

GİRİŞ

Birbirimizle ve çevremizle iletişim şeklimizi değiştiren ve her geçen gün kullanıcı sayısı artan sosyal ağlar, yapılan paylaşım ve aktiviteler ile çok fazla verinin ortaya çıkmasını sağlayan ortamlardır. Sosyal ağ madenciliği ile bu ortamda farklı biçimlerde ve büyük miktardaki verinin, veri madenciliği yöntemleri ile analiz edilmesi ve anlamlandırılması mümkün olmaktadır. Bu çalışmada sosyal ağ madenciliği uygulaması kapsamında Twitter ve WhatsApp verileri üzerinde R programlama ortamında yapılan analiz süreci ve sonuçları açıklanmaktadır.

GERÇEKLEŞTİRİLEN UYGULAMALAR

Twitter Veri Analizi

Geliştirilen yazılım aracılığı ile belirli bir tarih aralığına, atılan tweet'in diline ve özel etiket (hashtag)'e göre arama yapılması.

Gerçekleştirilen uygulama ile 01.07.2016 ve 01.07.2017 tarihleri arasında paylaşılan, içinde "samsung" ve "apple" kelimelerinin geçtiği, 1000 adet tweet program aracılığı ile iki ayrı veri setine kaydedilmiştir. Tweet'lerden oluşturulan doküman koleksiyonları içerisinde küçük harfe dönüştürme işlemi, noktalama işaretlerinin kaldırılması sağlanmıştır.

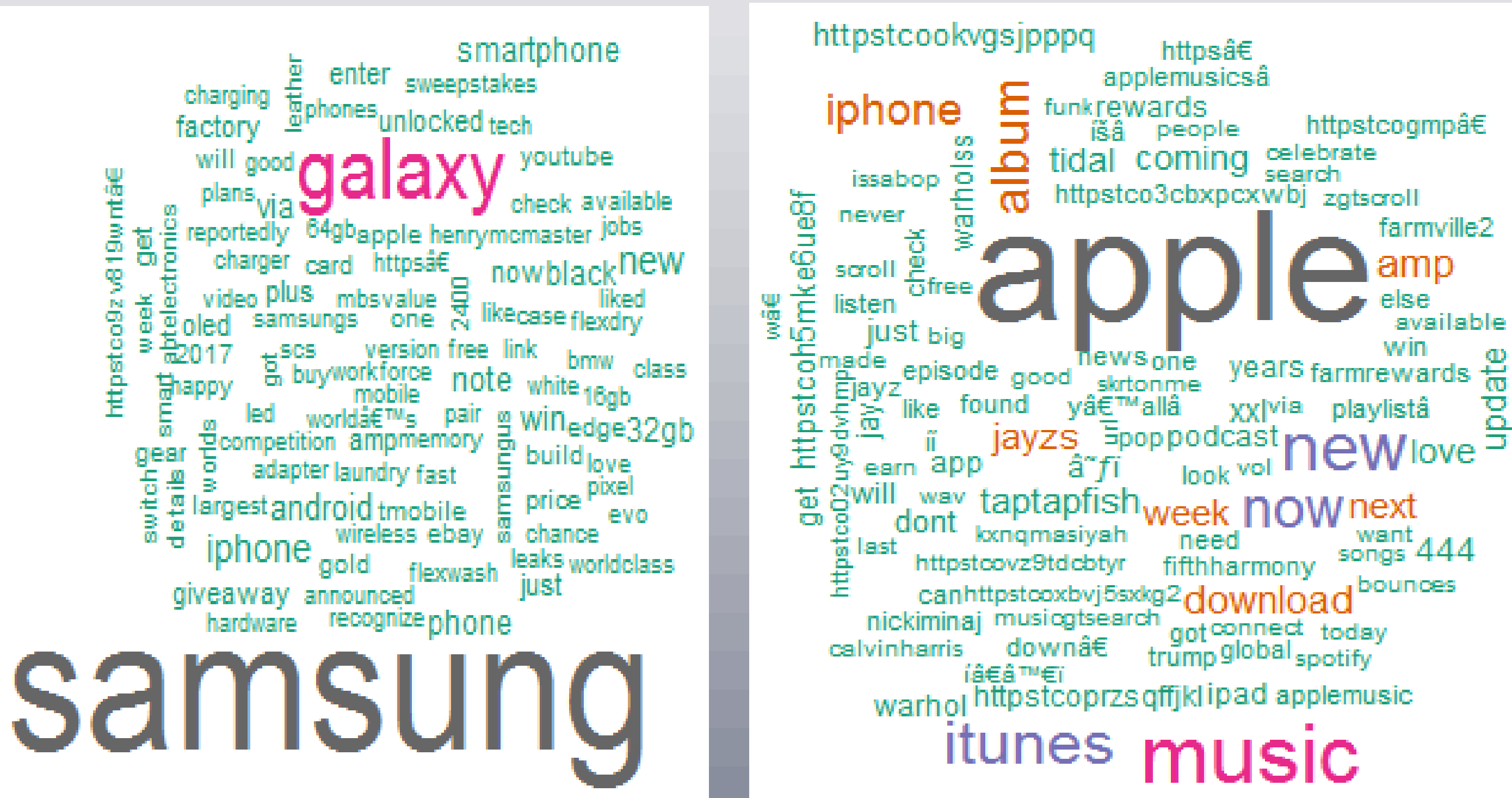
TermDocumentMatrix fonksiyonu ile tweet'lerden alınan veri seti için bir terim belge matrisinin oluşturulması sağlanır.

