

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



DOĞAL DİL ANLAMA MODELLERİNİN DİYALOG SORULARININ
CEVAPLANMASI İÇİN DEĞERLENDİRİLMESİ

YÜKSEK LİSANS TEZİ

Oğuzhan KARAHAN

Bilgisayar Mühendisliği Ana Bilim Dalı

Bilgisayar Mühendisliği Programı

Aralık, 2020

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



DOĞAL DİL ANLAMA MODELLERİNİN DİYALOG SORULARININ
CEVAPLANMASI İÇİN DEĞERLENDİRİLMESİ

YÜKSEK LİSANS TEZİ

Oğuzhan KARAHAN

Bilgisayar Mühendisliği Ana Bilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Dr. Öğr. Üyesi AHMET GÜRHANLI

Aralık, 2020

YEMİN METNİ

Yüksek Lisans tezi olarak sunduğum “Doğal Dil Anlama Modellerinin Diyalog Sorularının Cevaplanması İçin Değerlendirilmesi” adlı çalışmanın, tezin proje safhasından sonuçlanmasına kadarki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurulmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya’da gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim. (29/09/2020)

Oğuzhan KARAHAN

ÖNSÖZ

Doğal dil anlama modellerinin değerlendirilmesi, teknolojik ilerlemenin önemli bir parçasıdır. Yeni dil modelleri ve mimariler için dikkate alınacak pek çok faktör olsa da hangisinin seçileceğini belirten bir standart bulunmamaktadır. Her model ve mimari, test edilerek en uygun model ve mimari seçilmelidir. Bu durum, modeller ve mimariler arasında çoklu görev karşılaştırmalarını ortaya çıkarmaktadır.

Anlama modellerini ve mimarilerini değerlendirmek için diğer dillere (İngilizce ve Çince) nazaran Türkçe dili için kaynaklar çok azdır. Bu bitirme tezinde, amaç sınıflandırmasını state-of-the-art olan Transformer tabanlı mimari ile oluşturulmuş dört farklı dil modeli ile üç farklı domain veri setinde karşılaştırıyoruz. Bu veri kümeleri bankacılık, yatırım ve sosyal görevleri içermektedir. Karşılaştırılan modeller BERT, DistilBERT, ALBERT ve DIET'tir.

Bu tez çalışmam esnasında büyük bir sabır ve özveri ile bana destek olan tez danışman hocam Dr. Öğr. Üyesi Ahmet GÜRHANLI'ya teşekkür ederim. Ayrıca bu zorlu süreçte, tezimin her aşamasında bana destek olan, beni yalnız bırakmayan aileme ve arkadaşlarıma da sonsuz sevgi ve saygılarımı sunarım.

Eylül 2020

Oğuzhan KARAHAN

İÇİNDEKİLER

Sayfa

ÖNSÖZ	iv
İÇİNDEKİLER	v
KISALTMALAR	vi
ÇİZELGE LİSTESİ	vii
ŞEKİL LİSTESİ	viii
ÖZET	ix
ABSTRACT	x
1. GİRİŞ ve AMAÇ	1
1.1 Tezin Amacı	2
2. KAYNAK ARAŞTIRMASI	4
3. MATERYAL VE YÖNTEM	8
4. TEORİK ESASLAR	10
4.1 Transformer Mimarisi	10
4.1.1 Kodlayıcı (Encoder) ve kod çözücü (Decoder) yığınları	12
4.1.2 Dikkat işlevi (Attention)	13
4.1.2.1 Ölçekli nokta-ürün dikkat (Scaled Dot-Product Attention)	14
4.1.2.2 Çok başlı dikkat (Multi-Head Attention)	14
4.1.2.3 Modeldeki dikkat uygulamaları	15
4.1.3 Pozisyona yönelik ileri beslemeli ağlar	16
4.1.4 Embeddings ve Softmax	16
4.1.5 Konumsal kodlama	16
4.2 Modeller	17
4.2.1 BERT: Bidirectional Encoder Representations from Transformers	17
4.2.2 DistilBERT, a distilled version of BERT.....	19
4.2.3 ALBERT: A Lite BERT	20
4.2.4 DIET: Dual Intent and Entity Transformer.....	20
5. GÖREVLER VE DEĞERLENDİRME	21
5.1 Görevler.....	21
5.1.1 BNKD	21
5.1.2 INVD.....	21
5.1.3 SCLD	22
5.2 Değerlendirme	22
5.3 Transformer Tabanlı Modellerin Karşılaştırılması.....	22
6. SONUÇLAR VE ÖNERİLER	24
6.1 Sonuçlar.....	24
6.2 Öneriler.....	29
KAYNAKLAR	30
EKLER	33
ÖZGEÇMİŞ	36

KISALTMALAR

ALBERT	: A Lite Version of BERT
BERT	: Bidirectional Encoder Representations from Transformers
DistilBERT	: A Distilled Version of BERT
DIET	: Dual Intent and Entity Transformer
GLUE	: General Language Understanding Evaluation Benchmark
NLP	: Natural Language Processing
NLU	: Natural Language Understanding
MLM	: Masked Language Model
NSP	: Next Sentence Prediction
SentEval	: An Evaluation Toolkit for Universal Sentence Representations
decaNLP	: The Natural Language Decathlon
XNLI	: Evaluating Cross-lingual Sentence Representations
MNLI	: Multi-Genre Natural Language Inference
SQuAD	: Stanford Question Answering Dataset
XQuAD	: Cross-lingual Question Answering Dataset
BNKD	: Dataset of Banking Domain
INVD	: Dataset of Investment Domain
SCLD	: Dataset of Social Domain
AVG	: Average
IR	: Information Retrieval
BNF	: Backus-Naur Form
CFG	: Context-Free Grammar
LALR	: Look-Ahead parser with Left-to-right processing and Rightmost derivation
NN	: Neural Networks

ÇİZELGE LİSTESİ

Sayfa

Çizelge 5.1: Veri kümelerine genel bakış. Yalnızca niyet sınıflandırmadan oluşur. 21	
Çizelge 5.2: Transformer tabanlı modellere genel bakış. 23	
Çizelge 6.1: Transformer tabanlı modellerde temel değerlendirme. 25	
Çizelge 6.2: Bankacılık veri setinin (BNKD) modeller üzerindeki histogramı. 26	
Çizelge 6.3: Yatırım veri setinin (INVD) modeller üzerindeki histogramı. 27	
Çizelge 6.4: Sosyal veri setinin (SCLD) modeller üzerindeki histogramı. 28	

ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 3.1: Transformer tabanlı dil modellerinin genel mimarisi.....	9
Şekil 4.1: Transformer genel mimari şablonu	12
Şekil 4.2: Ölçekli Nokta-Ürün Dikkat ve Çok Başlı Dikkat, paralel işleyen birkaç dikkat katmanından oluşur.....	13
Şekil 4.3: Çıktıların matris formülü.....	14
Şekil 4.4: Çok başlı dikkat formülü.....	15
Şekil 4.5: Pozisyona yönelik ileri beslemeli ağ formülü.	16
Şekil 4.6: Farklı frekansların sinüs ve kosinüs fonksiyonları.....	17
Şekil 4.7: BERT için ön eğitim ve ince ayar prosedürlerinin genel mimarisi.....	18
Şekil 4.8: DIET mimarisinin şematik bir temsili.....	20
Şekil 6.1: Bankacılık veri setinin (BNKD) modellere göre sonuçları ve grafiği.....	26
Şekil 6.2: Yatırım veri setinin (INVD) modellere göre sonuçları ve grafiği.....	27
Şekil 6.3: Sosyal veri setinin (SCLD) modellere göre sonuçları ve grafiği	28

DOĐAL DİL ANLAMA MODELLERİNİN DİYALOG SORULARININ CEVAPLANMASI İÇİN DEĐERLENDİRİLMESİ

ÖZET

Teknolojinin gelişmesi ve Transformer tabanlı modellerin oluşturulmasıyla doğal dil anlama (NLU) alanında faydalı iyileştirmeler sağlanmıştır. Doğal dil anlayışının gelişmesi için en önemli ve birincil görev okuduđunu anlamadır. Adlandırılmış varlık tanıma ve amaç sınıflandırması için veri kümelerinden kabul edilen çeşitli yerleşik görevler içerir. Aynı zamanda, hayatın içinde çokça yer alan bankacılık, yatırım ve sosyal alanlar için veri setleri sunuyoruz. Paylaşılan değerlendirme şemasını garanti etmek için, farklı NLU görevlerini basitleştiren, farklı alanlardan ve uygulamalardan veri kümeleri içeren modeller oluşturduk. Sonuç olarak, Türkçe'nin Transformer tabanlı modellerinde bazı standart taban çizgileri ve geleceđi ile genel bir tahmin sunuyoruz.

Anahtar Kelimeler : *Türkçe, doğal dil anlama, diyaloglara cevap verme, derin öğrenme, değerlendirme*

THE EVALUATION OF NATURAL LANGUAGE UNDERSTANDING MODELS FOR ANSWERING DIALOGUE QUESTIONS

ABSTRACT

With the development of technology and the created of Transformer-based models, beneficial improvements have been achieved in the field of natural language understanding (NLU). The most important and primary task for the development of natural language understanding is reading comprehension. Contains of a varied established of tasks, accepted from datasets for named entity recognition and intent classification. At the same time, we offer datasets for the banking, investment and social domains that are very much in life. To guarantee shared evaluation scheme, we created models that simplify to different NLU tasks, contains datasets from diverse domains and applications. In conclusion, we offer a general estimation with some standard baselines and future in Transformer-based models of Turkish language.

Keywords: *Turkish, natural language understanding, answering dialogues, deep learning, evaluation*

1. GİRİŞ ve AMAÇ

Doğal dil anlama alanı (NLU), bilgisayarla görme alanında iyi bilinen bir fenomen olan bilginin yeniden kullanılabilirliğine ve aktarım öğrenmesine doğru büyük bir değişim yaşadı. Böyle bir değişikliğe, parametrik, sağlıklı, genel amaçlı modellerin yakın zamanda gözden geçirilmesi ortam sağladı. BERT (Devlin ve diğerleri, 2019), DistilBERT (Sanh ve diğerleri, 2020), ALBERT (Lan ve diğerleri, 2020) ve DIET (Bunk ve diğerleri, 2020) gibi.

Yayınlanan bu modeller, NLU alanında birçok iyileştirme sağlamış ve en son gelişmeleri hızlandırmıştır. Bu gelişmelerin ışığında, ilerleme önemli ölçüde hızlandı ve en son teknolojiyi daha da geliştirmek için Transformer tabanlı farklı modeller sık sık piyasaya sürüldü.

Standartlaştırılmış ve yaygın olarak kullanılan NLU değerlendirme standartları olmadan bir ilerleme olması muhtemel değildir. En popüler NLU standardı, çok yakın zamanda tanıtılan GLUE'dir (Wang ve diğerleri, 2019a). GLUE temel olarak soruları cevaplama, farklı metinler arasında ilişki bulma ve duyguları analiz etme gibi becerilere sahiptir. Bazı görevlerin çok sayıda eğitim örneğine ihtiyaç duyduğu, bazılarının ise sınırlı eğitim verisine sahip olduğu görülmüştür. Ayrıca eğitim seti ve test seti bazı görevler için farklı alanlar göstermektedir. Bu, modelleri evrensel dil temsillerini öğrenmeye teşvik eder ve çeşitli görevler ve alanlar arasında bilgi aktarımı gerçekleştirir. GLUE, mevcut veri kümeleri üzerine inşa edilmiştir ve GLUE'nin ana destekleri, çevrimiçi bir değerlendirme platformu ve bir lider panosunun yanı sıra dikkatli görev seçimidir.

