

Kocaeli Üniversitesi Teknoloji Fakültesi Bilişim Sistemleri Mühendisliği: Türkçe Şiirleri Konularına Göre Sınıflandırma

Tuğçe Gül
Bilişim Sistemleri Mühendisliği
Kocaeli Üniversitesi
Kocaeli, Türkiye
221307036@kocaeli.edu.tr

Bu projede, Türkçe şiirlerin içeriklerine göre otomatik olarak sınıflandırılması hedeflenmiştir. Şiirler; Aşk, Din, Doğa, Kahramanlık ve Gurbet olmak üzere beş ana tema altında etiketlenmiş ve bu tematik sınıflandırma işlemi için doğal dil işleme (NLP) ve makine öğrenmesi yöntemleri kullanılmıştır. Veri seti, çeşitli dijital şiir platformlarından web kazıma (web scraping) yöntemleriyle toplanmıştır. Toplanan veriler üzerinde küçük harfe çevirme, noktalama işaretlerinin kaldırılması, stopword temizleme ve kök bulma gibi ön işleme adımları uygulanmıştır. Modelleme sürecinde, günümüzde metin sınıflandırma alanında yaygın olarak kullanılan Transformer tabanlı derin öğrenme modelleri tercih edilmiştir. Bu kapsamda; BERT, RoBERTa, DeBERTa, ALBERT ve GPT modelleri kullanılmış ve her birinin performansı doğruluk, precision, recall ve F1-score gibi metriklerle karşılaştırılmıştır. Modellerin şiir gibi anlam yoğunluğu yüksek, sanatsal metinlerde gösterdiği başarı düzeyi değerlendirilmiş ve en etkili yaklaşım belirlenmiştir. Bu çalışma, Türkçe şiirlerin konularına göre otomatik olarak sınıflandırılmasına yönelik bir yöntem sunmakta ve edebi içeriklerin yapay zeka destekli analizine katkı sağlamayı amaçlamaktadır.

I. GİRİŞ

Şiir, bireylerin duygularını, düşüncelerini ve yaşantılarını estetik bir biçimde ifade ettikleri güçlü bir edebi türdür. Türk edebiyatında ise aşk, doğa, din ve vatan gibi temalar şiirlerde en sık işlenen konular arasında yer almaktadır. Ancak dijital ortamda yer alan binlerce şiirin konu bazlı sınıflandırılması, manuel olarak yapılması durumunda oldukça zaman alıcı ve subjektif bir sürece dönüşmektedir. Bu nedenle, metin verilerinin konuya göre otomatik sınıflandırılması, yapay zeka temelli yöntemlerle çözülmesi gereken önemli bir problem haline gelmiştir. Bu proje kapsamında, Türkçe şiirlerin içeriklerine göre konularına göre otomatik olarak sınıflandırılması amaçlanmıştır. Şiirler dört ana tema altında etiketlenmiş; ardından doğal dil işleme (NLP) teknikleri kullanılarak bu metinlerin sınıflandırılması gerçekleştirilmiştir. Web scraping yöntemiyle çeşitli dijital şiir platformlarından elde edilen veriler, metin temizleme, stopword kaldırma ve kök bulma gibi ön işleme adımlarından geçirilmiştir. Modelleme sürecinde, günümüzde metin sınıflandırma problemlerinde başarıyla kullanılan Transformer

tabanlı derin öğrenme modelleri tercih edilmiştir. Bu doğrultuda; BERT, RoBERTa, DeBERTa, ALBERT ve GPT gibi güçlü modeller kullanılarak eğitim ve doğrulama süreçleri gerçekleştirilmiş, her bir modelin başarı düzeyi çeşitli metriklerle değerlendirilmiştir. Bu modellerin şiir gibi anlam yoğunluğu yüksek, sanatsal metinler üzerinde gösterdiği performanslar karşılaştırılarak en uygun yaklaşım belirlenmiştir. Bu raporda, şiir sınıflandırma süreci; veri toplama, ön işleme, modelleme ve değerlendirme adımlarıyla sistematik biçimde sunulmakta, elde edilen sonuçlar doğrultusunda yorumlar ve öneriler paylaşılmaktadır. Bu çalışmanın amacı, edebi metinler üzerinde yapay zeka uygulamalarının potansiyelini ortaya koymak ve dijital şiir analizine katkı sunmaktır.

II. PROJE SÜRECİ

"Bu proje, Türkçe şiirlerin konularına göre sınıflandırılmasına yönelik çok aşamalı bir doğal dil işleme çalışmasıdır. İlk aşamada, farklı temalara ait şiirlerin veri olarak toplanması ve ön işleme süreçleri gerçekleştirilmiştir. Bu rapor ise projenin ikinci aşamasını, yani sınıflandırma modelinin eğitilmesi ve test edilmesi sürecini kapsamaktadır."

- Veri Toplama Süreci:** İlk aşamada 5 farklı şiir kategorisi belirlenip her kategoriye ait (aşk, pastoral, kahramanlık, gurbet, tasavvuf) web scrabing ile yaklaşık 5000 şiir toplanmıştır.
- Veri temizliği ve artırımı:** Veri ön işleme sürecinde, şiir metinleri üzerinde tokenizasyon uygulanarak kelimeler ayrıştırılmış, ardından anlam taşımayan kelimeler (stopword'ler) metinden çıkarılmıştır. Daha sonra her kelimenin kök ya da temel hâline ulaşmak amacıyla lemmatizasyon işlemi gerçekleştirilmiştir. Ama projenin ilerleyen aşamalarında şiirlere göre sınıflandırma olduğu için