Gerçekleştirilen uygulamada hiyerarşik kümeleme kullanılmıştır. Hiyerarşik kümeleme işlemini gerçekleştirmek için hclust fonksiyonu kullanılmış ve kümeleme sonucunda elde edilen gözlemler Şekil 1'deki gibi dendrogram üzerinde gösterilmiştir.



Şekil 1. Samsung Kelimesi İçin En Sık Kullanılan Kelimelerin Küme Dendrogramı

Veri seti içerisindeki kelimelerin frekans değerlerine göre büyük, küçük ve renkli gibi çeşitli niteliklerin kullanılması ile kelimelerin bir bulut olarak görselleştirilmesi işlemine kelime bulutu çıkarım işlemi denir. Kelime bulutu sayesinde elverişsiz ve ham veriyi daha anlamlı bir hale getirmek mümkündür. R programlama ortamında Wordcloud komutu ile belge matrisinde frekans değeri istenilen değerin üzerinde olan kelimelerin yer alacağı, belirlenen sayıda kelime için kelime bulutunun oluşturulması sağlanabilmektedir. Şekil 2 (a)'da Samsung kelimesi için ve Şekil 2 (b)'de ise Apple kelimesi için minimum frekansı 3 olan 100 adet olacak şekilde oluşturulan kelime bulutlarına ait ekran görüntüleri görülmektedir.



Şekil 2. Samsung ve apple Kelimeleri İçin Kelime Bulutu

Bu Poster aşağıda belirtilen dergide makale olarak yayınlanmıştır :

DOĞAN, Buket; SARIKAYA, Ufuk; AKTAŞ, Abdulsamet. R ile Sosyal Ağ Madenciliği. Marmara Fen Bilimleri Dergisi, 2017, 29.3: 94-101.