NLU'daki en büyük gelişme İngilizce ve Çince'dir. Çünkü bu diller hem önceden eğitilmiş modeller hem de değerlendirme kriterleri açısından diğer dillere göre çok daha zengindir. Bu tezde, GLUE görevlerinde Türkçe dil desteği olmadığı için oluşturduğumuz veri setleri üzerinden veri oluşturma ve anlama ölçümü sağlanmış ve yatırım, bankacılık ve sosyal alanlar için veri setleri oluşturulmuştur. GLUE görevleri, düşünülen dilin farklı özelliklerini

kapsayacak şekilde dikkatlice seçilmiştir. GLUE modelini takiben, model değerlendirme aşamalarını azaltmak için görevleri metin sınıflandırmasına uyacak şekilde ayarladık.

Bu çalışma için başlıca katkılarımız:

- Model sonuçlarını kontrol panelinde değerlendirmek ve sunmak için bir platform geliştiriyoruz,
- INVD: Yatırım alanı için yeni bir niyet sınıflandırması sunuyoruz,
- BNKD: Bankacılık alanı için yeni bir niyet sınıflandırma görevi sunuyoruz,
- SCLD: Sosyal alan için yeni bir niyet sınıflandırma görevi sunuyoruz,
- Temel değerlendirme, Türkçe için Transformer tabanlı modellerdir.

İçerik; çalışmayı 1.1'de tanımladık. İkinci bölümde doğal dil anlamının tarihçesini ve günümüzdeki durumu açıkladık. Bölüm 3'te görevleri ve karşılaştırılacak model mimarilerini paylaştık. Bölüm 4'te temel yöntemlere atıfta bulunduk. Bölüm 5'te, tüm veri setlerini ve modelleri uyguluyor ve değerlendiriyoruz. Altıncı bölümde tez çalışmasının sonuçları üzerine genel bir değerlendirme yapılmıştır. Ayrıca çalışma ile ilgili öneriler de bu bölümde yer almaktadır.

Tez çalışmasının sonunda yararlanılan kaynaklar verilmiştir.

1.1 Tezin Amacı

Doğal dil anlama modellerinin değerlendirilmesi, teknolojik ilerlemenin önemli bir parçasıdır. Yeni dil modelleri ve mimariler için dikkate alınacak pek çok faktör olsa da hangisinin seçileceğini belirten bir standart bulunmamaktadır. Her model ve mimari, test edilerek en uygun model ve mimari seçilmelidir. Bu durum, modeller ve mimariler arasında çoklu görev karşılaştırmalarını ortaya çıkarmaktadır.

Bu standardın bir örneği SentEval'dir. Bu, on yedi görevden oluşan cümle yerleştirme kalitesini değerlendirmek için bir modeldir (Conneau ve Kiela, 2018). Ek olarak, cümle yerleştirme ile hangi dil özelliklerinin kullanıldığını belirlemek için 10 görev sağlanmaktadır. Her görevde, dil modelleri girdi olarak

tek veya çift cümle eklemeyi alır ve bir sınıflandırma veya regresyon problemini çözer. Bu standardın bir başka örneği de decaNLP'dir (McCann ve diğerleri, 2018). Bu model, önceden var olan 10 görevden oluşur ve görevlerin seçimi SentEval'e kıyasla çok daha çeşitlidir. Tüm görevler, NLU'nun temel görevi olan otomatik bir soru cevaplama biçimine dönüştürülür.

Son olarak, GLUE kıyaslama örneği (Wang ve diğerleri, 2019a). Bu model bir dizi mevcut, iyi yapılandırılmış görev önermektedir. Karşılaştırmanın SentEval'e benzememesi için daha çeşitli ve zor görevler seçilmelidir.

Yukarıda verilen tüm modeller ve kriterler İngilizce ile sınırlıdır. Birçok dil için destek sağlama çabaları arasında 14 dil desteği ile MNLI (Williams ve diğerleri, 2018) veri kümesi ve XNLI veri kümesi (Conneau ve diğerleri, 2018) bulunmaktadır. Benzer şekilde, 10 dil desteğine sahip SQuAD veri setinin (Rajpurkar vd., 2016) modeli XQuAD'dir (Artetxe vd., 2019).

Ne yazık ki bu çalışmaların hiçbiri Türkçe dilini kapsamıyor. Anlama modellerini ve mimarilerini değerlendirmek için diğer dillere (İngilizce ve Çince) nazaran Türkçe dili için kaynaklar çok azdır. Bu bitirme tezinde, amaç sınıflandırmasını state-of-the-art olan Transformer tabanlı mimari ile oluşturulmuş dört farklı dil modeli ile üç farklı domain veri setinde karşılaştırıyoruz. Bu veri kümeleri bankacılık, yatırım ve sosyal görevleri içermektedir. Karşılaştırılan modeller BERT, DistilBERT, ALBERT ve DIET'tir.

Tez çalışmasının akış şeması Şekil A.1'de gösterilmiştir.

2. KAYNAK ARAŞTIRMASI

NLP, yapay zeka ve dilbilimin kesişimi olarak 1950'li yıllarda başlamıştır. Büyük hacimli metinleri verimli bir şekilde indekslemek ve aramak için yüksek düzeyde ölçeklenebilir istatistik tabanlı teknikler kullanan metin bilgisi erişiminden (IR) farklı olarak NLP: IR'ye mükemmel bir giriş sağlamaktadır. Ancak zamanla, NLP ve IR birleşmiştir (Nadkarni PM vd., 2011). Şu anda, NLP, günümüzün NLP araştırmacılarının ve geliştiricilerinin zihinsel bilgi tabanlarını önemli ölçüde genişletmelerini gerektiren çok çeşitli alanlardan beslenmektedir.

İlk basit yaklaşım olarak, kelimesi kelimesine Rusça'dan İngilizce'ye makine çevirisiyle, birden çok anlam ve metafor içeren homografiler, tesadüfi olarak hecelenen kelimeler tarafından daha iyi çevirilmiştir. İncil'in doğruluğu şüpheli olan öyküsü, "ruh isteklidir, ancak beden zayıftır", "votka hoştur, ancak et şımarık" olarak tercüme edilmiştir.

Chomsky'nin 1956 dil gramerlerinin teorik analizi, Backus-Naur Form (BNF) notasyonunun oluşturulmasını (1963) etkileyerek problemin zorluğunun bir tahminini sağlamıştır. BNF, "bağlamdan bağımsız dilbilgisi" (CFG) belirtmek için ve genellikle programlama dili sözdizimini temsil etmek için kullanılmaktadır. Bir dilin BNF spesifikasyonu, program kodunu sözdizimsel olarak doğrulayan bir türetme kuralları kümesidir (Nadkarni PM vd., 2011). (Buradaki "Kurallar", mutlak kısıtlamalardır, uzman sistemlerin buluşsal yöntemi değildir.) Chomsky ayrıca, metin arama modellerini belirtmek için kullanılan normal ifadelerin temeli olan, daha da kısıtlayıcı "normal" dilbilgileri belirlemiştir. Kleene (1956) tarafından tanımlanan normal ifade sözdizimi ilk olarak Ken Thompson'ın UNIX üzerindeki grep yardımcı programı tarafından desteklenmiştir (Nadkarni PM vd., 2011).

Daha sonra (1970'ler), sözcüksel analizci (lexer) üreteçleri ve lex/yacc kombinasyonu gibi ayrıştırıcı üreteçleri gramerleri kullanmıştır. Bir lexer, metni

belirteçlere dönüştürür; ayrıştırıcı bir belirteç dizisini doğrular. Lex/yacc üreteçleri, düzenli ifade ve BNF belirtimlerini girdi olarak alır ve sözcük oluşturma/ayırıştırma durumunu belirleyen kodu ve arama tablolarını oluşturup programlama dilini büyük ölçüde basitleştirir (Nadkarni PM vd., 2011).

CFG'ler teorik olarak doğal dil için yetersizdir fakat pratikte hala NLP için sıklıkla kullanılmaktadır (Nadkarni PM vd., 2011). Programlama dilleri, uygulamayı basitleştirmek için tipik olarak kısıtlayıcı bir CFG varyantı, bir LALR dilbilgisi (LALR, Soldan sağa işleme ve En Sağdan (aşağıdan yukarıya) türetme ile İleri Bakma ayrıştırıcısı) ile bilinçli olarak tasarlanmıştır. Bir LALR ayrıştırıcısı, metni soldan sağa tarar, aşağıdan yukarıya doğru çalışır (yani, daha basit olanlardan bileşik yapılar oluşturur) ve ayrıştırma kararları vermek için tek bir simgenin ileriye bakmasını kullanır.

Prolog dili (1970), ilk olarak NLP uygulamaları için oluşturulmuştur (Nadkarni PM vd., 2011). Sözdizimi özellikle gramer yazmak için uygundur, ancak en kolay uygulama modunda (yukarıdan aşağıya ayrıştırma), kurallar yacc tarzı bir ayrıştırıcı için amaçlananlardan farklı bir şekilde (yani sağa-yinelemeli olarak) ifade edilmelidir. Yukarıdan aşağı ayrıştırıcılar, aşağıdan yukarı ayrıştırıcılara göre daha kolaydır (oluşturuculara ihtiyaç duymazlar), fakat çok daha yavaştır (Nadkarni PM vd., 2011).

Doğal dilin çok büyük boyutu, kısıtlayıcı olmayan doğası ve belirsizliği, tamamen sembolik, el yapımı kurallara dayanan standart ayrıştırma yaklaşımları kullanılırken iki sorun ortaya çıkmaktadır.

- NLP nihayetinde metinden anlam çıkarmalıdır: metin birimleri arasındaki ilişkiyi belirleyen biçimsel gramerler - isimler, fiiller ve sıfatlar gibi konuşma bölümleri - öncelikle sözdizimini ele almaktadır. Dilbilgisi, ek kurallar / kısıtlamalarla alt kategorileri büyük ölçüde genişleterek doğal dil anlambilimini ele alacak şekilde genişletilebilir. Maalesef, kurallar artık çok sayıda olduğundan yönetilemez hale gelmektedir (bir kelime dizisinin birden fazla yorumu mümkündür).
- El yazması kurallar, "dilbilgisi kurallarına uymayan" sözlü düzyazı ve (tıbbi bağlamlarda) hastane içi ilerlemenin oldukça kısa ve öz

düzyazısını çok zayıf bir şekilde ele almakta, bu tür düzyazılar insanlar tarafından anca anlaşılabilir.

1980'ler, Klein tarafından özetlenen temel bir yeniden yönlendirme ile:

- Derin analizin yerini basit, sağlam yaklaşımlar aldı.
- Değerlendirme daha katı hale geldi.
- Olasılıkları kullanan makine öğrenimi yöntemleri öne çıktı. (Chomsky'nin Sözdizimsel Yapılar (1959) adlı kitabı, olasılıkçı dil modellerinin yararlılığı konusunda şüpheci olmuştur).
- Makine öğrenmesi algoritmalarını eğitmek için büyük, korpuslar kullanıldı ve değerlendirme için standartlar sağladı (Nadkarni PM vd., 2011).

Bu yönlendirme sayesinde istatistiksel NLP ortaya çıkmıştır (Nadkarni PM vd., 2011). Örneğin, istatistiksel ayrıştırma, olasılıksal CFG'ler aracılığıyla ayrıştırma kuralı çoğalmasını ele almaktadır: ayrı kuralların, makine öğrenimi yoluyla belirlenen ilişkisel olasılıkları bulunmaktadır. Bu nedenle, daha az sayıda, daha geniş kural seti, belirsizliği gidermek için aranan istatistiksel frekans bilgisiyle çok sayıda ayrıntılı kuralın yerini almaktadır. Diğer yaklaşımlar, özellik vektör verilerinden karar ağaçları oluşturan C4.5 gibi makine öğrenimi algoritmalarına benzer korpuslardan olasılıklı "kurallar" oluşturmaktadır. Her durumda, istatistiksel ayrıştırıcı, bir cümlenin en olası ayrıştırmasını belirlemektedir (Nadkarni PM vd., 2011). Manning ve Scheutze'nin çalışması, istatistiksel NLP'ye giriş sağlamaktadır.

