

2209/A

ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEK
PROGRAMI

SONUÇ RAPORU

**PROJE BAŞLIĞI: MAKİNE ÖĞRENİMİ İLE OTEL YORUMLARI
ÜZERİNDE DUYGU ANALİZİNİN GERÇEKLEŞTİRİLMESİ VE
KATEGORİ DÜZEYİNDE ÖZETLENMESİ**

PROJE YÜRÜTÜCÜLERİN ADI: Semiha ÖZCAN,Beyza ÇELİK

DANIŞMANININ ADI: Asst. Prof. Üyesi Muhammed Yasin PAK

GENEL BİLGİLER

PROJENİN KONUSU	Makine Öğrenimi ile Otel Yorumları Üzerinde Duygu Analizinin Gerçekleştirilmesi ve Kategori Düzeyinde Özetlenmesi
PROJE YÜRÜTÜCÜSÜNÜN ADI	Beyza ÇELİK
DANIŞMANIN ADI	Asst. Prof. Üyesi Muhammed Yasin PAK
PROJE BAŞLANGIÇ VE BİTİŞ TARİHLERİ	01.10.2023-01.10.2024

1. Giriş

Günümüzde hızla büyüyen turizm endüstrisi, seyahat edenlerin konaklama ihtiyaçlarını karşılamak amacıyla geniş bir otel hizmetleri yelpazesi sunmaktadır. Seyahat edenler için otel seçiminde önemli bir faktör olan otel yorumları, kullanıcı deneyimlerini ve memnuniyet düzeylerini ifade etmek için yaygın olarak kullanılmaktadır. Bu yorumlar hem otel işletmecileri hem de seyahat acenteleri için değerli bir geri bildirim kaynağı oluştururken, seyahat planlaması yapan müşterilere de önemli bir rehberlik sunmaktadır. Ancak, yüksek miktarda üretilen bu yorumların manuel olarak işlenmesi ve analiz edilmesi oldukça zor ve zaman alıcı bir süreçtir. Bu nedenle, yorum analizlerinin otomatik bir şekilde yapılması gerekliliği ortaya çıkmaktadır.

Günümüzde, metin madenciliği ve doğal dil işleme (NLP) yöntemleri kullanılarak bu tür analizlerin gerçekleştirilmesi mümkündür. Özellikle duygu analizi, farklı kaynaklardan elde edilen yorumlardaki duygusal öğelerin analiz edilmesi amacıyla kullanılmaktadır. Otel yorumlarının analizi için popüler bir yöntem olan duygu analizi, bu yorumların içerdiği olumlu ve olumsuz duyguları belirleyerek müşteri geri bildirimlerinin daha derinlemesine incelenmesine olanak tanımaktadır.

Duygu analizi, farklı düzeylerde gerçekleştirilebilmektedir. Yorumdaki genel duyguyu tespit etmeyi amaçlayan doküman düzeyinde analiz, gerçekleştirilmesi daha kolay ancak daha yüzeysel bir sonuç sunar. Özellik düzeyinde analiz ise, yorumdaki her bir özellik (örneğin, otelin yemekleri) hakkındaki duyguları ayrı ayrı tespit etmeyi amaçlar ve bu nedenle daha detaylı ancak daha karmaşık bir yaklaşımdır. Özellik düzeyinde yapılan analiz sonuçları özetlenirken, bu özelliklerin kategoriler altında birleştirilmesi yaygın bir yöntemdir. Özellik düzeyine girmeden gerçekleştirilebilen bir diğer analiz türü ise, yorumların kategori düzeyinde analiz edilmesidir. Bu yaklaşım, doküman düzeyinde analizden daha detaylı sonuçlar sunarken, özellik düzeyine göre daha uygulanabilir bir çözüm sunmaktadır.