3. **Veri Seti Hazırlığı:** Eğitim sürecinde, veri seti eğitim ve test verisi olarak %70 e %20 oranında adil bir şekilde ayrılır.
4. **Model Eğitimi:** Veri işleme sürecinin tamamlanmasının ardından, transörmör tabanlı modellerle model eğitimi sürecine geçilmiştir. Model eğitim süreci, Google Colab platformu üzerinde gerçekleştirilmiştir. Bu platform, hızlı işlem gücü ve GPU desteği ile büyük veri setlerinin işlenmesi ve modellerin eğitilmesi için etkili bir çözüm sunmuştur. Eğitim aşamasında, aşağıdaki transörmör tabanlı derin öğrenme modelleri kullanılmıştır:

- "albert-base-v2"
- "microsoft/deberta-base"
- "xlm-roberta-base"
- savasy/bert-base-turkish-sentiment-cased
- dbmdz/electra-base-turkish-cased-discriminator

5. **Test Süreci Ve Değerlendirme:** Model eğitimi tamamlandıktan sonra, sınıflandırma modellerinin performansını değerlendirmek için çeşitli testler gerçekleştirilmiştir. Her modelin başarımı, aşağıdaki performans metrikleri ve analiz yöntemleri kullanılarak incelenmiştir:

Performans Metrikleri:

- Accuracy (Doğruluk): Modelin genel sınıflandırma başarısını ölçmek için kullanılmıştır.
- Recall (Duyarlılık): Modelin pozitif örnekleri doğru şekilde tespit etme oranı değerlendirilmiştir.
- Precision (Kesinlik): Modelin pozitif tahminlerinin doğruluğu ölçülmüştür.
- Sensitivity (Hassasiyet): Pozitif sınıf tespiti için modelin başarımı analiz edilmiştir.
- Specificity (Özgüllük): Negatif sınıfta tespiti doğruluğu hesaplanmıştır.
- F1-Score: Precision ve Recall'un dengeli bir ölçüsü olarak değerlendirilmiştir.
- AUC (Area Under Curve): ROC eğrisi altındaki alan hesaplanarak modelin genel performansı analiz edilmiştir.

- a) Karmaşıklık Matrisi:

- Her bir hastalık sınıfı için doğru ve yanlış sınıflandırma sayıları görselleştirilmiştir. Bu matris, modelin hangi sınıflarda daha başarılı veya başarısız olduğunu belirlemek için kullanılmıştır.

- b) ROC Eğrisi(ReceiverOperatingCharacteristic):

- Sınıflandırma modelleri için ROC eğrileri çıkarılarak, modelin pozitif ve negatif sınıfları ayırtma başarımı analiz edilmiştir.

- c) Eğitim ve TestGrafikleri:

- Her bir modelin eğitim ve test veri kümeleri için epoch başına loss değişim grafikleri aynı eksen üzerinde çıkarılmıştır. Bu grafikler, modelin öğrenme sürecini ve genel eğilimlerini gözlemlemek için kullanılmıştır.

- I. Zaman Analizi:

- Her modelin eğitim süresi (training time) ve çıkarım süresi (inference time) ölçülerek karşılaştırılmıştır.

III.KULLANILAN MODELLER VE TEST SONUÇLARI

Çalışma öncesinde, model eğitimi sürecinde karşılaşılan aşırı öğrenme (overfitting) problemini önlemeye yönelik yöntemler araştırılmış; bu doğrultuda, geliştirilen kod yapısı ilgili önlemleri içerecek şekilde tasarlanmıştır.

- **Weight Decay (Ağırlık Cezalandırma):** Modelin ağırlıklarının büyüklüğünü cezalandırarak, öğrenilen parametrelerin aşırı büyümesini engeller. Bu sayede modelin veriye ezberleme yapmadan genellebilir öğrenme gerçekleştirmesi hedeflenmiştir.
- **Early Stopping (Erken Durdurma):** Eğitim sırasında doğrulama (validation) kaybı belirli bir süre boyunca iyileşmediğinde eğitim süreci otomatik olarak sonlandırılır. Bu mekanizma, modelin eğitim verisine aşırı uyum sağlamadan durmasını sağlar.
- **Warmup Ratio:** Öğrenme oranı başlangıçta küçük değerlerden başlayarak kademeli olarak artırılır. Bu sayede model, eğitimin başlarında ağırlıkları ani değişikliklerle güncelleyerek ezberleme eğilimine girmez.
- **Gradient Accumulation (Gradyan Birikimi):** Bellek kısıtlarından dolayı küçük batch'ler ile eğitim yapılırken, birden fazla küçük batch'in gradyanları biriktirilerek daha büyük etkin batch boyutları elde edilir. Bu, eğitim sürecinin daha stabil ilerlemesini sağlar ve modelin aşırı uyum göstermesini dolaylı olarak azaltabilir.
- **Label Smoothing (Etiket Yumuşatma):** Hedef etiketlerin kesin (keskin) yerine yumuşatılmış değerler kullanılması, modelin çıktıları aşırı güvenle tahmin etmesini engeller. Bu sayede modelin genelleme kapasitesi artar.
- **Öğrenme Oranı Planlayıcısı (Learning Rate Scheduler):** Öğrenme oranı eğitim süreci boyunca kademeli olarak azaltılır. Bu, modelin eğitim verisine

aşırı uyum göstermesini engelleyerek daha dengeli bir öğrenme sağlar.

- **Dropout:**Eğitim sırasında belirli nöronlar rastgele devre dışı bırakılarak (drop) modelin bağımlılıkları azaltılır. Bu sayede modelin genel performansı artar ve aşırı öğrenme riski azalır.
- **Veri Artırma (Data Augmentation):**Eğitim verisi üzerinde çeşitli dönüşümler (örneğin döndürme, kırpma, renk değişiklikleri) uygulanarak veri çeşitliliği artırılmıştır. Bu, modelin sadece sınırlı sayıda örneği ezberlemesini değil, genellemeyi öğrenmesini sağlar.