İstatistiksel yaklaşımlar pratikte iyi sonuçlar vermektedir, çünkü bol miktarda gerçek verilerle öğrenerek en yaygın durumları kullanırlar: veriler ne kadar bol ve temsili güçlü olursa, o kadar iyi sonuç vermektedir. Ayrıca bilinmedik / hatalı girişlerle daha az kırılabilir olurlar. Bununla birlikte, el yazması kurallarıyla istatistiksel yaklaşımların, birbirini tamamlayıcı olduğu açıkça ortaya koymaktadır. Bu konuda en iyi sonuçlar son zamanlarda sıklıkla kullanılan transformer tabanlı mimarilerden gelmektedir (Vaswani ve diğerleri, 2017).

Bu şekilde geçmişten günümüze kadar doğal dil anlamada birçok yöntem kullanılmıştır. Çalışmalarda görüldüğü gibi kullanılan yöntemler arasında state-

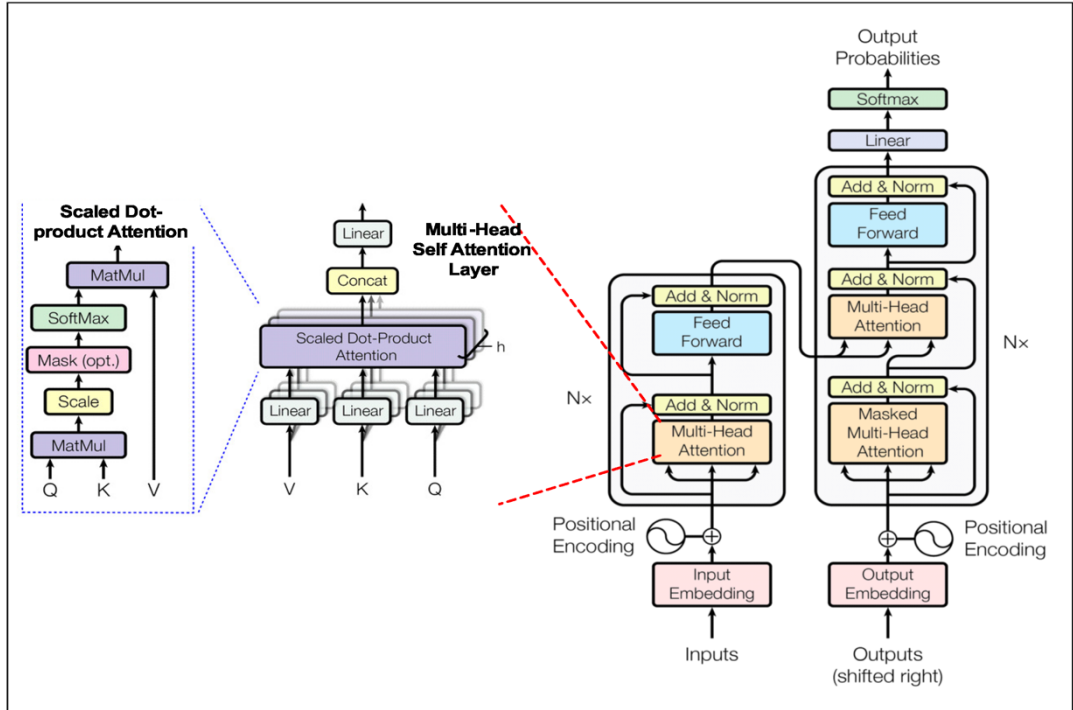
of-the-art olan transformer tabanlı modeller kullanılmıştır. Bu tezde, Türkçe için önceden eğitilmiş BERT modelleri ve yine BERT tabanlı DIET modeli ile yapılan doğal dil anlama çalışmaları incelenmiş ve farklı domainlerden oluşturulan verilerle karşılaştırması yapılmıştır. Ayrıca Türkçe'deki doğal dil anlama modelleri de incelenmiş ve uygun görülen bilgiler bu tezde kullanılmıştır.

3. MATERYAL VE YÖNTEM

Doğal dil anlama için önerilen ve kullanılan birçok yöntem vardır. Bu çalışma başlatılırken birçok çalışmanın avantaj ve dezavantajı göz önünde bulundurularak, en iyi yöntem ve mimari uygulanmaya çalışılmıştır. Bu tez çalışmasında, 3 farklı domain’de veri seti oluşturulup, transformer mimarisinde Türkçe için belirlenen dil modelleri ile eğitilerek niyet sınıflandırma işlemi yapılmaktadır.

Türkçe için önceden eğitilmiş BERT tabanlı dil modelleri iki temel aşamadan oluşmaktadır; ön eğitim ve ince ayar. Ön eğitim esnasında model, farklı eğitim öncesi görevler ve etiketlenmemiş veriler üzerinde eğitilir. Dil modeli, ince ayar için ilk olarak önceden eğitilmiş parametrelerle başlatılır ve tüm parametreler, aşağı akış görevlerinden etiketli veriler kullanılarak ince ayar gerçekleştirilir (Ashish V. vd., 2017). Önceden eğitilmiş aynı parametrelerle başlatılmış olsalar bile, her aşağı akış görevinin ayrı ince ayarlı modelleri vardır. Yalnızca DIET modeli birkaç anahtar parçadan oluşmaktadır. Şekil 3.1'deki mimari, transformer tabanlı mimarinin genel çalışma şeklini göstermektedir.

Doğal dil anlama için niyet sınıflandırması yapılmış ve transformer tabanlı 4 dil modeli, oluşturulan 3 farklı domain veri seti üzerinde doğru niyet sınıflandırması baz alınarak karşılaştırılmıştır.



Şekil 3.1: Transformer tabanlı dil modellerinin genel mimarisi

4. TEORİK ESASLAR

Bu bölümde doğal dil anlama için kullanılan transformer tabanlı mimariler ile ilgili teorik esaslar, özellikler ve karşılaştırılan dil modelleri anlatılmaktadır.

4.1 Transformer Mimarisi

Transformer (Vaswani ve diğerleri, 2017), hem doğal dil anlayışı hem de doğal dil üretimindeki görevler için performansta evrişimli ve tekrarlayan sinir ağları gibi alternatif sinir modellerini geride bırakarak, doğal dil işleme için hızla baskın mimari haline gelmiştir. Mimari, eğitim verileri ve model boyutuyla ölçeklenmekle birlikte, paralel eğitimi kolaylaştırmakta ve uzun aralıklı sekans özelliklerini yakalayabilmektedir (T. Wolf vd., 2019).

Model ön eğitimi (McCann vd., 2017; Howard ve Ruder, 2018; Peters vd., 2018; Devlin vd., 2018), modellerin genel yapıya göre eğitilmesine ve sonra daha güçlü performansla belirli görevlere kolayca uyarlanmasına olanak tanımaktadır (T. Wolf vd., 2019). Transformer mimarisi özellikle büyük metin yapıları üzerinde ön eğitim için elverişlidir ve metin sınıflandırması (Yang ve diğerleri, 2019), dil anlama (Liu ve diğerleri, 2019b; Wang ve diğerleri, 2018) dahil olmak üzere görevlerdeki doğruluklarda önemli kazanımlar sağlamaktadır. Dil anlama (Liu ve diğerleri, 2019b; Wang ve diğerleri, 2018, 2019), makine çevirisi (Lample ve Conneau, 2019a), çekirdek referans çözümü (Joshi ve diğerleri, 2019), sağduyu çıkarımı (Bosselut ve diğerleri, 2019) ve özetleme (Lewis ve diğerleri, 2019) alanlarından kullanılmaktadır.

Bu ilerleme, bu modellerin geniş kapsamda kullanılabilmesi için ele alınması gereken çok çeşitli zorlukları ortaya çıkarmaktadır (T. Wolf vd., 2019). Transformer'in her yerde kullanılması, sistemleri çeşitli platformlarda eğitmek, analiz etmek, ölçeklendirmek ve büyütme için sistemleri zorlamaktadır. Mimari, giderek daha karmaşık hale gelen uzantılar ile birlikte hassas deneyler tasarlamak için bir yapı taşı olarak kullanılmaktadır (T. Wolf vd., 2019). Ön

eđitim yöntemlerinin yaygın bir şekilde benimsenmesi, topluluk tarafından kullanılan önceden eğitilmiş çekirdek modellerin dağıtılması, ayarlanması ve sıkıştırılması ihtiyacını doğurmuştur.

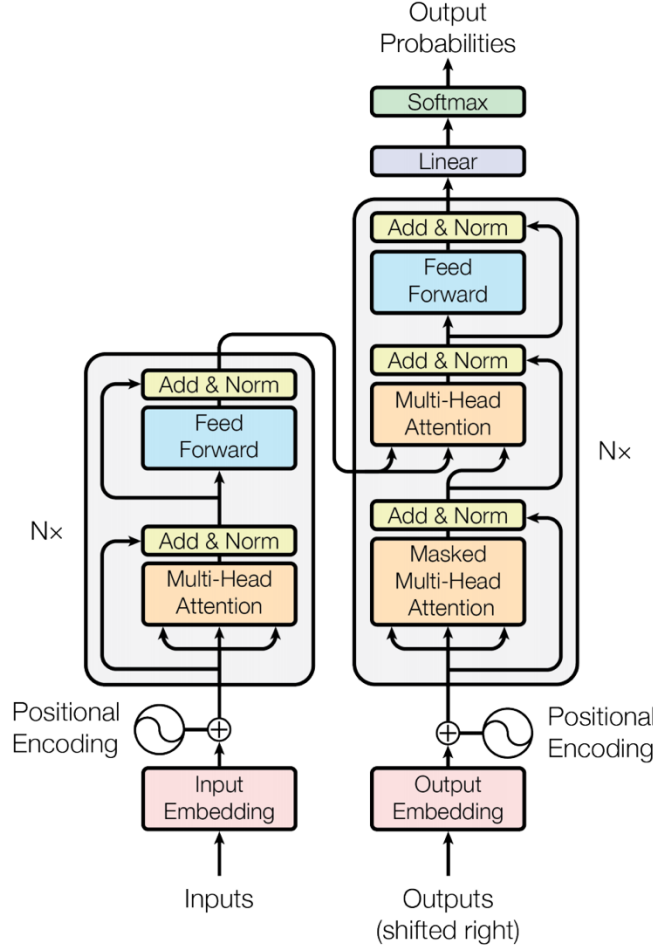
Transformer kütüphanesinin merkezinde, hem araştırma hem de üretim için tasarlanmış bir Transformer uygulaması vardır. Amaç olarak, okuması, genişletmesi ve dağıtması kolay popüler model varyasyonlarının endüstriyel güçteki uygulamalarını desteklemektir. Bu temelde kütüphane, merkezi bir model hub'ında çok çeşitli, önceden eğitilmiş modellerin dağıtımını ve kullanımını destekleyecek bir ortam sunmaktadır. Kullanıcıların farklı modelleri aynı mantıkla karşılaştırmasını ve çeşitli görevler üzerinde paylaşılan modelleri denemesini desteklemektedir (T. Wolf vd., 2019).

Transformers, Hugging Face'teki mühendis ve araştırmacılardan oluşan ve 400'den fazla dış katılımcının oluşturduğu canlı bir topluluğun desteğiyle sürdürülen bir platformdur. Kütüphane, Apache 2.0 lisansı ile yayınlanmakta olup, GitHub'da mevcuttur. Hugging Face'in web sitesinde ayrıntılı bilgiler ve belgeler mevcuttur (T. Wolf vd., 2019).