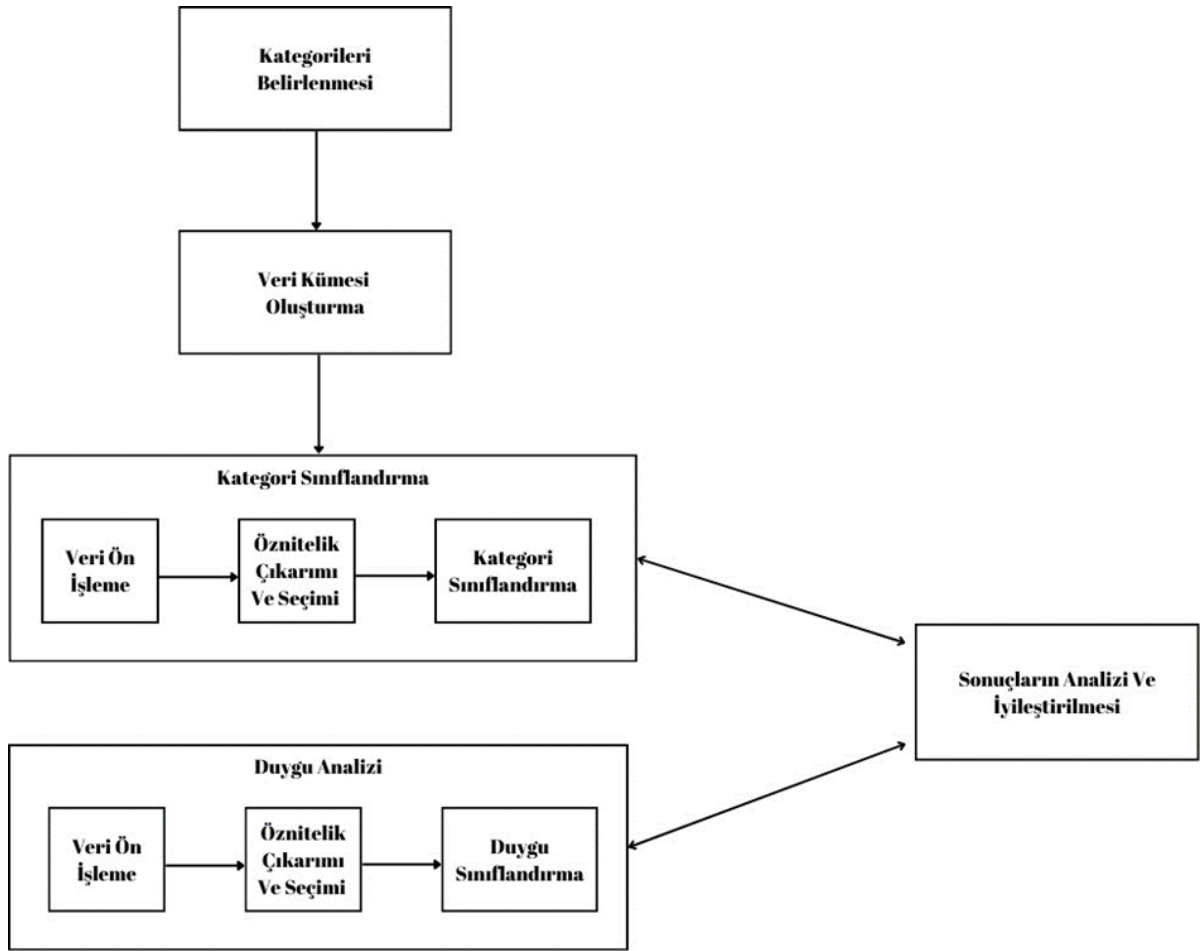
Yapılan araştırmalar, kategori düzeyinde analiz çalışmalarının az sayıda olduğunu göstermektedir. Otel yorumları üzerinde kategori düzeyinde bir duygu analizi yaklaşımının uygulanmasına ihtiyaç duyulmaktadır. Bu projede, öncelikle otel yorumlarından oluşan kapsamlı bir veri kümesinin oluşturulması amaçlanmaktadır. Bu veri kümesinin, gelecekteki çalışmalarda da kullanılabilecek şekilde yapılandırılması planlanmaktadır. İkinci olarak, makine öğrenimi tabanlı kategori düzeyinde bir duygu analizi sisteminin kurulabilmesi için iki aşamalı bir model

önerilmektedir. Önerilen sistemde, ilk aşamada otel yorumlarının genel kategorilere ve alt kategorilere ayrılması, ikinci aşamada ise yorumlardaki duygunun tespit edilmesi otomatik olarak gerçekleştirilecektir. Uygulanacak her bir adımda kullanılan algoritmalar ve yöntemlerin başarısı değerlendirilip, sistemin optimize edilmesi sağlanacaktır.