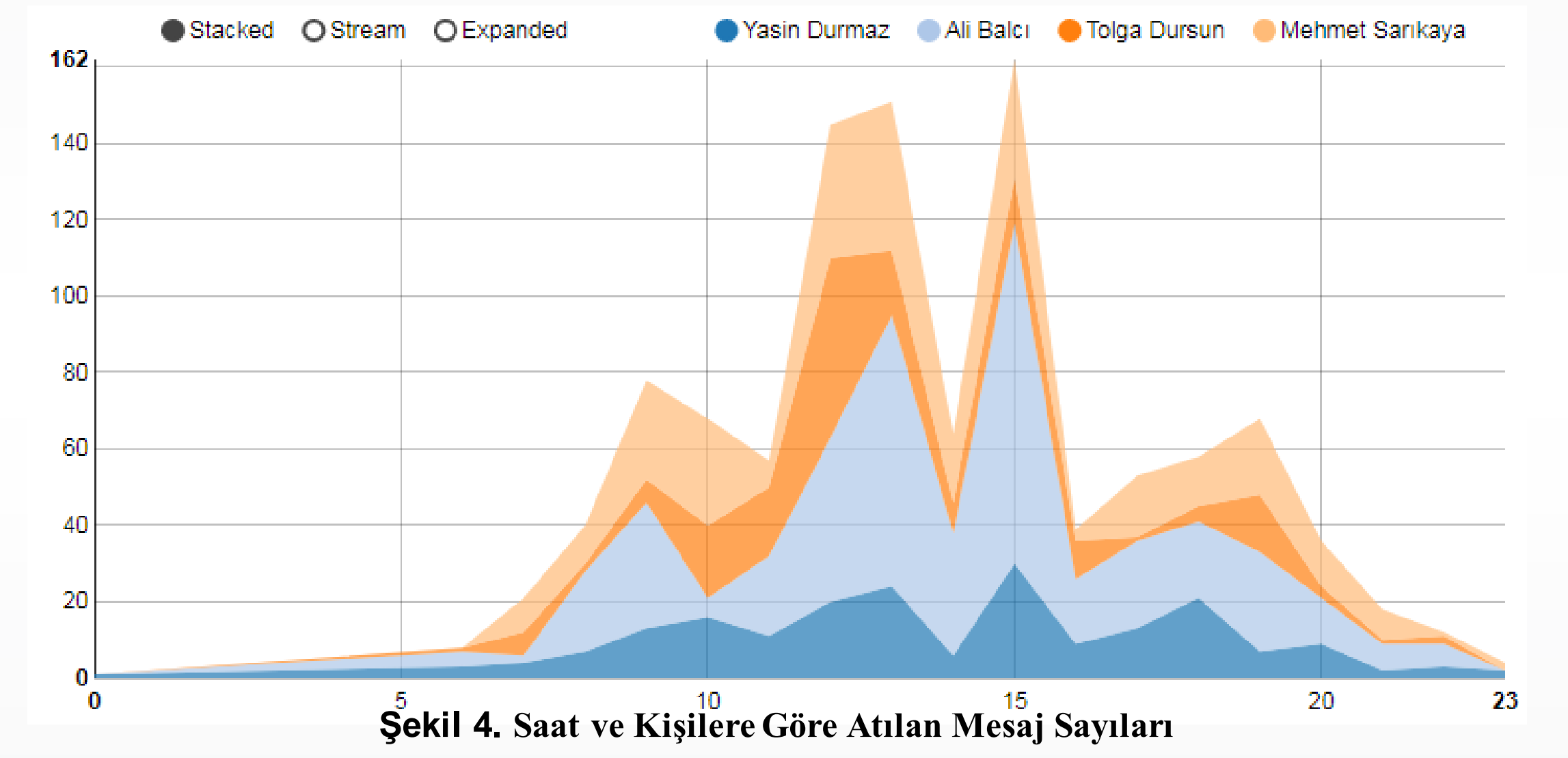