1.Albert-Base-v2 Modeli

Bu çalışmada, şiir metinlerinin konu başlıklarına göre sınıflandırılması amacıyla doğal dil işleme (NLP) alanında yaygın olarak kullanılan ve önceden eğitilmiş bir transformer modeli olan **ALBERT (A Lite BERT)** tercih edilmiş; model, çok sınıflı metin sınıflandırma görevine uygun şekilde yapılandırılarak eğitilmiştir

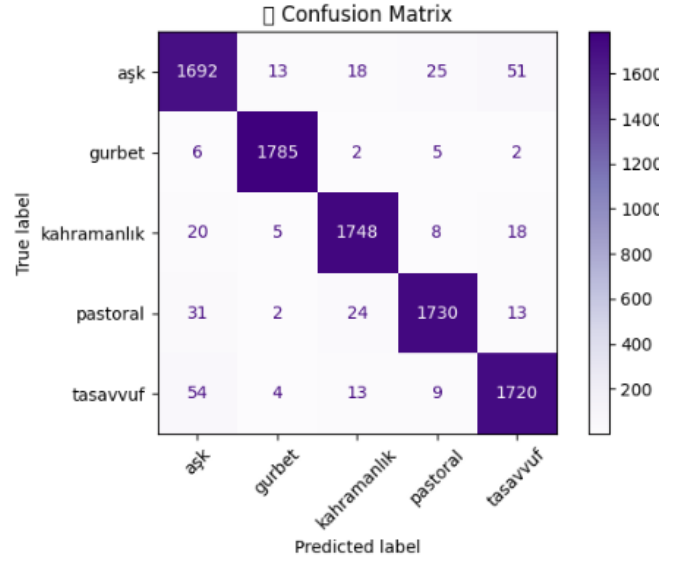
- **Model İçin Kullanılan Parametreler**

Parametre	Değer
model_ckpt	albert-base-v2
learning_rate	2e-5
num_train_epochs	50
per_device_train_batch_size	16
per_device_eval_batch_size	16
weight_decay	0.01
warmup_ratio	0.1
label_smoothing_factor	0.05
lr_scheduler_type	polynomial
load_best_model_at_end	True
metric_for_best_model	eval_loss
greater_is_better	False
early_stopping_patience	3
seed	42
report_to	[]
hidden_dropout_prob	0.1
attention_probs_dropout_prob	0.1

Modelin eğitim sürecinde kullanılan hiperparametreler yukarıda sunulmuştur. Bu parametreler, modelin öğrenme performansını artırmak, aşırı öğrenme (overfitting) riskini

azaltmak ve doğrulama verisi üzerindeki genelleme yeteneğini iyileştirmek amacıyla belirlenmiştir.Özellikle learning_rate, weight_decay, dropout oranları ve early_stopping_patience gibi değerler modelin öğrenme dengesini kontrol etmekte önemli rol oynamıştır. Ayrıca, label_smoothing_factor ve warmup_ratio gibi teknikler sayesinde modelin eğitim başında daha stabil öğrenmesi ve etiketlere aşırı güven duymaması sağlanmıştır. Eğitim sonunda en iyi doğrulama kaybına (eval loss) sahip model otomatik olarak kaydedilmiş ve değerlendirilmelerde bu model kullanılmıştır.

- **Karmaşıklık Matrisi Sonuçları**



Modelin sınıflandırma performansını daha ayrıntılı incelemek amacıyla oluşturulan **karmaşıklık matrisi**, her sınıfa ait doğru ve yanlış sınıflandırma sayılarını göstermektedir. Matriste diyagonal üzerinde yer alan değerler, modelin doğru tahminlerini temsil ederken; diyagonal dışındaki değerler yanlış sınıflandırmaları ifade etmektedir.

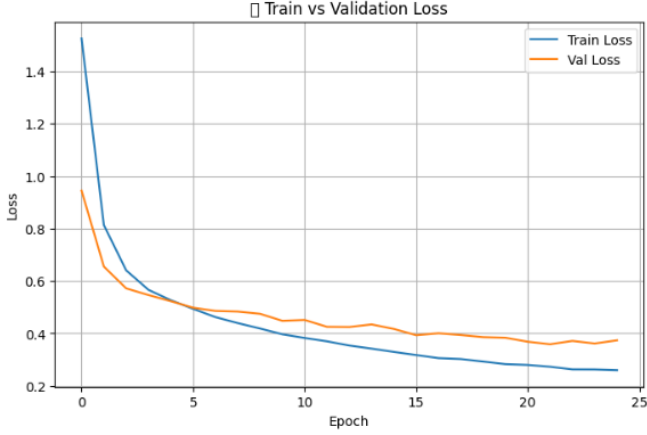
Elde edilen matrise göre, model beş farklı şiir teması olan **aşk**, **gurbet**, **kahramanlık**, **pastoral** ve **tasavvuf** sınıflarını yüksek doğrulukla ayırt edebilmiştir. Örneğin;

- "aşk" sınıfındaki 1692 örnek doğru sınıflandırılmış, yalnızca az sayıda örnek diğer sınıflarla karıştırılmıştır (örneğin 51 tanesi "tasavvuf" olarak sınıflanmıştır).
- "gurbet" sınıfı için modelin doğruluk oranı oldukça yüksektir; 1785 örnek doğru tahminlenmiş, yalnızca toplam 15 örnek başka sınıflara yanlış atanmıştır.
- Benzer şekilde "kahramanlık" sınıfında 1748 örnek doğru sınıflandırılmıştır ve hata sayısı çok düşüktür.