Yakın zamanda yapılan araştırmalar, GLUE modelinde en iyi sonuçların Bidirectional Encoder Representations (BERT) modelinden esinlenen Transformer tabanlı modellerle elde edildiğini göstermiştir. Bu modeller Maskeli Dil Modeli (MLM) kullanılarak önceden eğitilmiştir. Bu bölümde şu dil modellerini açıklayacağız: BERT (Devlin vd., 2019), DistilBERT (Sanh vd., 2020), ALBERT (Lan vd., 2020) ve DIET (Bunk vd., 2020). Bu makaleyi yazarken, DIET hariç, bunlar Türkçe metinle eğitilmiş mevcut Transformer tabanlı tek modellerdir.

Bu modelleri anlamak için her görev için farklı ayarlar yapmamız gerekiyor. Model eğitimi için Transformers (Wolf vd., 2019) kütüphanesi kullanılmıştır. Tüm modeller, parti boyutu 32 olan ve 100 yinelemeli eğitim ile 2×10^{-5} 'ten başlayan doğrusal olarak azalan bir öğrenme hızı şeması kullanılarak 50 iterasyon için eğitildi. Adam optimizer'ı şu parametrelerle kullanıyoruz: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-5}$ çalışmanın medyan performansını gösteriyoruz.

Şekil 4.1'de transformer modelinin mimarisi gösterilmektedir.



Şekil 4.1: Transformer genel mimari şablonu

Çoğu nöral sekans transdüksiyon modelinin bir kodlayıcı bir de kod çözücü yapısı bulunmaktadır. Kodlayıcı, sembol temsillerinin (x_1, \dots, x_n) bir giriş dizisini $z = (z_1, \dots, z_n)$ temsiller dizisine eşlemektedir. Z verildiğinde, kod çözücü, sembollerin bir çıkış dizisini (y_1, \dots, y_m) üretecektir. Her safhada, model auto-regressive olacak şekilde, bir sonraki safhayı oluştururken ek girdi olarak önceden oluşturulmuş sembolleri tüketmektedir (T. Wolf vd., 2019).

Transformer, sırasıyla Şekil 4.1'in sol ve sağ yarısında gösterilen, hem kodlayıcı hem de kod çözücü için yığılmış attention ve dot-product, wide convolution katmanlarını kullanarak bu genel mimariyi takip eder.

4.1.1 Kodlayıcı (Encoder) ve kod çözücü (Decoder) yığınları

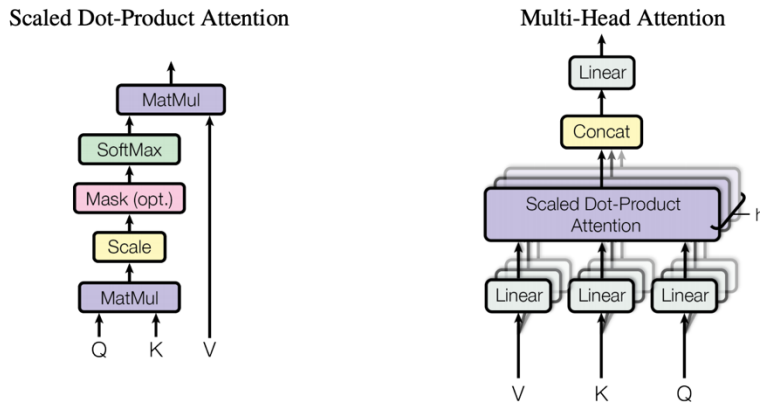
Kodlayıcı (Encoder): $N = 6$ özdeş katmandan oluşan bir yığından oluşur. Her katman, iki alt katmana sahiptir (T. Wolf vd., 2019). Birincisi, çok başlı bir öz- ilgi mekanizması, ikincisi ise basit, konumsal olarak tam bağlantılı ileri

beslemeli bir ağıdır. İki alt tabakanın da çevresinde bir artık bağlantı bulunup, ardından katman normalizasyonu kullanılmaktadır. Yani, her bir alt katmanın çıktısı LayerNorm ($x + \text{Sublayer}(x)$)'dur (T. Wolf vd., 2019). Buradaki artık bağlantıları kolaylaştırmak adına, modeldeki tüm alt katmanlar ile birlikte gömme katmanları da $d_{\text{model}} = 512$ boyutunun çıktılarını üretmektedir.

Kod çözücü (Decoder): Yığın olarak 6 özdeş katmandan oluşmaktadır. Kodlayıcı katmandaki iki alt katmanın yanında, kodlayıcı yığının çıktısı olarak çok başlı dikkat gerçekleştiren üçüncü bir alt katman eklemektedir. Kodlayıcıdakine benzer şekilde, alt katmanların her birinin çevresinde artık bağlantılar kullanılmakta, ardından katman normalizasyonu sağlanmaktadır (T. Wolf vd., 2019). Konumların sonraki konumlara gelmesini önlemek için kod çözücü yığınınındaki öz-ilgi alt katmanı da değiştirilir. Bu maskeleyme, çıktı yerleştirmelerinin bir konum kaydırılmış olması gerçeğiyle birleştiğinde, konum tahminlerinin yalnızca i 'den küçük konumlarda bilinen çıktılara bağlı olabilmelerini sağlar.

4.1.2 Dikkat işlevi (Attention)

Dikkat işlevinde görev, bir sorguyu ve bir dizi anahtar-değer çiftini bir çıktıya eşlemek olarak gösterilebilir; sorgu, anahtar, değer ve çıktının hepsi vektördür. Çıktı, tüm değerlerin ağırlıklı toplamı olarak hesaplanmaktadır. Her bir değere atanan ağırlık ise, her sorguya karşılık gelen anahtarla uyumluluk işlevi tarafından hesaplanmaktadır (T. Wolf vd., 2019).



Şekil 4.2: Ölçekli Nokta-Ürün Dikkat ve Çok Başlı Dikkat, paralel işleyen birkaç dikkat katmanından oluşur.

4.1.2.1 Ölçekli nokta-ürün dikkat (Scaled Dot-Product Attention)

Özel dikkat, "Ölçekli Nokta-Ürün Dikkati" şeklinde ifade edilmektedir (Şekil 4.2). Girdi, d_k boyutunun sorgularından ve anahtarlarından, d_v boyut değerlerinden oluşmaktadır (T. Wolf vd., 2019). Sorgunun nokta ürünlerini tüm anahtarlarla hesaplanmaktadır, her birini $\sqrt{d_k}$ 'ya bölüp, değerler üzerindeki ağırlıkları elde etmek için bir softmax işlevi uygulanır.

Pratikte, dikkat fonksiyonunu aynı anda bir sorgu kümesi üzerinde hesaplanır, bir Q matrisinde birlikte paketlenmektedir. Anahtarlar ve değerler aynı zamanda K ve V matrislerinde birlikte oluşturulmuştur (T. Wolf vd., 2019).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Şekil 4.3: Çıktıların matris formülü.

En sık kullanılan iki dikkat işlevi, toplamsal dikkat ve iç çarpım (çarpımsal) dikkattir. Nokta ürün dikkati, $\sqrt{1/d_k}$ ölçekleme faktörü dışında üstteki algoritmayla aynıdır. Ekstra dikkat, tek gizli katmana sahip olan ileri beslemeli bir ağ kullanarak uyumluluk işlevini hesaplamaktadır (T. Wolf vd., 2019). İkisi teorik karmaşıklık açısından benzer olsa da, nokta-ürün dikkati çok daha hızlı ve pratikte daha verimli, çünkü yüksek düzeyde optimize edilmiş matris çarpım kodu kullanılarak uygulanabilir.

Küçük d_k değerleri için iki mekanizma benzer şekilde çalışırken, ek dikkat, daha büyük d_k değerleri için ölçeklendirme olmaksızın nokta ürün dikkatinden daha iyi performans gösterir. Büyük d_k değerleri için, nokta ürünlerin büyüklük olarak büyüdüğünden ve softmax işlevini çok küçük gradyanlara sahip bölgelere ittiğinden yola çıkarak, bu etkiye karşı koymak için $\sqrt{1/d_k}$ şeklinde ölçeklendirme yapılmaktadır (T. Wolf vd., 2019).

4.1.2.2 Çok başlı dikkat (Multi-Head Attention)

Tek bir dikkat fonksiyonunu d_{model} boyutlu anahtarlar, değerler ve sorgularla gerçekleştirmek yerine, sorguları, anahtarları ve değerleri h kez farklı, öğrenilmiş doğrusal projeksiyonlarla sırasıyla d_k , d_k ve d_v boyutlarına doğrusal olarak yansıtmak daha faydalı bulunmuştur. Sorguların, anahtarların ve

değerlerin bu öngörülen sürümlerinin her birinde, daha sonra dikkat işlevini paralel olarak gerçekleştirerek d_v boyutlu çıktı değerleri elde edilir. Bunlar birleştirilir ve bir kez daha yansıtılır ve Şekil 4.2'de gösterildiği gibi nihai değerler elde edilir.

Çok başlı dikkatte, modelin farklı yerlerinde farklı temsil alt uzaylarından gelen verilerin birleştirilmesi sağlanır (T. Wolf vd., 2019). Tek bir dikkat ile, bu sağlanamaz.

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$
$$\text{where head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Şekil 4.4: Çok başlı dikkat formülü.

Bu çalışmada $h = 8$ paralel dikkat katmanları kullanılmaktadır. Her biri için $d_k = d_v = d_{\text{model}}/h = 64$ kullanılmaktadır. Her bir katmanın küçültülmüş boyutu sebebiyle, hesaplama maliyeti, tam boyutluluk ile tek katmanın dikkatine benzer şekildedir (T. Wolf vd., 2019).

4.1.2.3 Modeldeki dikkat uygulamaları

Transformer, çok başlı dikkati üç farklı şekilde kullanmaktadır:

- "Kodlayıcı-kod çözücü dikkat" katmanlarında, sorgular önceki kod çözücü katmanından gelmektedir. Bununla birlikte, bellek anahtarları ve değerleri kodlayıcının çıkışından gelmektedir. Bu, kod çözücüdeki her pozisyonun giriş dizisindeki tüm konumlarla birleştirilmesini sağlamaktadır (T. Wolf vd., 2019).
- Kodlayıcı, öz-dikkat katmanlarına sahiptir. Bir öz-dikkat katmanında tüm anahtarlar, değerler ve sorgular aynı yerden gelmektedir. Bu, kodlayıcıdaki önceki katmanın çıktısını oluşturmaktadır. Kodlayıcıdaki her pozisyon, önceki katmandaki tüm pozisyonlarla birleştirilebilir (T. Wolf vd., 2019).
- Kod çözücüdeki öz-dikkat katmanları, her bir pozisyonun, bu pozisyona kadar olan kod çözücüdeki tüm pozisyonlarla birleşmesini sağlar (T. Wolf vd., 2019). Otomatik gerileme özelliğini korumak için kod çözücüde sola doğru bilgi akışını önlemek gerekir. Bunu, softmax'ın

girişindeki geçersiz bağlantılara karşılık gelen tüm değerleri maskeleyerek ($-\infty$ 'a ayarlayarak) ölçeklenmiş nokta-ürün dikkatinin içine uygulanır. Şekil 4.2'de görülebilir.

4.1.3 Pozisyona yönelik ileri beslemeli ağlar

Dikkat alt katmanlarına ek olarak, kodlayıcıdaki ve kod çözücüdeki katmanların her biri, her bir konuma ayrı ayrı ve aynı şekilde uygulanan, tamamen bağlı bir ileri besleme ağı içerir. Bu, ReLU aktivasyonu olan iki doğrusal dönüşümden oluşmaktadır (T. Wolf vd., 2019).

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

Şekil 4.5 : Pozisyona yönelik ileri beslemeli ağ formülü.

Doğrusal dönüşümler farklı pozisyonlarda aynı olsa bile, katmandan katmana farklı parametreler kullanılmaktadır. Bunu tanımlamanın bir başka yolu da, çekirdek boyutu 1 olan iki evrişim katmanıdır. Girdinin ve çıktının boyutu, $d_{\text{model}} = 512$ 'dir. Ve iç katmanın boyutsallığı $d_f = 2048$ 'dir (T. Wolf vd., 2019).