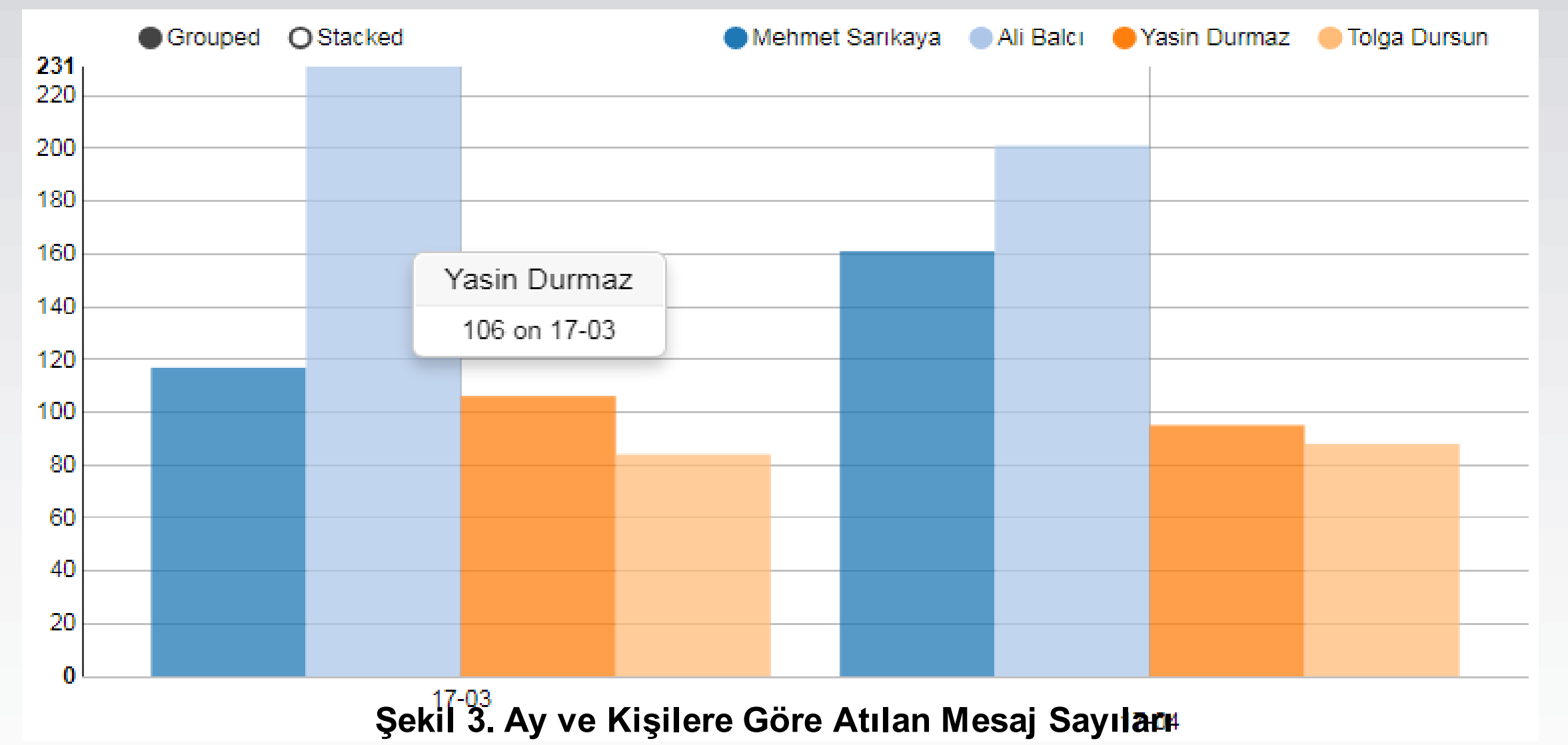
WhatsApp Veri Analizi

