

Makine Öğrenimi İle MBTI Kişilik Tipi Tahmini

MBTI Personality Prediction With Machine Learning

Kaan Sönmezöz
Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
sonmezozkaan@gmail.com

Özgür Uğur
Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
ozgurugur.ytu@gmail.com

Banu Diri
Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
diri@yildiz.edu.tr

Özetçe —Kişilik tipleri, davranışlarımızdan kariyer seçimlerimize kadar yaşamlarımızı sürekli olarak etkilemektedir. Bu kişilik tiplerinden yararlanarak daha başarılı öneri sistemlerinin tasarlanması ve daha etkili dijital pazarlama stratejilerinin geliştirilmesi mümkündür. Bu çalışmada sosyal medya paylaşımlarını kullanarak kullanıcıların MBTI kişilikleri tahmin edilmiştir. Türkçe dili kullanılarak gerçekleştirilen ilk çalışma olmasına rağmen elde edilen sonuçlar, kişilik tipi tespiti için Türkçe dilinde de uygulanabileceğini göstermektedir. En iyi sonuçlar MBTI kişilik tiplerinin boyutları tahmin edildiği zaman alınmıştır. Bir boyut hariç geri kalanların F-Skor değerleri %60 civarlarındadır.

Anahtar Kelimeler—Doğal dil işleme, makine öğrenimi, MBTI, kişilik tahmini, kişilik tipleri

Abstract—Personality traits, continuously affect our lives, from our behavior to our career decisions. It's possible to design more accurate recommendation systems and develop more efficient digital marketing strategies by the help of personality traits. In this work, people's MBTI personality traits were predicted according to their social media posts. Although it is the first study which uses Turkish language, the results show that personality type prediction can be applied in Turkish language as well. The best results were obtained when the MBTI dimensions were predicted. The F-Score values of those except one dimension are approximately 60%.

Keywords—Natural language processing, machine learning, MBTI, personality prediction, personality traits

I. GİRİŞ

Carl Jung içe dönük ve dışa dönük kişilik tiplerinden ilk bahseden kişi olmuştur ve Kişilik Teorisi alanının öncüleri arasında gösterilmektedir. Carl Jung'un ortaya attığı bu kavramlar geliştirilerek günümüzde çeşitli kişilik tipleri olarak karşımıza çıkmaktadır. Bunlardan en yaygın olarak MBTI (Myers-Briggs Type Indicator) [1] ve Beş Faktörlü Kişilik Kuramı [2] kullanılmaktadır.

Kişilik; insanların farklı ve belli bir düzen oluşturan duygu, düşünce ve davranışların bütünü olarak tanımlanabilir [3]. Kişilik tiplerinin, insanların mesleki başarılarında [4] ve maddi kararlarında da [5] etkili olduğu araştırmalarla tespit edilmiştir.

Kişilik tipleri günümüzde yaygın olarak kullanılmaktadır. Bunlara örnek olarak öneri sistemleri [6], [7], insan davranışlarının tahmin edilmesi [8]–[10], yasadışı içeriklerin saptanması [11], [12], şirketlerdeki işe alım süreçleri gösterilebilir.

Bir kişinin kullandığı dil ile kişilik tipinin ilişkili olduğunun araştırmalarla gösterilmesi [13]–[18] doğal dil işleme kullanılarak kişilik tipi tespiti için yapılabilmesinin önünü açmıştır. Bu alanda yapılan çalışmalarda kişilik tipi olarak MBTI ve Beş Faktörlü Kişilik Tipi kullanılmıştır. Her ne kadar MBTI ve Beş Faktörlü kişilik tipi arasında bir ilişki [19], [20] olsa da kişilik tiplerinin farklı olması yapılan çalışmaların doğrudan karşılaştırılmasının önüne geçmektedir.

