

**İnönü Üniversitesi**  
**Bilgisayar Mühendisliği Bölümü**  
**Bitirme Projesi (2022-2023 GÜZ)**

<b>Proje başlığı</b>	Trend Aramaların Sınıflandırılması
<b>Öğrenci(ler)</b>	Civan BOZAN - Hafta 10
<b>Proje özeti</b>	<p><b>Yöntem:</b> <u>Apache Spark</u> veri işleme çerçevesi ile arama motorlarında sıklıkla aratılan ifadelerin, <u>Spark Streaming</u> özelliği ile gerçek zamanlı olarak işlenmesi ve <u>Spark MLlib</u> makine öğrenmesi kütüphanesi kullanılarak bu verilerin sınıflandırılması (kategorilere ayrılması) sağlanıp uygun formatta (json, xml..) gerçek zamanlı olarak sunulması ve geçmişe yönelik analizlerin gerçekleştirilebilmesi açısından bu verilerin <u>Apache HBase</u> veritabanı ile tutulması ve ilgili analizlerin bir <u>web sitesi</u> aracılığıyla servis edilmesi.</p> <p><b>Amaç:</b> Kullanıcı eğilimlerinin tespit edilerek, ilgili kategoriler hakkında içerik üreten yazarlara ilham kaynağı olmak.</p>
<b>Tamamlanması gereken özellikler</b>	<ol style="list-style-type: none"><li>1. Spark ve ilgili kütüphanelerin incelenmesi.</li><li>2. HBase incelenmesi.</li><li>3. HBase ve Spark entegrasyonu</li><li>4. Arama motorlarından sıklıkla aratılan ifadelerin elde edilmesi.</li><li>5. Elde edilen ifadelerin spark ile gerçek zamanlı olarak işlenmesi.</li><li>6. Verilerin kategorilere ayrılması ve uygun formata dönüştürülmesi. CRISP-DM süreci<ol style="list-style-type: none"><li>a. İşin Anlaşılması</li><li>b. Verinin Anlaşılması</li><li>c. Verinin Hazırlanması</li><li>d. Modelleme</li><li>e. Değerlendirme</li><li>f. Konuşlandırma</li></ol></li><li>7. Spark ile işlenmiş tüm verilerin HBase'de tutulması.</li><li>8. Geçmiş verilerin web sitesi üzerinden sunulması.</li></ol>
<b>Sonradan eklenmesi gereken özellikler</b>	<ol style="list-style-type: none"><li>1. Verilerin bir API aracılığıyla sunulması.</li><li>2. Abonelik özelliği ile sadece ilgili kategorilere ait aramaların sunulduğu bir web sitesi.</li></ol>
<b>Bu hafta gerçekleştirilenler</b>	<ol style="list-style-type: none"><li>1. CRISP-DM süreci tamamlandı (konuşlandırma hariç).<ul style="list-style-type: none"><li>• Verinin Hazırlanması</li><li>• Modelleme</li><li>• Değerlendirme</li><li>• Konuşlandırma spark ile birlikte gerçekleştirilecek</li></ul></li></ol>
<b>Kaynaklar</b>	<ol style="list-style-type: none"><li>1. Veri Bilimi ve aşamaları:<ul style="list-style-type: none"><li>• Kelleher, John D. Veri Bilimi. Çeviren Onur Öztürk. İstanbul: QNB Finansbank, Ufuk Açıkan Yayınlar, 2019</li></ul></li><li>2. Veri hazırlama:<ul style="list-style-type: none"><li>• <a href="https://realpython.com/">https://realpython.com/</a></li></ul></li></ol>

**İnönü Üniversitesi**  
**Bilgisayar Mühendisliği Bölümü**  
**Bitirme Projesi (2022-2023 GÜZ)**

	<ul style="list-style-type: none"><li>• <a href="http://www.muratkarakaya.net">www.muratkarakaya.net</a></li><li>• <a href="http://www.analyticsvidhya.com">www.analyticsvidhya.com</a></li><li>• <a href="http://www.digitalocean.com">www.digitalocean.com</a> (<a href="http://www.journaldev.com">www.journaldev.com</a> )</li></ul> <p><b>3. Veri setleri:</b></p> <ul style="list-style-type: none"><li>• <a href="https://www.kaggle.com/datasets/savasy/multiclass-classification-data-for-turkish-tc32">https://www.kaggle.com/datasets/savasy/multiclass-classification-data-for-turkish-tc32</a></li><li>• <a href="https://starlangyazilim.com/kutuphanelerimiz/">https://starlangyazilim.com/kutuphanelerimiz/</a></li><li>• <a href="https://github.com/ncarkaci/tr-preprocessing/">https://github.com/ncarkaci/tr-preprocessing/</a></li><li>• <a href="https://github.com/stopwords-iso/stopwords-tr">https://github.com/stopwords-iso/stopwords-tr</a></li></ul> <p><b>4. Programlama dili ve kütüphaneler:</b></p> <ul style="list-style-type: none"><li>• <a href="https://docs.python.org/3/">https://docs.python.org/3/</a></li><li>• <a href="https://pandas.pydata.org/">https://pandas.pydata.org/</a></li><li>• <a href="https://scikit-learn.org/">https://scikit-learn.org/</a></li><li>• <a href="https://numpy.org/doc/">https://numpy.org/doc/</a></li></ul> <p><b>5. Diğer:</b></p> <ul style="list-style-type: none"><li>• <a href="https://www.geeksforgeeks.org/introduction-to-natural-language-processing/">https://www.geeksforgeeks.org/introduction-to-natural-language-processing/</a></li><li>• <a href="http://www.w3resource.com">www.w3resource.com</a></li><li>• <a href="https://stackoverflow.com/">https://stackoverflow.com/</a></li><li>• <i>İnternet (www)</i></li></ul>
--	---