

# SIU2023-NST - Nefret Söylemi Tespit Yarışması

## SIU2023-NST - Hate Speech Detection Contest

İnanç Arın<sup>1,2</sup>, Zeynep Işık<sup>1,2</sup>, Seçilay Kutal<sup>1,2</sup>, Somaiyeh Dehghan<sup>1,2</sup>, Arzucan Özgür<sup>3</sup>, Berrin Yanıkoğlu<sup>1,2</sup>

<sup>1</sup>Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey 34956

<sup>2</sup>Center of Excellence in Data Analytics (VERIM), Sabanci University, Istanbul, Turkey 34956

<sup>3</sup> Department of Computer Engineering, Bogazici University, Turkey

**Özetçe** —Sosyal medyada yayılan nefret söylemleri, özellikle din, kültür açısından belirli bir insan grubunu veya mültecilik gibi belirli bir sosyal durumu hedef aldığı anda, insanları ve toplumları güçlü bir şekilde etkileyebilir. Bu nedenle son yıllarda sosyal ağlarda nefret söyleminin tespiti ve ortadan kaldırılması konuları doğal dil işleme araştırmacılarının ilgisini çekmektedir. Bu bağlamda, SIU2023 konferansının Hesaplamalı Sosyal Bilimler özel oturumu çerçevesinde ödüllü Türkçe nefret söylemi tespit yarışması düzenlenmiştir. Bu yarışma dört farklı alt-problem üzerinde gerçekleştirilmiştir. Bu yarışmaya 20 takım kaydolmuş, sonuçta sekiz takım sonuç yollamıştır. En yüksek başarıya sahip olan takımların sistemlerinin ayrıntıları ve sonuçları bildiri şeklinde anlatılmıştır.

**Anahtar Kelimeler**—Nefret Söylemi Tespiti, Doğal Dil İşleme, Derin Öğrenme, Türkçe Dili

**Abstract**—Hate speech spread on social media can strongly affect people and societies, especially when it targets a specific group of people in terms of religion, culture, or a specific social situation, such as refugees. For this reason, the detection and elimination of hate speech in social networks have attracted the attention of natural language processing researchers in recent years. A competition is organized to benchmark progress in Turkish hate speech recognition, within SIU2023 Computational Social Sciences special session, with four different tasks. A total of 20 teams registered for the competition, while eight teams submitted results at the end. The details of the winning models and their results are explained in this paper.

**Keywords**—Hate Speech Detection, Natural Language Processing, Deep Learning, Turkish Language

## I. GİRİŞ

Bir kişiyi veya grubu; etnik köken, din, dil, cinsiyet, cinsel yönelim, yaş, engellilik gibi kimlik özellikleri temelinde ayırarak hedef alan ifadeler nefret söylemi olarak adlandırılmaktadır. Bu söylemler önyargıya dayalı, olumsuz ve saldırgan söz içerikli, yazılı veya görsel ifadelerdir [1].

Nefret söylemi tespiti, bir metin sınıflandırma görevi olarak, doğal dil işlemenin (NLP) ana uygulamalarından biridir. Nefret söylemi tespiti için geleneksel veya derin sınıflandırıcılara veya her iki yaklaşımın bir kombinasyonuna dayalı çeşitli yöntemler kullanılmaktadır. Bu yöndeki ilk adım, kelimeleri veya cümleleri vektör biçiminde temsil etmektir. Bunun için genellikle, geleneksel kelime torbası (BOW-Bag of Words) ve terim frekansı-ters doküman frekansı (TF-IDF) yöntemleri, Word2vec [2], Glove [3] ve Fasttext [4] gibi kelime gömme yöntemleri veya ELMo [5], USE [6], BERT [7], SBERT [8]

gibi dönüştürücü (Transformer) tabanlı dil modelleme yöntemleri kullanılmaktadır.

SIU'2023 kapsamında düzenlenen özel oturumun amacı, nefret söylemi konusundaki çalışmaların tartışılabileceği bir ortam yaratarak konu üzerinde sinerji yaratmaktır. Oturumda nefret söyleminin otomatik tespiti ve analizi için geliştirilen yöntemler değerlendirilerek, Türkçe nefret söylemi tespiti konusunda gelinen seviyenin tespit edilmesi de hedeflenmiştir. Bu amaçlar ışığında, Türkçe metinlerde nefret söylemi tespiti için yenilikçi yöntemlerin geliştirilmesinin özendirilmesi ve değerlendirilmesi için bir imce yarışması (shared task) düzenlenmiştir.