Beş Faktörlü Kişilik Tipi belirlenirken yararlanılan kişilik özellikleri: [21]

- **Dışadönüklük (Extraversion)**: Temelde hayat dolu, heyecanlı, neşeli, konuşkan, girişken ve sosyal olma gibi özellikleri içermektedir. Sosyalite ve insanlarla birlikte olmayı sevmek, eğlenmeyi sevmek, liderlik, güç, istekli olma ve arkadaşça davranma gibi özellikler de bu faktörü temsil etmektedirler.
- **Yumuşak Başlılık (Agreeableness)**: Nazik, ince, saygılı ve merhametli gibi özellikler bu boyutun özelliklerindendir.
- **Öz Denetim (Conscientiousness)**: Düzen, titizlik, başarı çabası, çalışma kararlılığı, disiplin özellikleri bu boyut ile ilişkilendirilebilir.
- **Nörotisizm (Neuroticism)**: Endişeli, güvensiz, kendiyle uğraşan, sinirli, kaygılı gibi özellikler bu boyut altında değerlendirilmektedir.
- **Gelişime açıklık (Openness)**: Analitik, karmaşık, meraklı, bağımsız, yaratıcı, liberal, geleneksel olmayan, hayal gücü kuvvetli, ilgileri geniş, değişikliği seven, açık fikirli gibi özellikler bu boyutla ilişkilendirilebilir.

MBTI'da ise kişiliği belirleyen dört boyut yer almaktadır. Bu boyutlar:

- **İçedönük/Dışadönük (Introversion/Extraversion)**: I (introversion) özelliği taşıyan kişiler içedönük bir karaktere, E (extraversion) özelliği taşıyanlar ise dışadönük bir karaktere sahiptirler.

- **Sağduyulu/Sezgisel** (*Sensing/Intuition*): Kişilerin bilgiyi ne şekilde aldıklarını temsil eder. S (sensing) kategorisindeki kişiler beş duyu organlarını kullanarak adım adım yani parçalar halinde bilgiyi işlediklerini, N (intuition) kategorisindekiler ise bilgiyi bir bütün olarak incelediklerini ifade eder.
- **Düşünen/Hisseden** (*Thinking/Feeling*): Bu boyut insanların karar verirken nasıl davrandıklarını göstermektedir. F (feeling) kategorisindekiler bir durum karşısında empati kurarak, olay içerisinde herkesin bakış açısından bakarak olayı anlamaya çalışırken; T'ler (thinking) kendilerinin olayla olan ilişkisini keserek mantıklı ve akılcı bir şekilde karar vermeye çalışırlar.
- **Yargılayan/Algılayan** (*Judging/Perceiving*): Bu grup, insanların olaylar karşısında tepki gösterirken nasıl davrandıklarıyla ilişkilidir. P (perceiving) grubundaki insanlar, bir aksiyon almadan önce olabildiğince bilgi toplayıp ondan sonra hareket ederlerken; J (judging) grubundakiler bir olay karşısındaki izlenimlerinden yola çıkarak herhangi bir veri toplamadan ekstrasdan, direkt harekete geçerler.

MBTI ile ilgili kişilik testleri çözen kişilerin; bu boyutların kombinyasyonları (INTP, ESTJ, INTJ gibi) ve bu kombinyasyonların ait olduğu sınıflarla (analizci, kaşif, diplomat, gözcü) kişilik tipleri ifade edilir. Analizci sınıfına ait olan kimseler N ve T; diplomat sınıfına ait olanlar N ve F; kaşif sınıfındakiler S ve P; gözcü ise S ve J özelliklerini taşımaktadırlar.

Bu alandaki çalışmalarda veri setinin oluşturulması en önemli aşamalardan bir tanesidir. Gizlilik, verilerin etiketlenme maliyetinin yüksek olması, metinleri oluşturan kişilerin anonim olmadıkları için gerçek duygu ve düşüncelerini yansıtmaması gibi problemler düzgün bir veri setinin oluşturulmasına engel olmaktadır [22].

Türkçe herhangi bir veri seti olmamasından ötürü bu çalışma kapsamında öncelikle sosyal medya aracılığıyla Türkçe veri seti elde edilmiş ardından da makine öğrenmesi teknikleriyle kullanıcıların MBTI kişilik tipleri tahmin edilmiştir.

