

# Kocaeli Üniversitesi

## Programlama Laboratuvarı Dersi 3.Proje Ödevi

1. Hannenur Dursun - 220202109  
Bilgisayar Mühendisliği 2.Öğretim  
Kocaeli Üniversitesi  
Umuttepe, İzmit, Kocaeli, Türkiye  
[hannedrsn@gmail.com](mailto:hannedrsn@gmail.com)

2. Ramazan Mert Atuş - 210202096  
Bilgisayar Mühendisliği 2.Öğretim  
Kocaeli Üniversitesi  
Umuttepe, İzmit, Kocaeli, Türkiye  
[mertatug93@gmail.com](mailto:mertatug93@gmail.com)

**Özetçe**—Bu proje, Twitter API aracılığıyla kullanıcı verilerini çekerek, hash tabloları ve graf analizleri kullanarak kullanıcıları benzer ilgi alanlarına göre eşleştirmeyi amaçlamaktadır. Veri yapıları, graf teorisi, hash tabloları, arama algoritmaları, API entegrasyonu, veri analizi ve ilişki yönetimi gibi konularda öğrencilere deneyim kazandırmayı hedefleyen projede, öğrencilerin problem çözme ve analitik düşünme becerilerini geliştirmeleri amaçlanmaktadır.

**Anahtar Kelimeler** — Veri Analizi, Hash Tabloları, Graf Analizi, Veri Yapıları, Arama Algoritmaları

### I. GİRİŞ

Bu projenin amacı, Twitter API aracılığıyla kullanıcı verilerini çekmek ve bu verileri analiz ederek, kullanıcıları benzer ilgi alanlarına göre eşleştirmektir. Twitter, günümüzde milyonlarca kullanıcı tarafından aktif olarak kullanılan bir sosyal medya platformudur. Bu platform üzerinden elde edilen veriler, çeşitli veri yapıları, graf teorisi, hash tabloları ve arama algoritmaları kullanılarak organize edilecek ve analiz edilecektir. Projenin genel hedefi, öğrencilere veri yapıları, graf teorisi, hash tabloları, arama algoritmaları, API entegrasyonu, veri analizi ve ilişki yönetimi gibi konularda deneyim kazandırmaktır. Bu sayede öğrenciler, teorik bilgilerini pratiğe dönüştürerek problem çözme ve analitik düşünme becerilerini geliştireceklerdir.

Projede kullanıcı verileri Twitter API aracılığıyla çekilecek ve her bir kullanıcı için belirli bilgiler içeren bir nesne oluşturulacaktır. Bu nesneler hash tabloları ile organize edilerek, veri yapıları kullanılarak işlemler gerçekleştirilecektir. Kullanıcılar arasındaki takipçi-takip edilen ilişkileri temsil etmek üzere bir graf oluşturulacak ve ilgi alanlarına göre eşleştirmeler yapılacaktır. İlgi alanına göre eşleme sürecinde hash tabloları ve arama algoritmaları kullanılarak benzer ilgi alanlarına sahip kullanıcılar belirlenecek ve bu kullanıcılar arasındaki ilişkiler analiz edilecektir. Ayrıca, belirli bir bölge ve dil için trend olan hashtagler ve konular listelenecek, tweet içeriklerindeki anahtar kelimeler ve hashtagler DFS algoritması kullanılarak tespit edilecektir.

Sonuç olarak, projenin giriş kısmında, Twitter API kullanarak elde edilen kullanıcı verilerinin nasıl organize edileceği ve ilgi alanlarına göre nasıl eşleştirmeler yapılacağı üzerinde odaklanılacaktır. Bu adımlar, öğrencilere veri analizi, graf analizleri, hash tabloları ve arama algoritmaları konularında pratik deneyim kazandıracaktır.

### II. YÖNTEM

Bu proje Visual Studio Code üzerinde Python programlama dili kullanılarak oluşturulmuştur.

