



## TÜRKÇE BİLGİSAYARLI DİL BİLİMİ ÇALIŞMALARINDA HİS ANALİZİ

Alaettin UÇAN\* - Ebru AKÇAPINAR SEZER\*\*

### Özet

Bilgisayarlı dil bilimi; sözlü ya da yazılı dili anlamayı, matematiksel olarak ifade etmeyi hedefleyen ve bu hedefe ulaşmak için yöntemler, modeller ve araçlar öneren disiplinler arası bir bilim dalıdır. Bilgisayarlı dil bilimi çalışmalarının bir araştırma alanı olan his analizi; ses, görüntü ya da metin içerisinde hangi hislerin ne oranda yer aldığını bulma işlemine verilen addır. İnternetin yaygınlaşması, sayısal içeriğin çoğalması, saklama ve hesaplama gücünün artması gibi gelişmeler hem otomatik his analizi yapmanın önünü açmış hem de his analizini bir gereklilik hâline getirmiştir.

Metinlerde his analizi konusunu Türk dili özelinde özetleyen bu çalışma, öncelikle his analizinin tarihçesi ve önemini dil bilim bakışı açısıyla açıklamayı hedeflemekte ve his analizinin güncel uygulama alanlarından kısaca bahsetmektedir.

**Anahtar Kelimeler:** Bilgisayarlı dil bilimi, metinlerde his analizi.

### EMOTION ANALYSIS IN TURKISH COMPUTATIONAL LINGUISTICS STUDIES

### Abstract

Computational linguistics is an interdisciplinary field that aims to understand the verbal or written language, to express it mathematically and suggest methods, models and tools to achieve these goals. Emotion

\* Arş. Gör., Hacettepe Üniversitesi Bilgisayar Mühendisliği Bölümü, [aucan@cs.hacettepe.edu.tr](mailto:aucan@cs.hacettepe.edu.tr), ORCID: 0000-0002-2493-4022.

\*\* Prof. Dr., Hacettepe Üniversitesi Bilgisayar Mühendisliği Bölümü, [ebruakcapinarsezer@gmail.com](mailto:ebruakcapinarsezer@gmail.com), ORCID: 0000-0002-9287-2679.



analysis, a research area of computational linguistics; is the process of finding which feelings taking place in what proportion in sound, image or text data. The developments such as the proliferation of the Internet, the increase of digital content, the increase of storage and computing power have both paved the way for automatic emotion analysis and made emotion analysis an important need.

This study, which summarizes the subject of emotion analysis, aims to explain the history and importance of emotion analysis from a linguistic perspective, and briefly introduce the current application areas of emotion analysis.

**Keywords:** Computational linguistics, emotion analysis in texts.

## Giriş

Bilgisayarlı (hesaplamalı) dil bilimi; sözlü ya da yazılı dili anlamayı, matematiksel olarak ifade etmeyi hedefleyen ve bu hedefe ulaşmak için yöntemler, modeller ve araçlar öneren disiplinler arası bir bilim dalıdır. Amerika Birleşik Devletleri'nin Rusça bilimsel makaleleri otomatik olarak İngilizceye çeviri ihtiyacıyla 1950'li yıllarda başlayan bilgisayarlı dil bilimi çalışmaları günümüzde Türkçenin de arasında olduğu birçok dil için sürdürülmektedir.

Alandaki çalışmalar, daha çok bilgisayar bilimlerinin bir alt dalı olan doğal dil işleme disiplininden araştırmacılar tarafından yapılmaktadır. Araştırmacılar bilgisayarla bir dili modellemek için ya da başka bir deyişle bilgisayara o dili öğretmek için uzmanlar tarafından yazılan kurallardan faydalanmanın yeterli olacağını düşünmüşler ancak dilin esnek ve canlı yapısı sebebiyle bu düşüncenin yetersiz olduğunu kısa sürede fark etmişlerdir. Araştırmacılar dili bilgisayara sabit kurallarla ezberletmek yerine, tıpkı insana öğretir gibi bilgisayara da öğretmenin bir yolunu bulmaya çalışmaktadırlar. İnsanlar öğrenme kabiliyeti olan tüm canlılar gibi, bir olguyu birçok kez deneyimleyerek öğrenir. Bilgisayarın bir dili öğrenmesi için o dilde birçok doğru örneğe ihtiyaç vardır. Bu sebeple alandaki ilk çalışmalar gelecek vadetse de uzunca bir süre gelişim sağlanamamıştır. İnternetin yaygınlaşması, sayısal içeriğin çoğalması, saklama ve hesaplama gücünün artması gibi gelişmeler sayesinde son yıllarda bilgisayarlı dil bilimi çalışmaları oldukça başarılı sonuçlar üretmektedir.

Bilgisayarlı dil bilimi çalışmalarında yaşanan gelişmelerin yansıması olarak; cümle sınırlarını belirleyebilen, ifadeleri ayırabilen, kelime kök ve gövdesini bulabilen, morfolojik analiz yapabilen, anlamsal analiz yapabilen, özet çıkarabilen, konu belirleyebilen veya başka bir dile çeviri yapabilen bir dil modeli oluşturulabilmektedir. Sonuç olarak araştırmacılar metnin anlamını bilgisayarla modellemekte oldukça başarılı sonuçlar elde edebilmişlerdir. Ancak nesnel içeriği modellemek için yeterli olan anlam bilgisi öznel içerik söz



konusu olduğunda yetersiz kalmaktadır. Çünkü ifadelerin taşıdıkları his bağlama göre değişim göstermektedir.

Öznel içeriğin bağlama ve yazarın hissiyatına göre değişmesi problemini çözmek isteyen bilim insanları ifadelerin anlamının ötesinde barındırdıkları duygu ve hisleri tespit etmeyi amaçlamışlardır (Turney, 2002). İlk başlarda kabaca ifadeleri “olumlu” ve “olumsuz” sınıflayabilen duygu analizi çalışmaları yapılmış (Pang, Lee ve Vaithyanathan, 2002), sonraki çalışmalarda “çok olumlu, olumlu, tarafsız, olumsuz, çok olumsuz” gibi daha ayrıntılı sınıflama (Wiebe, Wilson ve Cardie, 2005) yapılmıştır. Bununla da yetinmeyen araştırmacılar metinlerin barındırdığı “neşe, öfke, korku, üzüntü” gibi hisleri otomatik belirlemek amacıyla his analizi çalışmaları (Aman ve Szpakowicz, 2007) yapmışlardır.