Makalenin ikinci bölümde bu alandaki diğer çalışmalardan, üçüncü bölümde veri setinin nasıl oluşturulduğu ve özellikleri, dördüncü bölümde geliştirilen sistemden, beşinci bölümde deneysel sonuçlardan, altıncı bölümde ise tartışmadan ve sonuçlardan bahsedilmektedir.

II. LİTERATÜR TARAMASI

Türkçe dilinde MBTI kişilik tipi tespiti yapan başka bir çalışmaya rastlanılmamıştır. Türk araştırmacıların da Türkçe dili yerine İngilizce'yi tercih ettikleri gözlemlenmiştir [23].

Sunulan çalışmadan farklı bir amaca sahip kültürler arası Beş Faktör Kişilik tipini tahmin ederken Türkçe'yi kullanan bir araştırma bulunmuştur [24]. Ancak bu çalışma içerisinde herhangi bir sonuca yer verilmemiştir. Bu durum kişilik tespiti alanındaki çalışmalarla karşılaştırılamamasına neden olmaktadır.

Argamon vd. kelimelerin anlamsal kök özelliklerini (lexical features) kullanarak öğrenci kompozisyonlarından kişilik tipi tespiti yapmıştır [25]. Kompozisyonların, öğrencilerden bilinç akışlarını ve kendi analizlerini yaparak yazılması istenmiştir.

Bu yaklaşımın problemi, Kim Lucykx ve Walter Daelemans tarafından da belirtildiği üzere, öğrencilerin çalışmanın bir parçası olduklarını bilmelerinden ötürü, kendilerini gerçekten farklı göstererek kompozisyonları yazmış olabilmeye ihtimalidir [26].

Daelemans vd. kendi çalışmalarında bu sorunu giderebilmek için, 145 tane öğrenciye önce bir belgesel izletip belgesel hakkında 1400 kelimelik bir kompozisyon yazmalarını istemiştir. Ardından öğrenciler MBTI kişilik testi çözmüştür. TiMBL (Memory Based Learning) ile kişilik tipleri tahmin edilmiştir. Veri setinin kısıtlı bir konu üzerinde metinler içermesinden ötürü bu çalışmanın gerçek hayata uygulanabilirliği soru işareti oluşturmaktadır.

Giulio Carducci vd. kelime vektörlerini kullanarak makine öğrenmesi teknikleriyle Beş Faktör Kişilik tiplerini tahmin etmişlerdir [27]. Bharadwaj vd. TF-IDF, EmoSentNet, LIWC ve ConceptNet özelliklerini birleştirerek Naive Bayes, SVM ve Sinir Ağları kullanarak MBTI kişilik tiplerini tahmin etmiştir [28]. Verhoeven vd. kelime ve karakter n-gramları kullanarak 6 farklı dilde MBTI SVM ve logistic regression ile tahminde bulunmuşlardır [29].

III. VERİ SETİ

Bu alandaki çalışmalar incelendiğinde veri seti oluşturma metodlarını "test çözdürterek", "kendini kendini etiketleme" ve "uygun kişilik tipi ile etiketleyerek" olarak sınıflandırabiliriz.

Test çözdürme yönteminde öncelikle katılımcılar belirlenir. Ardından çevrimiçi kişilik tipi testleri (16personalities.com gibi) ya da uzmanlar yardımıyla hazırlanan testin çözülmesi ile kişilik tipleri belirlenir. Katılımcılardan metin oluşturmaları [25], [26] ya da sosyal medyadaki paylaşımları [30] istenir.

Kendi kendini etiketleme, sosyal medya üzerinden veri seti oluşturulduğunda sıkça tercih edilen bir yöntemdir. Bu yöntemde öncelikle sosyal medyada kişilik tiplerini belirten kullanıcılar bulunur. Ardından da onların paylaşımlarından veri seti oluşturulur [22], [29], [31].

