazıldığı ilgili dokümanı kontrol etmeniz, başka bir ekibin dokümanı kullanmadığından emin olmanız gerekmektedir**.
12. Tüm uyarılara ve verilen önbilgiye rağmen, **gözden kaçırma veya ihmalkarlık hasebiyle** veri setleri **aynı olan ekiplerin proje değerlendirmeleri 50 puan** üzerinden yapılacaktır. Bu noktada seçtiğiniz veri setinin daha önce başka bir grup tarafından seçilmiş olması tamamıyla öğrencilerin kontrolündedir.

**Proje Teslimi**

1. Projeniz size tahsis edilen eğitim öğretim haftasında sunulacak şekilde hazırlanacaktır.
2. Projenizi sadece hazırladığınız kod üzerinden de anlatabilirsiniz ancak slayt ile anlatmak isterseniz 10 slaytı geçmeyecek şekilde sunum hazırlamanız gerekmektedir.
3. Sunum bitimi sonrası kodunuzu açıklamanız ve çalıştırmanız gerekmektedir.
4. Proje ekibindeki her bir bireyin proje üzerindeki hakimiyeti ayrı ayrı değerlendirilecektir. Sunum esnasında ise katkılarınız değerlendirilecektir.
5. Projenizdeki;
   * + veri setini neden seçtiğinizi,
     + ilgili istatistiksel yöntemi/yöntemleri seçerken ki amacınızı,
     + proje tasarlamadan önceki beklentinizi,
     + projeyi yaparken karşılaştığınız durumları,
     + istatistiksel yöntemleri uyguladıktan sonra elde ettiğiniz analizi

detaylı şekilde anlatmanızı bekliyor olacağım.

1. Projenizde derste öğrendiğiniz teknikleri kullanıyor olmanız, puantaj olarak fayda sağlayacaktır. Ders kapsamında işlenmeyen ancak projenizde kullandığınız farklı yapılar ekstra puan sağlayacaktır.
2. Proje sürecinde yapılan ve açıklanan her işlem için ayrı ayrı puanlama yapılacaktır.
3. Projenizi ders esnasında sunuyor olmanın da bir öğrenme biçimi olduğunu önemle hatırlatmak isterim. Bu sebeple, derste arkadaşlarınızdan gelen soruları da cevaplayabiliyor olmanız gerekmektedir. Kısaca projeye sadece katılımcı olarak eşlik etmemeniz, her aşamasını açıklayabilen **uygulayıcı** olmanız gerekmektedir. Grup arkadaşlarınız proje sürecinde katkıda bulunmadıysa, bu durumu sunum esnasında beyan etmeniz ve değerlendirmeyi bu bilgiler ışığında yapmamı talep etmeniz gerekmektedir.
4. Projenizde kullanacağınız **veri setini** daha önceden paylaşılan dokümandaki veri seti adı kısmına **28 Mart 2025 saat 23:50 tarihine** kadar bildirmiş olmanız gerekmektedir.
5. Bu tarihe kadar veri seti seçimini yapmayan arkadaşların projeden not alması mümkün değildir. Dolayısı ile ilgili tarihte veri ve proje gruplarının belirlenme işlemi tamamlanmış olmalıdır. Bu tarihten itibaren yapılan başvurular **geçersiz sayılacaktır.**
6. İlgili tarihe kadar proje grubu oluşturamayan ancak listeye ismini yazan arkadaşlar 5’erli grup halinde rastgele belirlenecektir.
7. Projelerinizi **12 veya 13. Haftadan** itibaren sunuyor olacak şekilde hazırlık yapmanız gerektiğini önemle hatırlatmak isterim. Proje sunum tarihleri ilerleyen bir tarihte duyuru olarak paylaşılacaktır.

Başarılar Dilerim.