Bu proje sonucunda, otel işletmecileri ve kullanıcılar için belirli bir otel hakkında yapılan yorumların kategorilere ayrılmış bir şekilde pozitif ve negatif olarak değerlendirildiği özet bilgiler sunulması hedeflenmektedir. Elde edilen sonuçlar ve yapılacak analizler, bir makale ve bildiri aracılığıyla literatüre sunulacak olup, bu alanda gerçekleştirilebilecek pratik uygulamalar için bir zemin oluşturması beklenmektedir.

2. Rapor dönemlerinde yapılan çalışmalar

Yapılan çalışmanın deneysel akım şeması Şekil.1’de verilmiştir.



Şekil 1. Yapılan çalışmanın deneysel akım şeması.

Projede, otel yorumlarının kategori düzeyinde duygu analizi için çeşitli çalışmalar yürütülmüştür. Bu çalışmalar, kategorilerin belirlenmesi, veri kümesi oluşturulması ve analiz süreçlerini kapsamaktadır. Aşağıda, her bir iş paketinde gerçekleştirilen çalışmalar detaylı olarak açıklanmaktadır:

İş Paketi 1: Kategorilerin Belirlenmesi ve Veri Kümesi Oluşturma

Kategorilerin Belirlenmesi: Projede ilk adım olarak, otel yorumları için genel kategoriler ve alt kategoriler belirlenmiştir. Literatürde yapılan çalışmalar incelenerek, otel hizmetleriyle ilgili yaygın olarak kullanılan kategoriler (DENİZ, HAVUZ, HİZMET, KONUM VE ÇEVRE, ODA, PERSONEL, YEMEK) ve bu kategorilerin alt kategorileri (DENİZ-Deniz (Dalga durumu, zemin durumu, Temizliği), DENİZ-PLAJ-Plaj Temizliği, DENİZ-PLAJ-Plajın İmkanları (Snack bar, şezlong), HAVUZ-Havuzların Boyutu, HAVUZ-Havuzların sayısı, çeşitliliği, HAVUZ-Havuzların temizliği, HİZMET-Eğlence (Animasyon, Etkinlik Çeşitliliği, Canlı Müzik), HİZMET-spa-sauna-Fitness, KONUM VE ÇEVRE-Otel Bahçesi-Doğa güzellikleri, KONUM VE ÇEVRE-Otel Konumu-Ulaşım kolaylığı, ODA-oda boyutu , ODA-oda temizliği (banyo, tuvalet, yatak) , ODA-odanın Olanakları (Duş sorunu, Sıcak soğuk su sorunu, klima sorunu, Sineklik, koltuk) , ODA-Yatak (Yatak Kapasitesi, Yatak Kalitesi) , PERSONEL (Personel ilgisi ve hizmeti, Garson Hizmeti, Servis hızı, resepsiyon hizmetleri) , YEMEK-Fiyat-performans dengesi , YEMEK-Masa Düzeni ve Temizliği , YEMEK-Snack Bar- Bar , YEMEK-Yemek Çeşitliliği , Açık Büfe Çeşitliliği , YEMEK-Yemek Kalitesi, Yemek Lezzeti.) belirlenmiştir. Bu aşamada, kategorilerin belirlenmesi süreci titizlikle yürütülmüş ve otel yorumlarının analizine uygun olacak şekilde yapılandırılmıştır.

Veri Kümesi Oluşturma: Veri kümesinin oluşturulması için, Google Seyahat platformunda yer alan otel yorumları otomatik olarak çıkarılmıştır. Çıkarılan yorumlar, belirlenen kategorilere ve alt kategorilere göre etiketlenmiştir. Yorumların etiketlenmesi sürecinde, her bir yorumun pozitif veya negatif olarak sınıflandırılması sağlanmıştır. Objektif bir etiketleme süreci için, üç farklı kişi tarafından yapılan etiketlemeler birleştirilmiş ve tutarlılık sağlanmıştır. Bu aşama, yorumların kategorilere ayrılmış ve duygusal olarak etiketlenmiş bir veri kümesinin oluşturulmasıyla tamamlanmıştır.