Güncelliğini koruyan his analizi konusunu Türk dili özelinde özetleyen bu çalışma, öncelikle his analizinin tarihçesi ve önemini dil bilimi bakış açısıyla açıklamayı hedeflemekte ve his analizinin güncel uygulama alanlarından kısaca bahsetmektedir. Yazının devamında; his analizi yapmak için kullanılan doğal dil işleme ve öğrenme yöntemleri anlatılmaktadır. Örnek bir his analizi çalışması özetlenmekte ve Türkçe his analizi alanında yapılmış ve yapılması planlanan çalışmalardan bahsedilmektedir.

### **His Analizinin Tarihçesi ve Önemi**

His analizi ses, görüntü ya da metin içerisinde hangi hislerin ne oranda yer aldığını bulma işlemine verilen addır (Aman ve Szpakowicz, 2007). Metin içerisinde hangi hislerin ne oranda yer aldığı, bir hissin yer alıp almadığı, yazarın hangi hisle metni yazdığı, okuyucunun ne hissettiği, yazarın belirli bir konuda ne hissettiği, metinde yer alan örtülü hislerin ne olduğu, iç içe geçmiş hislerin neler olduğu gibi çok farklı sorulara cevap aramak için his analizi gerçekleştirilmektedir. İnsanlar arası iletişimde hisler iletişimin gidişatını şekillendiren ana unsurlardan birisidir. Hisler karşılıklı konuşma esnasında ses tonu, jest ve mimiklerden belirgin şekilde fark edilebilirken yazılı iletişimde yazarın hangi hislerle metni yazdığını belirlemek çoğu zaman bir insan için dahi zordur. Çünkü metinler bağlama göre anlamı ve yansıttığı duygusu farklı olan ifadelerden oluşmaktadır (Naderalvojud, Ucan ve Akcapinar Sezer, 2018). Bu kelime ve ifadelerin yansıttığı hisler belirlenebilirse cümlelerin, paragrafın ve bunlara bağlı olarak metnin yansıttığı hisler oransal olarak bulunabilir.

Metnin yazarı; seçmen, müşteri, okuyucu veya taraftar olduğunda metinde yer alan hisler okuyan için oldukça kıymetlidir. Bir siyasi aday, seçmenin; bir şirket, müşterisinin; bir yazar, okurunun memnuniyet oranını öğrenmek için anketlere ihtiyaç duyar. Oysa sosyal medyanın bu denli yaygınlaşmış olması ve internetin hayatın bir parçası hâline gelmesiyle insanlar görüş ve



hislerini sosyal medya üzerinden ifade etmektedirler. Doğal ortamında kullanıcının kendi isteğiyle yazdığı bir metnin yansıttığı hisler otomatik olarak belirlenebilmektedir. Bu sebeple metin içerisinde otomatik olarak his analizi yapmak oldukça değerlidir.

His analizi çalışmalarında his sınıfları belirlenirken hislerin evrensel hisler olmasına özen gösterilmektedir. Hislerin sınıflandırılması ve temel hislerin belirlenmesi için psikoloji ve sosyoloji alanında birçok çalışma yapılmıştır. Yapılan geniş çaplı çalışmaların (Ekman, 1972; Ekman ve Friesen, 1976) sonucunda temel hislerin yüz ifadelerinin (Şekil 1) öğrenilmediği, doğuştan bilindiği ve bu nedenle evrensel olduğu tespit edilmiştir. Ancak hangi durumda hangi hissin tetiklenebileceğini kültürel farkların belirlediği tespiti de eklenmiştir. Ekman (1992) yaptığı araştırmalar sonucunda temel hislerin kategorilere bölünebileceğini ve birbirine karışmayacağını öne sürmektedir. Ekman'ın "üzüntü, neşe, korku, öfke, tikslenme ve şaşırma" olmak üzere belirlediği 6 temel his kategorisi his analizi çalışmalarında sıklıkla kullanılmaktadır. Bu önermenin aksine Plutchik (1991) duyguların kategorilerden ziyade çarkın dışlıları şeklinde ifade edilebileceğini önermiştir. Araştırmacıya göre "neşe, güven, korku, şaşırma, üzüntü, tikslenme, öfke ve umut" olmak üzere 8 temel his vardır ve diğer tüm hisler bu temel hislerin karışımından oluşmaktadır. Örneğin aşk, güven ve neşe hislerinin birleşiminden oluşmaktadır. Teorisyenlerin kabul ettikleri başka yaklaşımlar ve his kategorileri olsa da his analizi özelinde en çok kabul edilen iki yaklaşım Ekman ve Plutchik modelleridir.



Şekil 1: Ekman Temel Hislerine Ait Örnek Yüz İfadeleri (Ekman ve Friesen, 1976)

Türkçe kelime anlamlarının yakın olması sebebiyle duygu ve his analizi birbirine karıştırılmaktadır. Duygu analizi bir metni olumlu, olumsuz ve tarafsız gibi sınıflarken his analizi metin içerisinde yansıtılan hislerin belirlenmesi işlemidir. Alandaki öncül çalışmalar genellikle İngilizce (Alm, Roth ve Sproat, 2005; Aman ve Szpakowicz, 2007) için yapılmıştır. Öte yandan Arapça (Abdul-Mageed, Alhuzli, Elhija, Diab ve Duaa'Abu Elhija, 2016), Fransızca (Abdaoui, Azé, Bringay ve Poncelet, 2017) ve Romence (Briciu ve Lupea, 2017) gibi birçok dil için yapılmış çalışmalar mevcuttur. Türkçe için his analizi yapılan ilk çalışmada (Boynukalin ve Karagoz, 2013) Türkçe çocuk masallarından etiketli cümleler ve ISEAR<sup>1</sup> (Wallbott ve Scherer, 1986) veri kümesinin çevirisi üzerinde “neşe, üzüntü, öfke ve korku” kategorilerinde 4.000 örnek üzerinde his analizi yapılmıştır. Makine öğrenimi yöntemlerinin uygulandığı çalışmanın sonucunda yaklaşık %80 oranında başarı elde edilmiştir. Türkçe diğer bir çalışmada (Demirci, 2014) ise sosyal medya ortamlarından Twitter üzerinde his analizi yapılmıştır. Kullanıcıların gönderilerinde kullandıkları etiketlere (*hashtag*) bakılarak “korku, öfke, tikslenme, neşe, üzüntü ve şaşırma” olmak üzere Ekman’ın 6 his kategorisinde toplanan her kategoride 1.000, toplam 6.000 *tweet* üzerinde makine öğrenmesi ve sınıflama yöntemleri kullanılarak yapılan his analizi sonucunda yaklaşık %70 başarı sağlanmıştır. Türkçe son çalışmada (Tocoglu ve Alpkocak, 2018) araştırmacıların daha önce oluşturdukları 26.000 örnek bulunan Türkçe his veri kümesi (Tocoglu ve Alpkocak, 2019) üzerinden öz nitelik seçme ve ağırlıklandırma yöntemleriyle bir his sözlüğü oluşturulmuş ve bu sözlük yardımıyla his analizi yapılmıştır. Ekman his kategorilerinde yapılan his analizi sonucunda yaklaşık %91 civarında başarı sağlanmıştır. Araştırmacılar (Tocoglu, Ozturkmenoglu ve Alpkocak, 2019) his kategorilerine ait anahtar kelimeleri içeren Twitter gönderilerini toplayarak oluşturdukları his veri kümesi üzerinde his analizi gerçekleştirmiş ve %73 başarı elde etmişlerdir. Görüleceği üzere, Türkçe his analizi alanında oldukça az çalışma yapılmış olup çalışmaların genelde metin sınıflama çalışmalarına öykünerek yapıldığı ve sözlüklü yöntemlerin de denendiği görülmüştür.