4.1.4 Embeddings ve Softmax

Diğer dizi dönüştürme modellerine benzer şekilde, giriş jetonlarını ve çıktı jetonlarını boyut d_{model} 'nin vektörlerine dönüştürmek için öğrenilmiş embeddingler kullanılır. Kod çözücü çıktısını, tahmin edilen sonraki olasılıklara dönüştürmek adına, olağan öğrenilmiş doğrusal dönüşüm ve softmax işlevi birlikte kullanılmaktadır. Modelde, iki embedding katman ile birlikte pre-softmax doğrusal dönüşümü arasında aynı ağırlık matrisi paylaşılmaktadır (T. Wolf vd., 2019). Embedding katmanlarında, bu ağırlıklar $\sqrt{d_{\text{model}}}$ ile çarpılır.

4.1.5 Konumsal kodlama

Model, yineleme ve evrişim içermediğinden, modelin dizinin sırasını kullanması için dizideki simgelerin göreceli veya mutlak konumu hakkında bazı bilgiler enjekte edilmelidir. Bu sebeple, kodlayıcı ve kod çözücü yığınlarının tabanındaki giriş yerleştirmelerine "konumsal kodlama" eklenmektedir. Konumsal kodlamalar, embeddinglerle aynı boyut d_{model} 'ine sahiptir ve

böylece ikisi toplanabilmektedir (T. Wolf vd., 2019). Konumsal kodlamanın öğrenilmiş ve düzeltilmiş birçok seçeneği vardır.

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}})$$
$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}})$$

Şekil 4.6 : Farklı frekansların sinüs ve kosinüs fonksiyonları.

Pos, konum ve i boyutu temsil etmektedir. Yani, konumsal kodlamanın her boyutu bir sinüzoide karşılık gelmektedir. Dalgaboyları, 2π ile $10000 \cdot 2\pi$ arasında geometrik bir ilerleme oluşturmaktadır (T. Wolf vd., 2019). Bu işlevin seçilme nedeni, modelin göreceli konumlara göre birleştirilmesini kolayca öğreneceği varsayılmıştır, çünkü herhangi bir sabit ofset k için $P Epos + k$, P Epos'un doğrusal bir işlevi olarak temsil edilebilmektedir (T. Wolf vd., 2019).

Bunun yerine öğrenilmiş konumsal yerleştirmeleri kullanmak neredeyse aynı sonuçlar üretmiştir. Modelin eğitim sırasında karşılaşılanlardan daha uzun dizi uzunluklarına ekstrapolasyon yapmasına izin verebileceği için sinüzoidal versiyon seçilmiştir.

4.2 Modeller

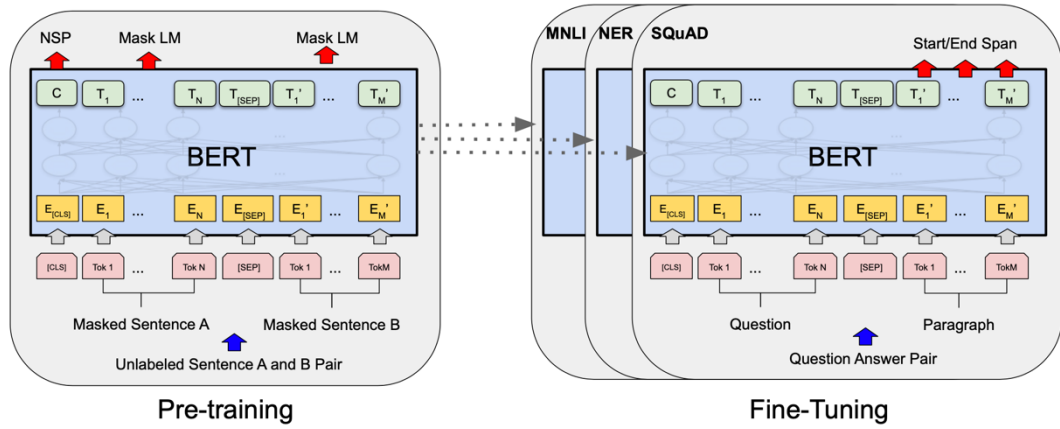
4.2.1 BERT: Bidirectional Encoder Representations from Transformers

Dil modeli için ön eğitiminin, birçok doğal dil işleme görevini iyileştirmede ve anlamada etkili olduğu gösterilmiştir (Dai ve Le, 2015; Peters ve diğerleri, 2018a; Radford ve diğerleri, 2018; Howard ve Ruder, 2018). Bu, cümleler arasındaki ilişkileri bütünsel olarak analiz ederek tahminlemeyi amaçlayan doğal dil çıkarımı (Bowman vd., 2015; Williams vd., 2018) ve anlama (Dolan ve Brockett, 2005) gibi cümle seviyesindeki görevleri içermektedir (Ashish V. vd., 2017).

Alt görevlere, önceden eğitilmiş dil temsillerini uygulamak için iki strateji bulunmaktadır: özellik tabanlı (feature based) ve ince ayarlı (fine tuning) stratejilerdir. ELMo (Peters ve diğerleri, 2018a) gibi özellik tabanlı yaklaşımda, ek özellikler olarak önceden eğitilmiş dil temsillerini içeren göreve/domaine özgü mimariler kullanılmaktadır. Üretken Önceden Eğitilmiş Transformatör (OpenAI GPT) (Radford ve diğerleri, 2018) gibi ince ayar yaklaşımında ise,

göreve/domaine özgü minimum parametreleri sunulur ve önceden eğitilmiş tüm parametrelerin ince ayarı yapılarak aşağı akış görevlerinde eğitilir. İki yaklaşımda da, genel dil temsillerini öğrenmek için önceden eğitilmiş dil modellerinin tek yönlü kullandıkları aynı amaç işlevi paylaşılmaktadır (Ashish V. vd., 2017).

Mevcut mimarilerin, özellikle ince ayar yaklaşımlarında bulunan önceden eğitilmiş temsillerin gücünü kısıtladığı iddia edilmektedir. Bu sınırlamaların en büyüğü, standart dil modellerinin tek yönlü olmasıdır. Ve bu durum, ön eğitim sırasında kullanılabilir mimari seçimini sınırlamaktadır. Örneğin, OpenAI GPT'de, kurguyu sağlayanlar soldan sağa bir mimari kullanmaktadır; buradaki her bir simge, Transformer'in öz-ilgi katmanlarında yalnızca önceki belirteçlere katılabilmektedir (Vaswani ve diğerleri, 2017). Bu tür kısıtlamalar cümle seviyesindeki görevler için yetersiz kalmaktadır. Ve her iki yönden de bağlamın birleştirilmesinin çok önemli olduğu soru yanıtlama gibi belirteç düzeyindeki görevlere, ince ayar tabanlı yaklaşımlar uygulanırken çok zararlı olabilmektedir (Ashish V. vd., 2017).



Şekil 4.7: BERT için ön eğitim ve ince ayar prosedürlerinin genel mimarisi.

BERT: Transformer'den Çift Yönlü Kodlayıcı Gösterimleri, ince ayar tabanlı yaklaşımları ve mimarileri iyileştirmektedir (Vaswani ve diğerleri, 2017). BERT, Cloze görevinden esinlenen bir "maskeli dil modeli" (MLM) eğitim öncesi hedefi kullanarak daha önce bahsedilen tek yönlü kısıtlamayı hafifletmektedir (Taylor, 1953). Maskeli dil modeli, girdiden bazı simgeleri rasgele maskeleymektedir. Amaç, maskelenmiş sözcüğün/sözcüklerin orijinal kimliğini yalnızca bağlamına göre tahmin etmektir. MLM'in soldan sağa dil

modelinde hedef, ön eğitiminden farklı olarak, temsilin sol ve sağ bağlamı birleştirmesini sağlamaktır. Bu da derin bir çift yönlü Transformer'i önceden eğitmeye olanak tanımaktadır. Ek olarak, metin-çifti temsillerini önceden eğiten "sonraki cümle/kelime tahmini" görevi de kullanılmaktadır (Ashish V. vd., 2017). BERT'in katkıları aşağıdaki gibidir:

- Dil temsilleri için çift yönlü ön eğitimin önemi ve gerekliliği gösterilmektedir. Ön eğitim için tek yönlü dil modelini kullanan BERT'te, önceden eğitilmiş çift yönlü temsilleri etkinleştirmek için maskelenmiş dil modellerini kullanılmaktadır. Bu, bağımsız olarak eğitilmiş soldan sağa ve sağdan sola dil modelleri için de kullanılır (Ashish V. vd., 2017).
- Önceden eğitilmiş temsillerin, büyük ölçekteki görevlere özgü tasarlanan mimarilere ihtiyacı azalttığı da gösterilmektedir. BERT, göreve özgü mimarilerden daha iyi performans gösterip, geniş bir cümle seviyesinde ve belirteç düzeyindeki görevler setinde son teknoloji performansa ulaşan ilk ince ayar tabanlı temsil modelidir (Ashish V. vd., 2017).
- BERT, on bir NLP görevi için son teknolojiyi geliştirir. Proje kodları ve önceden eğitilmiş modellerin detayları URL-1 adresinde mevcuttur.

4.2.2 DistilBERT, a distilled version of BERT

DistilBERT, BERT ile aynı genel mimariye sahiptir. Katman sayısı 2 kat azaltılmış olup, token tipi yerleştirmelerin yanında pooler de kaldırılmıştır. Transformer mimarisinde kullanılan işlemlerin çoğu (doğrusal katman ve katman normalizasyonu), modern doğrusal cebir şablonlarında optimize edilmiş olup, araştırmalarda da benzeri işlem uygulanmıştır. Tensörün son (gizli) boyutundaki varyasyonların, katman sayısı gibi diğer faktörlerdeki varyasyonlara göre hesaplama verimliliği (sabit parametreler maliyeti için) üzerinde etkisinin daha küçük olduğu görülmektedir. Böylece katman sayısının azaltılmasına odaklanılarak, işlem sağlanmıştır (V. Sanh vd., 2019).

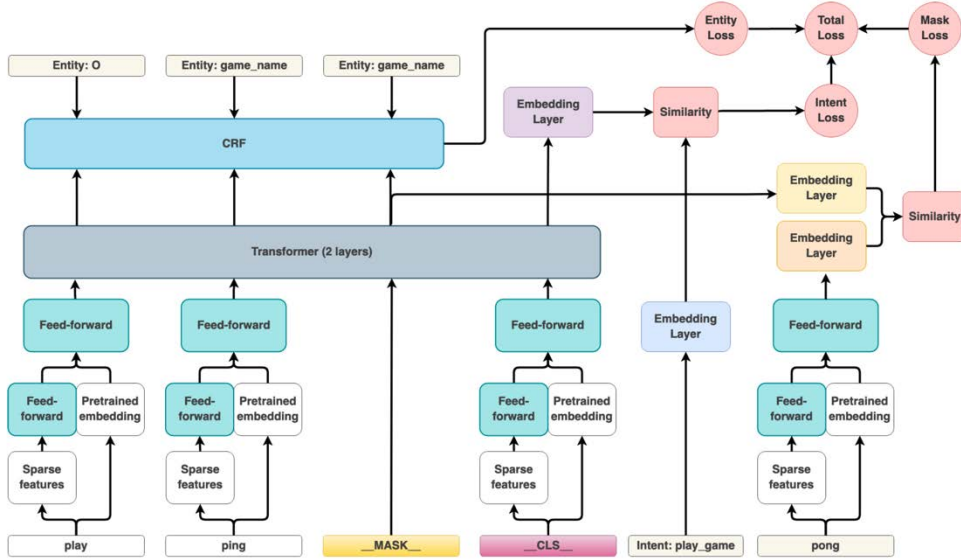
Damıtma, son zamanlarda önerilen BERT modelini eğitmek için uygulanmıştır. Bu nedenle DistilBERT, dinamik maskeleye kullanarak ve sonraki cümle tahmin hedefi olmadan gradyan birikiminden (parti başına 4K örneğe kadar) yararlanarak çok büyük partilerde damıtılır.