## II. PROBLEM TANIMLARI VE VERİ SETİ

Düzenlenen yarışma dört alt-problemden oluşmaktadır. Yarışmada kullanılan veri seti üç farklı konuda (İsrail-Filistin çatışması, Yunan karşıtlığı ve Göçmen/Mülteci karşıtlığı) toplanmış Türkçe tweet'leri içermektedir. Bu problemler II-A'da ve onlar için kullanılan veri setleri II-B'de anlatılmıştır.

### A. Problem Tanımları

1) *Problem-1: İsrail-Filistin Çatışması Konusunda Nefret Söylemi Kategorisinin (Beş-sınıf) Tespiti:* Bu problem kapsamında katılımcılardan konu özelindeki tweet'leri "Nefret Söylemi Yok", "Abartma/Genelleme/Çarpıtma", "Simgeleştirme", "Küfür/Hakaret/Aşağılama", "Düşmanlık/Saldırma/Yaralama Tehdidi" olmak üzere beş kategori üzerinden tasnifleyecek bir sınıflandırıcı geliştirmeleri beklenmiştir.

2) *Problem-2: Mülteci Karşıtlığı Üzerinde Nefret Söylemi İkili Sınıflandırması:* Bu problem kapsamında katılımcılardan konu özelindeki tweet'leri "Nefret Söylemi Yok" ve "Nefret Söylemi Var" olmak üzere iki sınıf üzerinden ayrıştırarak bir sınıflandırıcı geliştirmeleri beklenmiştir.

3) *Problem-3: İsrail-Filistin Çatışması Konusunda Nefret Söylemi Şiddetinin Tahmini:* Bu problem kapsamında katılımcılardan konu özelindeki tweet'lerdeki nefret söylemi şiddetini 0-4 aralığında tahminleyen bir regresyon modeli geliştirmeleri beklenmiştir.

4) *Problem-4: Mülteci Karşıtlığı Üzerinde Nefret Söylemi Şiddetinin Tahmini:* Problem-3'teki gibi, konu özelindeki tweet'lerdeki nefret söylemi şiddetini 0-4 aralığında tahminleyen bir regresyon modeli geliştirmeleri beklenmiştir.

### B. Eğitim Verisi

Eğitim verisi, Problem-1 için 2240, Problem-2,4 için 4683 ve Problem-3 için 2272 tweet verisinden oluşmaktadır. Tweet'ler Twitter kuralı gereği tweet id ve tweet'e karşılık gelen etiket şeklinde paylaşılmıştır. Katılımcılar verilen tweet id ile TwitterAPI kullanarak tweet verilerini elde etmişlerdir. Örneklem dağılımları TABLO I, II ve III'te gösterilmiştir.

TABLO I: İSRAİL-FİLİSTİN ÇATIŞMASI NEFRET SÖYLEMİ KATEGORİSİ VE ŞİDDETİ - EĞİTİM VERİSİ

Kategori	Veri Sayısı	Şiddet	Veri Sayısı
0:Nefret Söylemi Yok	1360	0	1438
1:Abartma/Genelleme/Çarpıtma	50	1	162
2:Simgeleştirme	231	2	255
3:Küfür/Hakaret/Aşağılama	301	3	314
4:Düşmanlık/Saldırma/Yaralama Tehdidi	298	4	103
Toplam	2240	-	2272

TABLO II: MÜLTECİ KARŞITLIĞI NEFRET SÖYLEMİ ŞİDDETİ VE VARLIĞI - EĞİTİM VERİSİ

Şiddet	Veri Sayısı	Kategori	Veri Sayısı
0	3493	0:Yok	3493
1	723	1:Var	1190
2	329		
3	110		
4	28		
Toplam	4683	-	4683

Yunan karşıtlığı konusunda bir problem tanımlanmamıştır. Bu veri seti eğitim aşamasında ihtiyaç olursa nefret söylemi kategorisinin ve şiddetinin tespiti problemlerinde kullanılmak üzere paylaşılmıştır. Yunan karşıtlığı veri seti nefret söylemi kategorisi ve şiddeti etiketlerine sahip 976 tweet verisinden oluşmaktadır.