### His Analizinin Uygulama Alanları

Günümüzde yapay zekâ uygulamaları; otonom olarak karar verebilmenin, problem çözebilmenin hatta sebepleri tespit edebilmenin ötesinde görevleri üstlenebilmektedir. İnsanlarla iletişime giren bir bilgisayar için karşısındaki insanı anlamak en önemli görevdir. İnsanlar iletişim kurarken düz ifadeler yerine hislerini yansıtan karmaşık ifadeler kullanırlar. Yapay zekânın bu karmaşık ve duygusal ifadeleri anlaması için, insanla duygusal bir ilişki kurması

<sup>1</sup> Öğrencilerin duygusal deneyimlerini ifade ettikleri cümlelerden oluşan popüler İngilizce his veri kümesi.



gerekmektedir. Tam da bu noktada ilişkiye his ve duygu analizi işlevleri eklenmektedir.

Örneğin bir firmanın ürünleri için sosyal medyada yapılan yorumları okuyup geri dönüşleri firma yetkililerine özetleyen bir itibar analizi sistemi, müşterilerin hangi üründe hangi kusurdan şikâyet ettiklerini bildirebilir (Bai, 2011). Mesela bir siyasi parti lideri verdiği demeçlere seçmenlerin tepkisini ölçebilir (Dwi Prasetyo ve Hauff, 2015). Bir izleyici seyretmek istediği film hakkında yapılan yorumların duygusal özetini elde edebilir (Akba, Uçan, Sezer ve Sever, 2014). Öte yandan yatırımcıların ve siyasilere sosyal medya gönderilerini analiz ederek borsa hareketlerini tahmin edebilen bir çalışma (Smailovic, Grcar, Lavrac ve Znidarsic, 2013) yapılmıştır. Günümüzde sıkça karşılaşılan toplum mühendisliği operasyonlarında toplumun nabzını tutmak için duygu ve his analizi teknikleri kullanılmaktadır. Sonuç olarak yaşamın her alanında yer almaya başlayan yapay zekânın hisleri algılayabilmesi henüz hayal bile edemediğimiz uygulamaların ortaya çıkmasını sağlayacaktır.

### **Kullanılan Doğal Dil İşleme Yöntemleri**

Doğal dil işleme sesli ya da yazılı doğal dili anlamak ve işlemek için hesaplamalı yöntemler kullanan bir bilim dalıdır. Birçok alt işlevi olan doğal dil işleme yöntemleri genellikle birlikte kullanılır. Ses tanıma, cümle ve kelimelere ayırma, morfolojik analiz, sentaktik analiz ve anlamsal analiz gibi alt işlevlere sahiptir. Ses tanıma alt işlevi girdinin ses olduğu durumda sesi metne dönüştürmek için analog ses verisini alır ve akustik öz niteliklerden oluşan bir vektör oluşturur. Sonra vektörler ile ön tanımlı seslere ait vektörler arasındaki fonetik benzeşimleri bulur ve sesi metne dönüştürür.

Metin üzerinde öncelikle normalizasyon işlevi gerçekleştirilir. Normalizasyon işlevi metin içerisindeki yanlış yazımları, kısaltmaları, hatalı harfleri, yanlışlıkla birleşik ya da ayrı yazılmış ifadeleri vb. anormallikleri düzeltir (Akın ve Akın, 2007). Sonrasında ise metin cümle ve ifadelere bölünür. İfadeler çoğu zaman bir kelimededen oluşmakla birlikte bazen birden fazla kelimededen oluşan bir eş dizimli ifade olabilmektedir. Cümlelere ve ifadelere bölünmüş hâldeki metin artık morfolojik analiz yapmaya hazırdır.

Her bir ifade kök, gövde ve eklerine bölünür ve ifadenin “fiil, isim, zarf, zamir, sıfat” gibi hangi türde olduğu morfolojik analiz (Oflazer, 1994) ile belirlenir. İfade üzerindeki ekler de bu aşamada bulunmaktadır. Bir sonraki aşamada kullanmak için çekim ekleri temizlenir. Bazen kök kullanmanın avantajlı olduğu durumlar olsa da genellikle ifadeler gövde hâllerine dönüştürülür.

Her bir ifadenin türü belirlenip ve gövde hâline dönüştürüldükten sonra söz dizimsel analiz (Eryigit, 2014) işlevi devreye girer. Söz dizimsel analiz kapsamında cümlenin öğeleri bulunmakta ve her bir ifadenin cümlede hangi



görevde kullanıldığı belirlenmektedir. Söz dizimsel analiz yapılırken ifadenin görev belirsizliği varsa istatistiksel olarak bu belirsizlik giderilmeye çalışılmaktadır (Hakkani-Tür, Oflazer ve Tür, 2002). Söz dizimsel analiz işlevine kimi zaman doğal dil işleme işlerinde ihtiyaç duyulurken kimi zaman ise ihtiyaç olmadığı için gerçekleştirilmez.