Tüm sınıflar özelinde yüksek doğru sınıflandırma oranları ve düşük çapraz sınıflandırma hataları, modelin **dengeyi**

koruyan ve genelleme yeteneği güçlü bir yapıya sahip olduğunu göstermektedir.

- **Eğitim Doğrulama Grafikleri ve Roc Eğrisi**



Grafikte, modelin eğitim süresince hem eğitim kaybının hem de doğrulama kaybının düzenli olarak azaldığı görülmektedir. Özellikle ilk epoch'larda her iki kayıpta da hızlı bir düşüş yaşanmış, bu da modelin veriden etkin şekilde öğrenmeye başladığını göstermektedir. Eğitim ilerledikçe eğitim kaybı istikrarlı şekilde azalmaya devam ederken, doğrulama kaybı da paralel bir seyir izlemektedir. 15. epoch'tan sonra doğrulama kaybında çok hafif dalgalanmalar gözlemlense de bu farklar oldukça küçüktür. Bu durum, modelin doğrulama verisi üzerinde genelleme yeteneğini koruduğunu ve belirgin bir aşırı öğrenme yaşanmadığını göstermektedir. Genel olarak, eğitim ve doğrulama kayıplarının birbirine yakın seyretmesi modelin dengeyi koruyarak iyi bir şekilde öğrendiğini ve overfitting riskinin düşük olduğunu göstermektedir.

Modelin performansını değerlendirmek amacıyla her bir sınıf için ROC (Receiver Operating Characteristic) eğrisi çizilmiştir. ROC eğrisi, modelin pozitif sınıfları ayırt etme kabiliyetini görselleştiren önemli bir metriktir. Eğri üzerindeki her bir nokta, farklı bir eşik değerine karşılık gelen True Positive Rate (TPR) ve False Positive Rate (FPR) değerlerini temsil eder. Grafikte her sınıf için ayrı ayrı çizilen ROC eğrileri, modelin sınıfları ne derece başarıyla ayırt edebildiğini göstermektedir. Bu değerlendirme, AUC (Area Under Curve) değeri ile desteklenmiştir. AUC değeri, ROC eğrisinin altında kalan alanı ifade eder ve modelin sınıf ayrımını ne kadar iyi yaptığı hakkında genel bir fikir verir. AUC değeri 1'e yaklaştıkça modelin ayırt edici gücü artar. Bu çalışma kapsamında elde edilen AUC değerleri; aşk sınıfı için 0.99, gurbet sınıfı için 1.00, kahramanlık sınıfı için 0.99, pastoral sınıfı için 0.99 ve tasavvuf sınıfı için 0.99 olarak hesaplanmıştır. Tüm sınıflarda AUC değerlerinin 0.99 ve üzeri olması, modelin sınıflar arasında yüksek ayırt edici başarıya sahip olduğunu göstermektedir. Özellikle "gurbet" sınıfı için elde edilen 1.00'lık AUC değeri, bu sınıfın model tarafından

hatasız olarak ayrıştırıldığını göstermektedir. Bu sonuçlara göre, modelin genel sınıflandırma başarısının oldukça yüksek olduğu ve overfitting (aşırı öğrenme) belirtisi göstermediği söylenebilir.

- **Perfonmans Göstergeleri Ve Metrikler**

Metrik	Değer
Accuracy	0.9641
Precision	0.9641
Recall	0.9641
F1 Score	0.9641
Sensitivity	0.9641
Specificity	0.991
AUC Macro	0.9909
Eğitim Süresi (sn)	10918.25
Inference Süresi (sn)	36.18

Modelin genel performans değerlendirmesi incelendiğinde oldukça başarılı sonuçlar elde edildiği görülmektedir. Modelin doğruluk (accuracy), kesinlik (precision), geri çağırma (recall), F1 skoru ve duyarlılık (sensitivity) değerlerinin tamamı 0.9641 olarak hesaplanmıştır. Bu durum, modelin hem doğru pozitif hem de doğru negatif sınıflandırmalarda dengeli ve istikrarlı bir performans gösterdiğini ortaya koymaktadır. Ayrıca özgüllük (specificity) değeri 0.9910 ile oldukça yüksek çıkmış, bu da modelin negatif sınıfları doğru şekilde tanımlamada oldukça başarılı olduğunu göstermektedir. AUC macro skoru ise 0.9909 olarak elde edilmiştir; bu değer, modelin sınıflar arasında genel ayırt edici gücünün yüksek olduğunu göstermektedir. Eğitim süresi toplamda 10918.25 saniye sürerken, modelin çıkarım (inference) süresi ise 36.18 saniye olarak kaydedilmiştir. Tüm bu metrikler değerlendirildiğinde, modelin hem eğitim hem de test aşamalarında yüksek doğrulukla çalıştığı ve aşırı öğrenme göstermeksizin genelleme yeteneğini koruduğu söylenebilir.

2. xlm-roberta-base Modeli

Bu çalışmada, şiir metnlerinin konu başlıklarına göre sınıflandırılması amacıyla doğal dil işleme (NLP) alanında yaygın olarak kullanılan ve önceden eğitilmiş bir transformer modeli olan **xlm-roberta-base** tercih edilmiş; model, çok sınıflı metin sınıflandırma görevine uygun şekilde yapılandırılarak eğitilmiştir.