Uygun kişilik tipi ile etiketleyerek oluşturulan veri setlerinde kişilik tipleri ünlü kişilerin hangi tipten olduğunu belirten forum siteleri (www.personality-database.com www.personalitycafe.com vb.) kullanılarak elde edilir [23], [32]. Bu tarz sitelerdeki tipler, tahmine dayanmaktadır. Kişilik tipleri, siteye üye olanlar tarafından oylanarak ya da uygun sınıf ile etiketlendirilerek belirlenmektedir. Daha sonrasında da bu kişilerin sosyal medya paylaşımları kullanılarak veri seti elde edilir.

Çalışma kapsamında kullanılan veri seti "kendi kendini etiketleme" yöntemi kullanılarak oluşturulmuştur. EksiSözlük (eksisozluk.com) sitesindeki 16personalities.com başlığında kullanıcılar, çözmüş oldukları testin sonuçlarını paylaşmışlardır. Paylaşımları kullanılarak öncelikle kullanıcı isimleri belirlenmiştir. Sonrasında ise EksiSözlük'teki paylaşımların indirilebilmesi için yazılan crawler ile metinler elde edilerek veri seti oluşturulmuştur.

Veri seti toplamda 647 kullanıcı 643912 metinden (entry) oluşmaktadır. İçedönük etiketli metin sayısı 320654, dışadönük 323268; sezgisel 462663, sağduyulu 181249; düşünen 343529, hisseden 300383; yargılayan 312543, kavrayan ise 331369'dur.

IV. SİSTEMİN MODELLENMESİ

Sistem altı aşamadan oluşmaktadır: Crawling, Eleme, Ön İşleme, Özellik Çıkarımı, Modelin Oluşturulması ve Modelin Değerlendirilmesi.

A. Crawling

Bu aşamada EkşiSözlük'ten daha önceden belirlenmiş olan kullanıcıların yazdıkları metin içerikleri indirilmektedir. Oluşturulan veri seti "Eleme" adımına girdi olarak verilmektedir. NodeJS ve Puppeteer kullanılarak yazılmıştır.

B. Eleme

bkz, *spoiler*, *#* gibi anahtar kelimeleri içeren metin içerikleri veri setinden çıkarılır. Bunun sebebi bu anahtar kelimeler başka olaylara, yazılara işaret etmek için kullanılır. Dolayısıyla bu tarz girdiler herhangi bir kişisel görüş bildirmeyip bir olaya atıfta bulunurlar ya da işaret ederler. Kişisel bir görüş içermemektedir. Kişilik tipinin belirlenmesinde önemli bir rol oynamadığı düşünülmüştür. Bu aşamanın çıktısı, ön işleme aşamasında girdi olarak kullanılmaktadır.

C. Ön İşleme

Bu aşamada üç farklı ön işleme gerçekleştirilerek üç farklı veri seti çıktısı elde edilir. Bu çıktılar hangi ön işleme adımlarının daha iyi sonuç verdiğinin anlaşılması için karşılaştırmalı olarak kullanılacaktır. Uygulanan ön işleme adımları aşağıda yer almaktadır.

- Ön İşleme-1: Büyük harflerin küçük harflere dönüştürülmesi; web sitesi linklerinin, noktalama işaretlerinin, rakamların, etkisiz kelimelerin (stop words) temizlenmesi
- Ön İşleme-2: İlk ön işlemeye ek olarak kök bulma (stemming) uygulanır. Python'da snowballstemmer kütüphanesinde yer alan TurkishStemmer kullanılmıştır. Stemming ile kelimelerin anlamsal açıdan mantıklı olması gerekmeyen dil bilgisi açısından doğru olan kökleri bulunmuştur.
- Ön İşleme-3: İlk ön işlemeye ek olarak Zemberek kullanılarak normalizasyon ve lemmatization işlemleri uygulanmıştır. Sosyal medyada kullanıcıların, dili düzgün kullanmayarak oluşturmuş oldukları içerikleri düzgün bir hale getirebilmek adına normalizasyon kullanılmıştır. Lemmatization ise anlamsal kök bulmak için kullanılmıştır.