Gerçekleştirilen çalışmada özellikle grup mesajlarında hangi kullanıcının, hangi tarihte ne sıklıkla mesaj gönderdiği ve aktif olduğu, mesajların zamana bağlı olarak gösterdikleri dağılımı incelemek için bir uygulama geliştirilmiştir.

Bu işlem için öncelikle sohbet geçmişinin metin dosyası biçimine dönüştürülerek kayıt edilmesi sağlanmıştır.

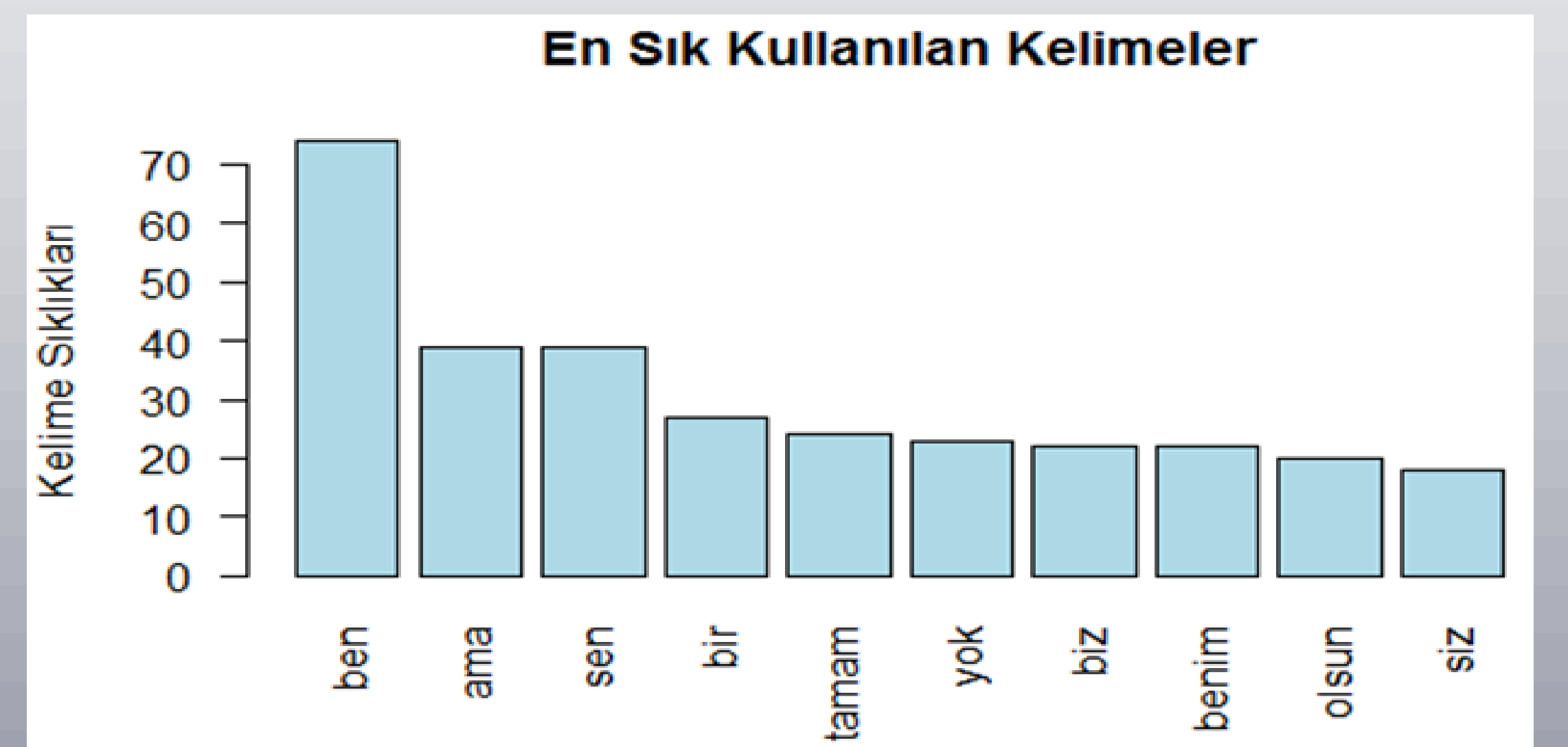
Ardından gönderilen metin dosyası üzerinden, sohbetlerde kimin ne kadar mesaj attığını, hangi ay, gün ve saat dilimlerinde ne sıklıkla sohbet edildiğini, duygu analizini, sohbette en sık kullanılan kelimelerin grafik olarak gösterim işlemi R programlama ile gerçekleştirilmiştir.

Metin analizi iki aşamada gerçekleştirilmektedir. Öncelikle verinin temizlenmesi ardından görselleştirme yapılması gerekmektedir.



Şekil 3'de ise grup sohbetindeki 4 kişinin gönderdikleri mesaj sayılarının yılın aylarına göre dağılımı çubuk grafik halinde gösterilmiştir.

Şekil 4'de ise grup sohbetindeki 4 kişinin gönderdikleri mesaj sayılarının günün saatlerine göre dağılımı yığın (stack) grafik halinde gösterilmiştir. Grafığe göre grup üyeleri arasındaki en çok mesajlaşmanın saat 15:00 civarında yapıldığı görülmektedir. Saat 15:00'dan sonra mesajlaşma sayısında ciddi bir düşüş gözlemlenmektedir. Şekil 5'deki grafikte, grup sohbetinde kullanılan kelimelerden en sık kullanılan ilk 10 kelime ve kullanılma sayıları yer almaktadır. Grafığe göre, en sık kullanılan kelime 74 defa ile "ben" kelimesi olmuştur.



Şekil 5. Mesajlarda En Sık Kullanılan Kelimeler ve Kullanılma Sıklıkları

SONUÇLAR

R programlama ortamı sosyal ağ analizi için etkin, hızlı ve esnek bir program geliştirme ortamı sağlamaktadır. Bu ortam aracılığı ile Twitter ve WhatsApp ortamlarından elde edilen verilerin analizi ve görselleştirilmesi bu çalışma kapsamında açıklanmakta ve sonuçları ortaya konmaktadır. Farklı zaman aralıkları, anahtar kelimeler ve amaçlar için bu ortamın kullanılması mümkündür. Sosyal ağlarda yer alan ürün yorumları kullanılarak buna göre satış ve pazarlama stratejilerinin geliştirilmesi sonraki çalışma planları arasında yer almaktadır.