- **Model İçin Kullanılan Parametreler**

Modelin eğitim sürecinde kullanılan hiperparametreler yukarıda sunulmuştur. Bu

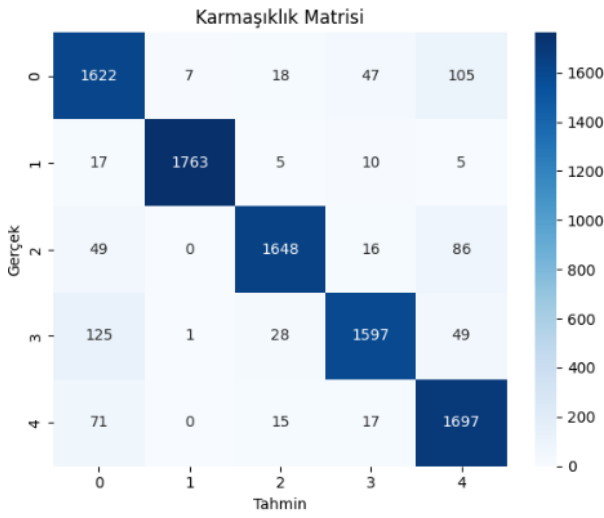
Parametre	Değer
model_checkpoint	xlm-roberta-base
learning_rate	3e-06
num_train_epochs	50
per_device_train_batch_size	8
per_device_eval_batch_size	8
weight_decay	0.1
warmup_ratio	0.1
label_smoothing_factor	0.1
lr_scheduler_type	polynomial
load_best_model_at_end	True
metric_for_best_model	eval_loss
early_stopping_patience	3
gradient_accumulation_steps	2
gradient_checkpointing	True
fp16	True
save_total_limit	1
hidden_dropout_prob	0.3
attention_probs_dropout_prob	0.3

parametreler, modelin öğrenme performansını artırmak, aşırı öğrenme (overfitting) riskini azaltmak ve doğrulama verisi üzerindeki genelleme yeteneğini iyileştirmek amacıyla belirlenmiştir. Özellikle learning_rate, weight_decay, dropout oranları ve early_stopping_patience gibi değerler modelin öğrenme dengesini kontrol etmekte önemli rol oynamıştır.

Ayrıca, label_smoothing_factor ve warmup_ratio gibi teknikler sayesinde modelin eğitim başında daha stabil öğrenmesi ve etiketlere aşırı güven duymaması sağlanmıştır.

Eğitim sonunda en iyi doğrulama kaybına (eval_loss) sahip model otomatik olarak kaydedilmiş ve değerlendirmelerde bu model kullanılmıştır

• Karmaşıklık Matrisi Sonuçları

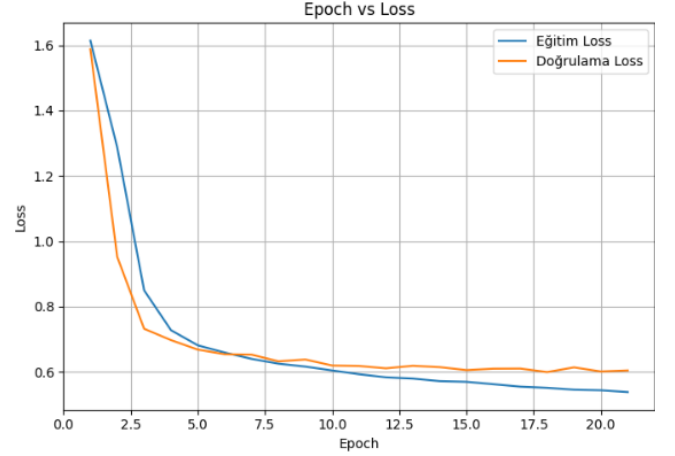


Modelin doğrulama verisi üzerindeki sınıflandırma performansı, aşağıda yer alan karmaşıklık matrisi ile detaylandırılmıştır. Matrisin çaprazındaki (köşegen) değerler, modelin doğru sınıflandırdığı örnek sayılarını, çapraz dışındaki değerler ise modelin hata yaptığı durumları göstermektedir.

Genel olarak, doğru sınıflandırma oranlarının oldukça yüksek olduğu gözlemlenmektedir. Özellikle **Sınıf 1**, 1763 doğru tahmin ile en başarılı sınıf olmuştur ve neredeyse hatasız sınıflandırılmıştır. **Sınıf 0** ve **Sınıf 4** arasında bazı karışıklıklar gözlemlense de bu karışıklıklar sınırlı düzeydedir ve genel performansı olumsuz etkilememektedir. **Sınıf 3**, 125 örneği yanlışlıkla Sınıf 0 olarak tahmin ettiği için en fazla hata yaptığı sınıf olmuştur, ancak yine de 1597 örneği doğru sınıflandırarak yüksek bir başarı sergilemiştir.

Bu sonuçlar, modelin tüm sınıflar arasında dengeli ve yüksek doğrulukla çalıştığını, şiir konularını doğru şekilde ayırt edebildiğini göstermektedir.

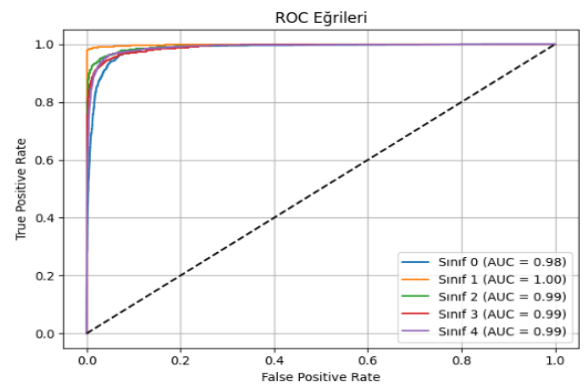
• Eğitim Doğrulama Grafikleri ve Roc Eğrisi



Modelin eğitim sürecine ait Epoch vs Loss grafiği incelendiğinde, eğitim (mavi çizgi) ve doğrulama (turuncu çizgi) loss değerlerinin ilk epoch'lerden itibaren hızlı bir şekilde düştüğü ve 5. epoch'tan sonra daha dengeli bir seyir izlediği gözlemlenmektedir. Bu durum, modelin öğrenmeye hızlı bir şekilde adapte olduğunu ve doğrulama verisinde de başarı sağladığını göstermektedir.