D. Özellik Çıkarımı

Ön işleme adımıyla oluşturulan veri setlerini kullanarak TF ve TF-IDF özellik vektörleri çıkarılmıştır.

E. Modelin Oluşturulması

Naive Bayes ve FastText kullanılarak model oluşturulur. FastText'i kullanarak eğitim setinden model eğitilir, sonrasında da test seti ile sonuçlar elde edilir.

F. Modelin Değerlendirilmesi

Karmışıklık matrisi oluşturularak duyarlılık (precision), anma (recall) sonuçları ve harmonik ortalaması alınarak F-Score hesaplanır.

V. DENEYSEL SONUÇLAR

Elde edilen en iyi sonuçlar "Ön İşleme-1" ile oluşturulan veri seti kullanılarak FastText'in varsayılan parametreleri ile oluşturulan model ile alınmıştır. "Ön İşleme-2" ile oluşturulan veri setinden Terim Frekansları (TF) özellik olarak çıkarılıp Naive Bayes tahmin edilmesi bir diğer başarılı yöntem olmuştur. Tablo 1'de Naive Bayes, Tablo 2'de FastText ile ilgili sonuçlar yer almaktadır.

Tablo 1: Naive Bayes Ön İşleme - 2 TF

	Duyarlılık	Anma	F-Ölçüm
Analizciler	0.4038	0.7723	0.5303
Diplomatlar	0.4837	0.4112	0.4445
Kaşıfçılar	0	0	0
Gözcüler	0.4825	0.0082	0.0161
I (İçedönük)	0.5471	0.7357	0.6275
E (Dışadönük)	0.6114	0.4057	0.4877
S (Sezgisel)	0.3888	0.0007	0.0013
N (Sağduyulu)	0.7193	0.9995	0.8365
T (Düşünen)	0.5773	0.8064	0.6728
F (Hisseden)	0.6219	0.3503	0.4481
J (Yargılayan)	0.5599	0.5139	0.5359
P (Kavrayan)	0.5693	0.6140	0.5908

Tablo 2: FastText Ön İşleme - 1

	Duyarlılık	Anma	F-Ölçüm
Analizciler	0.4631	0.5544	0.5046
Diplomatlar	0.4792	0.5246	0.5008
Kaşıfçılar	0.2240	0.1104	0.1479
Gözcüler	0.3049	0.2164	0.2531
I (İçedönük)	0.5869	0.5962	0.5915
E (Dışadönük)	0.6003	0.5910	0.5956
S (Sezgisel)	0.4049	0.2341	0.2966
N (Sağduyulu)	0.7405	0.8640	0.7974
T (Düşünen)	0.6126	0.6591	0.6349
F (Hisseden)	0.5914	0.5421	0.5656
J (Yargılayan)	0.5825	0.5733	0.5778
P (Kavrayan)	0.5981	0.6072	0.6026

FastText kullanıldığında Naive Bayes ile %0 F-Ölçüm değerine sahip olan "Kaşıfçılar" %14'e, F-Ölçüm değeri %0.01 olan "Gözcüler" %25'e ve "Sezgisel" %0.13'ten %29.66'ya yükselmiştir. FastText ile sonuçlar daha dengeli dağılmıştır. En doğru tahmin edilen etiket "Sağduyulu" iken, "Sezgisel" ise en düşük ölçüme sahip etikettir. Aynı boyutta yer alan birbirine zıt olan "İçedönük" ve "Dışadönük" etiketlerinin ölçümleri birbirine çok yakındır.

VI. SONUÇ

Bölüm 5'teki sonuçlar gözden geçirildiğinde FastText ile elde edilen sonuçların daha dengeli dağıldığı görülmektedir. Tablo 2'de görülebildiği üzere en iyi ölçüm N (sağduyulu) boyutundadır. N ve S (sezgisel) boyutlarının veri setinde dengeli dağılmamasından mı yoksa N'in S'ye göre daha ayırt edici bir özellik olmasından ötürü mü olduğu tam olarak anlaşılamamaktadır.