TABLO III: YUNAN KARŞITLIĞI NEFRET SÖYLEMİ KATEGORİSİ VE ŞİDDETİ - EĞİTİM VERİSİ

Kategori	Veri Sayısı	Şiddet	Veri Sayısı
0:Nefret Söylemi Yok	555	0	610
1:Abartma/Genelleme/Çarpıtma	145	1	40
2:Simgeleştirme	34	2	163
3:Küfür/Hakaret/Aşağılama	85	3	111
4:Düşmanlık/Saldırma/Yaralama Tehdidi	157	4	52
Toplam	976	-	976

### C. Test Verisi

Test verisi, Problem-1 için 571, Problem-2,4 için 1171 ve Problem-3 için 539 tweet verisinden oluşmaktadır. Örneklem dağılımları TABLO IV ve V'de gösterilmiştir.

TABLO IV: İSRAİL-FİLİSTİN ÇATIŞMASI NEFRET SÖYLEMİ KATEGORİSİ VE ŞİDDETİ - TEST VERİSİ

Kategori	Veri Sayısı	Şiddet	Veri Sayısı
0:Nefret Söylemi Yok	498	0	491
1:Abartma/Genelleme/Çarpıtma	4	1	9
2:Simgeleştirme	17	2	23
3:Küfür/Hakaret/Aşağılama	22	3	10
4:Düşmanlık/Saldırma/Yaralama Tehdidi	30	4	6
Toplam	571	-	539

TABLO V: MÜLTECİ KARŞITLIĞI NS ŞİDDETİ VE VARLIĞI - TEST VERİSİ

Şiddet	Veri Sayısı	Kategori	Veri Sayısı
0	873	0:Yok	873
1	181	1:Var	298
2	82		
3	28		
4	7		
Toplam	1171	-	1171

### D. Başarım Ölçüm Metrikleri

Sınıflandırma problemleri (Problem-1,2) başarımları Makro F-1 Puanı ile, regresyon problemleri (Problem-3,4) başarımları ise ortalama kare hata ölçütü ile değerlendirilmiştir.

## III. KATILIMCILAR VE METODLAR

SIU2023-NST yarışmasına çeşitli üniversite ve kurumlardan toplam 20 ekip kaydolmuş ve bunlardan sekizi alt-problemlerin biri veya daha fazlasına sonuç yüklemiştir. Dört problemden herhangi birinde ilk üç dereceye giren sistemlerin ayrıntılı açıklamaları aşağıda verilmiştir:

**Problem-1 ve Problem-3 Birincisi (Grup1):** Sistemin temelinde yapay öğrenme modeli olarak ön eğitilmiş BERTurk<sup>1</sup> modeli kullanılmış ve probleme adapte edilmiştir. Eğitim için *Huggingface* kaynakları kullanılmıştır. Model seçiminde beş katlı çapraz doğrulama yöntemi kullanılmış ve her bir kat için model üç devir eğitilmiştir. Çapraz doğrulama katlarından gelen olasılıkların ortalaması alınarak, topluluk yaklaşımı ile değerlendirilmiştir. En yüksek olasılığa sahip olan sınıf tahmin sonucu olarak atanmıştır. Bu sistem Problem-1 ve Problem-3'te birinci, diğerlerinde sırasıyla üçüncü ve ikinci olmuştur.

**Problem-1, -2 ve -3 İkincisi (Grup2):** Bu sistem, tamamen açık kaynaklı toplanan veri ile önce beş devir (epoch), daha sonrasında aynı veri ve yarışma için sağlanan Yunanistan verileri ile birleştirilip de iki devir eğitilmiştir. Bazı model olarak AçıkHack 2023 yarışması birincisinin paylaştığı model alınmıştır<sup>2</sup>. Daha sonrasında alt-problem özelinde eğitimler yapılmıştır. Bu eğitimde ise Yunanistan verisi ilgili probleme göre kategorize edilip iki devir eğitim yapılmıştır. Daha sonrasında eğitim verisi olarak sağlanan veriler yedi devir boyunca eğitilerek test veri seti üzerinde sonuçlar alınmıştır. Bu sistem ilk üç problemde ikinci, sonuncu problemde üçüncü sırada yer almıştır.