Her iki eğri arasındaki fark oldukça düşük olup doğrulama loss'unda ani artışlar görülmemektedir. Bu da modelin eğitim verisine fazla ezber yapmadan, genellenebilir bir öğrenme gerçekleştirdiğini göstermektedir. 10. epoch sonrasında loss değerleri yatay bir eğilim göstermekte, bu da modelin daha fazla eğitimle anlamlı bir gelişim göstermediğini ve erken durdurmanın devreye girmesiyle eğitim sürecinin dengelendiğini göstermektedir.

Sonuç olarak, bu grafik modelin **aşırı öğrenme (overfitting)** yaşamadığını ve eğitim sürecinin başarıyla tamamlandığını desteklemektedir.



Modelin sınıflar arası ayırt edicilik performansı, ROC (Receiver Operating Characteristic) eğrileri ile görselleştirilmiştir. ROC eğrileri, modelin pozitif ve negatif örnekleri ne kadar başarılı şekilde ayırdığını göstermekte

olup, bu eğrilerin altında kalan alan (AUC – Area Under Curve) değeri modelin sınıflandırma yeteneğini sayısal olarak ifade eder. Elde edilen sonuçlara göre, model tüm sınıflar için oldukça yüksek AUC değerlerine ulaşmıştır: Sınıf 0 için 0.98, Sınıf 1 için 1.00, Sınıf 2 için 0.99, Sınıf 3 için 0.99 ve Sınıf 4 için 0.99. Özellikle Sınıf 1’e ait AUC değerinin 1.00 olması, modelin bu sınıfı mükemmel bir şekilde ayırt ettiğini göstermektedir. Diğer sınıflar için de AUC değerlerinin 0.98 ve üzeri olması, modelin tüm sınıflarda dengeli ve yüksek performansla çalıştığını kanıtlamaktadır. Eğrilerin tamamı, rastgele sınıflandırmayı temsil eden diyagonal çizginin oldukça üzerinde yer almakta ve bu da modelin pozitif sınıfları başarıyla tahmin edebildiğini ortaya koymaktadır. Genel olarak, ROC eğrileri ve AUC skorları, modelin sınıf ayırma gücünün güçlü ve genellenebilir olduğunu doğrulamaktadır.

• Performans Göstergeleri Ve Metrikler

	18. Epoch
Training Loss	0.5515
Validation Loss	0.599677
Accuracy	0.925428
Precision	0.927681
Recall	0.925424
Specificity	0.981357
F1-score	0.925881
AUC	0.991257

Modelin eğitim süreci boyunca her epoch’ta hesaplanan metrikler incelendiğinde, en başarılı sonuçların 18. epoch’ta elde edildiği görülmüştür. Bu epoch’ta modelin eğitim kaybı 0.5515, doğrulama kaybı 0.5996 olarak gerçekleşmiş ve doğruluk (accuracy) oranı %92.54’e ulaşmıştır. Aynı zamanda precision %92.76, recall %92.54, F1 skoru %92.59 ve özgüllük (specificity) %98.13 olarak ölçülmüştür. ROC AUC skoru ise %99.12 ile oldukça yüksek bir değer göstermiştir. Erken durdurma mekanizması sayesinde eğitim süreci toplamda 21 epoch’ta tamamlanmış; ancak en düşük doğrulama kaybının elde edildiği 18. epoch’taki model performansı en iyi model olarak kaydedilmiştir. Bu sonuçlar, modelin hem eğitim hem de doğrulama verisi üzerinde dengeli bir şekilde öğrenme sağladığını göstermektedir.

3) Microsoft/deberta-base Modeli

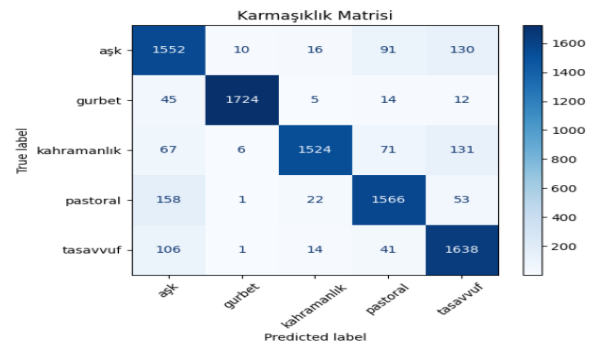
Bu projede, Türkçe metin sınıflandırma görevini yerine getirmek amacıyla, **Microsoft/deberta-base** modelini kullandık. DeBERTa (Decoding-enhanced BERT with disentangled attention), BERT tabanlı bir model olup, metin verilerini anlamada yüksek başarı sağlamak için geliştirilmiştir. Modelin güçlü dil anlama yetenekleri, özellikle

Türkçe gibi düşük kaynaklı dillerde başarılı sonuçlar elde etmek için önemli bir avantaj sunmaktadır.