Öncesindeki tüm alt işlevlerin çıktılarına ihtiyaç duyan anlamsal analiz (Oflazer ve Saraçlar, 2018) kapsamında birden fazla anlama sahip olan çok anlamlı ifadelerin anlamsal belirsizliği giderilmektedir. Kelime anlam belirsizliği giderimi (Ilgen, Adalı ve Tantug, 2016) işlemi, ifadenin sözlükteki anlamlarından hangisinin cümle içerisinde kullanıldığını belirlemeye çalışmaktır. Kelime anlam belirsizliği giderimi algoritmalarının temeli, cümlenin bağlamını belirleme işlemidir. Bağlam belirlemenin çok farklı yöntemleri olmakla birlikte, örnekler üzerinden anlam birlikteliklerine ait istatistiksel dağılımdan faydalanmak en çok başvurulan yöntemdir. İstatistiksel kelime anlam belirsizliği giderimi yöntemiyle, birlikte daha sık yer alan anlamlar belirlenerek tutarlı anlamlar zinciri tespit edilmektedir. Örneğin “sabahları yüz yıkama alışkanlığı yok” cümlesi içerisinde yer alan “yüz” çok anlamlı kelimesi; tek başına ele alındığında anlamı belirlenemezken birlikte yer aldığı kelimelere bakılarak “surat” anlamında kullanıldığı belirlenebilmektedir.



Şekil 2: Uygulanan Doğal Dil İşleme Adımları

Tüm bu doğal dil işleme alt işlevleri kendi içerisinde birçok alt başlığa bölünmekte ve her bir alt başlık için doğal dil işleme araştırmacıları yöntemler geliştirmektedir. Bu yöntemler istatistiksel olabildiği gibi makine öğrenimi yöntemleri de olabilmektedir.

### Kullanılan Öğrenme Yöntemleri

İnsanların öğrendiklerine tecrübe adı verilmektedir. İnsanlar birbirlerinin tecrübelerinden öğrenmek istediklerinde kurallara dayalı ifadeler ortaya çıkmaktadır. Bu ifadeler sebeple sonuç arasındaki ilişkiyi ortaya koymaktadır. Çoğu zaman ortaya çıkan kurallar “böyle olursa böyle olur” şeklinde olmaktadır. Bu kurallar doğrudan kullanılabildiği gibi bağlama göre değiştirilmesi de gerekebilir. Örneğin; “kadın uzun boyluysa, kemikleri kalındır” kuralında bahsedilen “uzun boy” kavramı kültüre göre değişim göstermektedir. Tam bu noktada, öğrenme; kuralın kültüre göre değişen bağlamlarda oluşmuş olaylar üzerinden tespit edilmesidir. Dolayısıyla bilgisayarla öğrenme yapılabilmesi

için öğrenilecek konuyu temsil eden; pozitif (olmuş) ve negatif (olmamış) örneklerinden oluşan bir veri uzayına ihtiyaç vardır. Bu kuralların sebep kısmında ifade edilen kısım vektörler, sonuç kısmı da öğrenmek istenilen bilgidir. His analizi probleminde bilgisayara metinler üzerinden his öğretilmek istenmektedir. Doğal olarak girdilerin; o metni oluşturan gövde, kelime, cümle, paragraf olması ve çıktılarının ise hisler olması beklenmektedir.

Günümüz dünyasında sosyal medyanın popüler olması ve insanların ürettikleri sayısal verinin çokluğu tüm doğal dil işleme yöntemlerinin büyük veri üzerinde yapılabilmesini zorunlu hâle getirmiştir. Üzerinde çalışılan veri evreni bu denli büyük olduğunda tüm senaryoları (ifade biçimlerini) kapsayan genelleştirilmiş kurallar oluşturmak ve kural tabanlı sistemler yordamıyla çalışmak makuliyetini yitirir ve ifadelerdeki saklı ilişkilerin tespitinde makine öğrenimi kullanışlı bir çözüm olarak öne çıkar. Makine öğrenimi yöntemleri her bir girdi ile tahmin edilecek (öğrenilecek) sınıf arasındaki ilişkiyi ortaya koymaktadır. Bu ilişkiyi bulmak için öz niteliklerden faydalanmaktadır. Öz nitelikler; kelimeler, kökler, gövdeler, karakter n-gramları (katarları), kelime n-gramları, özel noktalama işaretleri vb. metin özellikleri olabilmektedir.

Makine öğrenimi yöntemleri gözetimli veya gözetimsiz yapılabilir. İçerisinde yer alan girdilerin hangi sınıfa ait olduğu işaretlenmişse veri kümesi etiketli olarak adlandırılır. Eldeki verinin etiketli olması durumunda gözetimli makine öğrenmesi yapılabilir. Örneğin haber veri kümesi içerisinde yer alan her bir haberin hangi kategoriye ait olduğu önceden bilinecek sınıflama modeli geliştirilir. Eldeki verinin etiketlere sahip olmadığı durumda ise gözetimsiz bir yöntem olan kümeleme işlemi ile benzer ifadelerin gruplanması yapılır ancak grubun ya da kümenin bir adı ya da etiketi yoktur. Veri kümesindeki verilerin çok azı etiketli olduğunda yarı gözetimli bir yol izlenebilmektedir. Önce etiketli veriyi çoğaltmak için gözetimsiz kümeleme işlemi yapılır, devamında ise etiketli veri ile gözetimli sınıflama işlemi gerçekleştirilir.

Makine öğrenimi yöntemlerinde veri kümesi içerisinde yer alan her bir eleman öz niteliklerden oluşan bir vektör olarak temsil edilmektedir. Geleneksel makine öğrenmesi yöntemlerinde hangi öz niteliklerin bu temsil vektöründe yer alacağı çok önemlidir. Çünkü hem algoritmanın yapısı gereği hem de öğrenmenin gerçekleştirildiği bilgisayarın hesaplama gücü ve bellek kapasitesi gereği hesaba katılacak öz nitelik sayısının bir sınırı vardır. Genellikle veri kümesi içerisinden çıkabilecek öz nitelik sayısı bu sınırdan fazladır. Bu sebeple öz nitelik mühendisliği yapılması gerekmektedir. Önemli öz niteliklerin seçilmesi, seçilen öz niteliklerin önem derecesine göre ağırlıklandırılması ve içeriğe bağlı kalarak yeni öz niteliklerin çıkarılması işlemleri öz nitelik mühendisliği olarak adlandırılmaktadır. Geleneksel makine öğrenmesi gerçek-





leştirilirken verinin kalitesinden sonra sonucu etkileyen en önemli aşama öz nitelik mühendisliğidir.

Geleneksel yöntemlerle yapılan his analizi, verinin etiketli olduğu durumlarda gözetimli sınıflama yöntemleri kullanılarak yapılmaktadır. Öncelikle veriyi oluşturan her bir metin parçası seçilen öz niteliklerle vektörel olarak ifade edilir. Oluşturulan vektörlerin büyük bir kısmı öğrenme amacıyla ayrılır, bu kısım öğrenme kümesi olarak adlandırılır. Kalan kısım ile öğrenilen modelin doğruluğu ölçümlenir, bu doğrulama verileri ise test kümesi olarak adlandırılır. Öğrenme kümesi üzerinden Naive Bayes, Lojistik Regresyon, Karar Ağaçları, Yapay Sinir Ağları, En Yakın Komşu veya Destek Vektör Makinesi (*Support Vector Machine*) gibi bir öğrenme yöntemi ile model oluşturulur (Aggarwal ve Zhai, 2012). Oluşturulan modelin doğruluğu test kümesi ile ölçümlenir. Oluşturulan model saklanarak vektörel olarak ifade edilmiş canlı verileri analiz etmek için kullanılabilir.