4.2.3 ALBERT: A Lite BERT

Bu dil modelinin mimarisi, doğrusal olmayan bir GELU ve bir transformatör kodlayıcı kullanması bakımından BERT'inkine benzer. Buna ek olarak, ALBERT ayrıca önceden eğitilmiş modellerdeki bazı engelleri ortadan kaldıran iki parametre azaltma tekniğini kullanır.

4.2.4 DIET: Dual Intent and Entity Transformer

DIET (Dual Intent and Entity Transformer) dil modeli, amaç sınıflandırması ve varlık tanıma için çok görevli bir mimaridir. Bu dil modelinin en önemli özelliklerinden biri, kelimeleri önceden eğitilmiş dil modeliyle birleştirme yeteneğidir.



Şekil 4.8: DIET mimarisinin şematik bir temsili.

5. GÖREVLER VE DEĞERLENDİRME

5.1 Görevler

Bu bitirme tezinde, çok büyük veri setlerine sahip olmayan görevlere odaklandık. Veri setinin göreceli olarak küçük olması, modelin problemlerini çözmek için dışarıdan bilgi gerektirebilir. Ana hedefimiz Türkçe dil desteği için genel bir model oluşturmaktır. Tüm görevleri aşağıdaki bölümlerde sunuyoruz ve Çizelge 5.1'de özetliyoruz.

Çizelge 5.1: Veri kümelerine genel bakış. Yalnızca niyet sınıflandırmadan oluşur.

İsim	Eğitim Seti	Domain	Metrik
BNKD	6.5k	Bankacılık	F1-Score
INVD	3.9k	Yatırım	F1-Score
SCLD	2.4k	Sosyal	F1-Score

5.1.1 BNKD

Bankacılık alanından 6,5 bin veri kümesi çıkaran BNKD adlı yeni bir veri kümesi sunuyoruz. Buradaki amaç; niyet sınıflandırması görevidir. Her veri en az 6 kelime uzunluğundadır ve 419 niyete göre sınıflandırılır. Görev, belirli bir bankacılık alanı veri kümelerinin sınıflandırmasını tahmin etmektir.

5.1.2 INVD

Yatırım alanından 3,9 bin veri kümesini çıkaran INVD adlı yeni bir amaç sınıflandırma veri kümesi sunuyoruz. Her inceleme en az 2 kelime uzunluğundadır ve 134 niyete göre sınıflandırılır. Görev, belirli bir yatırım alanı veri kümelerinin sınıflandırmasını tahmin etmektir.

5.1.3 SCLD

Sosyal domainindeki 2,4 bin veri kümesini çıkararak SCLD adlı yeni bir amaç sınıflandırma veri kümesi sunuyoruz. Her bir inceleme en az 4 kelime uzunluğundadır ve 122 niyete göre sınıflandırılır. Görev, belirli bir sosyal alan veri kümelerinin sınıflandırmasını tahmin etmektir.

5.2 Değerlendirme

Modelleri karşılaştırmanın kesin bir ölçüsü olmadığını bilerek, modeller ortalama performanslarına göre karşılaştırılmalıdır. Modellerin zorluk seviyeleri farklı olduğu için, tüm modelleri ortalama performanslarıyla karşılaştırmak doğru olacaktır.

Üç farklı domain için transformer tabanlı BERT, ALBERT, DistilBERT ve DIET modelleri karşılaştırılacaktır. BERT, ALBERT ve DistilBERT içerisinde önceden eğitilmiş Türkçe dil modeli bulunmaktadır. DIET ise eğitim verisi üzerinden kendisi bir dil modeli oluşturarak sonuç vermektedir.

Eğitilen modeller için 3 kat çapraz doğrulama gerçekleştirilerek niyet sınıflandırmasının doğruluğu incelenecektir. Her niyetin, gitmesi gereken sadece bir sınıfı bulunmaktadır.

5.3 Transformer Tabanlı Modellerin Karşılaştırılması

Bu bölüm, niyet sınıflandırma işleminin yapıldığı ana bölümdür. Bu bölümde doğru sınıflandırma açısından 4 farklı transformer tabanlı dil modeli karşılaştırılmış, Python dili ve Tensorflow kütüphanesi kullanılmıştır. Bu algoritmalar BERT, ALBERT, DistilBERT ve DIET modelleridir. Modeller ile ilgili detaylı bilgi bölüm 4.3'te verilmiştir.

Belirlenen dört modelin, üç farklı domainde oluşturulmuş veri seti ile niyet sınıflandırma için 3 kat çapraz doğrulama karşılaştırılmıştır. Oluşturulan veri setlerinin detayları Çizelge 5.1'de verilmiştir. Belirlenen ve karşılaştırması yapılan modeller Çizelge 5.2'de verilmiştir.

Çizelge 5.2: Transformer tabanlı modellere genel bakış.

İsim	Corpora	Tokenizer	Pre-trained Model	
ALBERT	18k	WordPiece	Turkish ALBERT	Uncased
BERT	32k	WordPiece	Turkish Cased BERT	
DistilBERT	32k	WordPiece	Turkish DistilBERT	Cased
DIET	Veri Seti	LanguageModelTokenizer	LanguageModelFeaturizer	

6. SONUÇLAR VE ÖNERİLER

6.1 Sonuçlar

Modelleri karşılaştırmanın kesin bir ölçüsü olmadığını bilerek, modeller ortalama performanslarına göre karşılaştırılmalıdır. Modellerin zorluk seviyeleri farklı olduğu için, tüm modelleri ortalama performanslarıyla karşılaştırmak doğru olacaktır.

Önceden eğitilmiş BERT modellerine kıyasla, DIET modeli biraz daha iyi performans gösterir. Bunun istisnası, küçük eğitim veri kümeleridir. Genel olarak değerlendirilirse, Türkçe için önceden eğitilmiş BERT modellerinin ve muhtemelen küçük veri seti olan modellerin önemli eksikliklerini göstermektedir. Genel olarak, önceden eğitilmiş her BERT modeli, test edilen DIET modellerinden ortalama olarak hala daha kötüdür.

Eğitim içeriğinin, bir veya birden fazla domaindeki görevin yerine getirilmesinde çok önemli bir rol oynadığı görülmektedir. Bunun nedeni, soru cevaplama görevinin Wikipedia tabanlı olması ve BERT modelinin Wikipedia külliyatında eğitilmiş olmasıdır.

DIET, diğer Transformatör tabanlı modellere kıyasla oldukça rekabetçi sonuçlar elde eder. Ortalama olarak en iyi performansa sahiptir ve iki görevde, BNKD ve SCLD'de iyi sonuçlar elde etmektedir. Dahası, DIET, BNKD ve SCLD arasındaki en küçük performans açığına sahiptir, bu da alanlar arasında daha iyi genelleme yapılmasını sağlamaktadır. Ancak diğer Transformer tabanlı modellerle karşılaştırıldığında, INVD görevinde kötü performans göstermektedir.

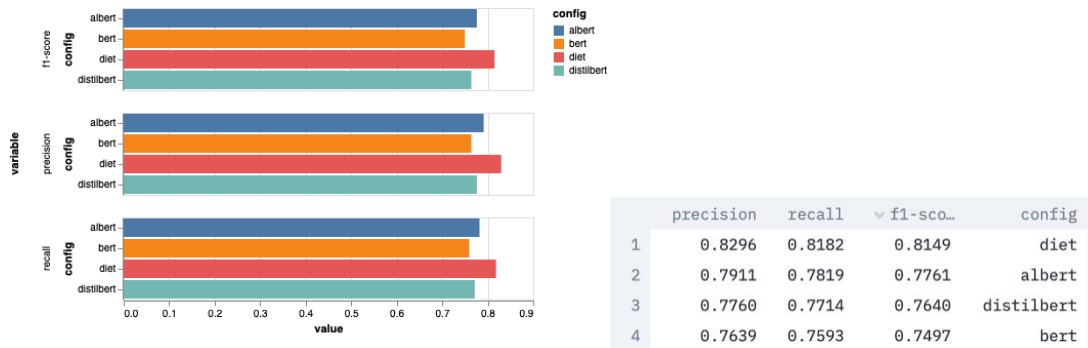
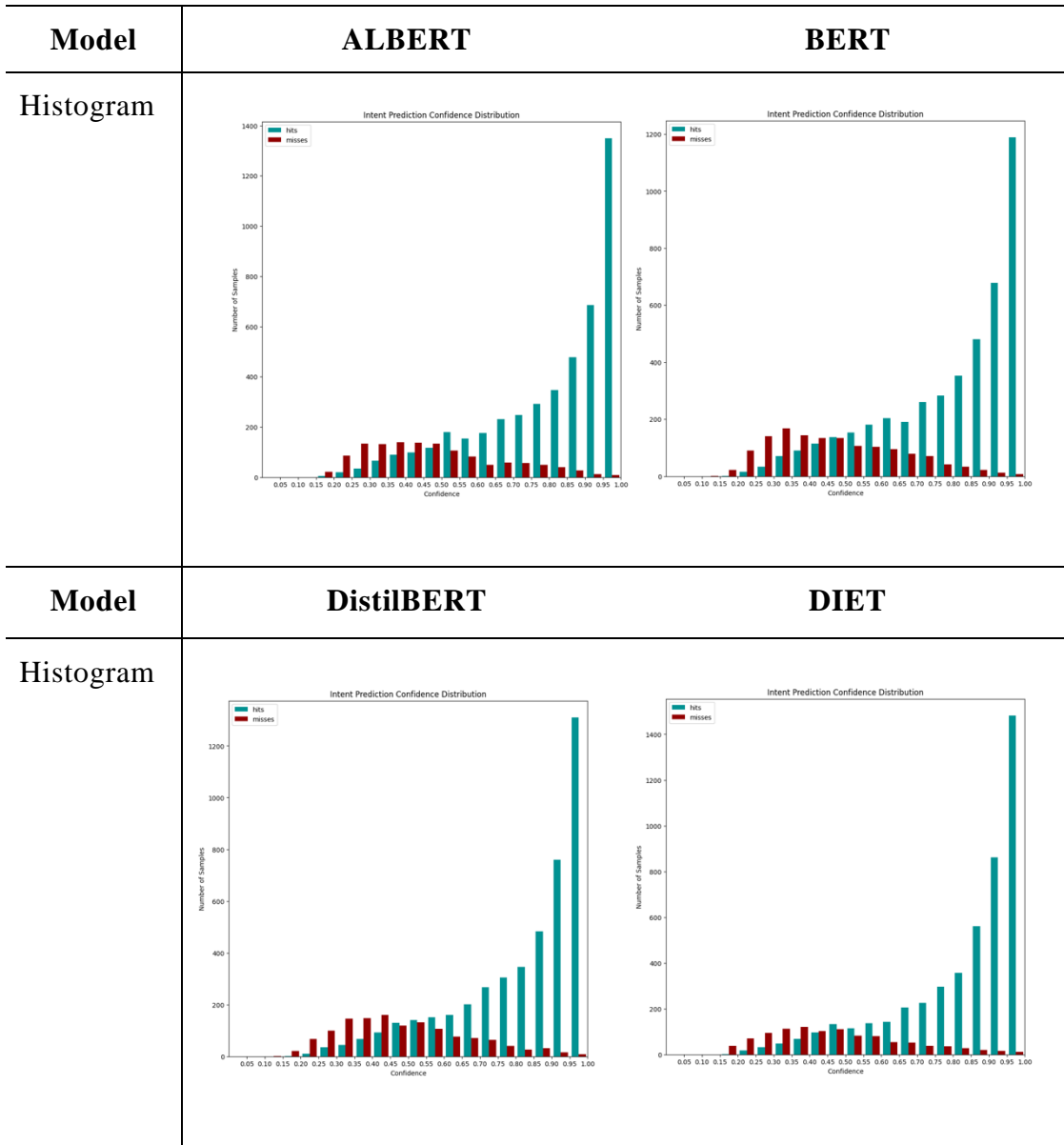
Ayrıca DIET modelini Türkçe için Transformer tabanlı önceden eğitilmiş BERT modelleriyle karşılaştırıp, ortalama olarak en iyi olduğunu ve iki görevde en yüksek puanları aldığını görüyoruz. DIET üzerindeki çalışmaya devam etmeyi ve gelişimine rehberlik etmesi için diğer alan veri setlerini kullanmayı ve oluşturmayı planlıyoruz.