Kişilik tiplerinin dört sınıftan hangi birine ait olduğunu tahmin etmeye çalışıldığında modelin başarı oranı düşmektedir. Bu durum dört ana sınıfın tahmin edilmesinin kolay olmadığını göstermektedir. Ön işlemenin direkt olarak sonuca etki ettiği gözlemlenmiştir. Ancak Türkçe dili için açık kaynaklı doğal dil işleme araçlarının kısıtlı olması, ön işlemede izlenebilecek adımları da kısıtlamaktadır. Türkçe dilinde MBTI kullanılarak kişilik tespiti daha önceden yapılmadığı için karşılaştırılabilecek herhangi bir çalışma bulunmamaktadır.

Kelime vektörleri ve derin öğrenme teknikleri kullanılarak elde edilen sonuçlar iyileştirilebilir. Özellikle vektörleri oluşturulurken çıkarılan özellikler içerisinde kelimelerin hangi duyguları taşıdığı ile ilgili bir vektör olması sonuçlara olumlu yönde etki edecektir. Kişiliklerin insanların duygularıyla paralellik göstermesi bu beklentinin altında yatan temel sebeptir. Bu amaçla kullanılan LIWC ve EmoSentNet diğer dillerde başarıya olumlu etki etmiştir [28]. EmoSentNet, bir kelimedenden sekiz duygu (kızgınlık, korku, sevgi, güven, şaşkınlık, hüznün, neşe, tikslenme) ve iki duygu durumu (pozitif ve negatif) için özellik çıkartmaktadır [33]. LIWC ise kullanılan kelimelerin düşünce, hissiyat ve kişilik ile ilgili taşıdıkları anlamların çıkartılmasında kullanılmaktadır [34].