Projenin başlangıcında tweeter veriler “twitter\_data.json” adlı dosyadan okunur. Okuma işlemi akabinde projenin isterleri doğrultusunda kullanıcı nesneleri oluşturulur. Her bir kullanıcı verileri HashTablosu adlı fonksiyon ile birlikte hash tablosu içerisinde tutulur. Bu işlemden sonra NetworkX kütüphanesi kullanılarak kullanıcılar arasındaki ilişkiler graf olarak modellenmiştir. Daha sonra kullanıcıların tweet içeriklerinden ilgi alanı analizi yapabilmek için Türkçe doğal dil işleme kütüphaneleri (nltk, snowballstemmer) indirilir ve kelimelerin frekans analiz işlemi gerçekleştirilir. Bunun için tüm tweetler birleştirilerek analiz işlemine başlanır. Kelimeler küçük harfe çevrilir. Verilen içinden stopwords’leri çıkartmak için HashTable2 yapısının içinde hash tablosunda stopwords’ler tutulur. Frekans analizi yapılan tweetlerden stopwords’ler çıkartılır, Türkçe stemleme işlemi uygulanır. Bu kelimeler üzerinden Türkçe doğal dil işleme ile sadece isim (NN) kategorisindeki kelimeler filtelenir. Kullanıcıların ilgi alanları belirlenir ve istatistikler çıkarılır. Hangi kullanıcı hangi ilgi alanıyla ilgilendiğini belirtmek için yeni bir hash tablosu oluşturulur ve ilgi alanları ile kullanıcı adları (username) eşleştirilir.

Bundan sonra sadece iki kullanıcı üzerinden işlemle gerçekleştirilir. İki kullanıcının ilgi alanı benzerlik ölçülür. Ortak ilgi alanları, takipçi ve takip ettikleri üzerinden benzerlik analizi yapılır. Kullanıcıların bölge ve dil bazında en çok kullandığı kelimeler belirlenir, ilgili istatistikler çıkarılır. Kullanıcıların tweet içeriklerinde belirli anahtar kelimeleri ve hashtagleri içeren tweetler DFS algoritması kullanarak listelenir ve graf üzerinde gösterilir. BFS algoritması kullanılarak, ilgi alanlarının benzerliği temel alınarak

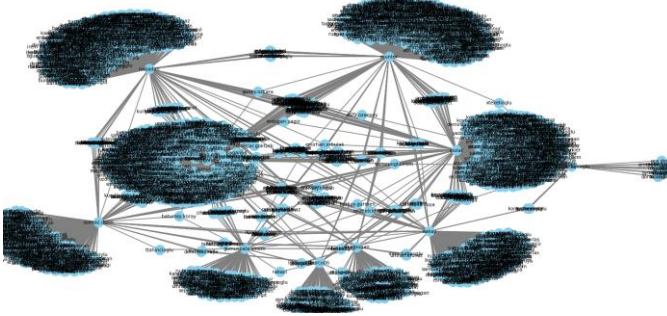
kullanıcılar arasında bir bağlantı olup olmadığı belirlenir ve BFS algoritmasının çıktısı arayüzde gösterilir. Minimum Spanning Tree algoritması kullanılarak, benzer ilgi alanlarına sahip kullanıcılar arasında en az bağlantı ile ağaç oluşturulur.

### III. SONUÇ

Bu projenin amacı, Twitter API aracılığıyla elde edilen kullanıcı verilerini kullanarak, benzer ilgi alanlarına sahip kullanıcıları eşleştirmek ve ilişkilendirmektir. Veri yapıları, graf analizleri, hash tabloları ve arama algoritmalarının kullanıldığı bu proje, öğrencilere geniş bir yelpazede bilgi ve deneyim sunmayı amaçlamaktadır. Projenin sonuçları, kullanıcıların benzer ilgi alanlarına dayalı olarak birbirleriyle olan ilişkilerini anlamak ve analiz etmek adına etkileyici bulgular ortaya koymaktadır. Kullanıcılar arasındaki takipçi-takip edilen ilişkileri graf olarak modelleyerek, bu ilişkileri derinlemesine incelemek ve benzer ilgi alanlarına sahip grupları belirlemek mümkün olmuştur. İlgi alanlarına göre eşleştirmelerin yapıldığı bu proje, sosyal medya platformlarında daha anlamlı ve nitelikli etkileşimlerin sağlanmasına katkı sağlayabilir. Belirli bölge ve dilde trend olan hashtaglerin listelenmesi, kullanıcıların tweet içeriklerindeki anahtar kelimelerin ve hashtaglerin analizi gibi detaylar, kullanıcıların ilgi alanlarına daha fazla ışık tutmaktadır. Sonuç olarak, bu projenin öğrencilere sağladığı deneyim, veri analizi, algoritmalar, graf teorisi ve ilişkisel veri yönetimi alanlarında derinlemesine bilgi edinmelerine ve gerçek dünya problemlerini çözmelerine olanak tanımaktadır. Sosyal medya verilerinin etkin bir şekilde kullanılması, kullanıcılar arasındaki ilişkilerin daha iyi anlaşılmasına ve sosyal medya platformlarının potansiyelinin daha iyi değerlendirilmesine yönelik gelecekteki çalışmalara ilham kaynağı olabilir.