Modellerin karşılaştırılması Çizelge 6.1’de verilmiştir. Domain bazında modellerin sonuçları bölüm altındaki çizelgelerde görülmektedir. Öncelikle domain bazında her modelin histogram grafikleri verilmiştir. Ardında domain bazında model grafiğiyle birlikte detaylı sonuç bilgisi verilmiştir.

Çizelge 6.1: Transformer tabanlı modellerde temel değerlendirme.

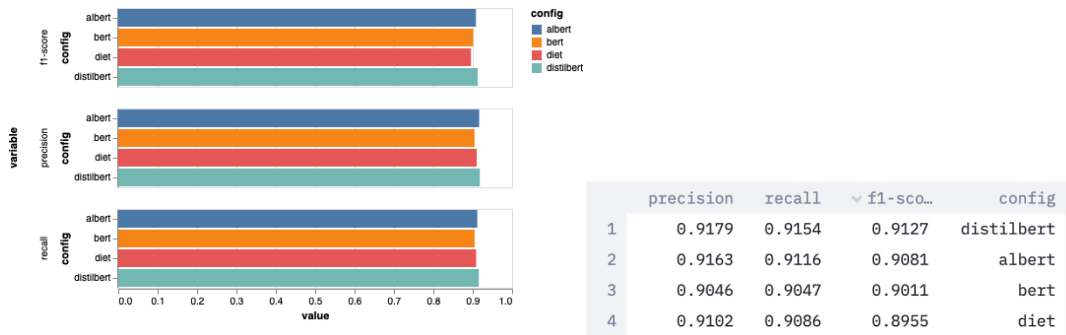
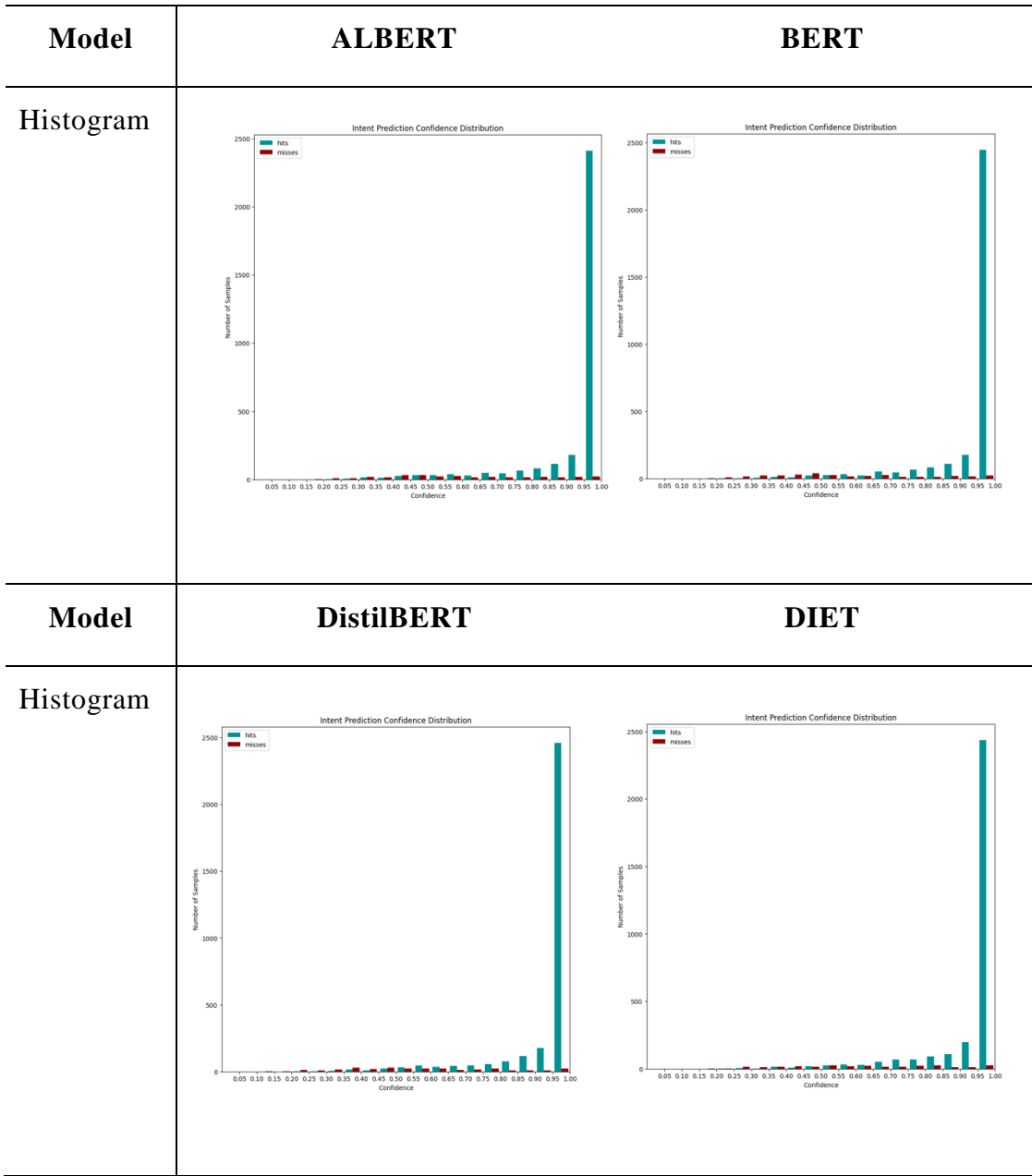
Model	AVG	BNKD	INVD	SCLD
BERT	81.4	74.9	90.1	79.4
DistilBERT	82.6	76.4	91.2	80.3
ALBERT	82.7	77.6	90.8	79.7
DIET	85.0	81.4	89.5	84.2

Çizelge 6.2: Bankacılık veri setinin (BNKD) modeller üzerindeki histogramı.



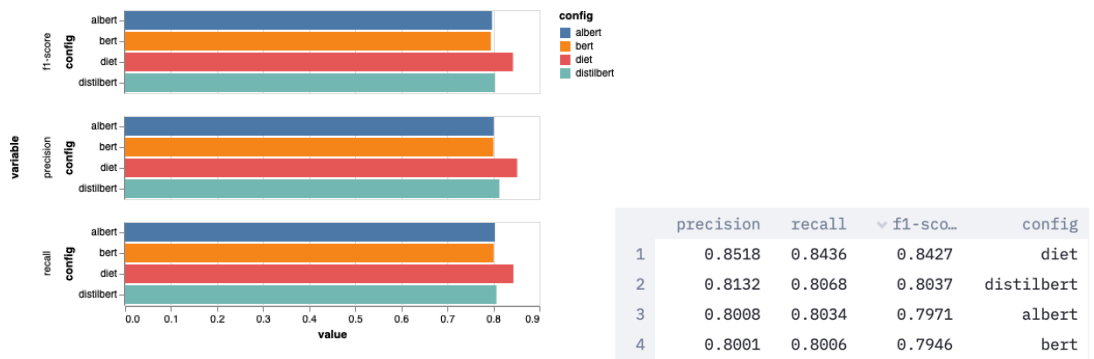
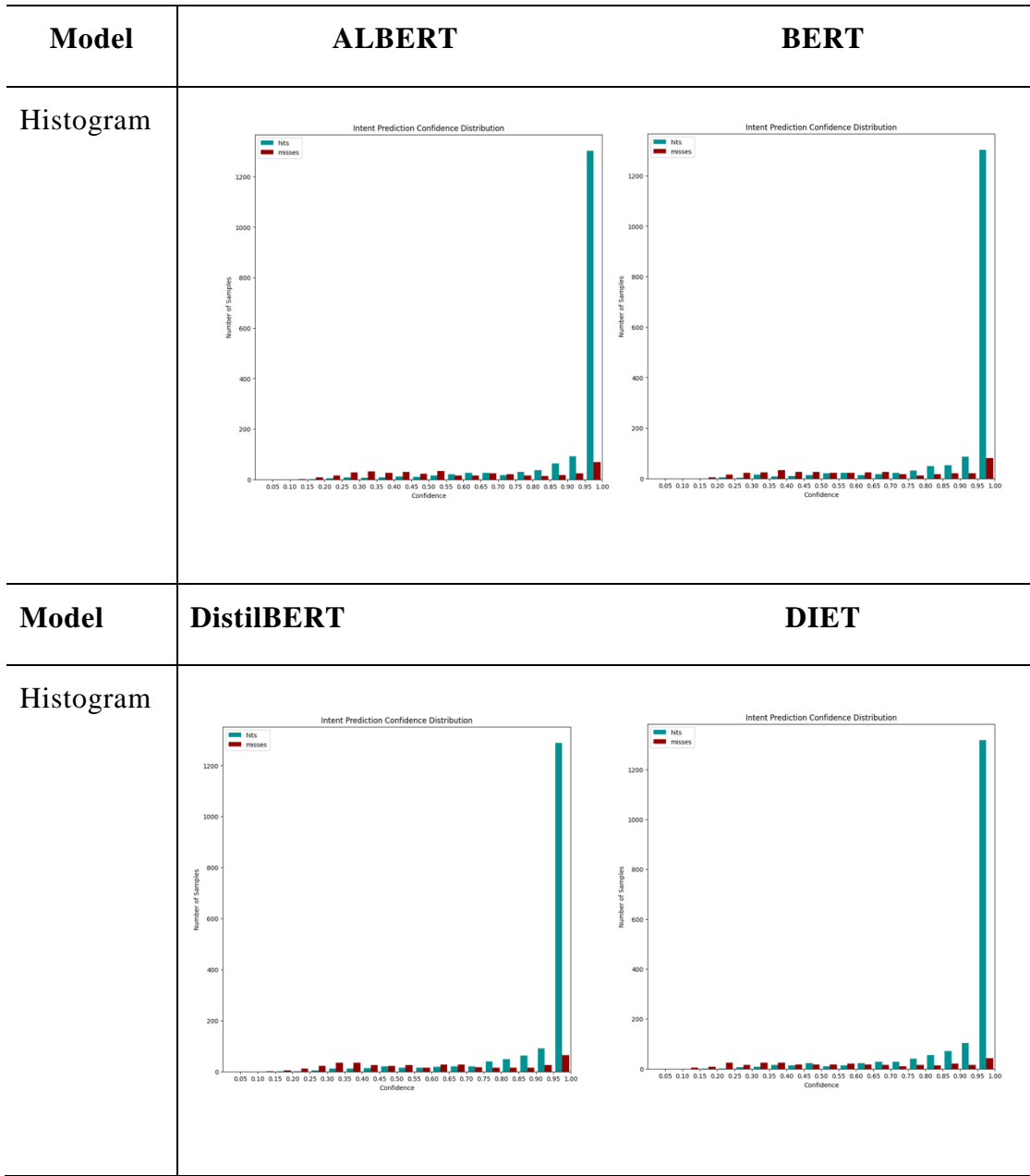
Şekil 6.1: Bankacılık veri setinin (BNKD) modellere göre sonuçları ve grafiği

Çizelge 6.3: Yatırım veri setinin (INVD) modeller üzerindeki histogramı.