İş Paketi 2: Veri Ön İşleme

Veri Ön İşleme: Veri ön işleme adımı, yorumların analize uygun hale getirilmesi için çeşitli işlemler gerçekleştirilmiştir. Bu işlemler arasında metinlerin normalizasyonu işlemi, duygu belirten emojilerin tespiti işlemi, noktalama işaretlerini kaldırma işlemi, fazla boşlukların kaldırma işlemi, kök bulma işlemi ,metinlerin küçük harfe dönüştürülmesi, özel karakterlerin ve gereksiz boşlukların kaldırılması yer almaktadır. Bu adımlar, metinlerin homojen bir yapıya kavuşmasını sağlamış ve analiz sürecinde daha tutarlı sonuçlar elde edilmesine olanak tanımıştır.

İş Paketi 3: Öznitelik Çıkarımı ve Seçimi

Öznitelik Çıkarımı ve Seçimi: Kategori sınıflandırma ve duygu analizi için öznitelik çıkarımı adımı, "Bag of Words" (Kelime Çantası) yaklaşımı kullanılmıştır. Kelime n-gram ve karakter n-gram yöntemleri uygulanarak, metinlerdeki kelimelerin ve karakterlerin sıklıkları öznitelik olarak kullanılmıştır. Kelime n-gram yöntemi, metinlerdeki kelime ilişkilerini anlamak ve metin sınıflandırmasında etkili sonuçlar elde etmek amacıyla kullanılmıştır. Karakter n-gram yöntemi ise, metinlerdeki ardışık

karakter dizilerini öznitelik olarak kullanarak, metinlerin içerdği anlamı daha derinlemesine analiz etmeye yardımcı olmuştur.

Ayırt Edici Özniteliklerin Seçilmesi: Özniteliklerin sınıflandırma performansına olan etkisini ölçmek amacıyla, "Chi-Square" yöntemi uygulanmıştır. Bu yöntem, yüksek bilgi kazanımı veya düşük p-değerine sahip özellikleri seçerek, sınıflandırma performansını iyileştirmeyi amaçlamıştır.

İş Paketi 4: Kategori Sınıflandırma ve Duygu Analizi

Kategori Sınıflandırma ve Duygu Analizi: Öznitelik çıkarımı ve seçimi adımlarının ardından, kategori sınıflandırma ve duygu analizi gerçekleştirilmiştir. Bu aşamada, SVM (Destek Vektör Makineleri), Lojistik Regresyon ve Random Forest makine öğrenimi algoritmaları kullanılarak, otel yorumları genel ve alt kategorilerde sınıflandırılmıştır. Makine öğrenimi algoritmaları, Python programlama dili ve scikit-learn kütüphanesi kullanılarak uygulanmıştır.

İş Paketi 5: Sonuçların Analizi ve İyileştirilmesi

Sonuçların Analizi ve İyileştirilmesi: Kategori sınıflandırma ve duygu analizi performansı çeşitli ölçütlerle değerlendirilmiştir. Doğruluk ve F-skor değerleri hesaplanmış, hata matrisi analiz edilerek iyileştirme alanları belirlenmiştir. Sonuçların analizi, model performansının optimize edilmesine ve daha doğru sınıflandırma sonuçları elde edilmesine yardımcı olmuştur.

Projenin ikinci aşaması olan duygu analizi aşamasında, kategori sınıflandırma aşamasında uygulanan adımlara benzer işlemler gerçekleştirilmiştir. Ancak, duygu analizi için özel karakterlerin ayrı bir şekilde ele alınması gibi bazı değişiklikler yapılmıştır. Bu aşamada, otel yorumlarının duygusal içeriğinin daha doğru bir şekilde analiz edilmesi sağlanmıştır.

Bu projede gerçekleştirilen çalışmalar, otel yorumlarının kategori düzeyinde duygu analizine yönelik önemli adımlar atılmış ve elde edilen veriler, literatüre ve pratik uygulamalara değerli katkılar sunmuştur.

3. Sonuç

Bu çalışmada proje planının tüm kısımları gerçekleşmiştir. Yapılan çalışmada elde edilen sonuçlar;

1)Genel veri setinin Performans değeri:

	Mean Accuracy	Mean F1 Score
SVM	0.7380781511017624	0.729249552226825
Random Forest	0.7993938061976245	0.7986354418206052
Logistic Regression	0.844497626536173	0.8427158751601574

- **Logistic Regression** en yüksek doğruluk ve F1 skorlarına sahip, ardından **Random Forest** gelirken, **SVM** en düşük performansı sergilemiştir.