**Problem-1 Üçüncüsü (Grup3):** Sistemin temelinde iki farklı çözüm tekniği kullanılmıştır. İlk çözüm yöntemi kelime ve karakter tabanlı kelime torbası olup, XGBoost [9] ve CatBoost [10] modelleri eğitilmiş ve beş katlı (k-fold) değerlendirme yaklaşımı benimsenmiştir. Sonuç olarak her bir kat için elde edilen olasılıkların ortalaması alınmıştır. İkinci çözüm yönteminde ise BERTurk-uncased<sup>3</sup> algoritması kullanılmış, beş katlı değerlendirme tekniği ile değerlendirilmiş, ve son aşamada bu modellerden gelen sonuçlar, performansları baz alınarak hesaplanmış ağırlıklar ile ağırlıklandırılarak, bir komite yaklaşımı benimsenmiştir.

<sup>1</sup><https://huggingface.co/dbmdz/bert-base-turkish-cased>

<sup>2</sup><https://huggingface.co/nanelimon/bert-base-insult-model/tree/main>

<sup>3</sup><https://huggingface.co/dbmdz/bert-base-turkish-uncased>

**Problem-2 ve Problem-4 Birincisi (Grup4):** Sisteminin temelinde yapay öğrenme modeli olarak ön eğitilmiş BERTurk modeli kullanılmıştır. BERTurk modeli ince ayar yapılarak ilgili probleme uyumlu hale getirilmiştir. Mülteci karşıtlığı veri seti üzerinde herhangi bir ön işlem ve öznitelik çıkarımı yapılmadan, veri, modele ait kelime bölücü ile sayısallaştırılıp modele verilmiştir. Sonrasında ince ayar ile model, mevcut probleme adapte edilmiştir. Veri seti, etiketlerin dağılımına göre dengeli bir şekilde 10 parçaya bölünüp, çapraz doğrulama yapılmıştır. Deneyler sürecinde en yüksek F-1 puanına sahip olan model ile test veri seti üzerinde sonuçlar alınmıştır. Bu sistem Problem-3'de ise 3. olmuştur.

**Problem-3 Birincisi (Grup1):** İsrail-Filistin veri setinde beş sınıflı sınıflandırma probleminde birinci olan grup, aynı veri setinde nefret söylemi şiddetinin tahmininde de birinci olmuştur. Burada tweet metinlerinden elde edilen TF-IDF öznitelikleri ve yapay öğrenme modeli olarak LightGBM [11] (Gradient Boosting Machine) kullanılmıştır. Çapraz doğrulama sonuçları baz alınarak optimize edilen sistemde, çapraz doğrulamadan gelen tahminlerin ortalaması sınıflandırıcı topluluğu çıktısı olarak kullanılmıştır.

**Problem-4 Birincisi (Grup4):** Mülteci Karşıtlığı veri setinde iki sınıflı sınıflandırma probleminde birinci olan grup, aynı veri setinde nefret söylemi şiddetinin tahmininde aynı yaklaşımı kullanmış ve birinci olmuştur. Farklı olarak, veri seti etiketlerin dağılımına göre dengeli bir şekilde sekiz parçaya bölünüp, çapraz doğrulama yapılmıştır. Deneyler sürecinde en düşük ortalama kare hataya sahip olan model ile test veri seti üzerinde sonuçlar alınmıştır.

#### IV. SONUÇLAR

Yarışmanın tüm alt-problemlerindeki ilk üç sonuç Tablo VI'de verilmiştir.

Bu tablonun ilk satırında görüleceği üzere, beş sınıflı nefret söylemi kategori tespiti zor bir problem olarak ortaya çıkmaktadır. Bu alt-problemde birinci olan Grup1 0.5758 makro F-1 değeri elde etmiş, ikinci olan Grup2, çok az farklı ikinci olmuştur. Karşılaştırma için tüm kategori tahminlerinin sıfır olarak yapıldığı bir referans sistemi oluşturulduğunda ise 0.1863 makro F-1 değeri elde edilmektedir. Dolayısıyla, dereceye giren modellerin referans modelden açık ara daha iyi olduğu görülmektedir. Bu alt-problemde birinci olan sistemin hata matrisi TABLO VII'de görülebilir. Görüleceği üzere, veri setinde nefret söylemi olmayan tweetler büyük çoğunlukta olsa da, model uç kategorileri (üç ve dört) de nispeten başarı ile sınıflandırabilmiştir.

TABLO VI: Yarışmanın Sonuçları. Problem-1 ve Problem-2 için F-1 puanı, Problem-3 ve Problem-4 için Ortalama Kare Hata Ölçüm Metrikleri Olarak Kullanılmıştır.

