

## İçerik

| Ders Kodu | Dersin Adı               | Yarıyıl | Teori | Uygulama | Lab | Kredisi | AKTS |
|-----------|--------------------------|---------|-------|----------|-----|---------|------|
| ISI 525   | Açıklanabilir Yapay Zeka | 2       | 3     | 0        | 0   | 3       | 6    |

|                       |  |
|-----------------------|--|
| Ön Koşul              |  |
| Derse Kabul Koşulları |  |

|               |   |
|---------------|---|
| Dersin Dili   | İngilizce   |
| Türü          | Seçmeli   |
| Dersin Düzeyi | Yüksek Lisans   |
| Dersin Amacı  | Bu ders, yapay öğrenme sistemlerinin kararlarını açıklama ve yorumlama üzerine odaklanmaktadır. Dersin temel amacı, öğrencileri açıklanabilir yapay zeka (XAI) yöntemleriyle tanıştırmak ve bu yöntemlerin çeşitli alanlarda nasıl kullanıldığını pratik uygulamalar aracılığıyla göstermektir.   |
| İçerik        | Bu ders, yapay zeka tabanlı sistemlerin kararlarını, tahminlerini veya çıkarımlarını anlamlandırmak ve bu çıktıların mevcut algoritmalar tarafından neden ve nasıl hesaplandığını takip edebilmek için kullanılan yöntemleri açıklamayı hedefler. Ders kapsamında, sağlıktan finansa farklı alanlarda kullanılan ve "kara kutu" olarak tabir edilen yapay öğrenme modellerinin kararlarının yorumlanması ve bu modellerin güvenilir, şeffaf ve etik normlara uyan yapay zeka sistemleri geliştirmenin kritik yönlerine ilişkin genel bir bakış açısı sunulur. Öğrenciler, derste anlatılan yöntemleri Python programlama dili kullanarak uygulayacak ve elde edilen sonuçlar üzerinde tartışmalar yürütecektir. |
| Kaynaklar     | <ul style="list-style-type: none"><li>- Mehta, M., Palade, V., &amp; Chatterjee, I. (Eds.). (2023). Explainable AI: Foundations, methodologies and applications (Vol. 232, p. 273). Springer.</li><li>- Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., &amp; Müller, K. R. (Eds.). (2019). Explainable AI: interpreting, explaining and visualizing deep learning (Vol. 11700). Springer Nature.</li><li>- Molnar, C. (2020). Interpretable machine learning.</li><li>- Hsieh, W., Bi, Z., Jiang, C., Liu, J., Peng, B., Zhang, S., ... &amp; Liu, M. (2024). A comprehensive guide to explainable AI: from classical models to LLMs. arXiv preprint arXiv:2412.00800.</li></ul>                          |

## Teori Konu Başlıkları

| Hafta | Konu Başlıkları   |
|-------|---|
| 1     | Temel Kavramlar: Açıklanabilirlik, Şeffaflık, Yorumlanabilirlik ve Adalet, Açıklanabilir Yapay Zeka |
| 2     | Açıklanabilir Yapay Zekanın Teorik Temelleri  |
| 3     | Geleneksel Makine Öğrenmesi Modellerinin Yorumlanması   |
| 4     | Derin Öğrenme Modellerinin Yorumlanması   |
| 5     | Açıklanabilir Yapay Zeka Teknikleri   |
| 6     | Öznitelik Atama Yöntemleri  |
| 7     | Görselleştirme Yöntemleri   |
| 8     | Ara Sınav   |
| 9     | Zaman ve Sıralı Veriler için Yöntemler  |
| 10    | Multimodal Açıklanabilirlik   |
| 11    | Açıklanabilir Yapay Zeka Uygulamalarına Örnekler I  |
| 12    | Açıklanabilir Yapay Zeka Uygulamalarına Örnekler II   |
| 13    | Karşılaşılan Zorluklar  |

| Hafta | Konu Başlıkları   |
|-------|-------------------|
| 14    | Öğrenci Projeleri |