• Model İçin Kullanılan Parametreler

Hiperparametre	Değer
Model	microsoft/deberta-base
Epoch Sayısı	100
Erken Durdurma	3 epoch
Batch Size (Train)	8
Batch Size (Eval)	8
Learning Rate	3e-6
Weight Decay	0.1
Label Smoothing	0.1
Warmup Ratio	0.1
Dropout Oranı	0.3
Gradient Accumulation	2
Metric for Best Model	eval_loss
Seed	42
Label Encoder	scikit-learn
Max Token Length	256

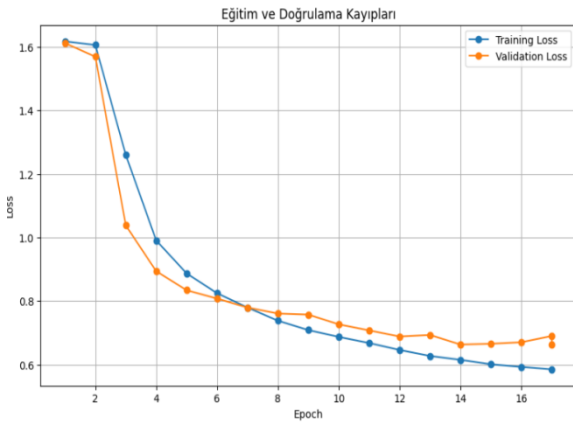
Yukarıdaki tabloda, modelin eğitimi sırasında kullanılan temel hiperparametreler özetlenmiştir. Bu parametrelerin her biri, modelin genel performansını doğrudan etkilemektedir. Eğitim sürecinde microsoft/deberta-base adlı önceden eğitilmiş bir transformer modeli tercih edilmiştir. Bu model, Türkçe şiir verisinde çok sınıflı bir metin sınıflandırma görevi için yeniden eğitilmiştir. Eğitim, maksimum 100 epoch boyunca sürdürülmüş olup, erken durdurma (early stopping) mekanizması sayesinde doğrulama kaybı belirli bir süre iyileşmediğinde eğitim 17. epochta sonlandırılmıştır. Aşırı öğrenmeyi önlemek amacıyla weight decay, dropout ve label smoothing gibi düzenleme (regularization) teknikleri uygulanmıştır. Öğrenme oranı düşük (3e-6) seçilerek modelin daha istikrarlı ve kontrollü şekilde öğrenmesi sağlanmıştır. Ek olarak, eğitim başlangıcında öğrenme oranını kademeli artırmak için warmup ratio değeri %10 olarak belirlenmiştir. Eğitim sırasında hesaplanan en iyi model, doğrulama kaybı (eval_loss) temel alınarak load_best_model_at_end=True ayarıyla otomatik olarak kaydedilmiştir. Belirlenen bu hiperparametreler, modelin hem öğrenme kapasitesini hem de genelleme yeteneğini dengelemek amacıyla özenle seçilmiş olup, deneysel gözlemler sonucunda optimize edilmiştir.



• Karmaşıklık Matrisi Sonuçları

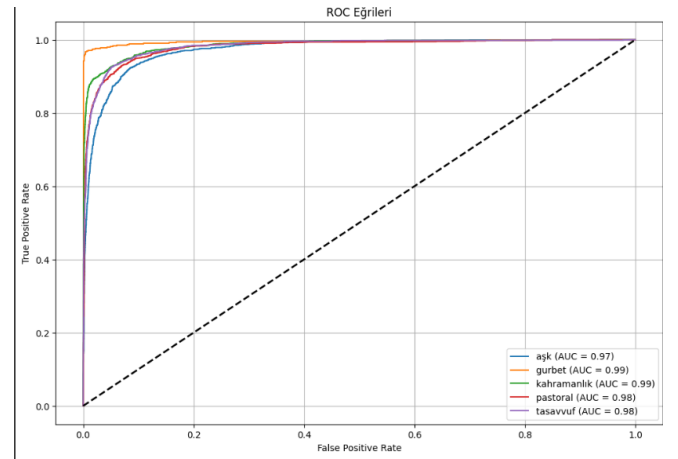
Modelin doğrulama veri kümesindeki performansını görselleştiren karmaşıklık matrisi, sınıflar arası tahmin başarısını detaylı biçimde ortaya koymaktadır. Matrisin diyagonalinde yer alan yüksek değerler, modelin genel olarak sınıfları doğru sınıflandırma yeteneğine sahip olduğunu göstermektedir. Örneğin, "aşk" sınıfına ait 1.552 örnek doğru tahmin edilmişken, 130 örnek "tasavvuf" ve 91 örnek "pastoral" sınıfına yanlış atanmıştır. Bu durum, tematik benzerliklerden kaynaklanan karışıklıklara işaret etmektedir. Benzer şekilde, "gurbet" sınıfı yüksek doğrulukla (1.724 doğru tahmin) sınıflandırılmış ve hata oranı oldukça düşük kalmıştır. "Kahramanlık" sınıfında 1.524 doğru tahmine karşın 131 örnek "tasavvuf" ve 71 örnek "pastoral" sınıfıyla karıştırılmıştır. Bu karışıklıklar, sınıflar arası bağlamsal veya sözcüksel örtüşmelerin model üzerinde etkili olabileceğini göstermektedir. "Pastoral" sınıfı için model 1.566 doğru tahmin üretmiş, özellikle 158 örneği "aşk" sınıfıyla karıştırmıştır. Bu da pastoral anlatıların duygu yüklü diliyle aşk temasına yakınlığını göstermektedir. "Tasavvuf" sınıfı ise 1.638 örnekle başarılı şekilde sınıflandırılmış olup, 106 örneğin "aşk", 41 örneğin ise "pastoral" sınıfına kayması, bu sınıflar arası semantik benzerliğe işaret etmektedir. Genel olarak, modelin ayırt edici gücü yüksek olsa da, bazı sınıflar arasındaki anlamsal yakınlıklar nedeniyle sınıflar arası karışıklıklar meydana gelmiştir. Bu durumun azaltılması için eğitim verisinin tematik çeşitliliği artırılabilir, sınıflar arasındaki sınırlar daha net çizilebilir ve modelin bağlam analizi yeteneği daha güçlü yapılarla desteklenebilir.