Kural tabanlı yöntemler kullanılarak his analizi yapmak için genellikle sözlükler (Mohammad ve Turney, 2013; Tocoglu ve Alpkocak, 2019) kullanılmaktadır. Bu sözlükler hazırlanırken anket veya uzman görüşünden faydalanılmaktadır. Anket ve uzman görüşünün alınmadığı durumlarda ise etiketli veriler üzerinden otomatik olarak oluşturulabilmektedir. His sözlükleri sayesinde hızlı ve hedefe yönelik his analizi yapılmaktadır. Veri kümesindeki her bir metin parçası içerisinde his sözlüğünde yer alan ifadeler belirlenir. Sözlükte yer alan his skorları alınıp metin parçasında ağırlıklı olarak hangi hissin yer aldığı belirlenmektedir.

Genellikle daha başarılı his analizi sonuçları elde etmek amacıyla hem kural tabanlı hem de makine öğrenmesi yöntemleri birlikte kullanılmaktadır (Bandhakavi, Wiratunga, Padmanabhan ve Massie, 2017) Veri kümesindeki her bir metin parçası, içerisinde yer alan öz niteliklerle vektörel olarak ifade edilmektedir. Bu temsil vektörleri içerisinde his sözlüğünden faydalanarak hislere ait bazı yeni öz nitelikler eklenmektedir. Hislerin de dâhil olduğu vektörlerle yapılan makine öğrenimi yöntemleri, görece daha başarılı sonuçlar üretmektedir.

Geleneksel makine öğrenimi yöntemlerinin aksine öz nitelik mühendisliği yapılmaksızın, elde edilebilecek tüm öz nitelikler temsil vektörlerine eklenebilmekte ve derin öğrenme algoritmaları yardımıyla his analizi yapılabilir (Naderalvojud ve diğerleri, 2018; Tocoglu ve diğerleri, 2019). Derin öğrenme algoritmaları geleneksel makine öğrenimi yöntemlerinden farklı olarak girdileri gruplara ayırarak almakta ve bu sayede neredeyse sınırsız sayıda öz nitelikle çalışabilmektedir. Aslında derin öğrenme bir tür yapay sinir ağı algoritmasıdır. Basitçe yapay sinir ağı algoritmaları; girdi katmanı, çıktı katma-



nı ve gizli katmanların birbirlerine belli bir fonksiyonla bağlanmasıyla çalışan algoritmalarıdır. Yapay sinir ağları insan beynindeki nöronlar arasındaki iletim ve beyin öğrenme şekline öykünerek geliştirilmiştir. Yapay sinir ağlarının ilk ortaya atılışı 50 yıl önceye dayansa da 90'lı yıllarda uygulanmaya başlamış ancak bellek kapasitesi ve hesaplama gücünün yetersiz olması sebebiyle 10 yıl öncesine kadar etkin şekilde kullanılamamıştır. Grafik kartları üzerindeki işlemci çekirdeklerinin artması ve hesaplamada kullanılabilir hâle gelmesiyle yapay sinir ağları popüler hâle gelmiştir. Bilim insanları gizli katmanların sayısını artırarak daha derin ağlar oluşturmuş ve derin öğrenme olarak adlandırmışlardır.

Derin öğrenme uygulayan bir araştırmacı bilinen ağları kullanabileceği gibi isterse kendi ağını tasarlayabilmektedir. Bunlara ilaveten makine öğrenmesine girdi olarak verilen vektörlerin seyrek (*sparse*) yapıdan yoğun (*dense*) yapıya dönüşmesi derin öğrenmenin başarılı olması için gerekli hâle gelmiştir. Yapılan çalışmada (Mikolov ve diğerleri, 2013) araştırmacılar girdileri temsil eden seyrek matrisi CBOW ve SkipGram yöntemleriyle sabit boyutlu yoğun bir matris hâline dönüştürmüş ve oluşan yapıya Word2Vec Kelime Öz Yerleşikleri (*Word Embedding*) adını vermişlerdir. Çalışmada öğrenilen vektörlerin analizi ile öğrenilen temsiller üzerinde yapılabilecek matematiksel işlemlere de yer verilmiştir. Örneğin erkek – kadın ilişkisi otomatik olarak öğrenilmiş ve vektörel işlemlerle yapılan “Kral – (eksi) Erkek + (artı) Kadın” işleminin sonucunun “Kraliçe” vektörüne oldukça yakın olduğu bildirilmiştir. Elde edilen temsillerin söz dizimsel ve anlamsal özellikleri başarılı şekilde yansıtmakta olduğu belirlenmiştir. Metin işleme alanında yapay sinir ağları tabanlı öğrenme algoritmaları için Word2Vec veya benzeri bir ön tanımlı (*pretrained*) kelime öz yerleşigi kullanmak bir standart hâline gelmiştir.

Derin öğrenmenin geleneksel makine öğrenmesinden temel farkı büyük veri üzerinde öz nitelik mühendisliği yapmadan kullanılabilir olmasıdır. Derin öğrenme kendisi hangi öz niteliğin önemi fazlaysa o öz niteliğin ağırlığını artırmaktadır. Bu sayede araştırmacının öz nitelikler üzerinde işlem yapmasına gerek kalmamaktadır. Ayrıca ön tanımlı kelime öz yerleşikleri sayesinde geleneksel öğrenme yöntemlerinden farklı olarak metin hem anlamsal hem de söz dizimsel olarak daha başarılı ifade edilmektedir. Ancak derin öğrenme yapısı oluşturulurken ağırlık yapısı, geçiş fonksiyonları, hata fonksiyonları, grup büyüklüğü (*batch size*) vb. hiper (*hyper*) parametrelerin araştırmacı tarafından optimize edilmesi gerekmektedir. Bu sebeple araştırmacı geleneksel makine öğrenmesi yöntemlerinde öz nitelik mühendisliği yaparken derin öğrenmede ise hiper parametre mühendisliği yapmaktadır.