SIU2023-NST Alt-Problemleri	Birinci Model	İkinci Model	Üçüncü Model
Problem-1- İsrail-Filistin Çatışması Nefret Söylemi Kategorisi Tespiti (5-sınıf)	0.5758	0.5750	0.5682
Problem-2- Mülteci Karşıtlığı Nefret Söylemi Kategorisi Tespiti (2 sınıf)	0.7687	0.7562	0.7280
Problem-3- İsrail-Filistin Çatışması Nefret Söylemi Şiddetinin Tahmini	0.3161	0.3520	0.3547
Problem-4- Mülteci Karşıtlığı Nefret Söylemi Şiddetinin Tahmini	0.4470	0.4715	0.5448

TABLO VII: Problem-1'de Birinci Olan Sistemin Hata Matrisi

	Gerçek Değer				
	0	1	2	3	4
Tahmin Edilen 0	483	0	7	2	6
1	3	1	0	0	0
2	12	0	4	1	0
3	4	1	4	13	0
4	9	0	1	0	20

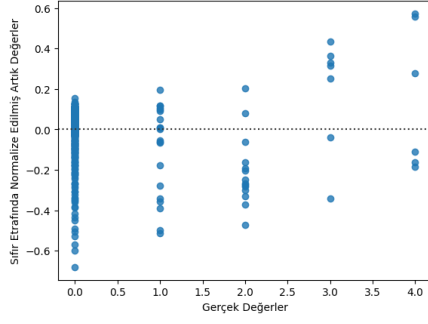
TABLO VIII: Problem-2'de Birinci Olan Sistemin Hata Matrisi

	Gerçek Değer	
	0	1
Tahmin Edilen 0	779	94
1	111	187

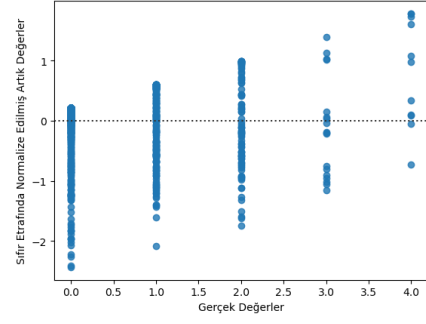
TABLO VI'nın ikinci satırı, görece basit olan iki sınıflı nefret söylemi kategori tespitine aittir. Bu alt-problemde birinci olan Grup4 0.7687 makro F-1 değeri elde etmiştir. Elde edilen skor, nefret söylemi tespitinin (var/yok) kabul edilebilir başarı oranı ile yapılabildiğini göstermektedir. Bu modelin hata matrisi TABLO VIII'de verilmiştir.

Üçüncü ve dördüncü alt-problemler nefret söylemi şiddetini tahmin etmeye yöneliktir. Üçüncü satırda gösterilen İsrail-Filistin alt-problemine ait sonuçlarda, birinci olan Grup1 0.3161 ortalama kare hata değeri elde etmiştir. Bu sonuca ait sıfır etrafında normalize edilmiş artık hata (residual) dağılımı Şekil 1'de verilmiştir. Buna ek olarak Şekil 2'deki QQ grafiğinde hataların normal bir dağılımdan uzaklaştığı görülmektedir. Bu sonuç, verideki bazı sistematik kalıp ve ön yargıların model tarafından kaçırılmış olabileceğine dair işaret olarak düşünülebilir. Bu veri setinde şiddet değeri sıfır olan çok sayıda örnek olduğundan, tüm tahminlerin sıfır olarak yapıldığı referans bir sistem oluşturulduğunda ise 0.4360 ortalama kare hata skoru elde edilmiştir. Bu sonuç ilk üç sıradaki modellerin referans modelden oldukça daha iyi olduğunu göstermektedir.

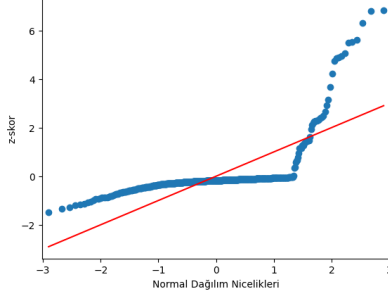
Tablo VI'nın dördüncü satırında mülteci karşıtlığı nefret söylemi şiddet tahmini sonuçları verilmiştir. Bu alt-problem bir önceki alt-probleme göre veri açısından nispeten daha dengeli dağılım göstermiş, birinci olan Grup4 0.4470 ortalama kare hata değeri elde etmiştir. Bir önceki alt-problemde olduğu gibi artık hata dağılımları Şekil 3'te gösterilmiştir. Buna ek olarak Şekil 4'te normalize edilmiş dağılımların QQ grafiği verilmiş ve normal dağılıma önceki probleme göre nispeten daha yakın olduğu gözlemlenmiştir. Aynı şekilde tüm tahminlerin sıfır olarak yapıldığı referans bir sistem oluşturulduğunda ise 0.7318 ortalama kare hata skoru elde edilmiş ve bu sonuç önceki alt-problemdeki gibi derece yapan modellerin referans modelden çok daha iyi olduğunu göstermektedir.