### IV. DENEYSEL SONUÇLAR

#### 1. 10 kullanıcı için takip-takipçi grafi



#### 2. İki kullanıcının ortak takip ettikleri

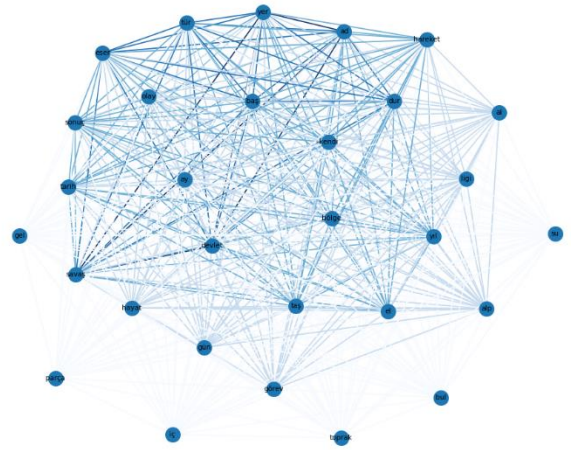
```
burcu06 ve eakay kullanıcılarının ortak takip ettikleri:  
tutuncu.kutay  
efe.dogan  
ebru.cetiner  
kutay.sozeri  
ece34
```

Takip benzerlik yüzdesi= %5.556

#### 3. İlk 10 ilgi alanları

```
Tüm kullanıcıların en yüksek frekanslı ilk 10 ismi  
ak: 34  
bölge: 31  
ad: 31  
savaş: 31  
yer: 30  
dil: 30  
tür: 29  
yıl: 29  
grup: 29  
eser: 29  
birlik: 28  
baş: 28  
hareket: 28  
merkez: 27  
film: 27  
motor: 27  
kitap: 27  
model: 27  
devlet: 26  
iş: 26
```

#### 4. İlgi alanlarının DFS algoritması ile gösterimi

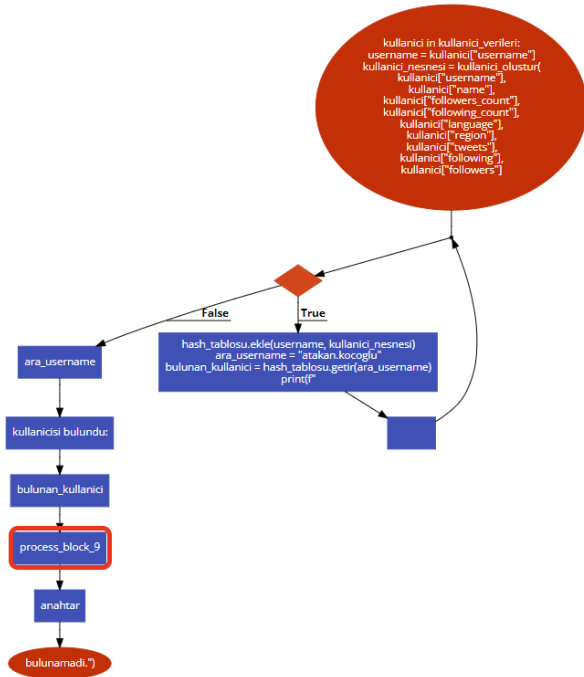


## KAYNAKLAR

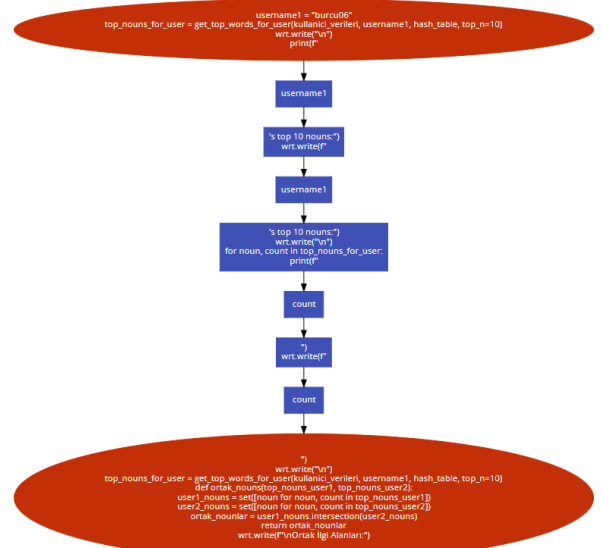
- [1] <https://burhanbilen.medium.com/python-nltk-kütüphanesi-kullanımı-7ab91b19f201>.
- [2] <https://medium.com/@ayhanbzkr/đođal-dil-iřleme-nlp-natural-language-processing-8a9493a4cfee>
- [3] <https://ichi.pro/tr/python-da-metin-madenciligi-adimlar-ve-ornekler-132714371846459>
- [4] <https://raw.githubusercontent.com/stopwords-iso/stopwords-tr/master/stopwords-tr.json>
- [5] <https://wmaraci.com/nedir/stop-words>
- [6] <https://medium.com/@Kolay.Zeka/đođal-dil-iřleme-nltk-spacy-ve-metin-sınıflandırma-metin-analizi-dil-modelleri-adlı-blog-689dfce21e02>