2)Tüm alt kategoriler için performans değerleri:

2.1) ODA ALT KATEGORI İçin Performanslar:

	Mean Accuracy	Mean F1 Score
SVM	0.6577460600654177	0.5491865471099479
Random Forest	0.7303597977995837	0.6722930410996424
Logistic Regression	0.7305976806422837	0.6732645389433997

- **Random Forest** ve **Logistic Regression** en yüksek doğruluk ve F1 skorlarına sahiptir, **SVM** ise en düşük performansı göstermiştir.

2.2) HİZMET ALT KATEGORI İçin Performanslar:

	Mean Accuracy	Mean F1 Score
SVM	0.7988385598141695	0.7463768875630501
Random Forest	0.8751451800232288	0.8552480627006578
Logistic Regression	0.8656213704994192	0.8434995476940911

- **Random Forest** en yüksek doğruluk ve F1 skorlarına sahiptir, ardından **Logistic Regression** gelirken, **SVM** en düşük performansı sergilemiştir.

2.3) HAVUZ ALT KATEGORI İçin Performanslar:

	Mean Accuracy	Mean F1 Score
SVM	0.5737728937728938	0.41906846776412
Random Forest	0.7203663003663003	0.661774339572659
Logistic Regression	0.6486446886446886	0.5490393300930385

- **Random Forest** en yüksek doğruluk ve F1 skorlarına sahiptir, **Logistic Regression** onu takip ederken, **SVM** en düşük performansı göstermiştir.

2.4) DENİZ ALT KATEGORİSİ İçin Performanslar:

	Mean Accuracy	Mean F1 Score
SVM	0.7084330484330484	0.6424878214959187
Random Forest	0.7315327635327635	0.6771133628469502
Logistic Regression	0.7700170940170941	0.7355263312795265

- **Logistic Regression**, en yüksek doğruluk ve F1 skorlarına sahiptir, **Random Forest** onu takip ederken, **SVM** en düşük performansı göstermiştir.

2.5) KONUM VE CEVRE ALT KATEGORI İçin Performanslar:

	Mean Accuracy	Mean F1 Score
SVM	0.7898987854251013	0.7054909735698329
Random Forest	0.7898987854251013	0.7054909735698329
Logistic Regression	0.8105533063427799	0.7486865762017557

- **Logistic Regression**, en yüksek doğruluk ve F1 skorlarına sahiptir, **SVM** ve **Random Forest** benzer performans göstermiştir.

2.6) YEMEK ALT KATEGORI İin Performanslar:

	Mean Accuracy	Mean F1 Score
SVM	0.6106842105480685	0.5201650543955392
Random Forest	0.7090777205536953	0.670736625434501
Logistic Regression	0.7167652784419325	0.6767044839223458

- **Logistic Regression**, en yksek doęruluk ve F1 skorlarına sahiptir, **Random Forest** onu takip ederken, **SVM** en dřk performansı gstermiřtir.

3) Tm alt kategoriler İin Performanslar (Personel eklenmiř hali):

	Mean Accuracy	Mean F1 Score
SVM	0.6129949551603548	0.5616602749367596
Random Forest	0.7117676540476368	0.6830356153663334
Logistic Regression	0.7476847393219567	0.7198995854941417

- **Logistic Regression** en yksek doęruluk ve F1 skorlarına sahiptir, ardından **Random Forest** gelirken, **SVM** en dřk performansı sergilemiřtir.

Genel Deęerlendirme:

Genel olarak, Logistic Regression tm alt kategorilerde ve personel eklenmiř haliyle en yksek performansı gstermiřtir. Random Forest ise genellikle ikinci en iyi performansı sergilemiřtir, zellikle deniz alt kategorisinde belirgin bir bařarı gstermiřtir. SVM ise genellikle en dřk performansı sergilemiřtir, zellikle oda alt kategorisi iin belirgin bir řekilde geride kalmıřtır.

4. ıktılar (Yayınlar,sunumlar v.b.)