Geleneksel makine öğrenmesi yöntemlerinde olduğu gibi derin öğrenme yöntemlerine de girdi olarak sözlüksel bilgiler eklenebilmekte ve daha başarılı



hibrit yöntemler oluşturulabilmektedir. Günümüzde hem geleneksel makine öğrenmesi hem kural tabanlı sistemler hem de derin öğrenmeli yöntemler kullanılarak his analizi yapılmaktadır.

### Bir His Analizi Örneği

Metin üzerinde his analizinin; doğrudan hislerin belirlenmesi, dolaylı hislerin belirlenmesi veya gizli hislerin belirlenmesi gibi türleri vardır. Doğal dilde yazılmış metinler içerisinde hisler çoğu zaman dolaylı olarak ifade edilmekte ve asıl hissedilen gizlenmektedir. Bazı kişiler hislerini “mutluyum” şeklinde doğrudan ifade etmek yerine, “sanki bulutların üzerinde uçuyorum” şeklinde örtülü olarak anlatırlar.

Doğal dil işleme alanında 2018 yılında “Doğal Dil İşlemede Ampirik Yöntemler” (EMNLP) konferansı sırasında düzenlenen “Öznellik, Duygu ve Sosyal Medya Analizi Konusunda Hesaplamalı Yaklaşımlar Çalıştayı”nda (WASSA) “Örtülü Hisler Görev Paylaşımı” (IEST<sup>2</sup>) etkinliği yapılmıştır. IEST etkinliğinin amacı metin içerisindeki örtülü hislerin doğru olarak belirlenmesidir. Etkinlik için; içerisinde Ekman temel his kategorileri olan “üzüntü, neşe, korku, öfke, tikslenme ve şaşırma” kelimelerini ve eş anlamlılarını barındıran yaklaşık 200 bin *tweet* toplanmıştır. Sonrasında toplanan *tweet*’ler içerisinde yer alan bu his ifadeleri silinmiş, silindiği yer [#TARGETWORD#] şeklinde işaretlenmiş ve her bir *tweet*’in his etiketi kaydedilmiştir. Örneğin “So [#TARGETWORD#] when lovely couples separate. Hay” – Etiket: “Sad”. Veri kümesi içerisinde yer alan yaklaşık 150 bin *tweet* öğrenme kümesi olarak, yaklaşık 10 bin *tweet* geliştirme kümesi ve yaklaşık 30 bin *tweet* ise test kümesi olarak yayımlanmıştır.

IEST etkinliğinde öncelikle örnek veriler ve yarışmanın tanımı yapılmıştır. Devamında öğrenme ve geliştirme kümeleri yayımlanmış ve yarışma gününe kadar test kümesi yayımlanmamıştır. Öğrenme ve geliştirme aşamasından sonra test aşaması günleri geldiğinde test kümesi etiketsiz şekilde yayımlanmış ve yarışmacılardan etiketleri tahmin ederek test aşaması sonuna kadar sisteme yüklemeleri istenmiştir. Yarışmaya 30 takım katılmıştır.

Hacettepe Üniversitesi Bilgisayar Mühendisliği Çoklu Ortam Bilgi Erişim Laboratuvarı (HUMIR) takımı IEST etkinliğine katılmıştır. HUMIR takımı sonuçları başarılı şekilde tahmin eden ilk 6 takım arasına girmiştir (Naderalvojud ve diğerleri, 2018). Takım, problemin kelime anlam belirleme problemine benzediğini vurgulamıştır. Kelime anlam belirleme yöntemi bir kelimenin cümle içerisinde hangi anlamda kullanıldığını bulmayı amaçlamaktadır. Araştırmacılar veri seti içerisinde silinen kelimelerin yerine konan [#TARGETWORD#] ifadesinin sentetik bir kelime olduğunu ve bu sentetik

<sup>2</sup> <http://implicitmotions.wassa2018.com/>



kelimenin altı farklı anlamı olduğunu varsaymıştır. Bu varsayımda anlamlar Ekman temel hislerine işaret etmektedir. Yapılan uygulama cümle ve bağlama bağlı olarak sentetik kelimenin hangi his anlamında kullanıldığını başarıyla bulmuştur.

Araştırmacılar makine öğrenimi yöntemlerinden “Multi-Layer Perceptron” (MLP) ve 6 farklı derin öğrenme yönteminin sonuçlarını birleştirerek sonuç üretmişlerdir. Bu yöntemlere girdi olarak *tweet*’lerdeki kelimeleri temsil etmek için; 200 boyutlu, 2 milyar *tweet* üzerinde eğitilmiş TwitterGloVe<sup>3</sup> ön tanımlı kelime öz yerleşikleri kullanılmıştır. Öncelikle veri kümesinde yer alan her *tweet* için; URL silme, kullanıcı adı silme, satır sonu silme işlevlerinden oluşan kısıtlı bir normalizasyon işlemi yapılmıştır. Sonrasında *tweet*’ler NLTK<sup>4</sup> araç seti ile kelimelere bölünmüştür. Veri kümesi içerisi içerisinde yer alan 25 bin biricik kelimenin 3.500 kadarı TwitterGloVe içerisinde yer almaktadır. Bu kelimeler yerine ortalamayı bozmayacak şekilde rastgele sayılardan vektörler oluşturulmuştur. Sonuç olarak veri kümesi içerisinde yer alan tüm *tweet*’ler vektörler şeklinde temsil edilmiştir. Vektörlere öz yerleşiklerden gelen 200 boyutun yanı sıra (Mohammad ve Bravo-Marquez, 2017) makalesinde ayrıntılı şekilde anlatılan 45 sözlüksel öz nitelik eklenmiştir. Ayrıca her bir *tweet*’in her bir his kategorisine istatistiksel olarak ne ölçüde yakın olduğu bir tür metrikle hesaplanarak vektöre eklenmiştir. Bu sayede her bir *tweet* hem anlamsal hem sözlüksel hem de istatistiksel olarak temsil edilmeye çalışılmıştır.