## V. AKIř DİYAGRAMLARI

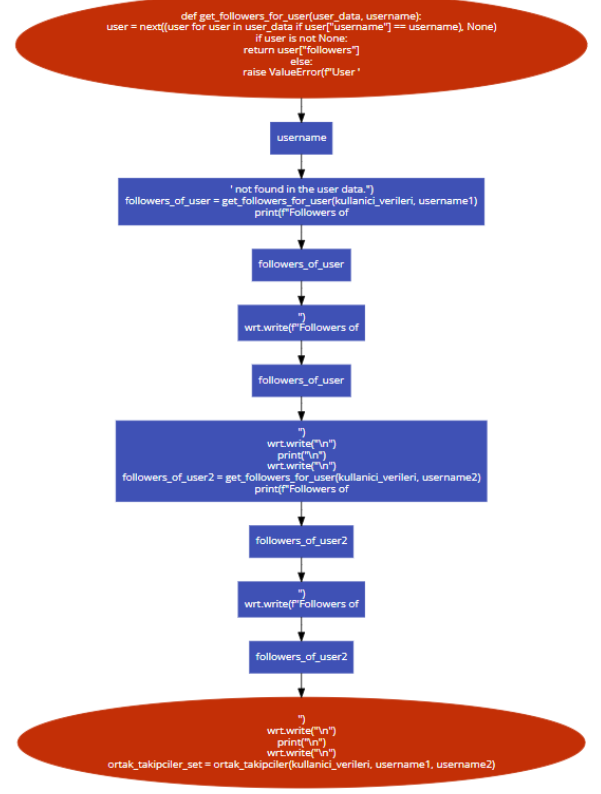
### 1. Kullanıcı ismini hash tablosundan alınması



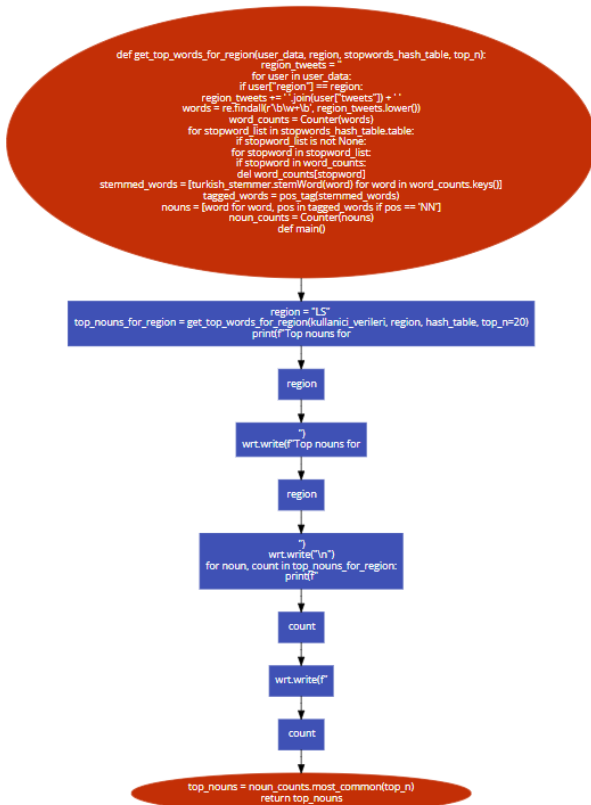
### 2. Kullanıcının ilgi alanlarını bulma



### 3. Kullanıcının takipçi listesini alma



4. Belirli bir bölgenin ilgi alanlarını listeleme



## UML DİYAGRAMI