Bu proje kapsamında makine ęrenimi ile otel yorumları zerinde duygu analizinin gerekleřtirilmesi ve kategori dzeyinde zetlenmesi gerekleřtirilerek bařarılı sonular elde edildięinden dolayı 2023 / 1 dnemi TBİTAK 2209 - niversite ęrencileri Arařtırma Proje Yarıřmalarına bařvuru yapılmıřtır.

5. Proje ile ilgili harcama kalemleri hakkında ayrıntılı bilgi

2209/A Üniversite Öğrencileri Araştırma Projeleri Destek Programı kapsamında proje desteği ile;

1)Çalışmanın gerçekleştirilmesi kırtasiye malzemeleri ve lazer kumanda alımı yapılmıştır. Bu malzemeler şunlardır:

- Kalem
- Ajanda
- Lazer kumanda

Toplam harcama **3.000,00 ₺** olarak gerçekleştirilmiştir.

2)Alımı yapılan ekipmanların ve kırtasiye malzemelerinin EK-1, EK-2’te faturaları sunulmuştur.

EK-1

(buraya gelecek)

EK-2

(buraya gelecek)

6. KAYNAKLAR

- [1]. Arıca, R., & Çorbacı, A. (2019). Turizm sektöründe müşterilerin bilgi üretici rolü: Adıyaman'daki turistik çekiciliklere yönelik Tripadvisor sitesinde yer alan yorumlar üzerine bir araştırma. *Seyahat ve Otel İşletmeciliği Dergisi*, 16(3), 437-455.
- [2]. Ekinci, E., & Omurca, S. İ. (2017). Ürün özelliklerinin konu modelleme yöntemi ile çıkartılması. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 9(1), 51-58.
- [3]. İnan, H. E. (2019). Otel Yorumlarının Sınıflandırılmasında Makine Öğrenmesi Algoritmalarının Karşılaştırılması ve Eskişehir Otellerine Yönelik Yorumların Duygu Analizi. *20. Ulusal ve 4. Uluslararası Turizm Kongresi*, 16-19.
- [4]. Kaynar, O., Görmez, Y., Yıldız, M., & Albayrak, A. (2016, September). Makine öğrenmesi yöntemleri ile Duygu Analizi. In *International Artificial Intelligence and Data Processing Symposium (IDAP'16)* (Vol. 17, No. 18, pp. 17-18).
- [5]. Özyurt, B., & Akçayol, M. A. (2018). Fikir Madenciliği ve Duygu Analizi, Yaklaşımlar, Yöntemler Üzerine Bir Araştırma. *Selçuk Üniversitesi Mühendislik, Bilim ve Teknoloji Dergisi*, 6(4), 668-693.

[6]. Polat, H., & Yılmaz, A. Ğ. C. A. (2022). Tripadvisor kullanıcılarının türkçe ve ingilizce yorumları kapsamında duygu analizi yöntemlerinin karşılaştırılmalı analizi. *Abant Sosyal Bilimler Dergisi*, 22(2), 901-916.

[7]. Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.

[8]. Sezgin, M., & Duman, A. (2023). Elektronik Ağızdan Ağıza Pazarlama Kapsamında Konaklama İşletmelerine Yönelik Çevrimiçi Yorumların Duygu Analizi Yöntemiyle İncelenmesi: Alanya Örneği. *Türk Turizm Araştırmaları Dergisi*, 7(2), 244-265.

[9]. Şahin, H., Çolakoğlu, O. E., Kılıçdere, S., & Hakan, A. T. A. Y. (2020). Kuşadası'nda bulunan 5 yıldızlı otel işletmelerine yönelik yapılan sosyal medya yorumlarının incelenmesi. *Tourism and Recreation*, 2(Ek 1), 103-110.

[10]. Uslu, O. ve Akyol, S. (2021). Türkçe Haber Metinlerinin Makine Öğrenmesi Yöntemleri Kullanılarak Sınıflandırılması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 2(1), 15-20.

PROJE YÜRÜTÜCÜSÜNÜN ADI – SOYADI- İMZA	DANIŞMANIN ADI – SOYADI- İMZA
Beyza ÇELİK	Asst. Prof. Üyesi Muhammed Yasin PAK
Semiha ÖZCAN	Asst. Prof. Üyesi Muhammed Yasin PAK

Tarih:29.05.2024