Hem klasik makine öğrenmesi hem de derin öğrenme yöntemleriyle his analizi yapılmıştır. Klasik makine öğrenmesi için MLP algoritmasının kullanıldığı deneylerde, derin öğrenme için Keras<sup>5</sup> kütüphanesi kullanılmıştır. Derin öğrenme ağı kurgulanırken tüm *tweet*’in tek bir cümle olarak alındığı modelin yanı sıra sentetik kelimenin sağ ve solunun ayrı birer cümle olarak alındığı modeller de kullanılmıştır. Bağlamı ifade etmek için tasarlanmış bir derin öğrenme ağ çeşidi olan Uzun Kısa Süreli Hafıza (*Long Short Term Memory*- LSTM) (Hochreiter ve Schmidhuber, 1997) ve çift yönlü Bi-LSTM kullanılarak tasarlanmıştır. LSTM zincir şeklinde bağlılıkların etkilerini cümlelerin elemanlarına taşımaktadır. Bu sayede cümle içerisindeki her bir kelimenin öncesinde, daha öncesinde, sonrasında ve daha sonrasında yer alan kelimelere ait birliktelik bilgileri derin ağı yansıtılabilmektedir. Hiper parametrelerin ve cümleyi ele alış şekillerinin değiştirildiği 6 farklı derin öğrenme modeli 50 kez (*epoch*) çalıştırılarak modeller eğitilmiştir. Tüm modellerde girdi vektör çıktı ise sınıf bilgisidir. Tüm modellerin sonuçları eşit ağırlıkta kabul edilerek

<sup>3</sup> <http://nlp.stanford.edu/projects/glove/>

<sup>4</sup> <https://www.nltk.org/>

<sup>5</sup> <https://keras.io/>



ortalaması alınmış ve nihai sınıf bilgisi üretilmiştir. Nihai sistem, veri kümesi üzerinde %68,8 başarıyla his analizi yapmıştır.

### Sonuç

Bir metnin barındırdığı “neşe, üzüntü, öfke, şaşırma, tikslenme ve korku” gibi hisleri belirleme işlemine his analizi adı verilmektedir. İnsan bilgisayar etkileşiminde hislerin rolü büyüktür çünkü insanlar çoğu zaman hislerini belirtmek için iletişim kurarlar ve hislerini dolaylı olarak anlatırlar. Görüntü ve sesin olmadığı yazılı ortamda hangi hislerin yer aldığını bulmak oldukça zor bir doğal dil işleme problemidir.

Anketlerin yerini almaya aday olan sosyal medya verileri üzerinden duygu ve his analizi sayesinde; yazarın belli bir konu hakkında ne fikirde olduğu, yönelimi ya da siyasi eğilimi yüksek doğrulukta bulunabilmektedir. İnternet ve sosyal medya kullanımının toplum genelinde oldukça yaygın olduğu göz önünde bulundurulduğunda, otomatik his analizi yapmanın oldukça değerli olduğu görülmektedir. Ayrıca bilgisayarların koştur işleme ve saklama becerilerinin artması da metin üzerinde his analizinin popüler olmasını desteklemiştir. Tüm dünyada olduğu gibi ülkemizde de his analizinin önemi anlaşılmaya başlamış ve birtakım araştırmacılar tarafından çalışılmaya başlamıştır. Türkçe etiketli his veri kümelerinin ve doğal dil işleme altyapılarının erişilebilir olmasıyla birlikte Türkçe his analizi çalışmaları ivme kazanmıştır.

Şimdiye kadar Türkçe için; his sınıflama, sosyal medya verilerinde his sınıflama, his sözlüğü ile kural tabanlı his analizi ve hibrit yöntemler ile his analizi yapılmıştır. Ancak metin içerisinde yer alan hislerin oranı, yazar veya okuyucu bakış açısıyla his analizi ya da örtülü his analizi konularında Türkçe çalışma yapılmamıştır. Ayrıca derin öğrenme ve ön tanımlı kelime öz yerleşikleri kullanarak Türkçe his analizi henüz gerçekleştirilmemiştir. Temel his veri kümelerinin çeşitliliğin artırılması, derin öğrenmeyle Türkçe his analizinin denenmesi ve bağlama göre his analizinin gerçekleştirilmesi planlanmaktadır.

### Kaynakça

- Abdaoui, A., Azé, J., Bringay, S. ve Poncelet, P. (2017). FEEL: a French expanded emotion lexicon. *Language Resources and Evaluation*, 51(3), 833–855. doi:10.1007/s10579-016-9364-5
- Abdul-Mageed, M., Alhuzli, H., Elhija, D. A., Diab, M. ve Duaa' Abu Elhija, M. D. (2016). DINA: A multidialect dataset for arabic emotion analysis. *The 2nd workshop on Arabic corpora and processing tools* içinde (s. 29).
- Aggarwal, C. C. ve Zhai, C. (2012). A survey of text classification algorithms. *Mining text data* içinde (ss. 163–222). Springer.



- Akba, F., Uçan, A., Sezer, E. ve Sever, H. (2014). Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews. *8th European Conference on Data Mining 2014*, 191 (2002), 180–184.
- Akın, A. A. ve Akın, M. D. (2007). Zemberek, an open source Nlp framework for Turkic Languages. *Structure*, 10, 1–5. doi:10.1.1.556.69
- Alm, C. O., Roth, D. ve Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. *Proceedings of the conference on human language technology and empirical methods in natural language processing* içinde (ss. 579–586).
- Aman, S. ve Szpakowicz, S. (2007). Identifying expressions of emotion in text. *International Conference on Text, Speech and Dialogue* içinde (ss. 196–205).
- Bai, X. (2011). Predicting consumer sentiments from online text. *Decision Support Systems*, 50(4), 732–742.
- Bandhakavi, A., Wiratunga, N., Padmanabhan, D. ve Massie, S. (2017). Lexicon based feature extraction for emotion text classification. *Pattern Recognition Letters*, 93, 133–142. doi:10.1016/j.patrec.2016.12.009
- Boynukalin, Z. ve Karagoz, P. (2013). Emotion analysis on Turkish texts. *Information Sciences and Systems 2013* içinde (ss. 159–168). Springer.
- Briciu, A. ve Lupea, M. (2017). RoEmoLex - a Romanian emotion lexicon. *Studia Universitatis Babeş-Bolyai Informatica*, 62(2), 45–56. doi:10.24193/subbi.2017.2.04
- Demirci, S. (2014). *Emotion analysis on Turkish tweets*. Middle East Technical University.
- Dwi Prasetyo, N. ve Hauff, C. (2015). Twitter-based election prediction in the developing world. *Proceedings of the 26th ACM Conference on Hypertext & Social Media* içinde (ss. 149–158).
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation* içinde (C 19, ss. 207–282).
- Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3–4), 169–200.
- Ekman, P. ve Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1), 56–75. doi:10.1007/BF01115465
- Eryigit, G. (2014). ITU Turkish NLP web service. *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics* içinde (ss. 1–4).
- Hakkani-Tür, D. Z., Oflazer, K. ve Tür, G. (2002). Statistical morphological disambiguation for agglutinative languages. *Computers and the Humanities*, 36(4), 381–410.
- Hochreiter, S. ve Schmidhuber, J. J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1–32.