Şekil 6.2: Yatırım veri setinin (INVD) modellere göre sonuçları ve grafiği

Çizelge 6.4 : Sosyal veri setinin (SCLD) modeller üzerindeki histogramı.



Şekil 6.3 : Sosyal veri setinin (SCLD) modellere göre sonuçları ve grafiği

6.2 Öneriler

Farklı modellerin farklı görevleri için ortak bir değerlendirme sağlamasını amaçlıyoruz. Bu amaçla, birçok kaynağın uyarlanması ve farklı alanlarda veri kümelerinin oluşturulması gerekmektedir.

Bu tür model karşılaştırmalarında, her modelin her problem için farklı sonuçlar vereceği bilinerek model performansına odaklanılmalıdır. Örneğin; eğitim verileri, hız veya model parametreleri gibi. Katılımcı modellerden belirli bir verimlilik düzeyi talep ederek ek kıyaslamalar elde etmek makul görünmektedir. Onu gelecekteki iş olarak görüyoruz.

DIET, değerlendirilen modeller arasında kısmen iyi olsa da; farklı modeller farklı görevlerde daha iyi performans gösterebilmektedir. Farklı alanların çözülmekten uzak olduğu ve gelecekteki modelleri değerlendirmek ve karşılaştırmak için kullanılabileceğini görülmektedir. Farklı domainler için az sayıda veri kümesiyle iyi sonuçlar elde etmenin yolu, domain bazlı embedding'ler kullanmak veya oluşturmak olabilir.

KAYNAKLAR

- [1] **M. Artetxe, S. Ruder, and D. Yogatama.** 2019. *On the cross-lingual transferability of monolingual representations.*
- [2] **P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov.** 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606.*
- [3] **Nadkarni PM., Ohno-Machado L., Chapman WW.** Natural language processing: an introduction . *J Am Med Inform Assoc* . 2011; 18 (5): 544 – 51.
- [4] **Ashish V., Noam S., Niki P., Jakob U., Llion J., Aidan N.G., Lukasz K., and Illia P.** 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 6000–6010.
- [5] **A. Conneau and D. Kiela.** 2018. Senteval: An evaluation toolkit for universal sentence representations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018).*
- [6] **A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes.** 2017. Supervised learning of universal sentence representations from natural language inference data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 670–680, Copenhagen, Denmark. Association for Computational Linguistics.
- [7] **J. Devlin, M. Chang, K. Lee, and K. Toutanova.** 2019. 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, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- [8] **E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov.** 2018. Learning word vectors for 157 languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018).*
- [9] **J. Howard and S. Ruder.** 2018. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 328–339.
- [10] **D.P. Kingma and J. Ba.** 2014. Adam: A method for stochastic optimization. Cite *arxiv: 1412.6980* Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

- [11] **G. Lample and A. Conneau.** 2019. Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291*.
- [12] **P. Lison and J. Tiedemann.** 2016. Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. Euro- pean Language Resources Association (ELRA).
- [13] **B. McCann, N.S. Keskar, C. Xiong, and R. Socher.** 2018. The natural language decathlon: Multitask learning as question answering. *arXiv preprint arXiv:1806.08730*.
- [14] **B.H.R. Sennrich and A. Birch.** 2016. Neural machine translation of rare words with subword units. In *Association for Computational Linguistics (ACL)*, pages 1715–1725.
- [15] **A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S.R. Bowman.** 2019a. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In the Proceedings of ICLR.
- [16] **A. Wang, I.F. Tenney, Y. Pruksachatkun, K. Yu, J. Hula, P. Xia, R. Pappagari, S. Jin, R.T. McCoy, R. Patel, Y. Huang, J. Phang, E. Grave, H. Liu, N. Kim, P.M. Htut, T. F’evry, B. Chen, N. Nangia, A. Mohanane, K. Kann, S. Bordia, N. Patry, D. Benton, E. Pavlick, and S.R. Bowman.** 2019b. jiant 1.2: A software toolkit for research on general purpose text understanding models.
- [17] **A. Williams, N. Nangia, and S. Bowman.** 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1112–1122. Association for Computational Linguistics.
- [18] **P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang.** 2016. SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas. Association for Computational Linguistics.
- [19] **P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang.** 2016. Squad: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392.
- [20] **A. Conneau, R. Rinott, G. Lample, A. Williams, S.R. Bowman, H. Schwenk, and V. Stoyanov.** 2018. Xnli: Evaluating cross-lingual sentence representations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- [21] **T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, and J. Brew.** 2019. Huggingface’s transformers: State-of-the-art natural language processing. *ArXiv*, abs/1910.03771.

- [22] **A. Williams, N. Nangia, and S. Bowman.** 2018. A broad-coverage challenge corpus for sentence understanding through inference. In NAACL.
- [23] **Y. Wu, M. Schuster, Z. Chen, Q.V Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al.** 2016. Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.
- [24] **T. Bunk, D. Varshneya, V. Vlasov, and A. Nichol.** 2020. DIET: Lightweight language understanding for dialogue systems. *CoRR, abs/2004.09936*.
- [25] **Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut.** Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942, 2019*.
- [26] **V. Sanh, L. Debut, J. Chaumond, and T. Wolf.** DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108, 2019*.

İnternet Kaynakları:

URL-1 Github BERT <https://github.com/google-research/bert> adresinden alındı.
Erişim Tarihi: 28.09.2020

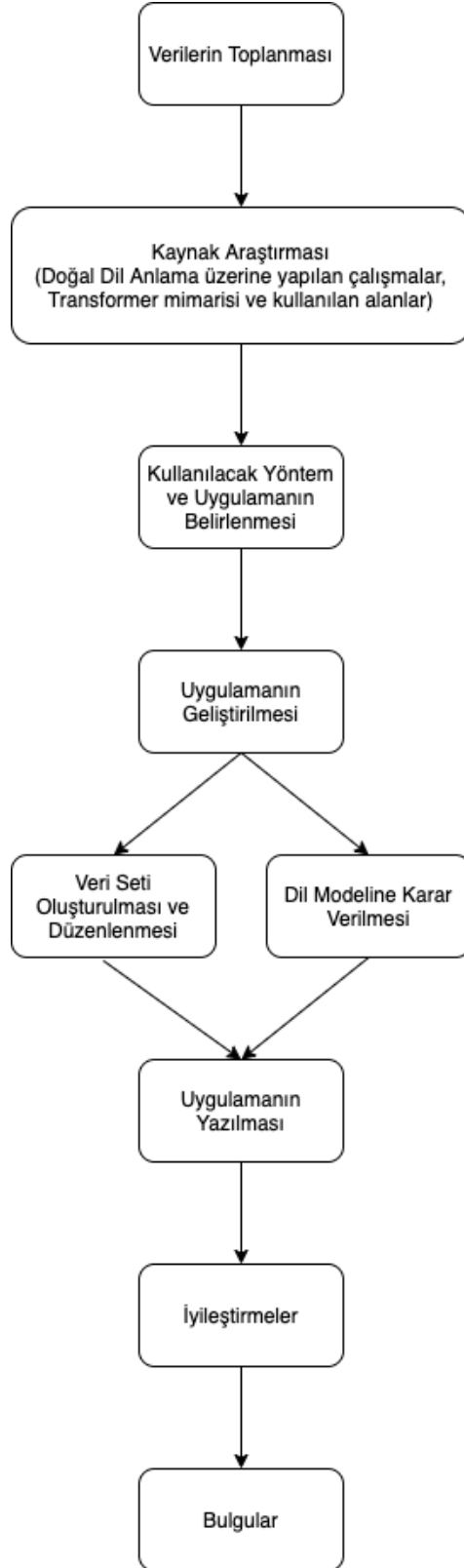
URL-2 Bitirme tezi
https://github.com/oguzhankarahan/Evaluation_of_NLU_PAPER.git
adresine eklendi. Oluşturma Tarihi: 28.09.2020

EKLER

EK A : Şekiller

EK B : Kodlar

EK A Şekiller



Şekil A.1 : Tez Çalışması Akış Diyagramı

EK B. Transformer Tabanlı Dil Modellerinin Karşılaştırılması için Kullanılan
Python Dilinde Yazılmış Kod
URL-2

ÖZGEÇMİŞ



Ad-Soyad : Oğuzhan KARAHAN
Doğum Tarihi : 02.05.1992
Doğum Yeri : İslahiye/Gaziantep
E-posta : oguzhankarahan.tr@gmail.com

Öğrenim Durumu

- **Lise** : 2010, Gaziantep Abdulkadir Konukoğlu Fen Lisesi
- **Lisans** : 2015, Anadolu Üniversitesi, İşletme Fakültesi, İşletme
- **Lisans** : 2015, İnönü Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği
- **Yüksek Lisans** : İstanbul Aydın Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği

Beceri/Yetenek

- Android, IOS, Hybrid
- Java, .NET, Python, C++
- Javascript
- SQL, NoSQL
- Haskell
- Route Planning/Optimization
- AI, ML
- Data Science
- Big Data
- Software Architecture

Mesleki Deneyim

- **Yazılım Mimarı - AI/ML: 05/2017 – Devam Ediyor**
Softtech A.Ş., İstanbul
 - Robotik ve yapay zeka projelerinde yer almaktayım. İnsansı robot, veri bilimi, büyük veri, doğal dil işleme ve makine öğrenmesi projelerinde çalışıyorum. Aynı zamanda ekibin yazılım mimarlığı görevini yürütmekteyim.

Projeler:

- Şube içi mobil ve web uygulamalar, dijital onay, biyometrik imza, rota optimizasyonu ve planı, insansı robot Pepper, doğal dil işleme ve yapay zeka.

Teknolojiler:

- Python, Haskell, C++, Java, React/Angular/Vue, Node.js, SQL/NoSQL, React Native/Android Java/Kotlin/Ionic.

- **Yazılım Geliştirme Danışmanı: 08/2015 – 05/2017**
Manim Finans Teknolojileri, İstanbul
 - Terralabs'ın izniyle, cumartesi günleri danışman olarak çalıştığım bir yazılım şirketi. Burada bankacılık entegrasyonu ve farklı uygulamalara danışmanlık desteği verdim.

Projeler:

 - Bankacılık ve Finans mobil, web ve masaüstü uygulamaları (Lookin2me – Finans Platformu, Manim, BankFIX), Mulfie – Çoklu Selfie, Cast Application.

Teknolojiler:

 - Angular, Web API, WCF, MongoDB, Ionic Framework, Node.js, Electron.js, React Native.
- **Ekip Lideri: 04/2015 – 05/2017**
Terralabs Innovative Solutions, İstanbul
 - Terralabs bünyesine yazılım geliştirme uzmanı olarak katıldım. Farklı alanlarda birçok projede çalıştım.

Projeler:

 - Şirket içi yönetim uygulamaları (TCRM), saha takip uygulaması (Bulk Delivery Application - Android), çağrı merkezi uygulaması (TCallCenter), saha takip mobil ve web uygulaması (TKATO), konum bazlı sistem geliştirme (MAKS Projesi), saha takip mobil ve web uygulamaları (TKMobile), rota optimizasyon entegrasyonu (TKRoute).

Teknolojiler:

 - .Net, Java, Angular 1.x, Web API, WCF, Windows Service, MSSQL/Oracle/MongoDB/CouchDB, Ionic Framework, Android Java, Node.js, GMap, Leaflet, Graphhoppers.
- **Yazılım Uzmanı: 07/2014 – 04/2015**
Leonardo Travel, İstanbul
 - Leonardo Travel bünyesinde önce uzman yardımcısı sonrasında ise uzman olarak çalıştım.

Projeler:

 - Yönetim paneli (Iceberg Management Systems), mobil uygulama (Galahotels).

Teknolojiler:

 - .Net, Angular 1.x, Web API

Yabancı Dil

- İngilizce