KAYNAKLAR

- [1] I. B. Myers and P. B. Myers, "Gifts differing: Understanding personality type." Davies-Black Publishing, 2002.
- [2] L. R. Goldberg, "An alternative "description of personality": The big-five factor structure," *Journal of Personality and Social Psychology*, vol. 59, no. 6, p. 1216–1229, 1990.
- [3] *The Cambridge Handbook of Personality Psychology*, ser. Cambridge Handbooks in Psychology. Cambridge University Press, 2009.
- [4] M. R. Barrick and M. K. Mount, "The big five personality dimensions and job performance: A meta-analysis," *Personnel Psychology*, vol. 44, no. 1, pp. 1–26, 1991.
- [5] K. B. Ford, "Brands laid bare: Using market research for evidence-based brand management," 2005.
- [6] R. Dong, M. P. O'Mahony, M. Schaal, K. McCarthy, and B. Smyth, "Sentimental product recommendation," *Proceedings of RecSys '13: 7th ACM conference on Recommender systems*, 2013.
- [7] H. Feng and X. Qian, "Mining user-contributed photos for personalized product recommendation," *Neurocomputing*, vol. 129, pp. 409–420, 2014.
- [8] Adar, Eytan, Weld, D. S., Bershad, B. N., Gribble, and S. S., "Why we search: visualizing and predicting user behavior," 2007.
- [9] X. Shi, J. Zhu, R. Cai, and L. Zhang, "User grouping behavior in online forums," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. Association for Computing Machinery, 2009, p. 777–786.
- [10] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang, "Modeling user posting behavior on social media," in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '12. Association for Computing Machinery, 2012, p. 545–554.
- [11] R. Y. K. Lau, Y. Xia, and Y. Ye, "A probabilistic generative model for mining cybercriminal networks from online social media," *IEEE Computational Intelligence Magazine*, vol. 9, no. 1, pp. 31–43, 2014.
- [12] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from twitter posts," in *Proceedings of the 5th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction*, ser. SBP'12. Springer-Verlag, 2012, p. 231–238.
- [13] A. J. Gill, "Personality and language: the projection and perception of personality in computer-mediated communication," 2004.
- [14] A. J. Gill and J. Oberlander, "Taking care of the linguistic features of extraversion," 2002.
- [15] R. S. Campbell and J. W. Pennebaker, "The secret life of pronouns: Flexibility in writing style and physical health," *Psychological Science*, vol. 14, pp. 60–65, 2003.
- [16] E. Mergenthaler, "Emotion–abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes," *Journal of Consulting and Clinical Psychology*, vol. 64, no. 6, pp. 1306–1315, 1996.
- [17] J. W. Pennebaker and L. A. King, "Linguistic styles: Language use as an individual difference," *Journal of Personality and Social Psychology*, vol. 77, no. 6, p. 1296–1312, 1999.
- [18] S. Argamon, S. Dhawle, M. Koppel, and J. W. Pennebaker, "Lexical predictors of personality type," in *In Proceedings Of The Joint Annual Meeting Of The Interface And The Classification Society Of North America*, 2005.
- [19] R. R. McCrae and P. T. Costa Jr., "Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality," *Journal of Personality*, vol. 57, no. 1, pp. 17–40, 1989.
- [20] A. Furnham, "The big five versus the big four: the relationship between the myers-briggs type indicator (mbti) and neo-pi five factor model of personality," *Personality and Individual Differences*, vol. 21, no. 2, pp. 303 – 307, 1996.
- [21] O. Somer, M. Korkmaz, and A. Tatar, "Beş faktör kişilik envanteri'nin geliştirilmesi i: Ölçek ve alt Ölçeklerin oluşturulması," *Turkish Journal of Psychology*, vol. 17, pp. 21–33, 01 2002.
- [22] M. Gjurmović and J. Šnajder, "Reddit: A gold mine for personality prediction," in *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*. Association for Computational Linguistics, 2018, pp. 87–97.
- [23] K. Bastem and Şadi Evren Şeker, "Veri madenciliği yöntemleri ile twitter Üzerinden mbti kişilik tipi analizi," *YBS Ansiklopedisi*, vol. 4, no. 2, pp. 1–11, 2017.
- [24] C. Catal, M. Song, C. Muratlı, H.-J. Kim, M. Tosuner, and Y. Kayıkci, "Cross-cultural personality prediction based on twitter data," *Journal of Software*, vol. 12, pp. 882–891, 2017.
- [25] S. Argamon, D. S. M. Koppel, and J. Pennebaker, "Lexical predictors of personality type," 2005.
- [26] K. Luyckx and W. Daelemans, "Personae: a corpus for author and personality prediction from text," in *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA), 2008.
- [27] G. Carducci, G. Rizzo, D. Monti, E. Palumbo, and M. Morisio, "Twitpersonality: Computing personality traits from tweets using word embeddings and supervised learning," *Information*, vol. 9, p. 127, 2018.
- [28] S. Bharadwaj, S. Sridhar, R. Choudhary, and R. Srinath, "Persona traits identification based on myers-briggs type indicator(mbti) - a text classification approach," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2018, pp. 1076–1082.
- [29] B. Verhoeven, W. Daelemans, and B. Plank, "Twisty: A multilingual twitter stylometry corpus for gender and personality profiling," in *Proceedings of the 10th edition of the Language Resources and Evaluation Conference*. European Language Resources Association (ELRA), 2016.
- [30] M. S. Salem, S. S. Ismail, and M. Aref, "Personality traits for egyptian twitter users dataset," in *Proceedings of the 2019 8th International Conference on Software and Information Engineering*, ser. ICSIE '19, 2019, p. 206–211.
- [31] B. Plank and D. Hovy, "Personality traits on twitter—or—how to get 1,500 personality tests in a week," *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2015.
- [32] A. Ma, "Neural networks in predicting myers brigg personality type from writing style," 2017.
- [33] L. Canales, C. Strapparava, E. Boldrini, and P. Martínez-Barco, "Intensional learning to efficiently build up automatically annotated emotion corpora," 2017.
- [34] Y. Tausczik and J. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, pp. 24–54, 2010.