- Ilgen, B., Adali, E. ve Tantug, A. C. (2016). Exploring feature sets for Turkish word sense disambiguation. *Turkish Journal of Electrical Engineering & Computer Sciences*, 24(5), 4391–4405.
- Mikolov, T., Corrado, G., Chen, K., Dean, J., Corrado, G. ve Dean, J. (2013). Efficient estimation of word representations in vector space. *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, 1–12.
- Mohammad, S. M. ve Bravo-Marquez, F. (2017). Emotion intensities in tweets. *SEM 2017: The Sixth Joint Conference on Lexical and Computational Semantics* içinde (ss. 65–77).
- Mohammad, S. M. ve Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436–465. doi:10.1111/j.1467-8640.2012.00460.x
- Naderalvojud, B., Ucan, A. ve Akcapinar Sezer, E. (2018). HUMIR at IEST-2018: Lexicon-sensitive and left-right context-sensitive bi-lstm for implicit emotion recognition. *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* içinde (ss. 182–188). Association for Computational Linguistics.
- Oflazer, K. (1994). Two-level description of Turkish morphology. *Literary and linguistic computing*, 9(2), 137–148.
- Oflazer, K. ve Saraçlar, M. (2018). *Turkish natural language processing*. Springer.
- Pang, B., Lee, L. ve Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10* içinde (ss. 79–86). Association for Computational Linguistics. doi:10.3115/1118693.1118704
- Plutchik, R. (1991). *The emotions*. University Press of America.
- Smailovic, J., Grcar, M., Lavrac, N. ve Znidarsic, M. (2013). Predictive sentiment analysis of tweets: A stock market application. *International Workshop on Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data* içinde (ss. 77–88).
- Tocoglu, M. A. ve Alpkocak, A. (2018). TREMO: A dataset for emotion analysis in Turkish. *Journal of Information Science*, 44(6), 848–860.
- Tocoglu, M. A. ve Alpkocak, A. (2019). Lexicon-based emotion analysis in Turkish. *Turkish Journal Of Electrical Engineering & Computer Sciences*, 27(2), 1213–1227.
- Tocoglu, M. A., Ozturkmenoglu, O. ve Alpkocak, A. (2019). Emotion analysis from Turkish tweets using deep neural networks. *IEEE Access*, 7, 183061–183069. doi:10.1109/ACCESS.2019.2960113
- Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* içinde (ss. 417–424). Association for Computational Linguistics. doi:10.3115/1073083.1073153



- Wallbott, H. G. ve Scherer, K. R. (1986). How universal and specific is emotional experience? Evidence from 27 countries on five continents. *Information (International Social Science Council)*, 25(4), 763–795.
- Wiebe, J., Wilson, T. ve Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2–3), 165–210.





### Extended Summary

Computational linguistics is an interdisciplinary field that aims to understand the verbal or written language, to express it mathematically and suggest methods, models and tools to achieve these goals. First of all, researchers tried to write rules to create a language model, however, they became unsuccessful. For this reason, they tried to find a way to learn instead of writing rules. Learning takes place by experiencing is the fact for most of times. For a computer to learn a language, many correct examples are needed in that language. The developments such as the proliferation of the Internet, the increase of digital content, storage and computing power, data required for learning has been provided.

Today, it is not difficult to create a language model that can determine sentence boundaries, distinguish expressions, find the root and body of the word, make morphological analysis, make semantic analysis, summarize, identify topics or translate into another language. Researchers have achieved very successful results in modeling the meaning of the text. However, even though learning the semantic model is sufficient for modelling the objective content, it is inadequate when it comes to subjective content since the feelings of expressions vary according to the context.

The researchers aimed to detect the emotions and sentiments that the expressions contain beyond its meaning in order to solve the problem of subjective content. At first, emotional analysis studies were conducted that can roughly classify expressions as “positive” and “negative”. Subsequent studies were made classifications are more detailed such as “very positive, positive, neutral, negative, very negative”. Ultimately, the researchers conducted emotion analysis studies to automatically determine the feelings inside the texts such as “joy, anger, fear, sadness”.

Emotion analysis stands for the process of finding which feelings are included in a sound, image or text. Textual emotion analysis is performed to find answers many different questions such as; what feelings are included in what proportion, whether there is a feeling, which feeling the writer narrates, what the reader feels, what the writer feels about a certain subject, what are the implicit feelings in the text, what are the nested feelings. While emotions can be clearly noticeable from the tone of the voice, gestures and facial expressions during a conversation, it is often difficult for even for a person to determine what feelings the author has written in written communication since texts are composed of expressions with different meanings and feelings according to the context. If the emotions reflected by these words and expressions can be determined, the feelings reflected by a sentence, a paragraph, and ultimately a text can be found proportionally.

In emotion analysis studies, universal emotions are preferred when determining emotion classes. Many studies have been conducted in the field of psychology and sociology to determine basic feelings. As a result of these studies, it has been determined that basic feelings are not learned, they are known naturally, therefore, are universal. At the same time, it has been determined that cultural differences determine which feelings can be triggered (Ekman, 1972). Research suggests that basic emotions can be divided into categories. The six main categories of emotions that Ekman (1992)



determined, namely “sadness, joy, fear, anger, disgust, and surprise”, are frequently used in emotion analysis. According to another study, there are eight basic emotions, namely “joy, trust, fear, surprise, sadness, disgust, anger and anticipation”, and all other emotions are a mixture of these basic feelings (Plutchik, 1991). Ekman and Plutchik basic emotion classes are frequently used in emotion analysis studies.

Few studies have been done in the field of Turkish emotion analysis, and it can be determined that these studies are generally carried out by simulating text classification methods and also trying the dictionary-based methods. Datasets which translated from the English emotion dataset, collected self-labelled posts from social media sources, and created by survey method were used for Turkish emotion analysis. Turkish emotion analysis studies gained momentum with the availability of Turkish emotion datasets and natural language processing infrastructures. Up to now emotion classification, emotion classification in social media data, rule-based emotion analysis with emotion dictionary and emotion analysis with hybrid methods were conducted for the Turkish language. However, there was no study in Turkish on the distribution of emotion in the text, emotion analysis from a writer or reader perspective, or implicit emotion analysis. In addition, Turkish emotion analysis using deep learning or pre-trained language models has not been performed.

This study, which summarizes the subject of emotion analysis, aims at explaining the history and importance of emotion analysis with a linguistic perspective and briefly introduce the current application areas of emotion analysis. In the rest of the article; the natural language processing and learning methods used to analyze emotion are described. A sample emotion analysis study is summarized and the studies conducted and planned to be carried out in the field of Turkish emotion analysis are mentioned.

