**2023-2024 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.

**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 algoritmaya göre çeşitli işlemlerin yapılması proje isterleri içerisinde değerlendirilecektir.
4. Veri setindeki sütün sayısı az olmamalıdır.
5. Veri seti kararının ardından tarafınızdan seçilecek ilgili istatistiksel yöntemler ve veri önişleme adımları derste bahsedilen şekilde veya farklı yöntemler kullanılarak titizlikle yapılmalıdır.
6. Veriye ait istatistiksel analizlerde mutlaka görselleştirme yöntemleri de kullanılmalıdır.
7. Yapılacak projede kullanılacak programlama dili bağımsızdır. Puanlamaya etki etmeyecektir.
8. Aynı veri kümesinin birden çok proje ekibi tarafından kullanılması durumunda, ilgili proje dokümantasyonuna ilk yazan ekibin veri seti kabul edilecek olup, diğer ekibin veri setini değiştirmesi beklenecektir.

**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.
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. Proje ekibi 2 kişi ve daha fazla olan arkadaşlardan daha kapsamlı bir proje bekliyor olacağım; örneğin daha büyük bir veri seti ve birden çok istatistiksel yöntemin detaylıca kıyaslanması vb.
4. 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.
5. Projenizde kullanacağınız **veri setini** daha önceden paylaşılan dokümandaki veri seti adı kısmına 05 Nisan 2024 saat 23:59 tarihine kadar bildirmiş olmanız gerekmektedir.
6. Projelerinizi 10 veya 11. Haftadan itibaren sunuyor olacak şekilde hazırlık yapmanız gerektiğini önemle hatırlatmak isterim.
7. Proje sunum tarihleri dokümanda ilerleyen bir tarihte paylaşılacaktır.