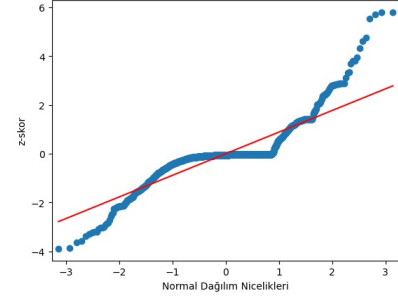
Şekil 1: Problem-3'te Birinci Olan Sistemin Gerçek ve Artık Hata Dağılımı



Şekil 3: Problem-4'te Birinci Olan Sistemin Gerçek ve Artık Hata Dağılımı



Şekil 2: Problem-3 Birincisi QQ Grafiği



Şekil 4: Problem-4 Birincisi QQ Grafiği

Düzenlenen yarışmalarla ilgili tüm sonuçlara ve her alt-problemin Kaggle linkine şu sayfadan erişilebilir: <https://myweb.sabanciuniv.edu/berrin/siu2023-nst/>

## V. ÖZET

Türkçe nefret söylemi tespiti konusundaki çalışmalarını değerlendirmek için SİU2023 konferansı Hesaplamalı Sosyal Bilimler özel oturumu çerçevesinde bir yarışma düzenlenmiştir. Yarışma dört farklı alt-problem üzerinde yapılmıştır. Problem-1 ve Problem-2'de nefret söylemi sınıflandırma; Problem-3 ve Problem-4'de ise nefret söylemi şiddet tahmini yapılmıştır.

Yarışmada ilk üçe giren sistemler dönüştürücü veya LightGBM modellerini kullanmış ve karşılaştırıldıklarında referans performans düzeyini rahatlıkla geçtikleri gözlenmiştir. Yine de, nefret söylemi tespiti probleminin henüz çözülmekten uzak olduğu söylenebilir. Bu açıdan hem daha büyük veri setlerine, hem de araştırmaya ihtiyaç duyulmaktadır.

## VI. TEŞEKKÜR

Bu proje Hrant Dink Vakfı tarafından yürütülen "Dijital Teknolojileri Kullanarak Nefret Söylemi ve Ayrımcılıkla Mücadele" (EuropeAid/170389/DD/ACT/Multi) isimli AB projesi ve 119E358 proje kodu ile TÜBİTAK tarafından desteklenmiştir.

## KAYNAKLAR

- [1] F. Beyhan, B. Çarık, I. Arın, A. Terzioğlu, B. Yanıkoğlu and R. Yeniterzi, "A Turkish Hate Speech Dataset and Detection System", In Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4177–4185, 20-25 June 2022.
- [2] T. Mikolov, K. Chen, G. Corrado, J. Dean, "Efficient estimation of word representations in vector space", *arXiv* 2013.

- [3] J. Pennington, R. Socher, C. Manning, "Glove: Global vectors for word representation", In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014.
- [4] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, "Enriching word vectors with sub word information", *Trans. Assoc. Comput. Linguist.*, pages 135–146, 2017.
- [5] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, "Deep contextualized word representations", In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, LA, USA, 1–6 June 2018.
- [6] D. Cer, Y. Yang, S. Kong, N. Hua, N. Limtiaco, R. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, "Universal Sentence Encoder for English", In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Brussels, Belgium, 31 October–4 November, pages 169–174, 2018.
- [7] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 3–5 June 2019; Volume 1, pages 4171–4186, 2019.
- [8] N. Reimers, I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert networks", In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019.
- [9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 785–794, 2016.
- [10] L. Prokhorenkova, G. Gusev, A. Vorobev, A. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features", *NeurIPS*, pages 6639–6649, 2018.
- [11] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "Lightgbm: A highly efficient gradient boosting decision tree", *Advances in Neural Information Processing Systems*, Volume 30, pages 3146–3154, 2017.