• Eğitim Doğrulama Grafikleri ve Roc Eğrisi



Yukarıda yer alan grafik, modelin eğitim sürecindeki eğitim (training loss) ve doğrulama (validation loss) kayıplarının epoch bazında değişimini göstermektedir. Eğitim süreci boyunca hem eğitim hem de doğrulama kayıplarında belirgin bir azalma gözlemlenmiştir. Bu durum, modelin öğrenme

sürecinde istikrarlı bir şekilde ilerlediğini ve başlangıçtaki yüksek kayıpların zamanla azaldığını göstermektedir. İlk birkaç epoch içerisinde özellikle hızlı bir düşüş gerçekleşmiş, ardından daha stabil bir azalma eğilimi gözlenmiştir. Eğitim kaybı ile doğrulama kaybı arasındaki fark, 10. epoch sonrasında belirginleşmekle birlikte, bu fark aşırı öğrenmeye (overfitting) işaret edecek düzeye ulaşmamıştır. Bu da, düzenleyici hiperparametrelerin (örneğin weight decay, dropout, label smoothing) etkin biçimde çalıştığını ve modelin genelleme yeteneğini koruduğunu göstermektedir. Eğitim, doğrulama kaybının iyileşmemesi üzerine 17. epochta erken durdurma (early stopping) mekanizması ile sonlandırılmıştır. Bu da modelin aşırı eğitilmesinin önüne geçerek genel performansın korunmasına katkı sağlamıştır. Sonuç olarak, eğitim ve doğrulama kayıplarının birlikte ve uyumlu şekilde azalması, modelin hem eğitim verisine hem de doğrulama verisine karşı tutarlı bir öğrenme gerçekleştirdiğini ortaya koymaktadır.



Yukarıdaki grafik, modelin çok sınıflı sınıflandırma performansını değerlendirmek amacıyla oluşturulan ROC (Receiver Operating Characteristic) eğrilerini ve her sınıfa ait AUC (Area Under Curve) değerlerini göstermektedir. ROC eğrileri, sınıflandırma modelinin farklı eşik değerlerinde ne ölçüde doğru pozitif oranı (True Positive Rate) ile yanlış pozitif oranı (False Positive Rate) arasında denge kurabildiğini ortaya koymaktadır. Grafikte yer alan eğrilerin tamamı, ideal sınıflandırmaya yakın olacak şekilde sol üst köşeye yakın konumlanmıştır. Bu durum, modelin yüksek ayırt edici güce sahip olduğunu göstermektedir. AUC değerleri incelendiğinde, modelin tüm sınıflarda 0.97 ile 0.99 aralığında oldukça yüksek performans sergilediği görülmektedir. "Aşk" sınıfı için AUC değeri 0.97 olarak hesaplanmış olup, bu sınıfta diğerlerine kıyasla görece daha düşük ama hâlâ yüksek bir doğruluk sağlanmıştır. "Gurbet" ve "Kahramanlık" sınıflarında AUC değeri 0.99 olarak elde edilmiştir; bu da modelin bu sınıflarda neredeyse kusursuz bir ayırım gerçekleştirdiğini göstermektedir. "Pastoral" ve "Tasavvuf" sınıflarında ise AUC değeri 0.98 olup, oldukça başarılı bir sınıflandırma

performansına işaret etmektedir.ROC eğrilerinin genel yapısı ve AUC değerlerinin yüksekliği, modelin hem duyarlılık (sensitivity) hem de özgüllük (specificity) bakımından dengeli ve güvenilir sonuçlar ürettiğini göstermektedir.Bu bulgular, eğitim sürecinde kullanılan hiperparametrelerin ve model mimarisinin başarılı şekilde optimize edildiğini desteklemektedir.

- **Perfonmans Göstergeleri Ve Metrikler**

	17. Epoch
Training Loss	0.585
Validation Loss	0.689924
Accuracy	0.889864
Precision	0.894701
Recall	0.889855
Specificity	0.972466
F1-score	0.890755
AUC	0.981788

Yukarıdaki tabloda, modelin eğitim sürecinde en yüksek performansı gösterdiği 17. epoch'a ait temel değerlendirme metrikleri özetlenmiştir.Bu epoch'ta training loss değeri 0.5850 ve validation loss değeri 0.6899 olarak gerçekleşmiştir.Eğitim ve doğrulama kayıplarının birbirine yakın ve düşük seviyelerde olması, modelin aşırı öğrenme (overfitting) yapmadığını ve genelleme kabiliyetinin yüksek olduğunu göstermektedir.Modelin sınıflandırma başarımı, temel ölçütler üzerinden değerlendirildiğinde oldukça güçlü bir performans ortaya koymaktadır.Doğruluk oranı (accuracy) %88.98 olup, modelin tüm sınıflar genelinde doğru tahmin yapma oranı yüksektir.Hassasiyet (precision) %89.47 ve geri çağırma (recall) %88.98 ile modelin pozitif örnekleri hem doğru şekilde tanıyabildiği hem de pozitif tahminlerinin büyük kısmının isabetli olduğu anlaşılmaktadır.F1-score değeri %89.07 ile precision ve recall arasında denge sağladığını göstermektedir.Ayrıca, duyarlılık (sensitivity) %88.98 ve özgüllük (specificity) %97.24 seviyesinde olup, modelin hem pozitif hem negatif sınıfları ayırt etme başarısı yüksek düzeydedir.Tüm bu metriklerin ötesinde, ROC eğrisi altındaki alan (AUC) değeri 0.9818 olarak hesaplanmıştır.Bu da modelin tüm sınıflar arasında ayırt edicilik kabiliyetinin oldukça güçlü olduğunu göstermektedir.Sonuç olarak, 17. epoch, hem doğrulama kaybı hem de performans metrikleri açısından modelin en dengeli ve başarılı olduğu dönem olarak belirlenmiştir.

4) savasy/bert-base-turkish-sentiment-cased modeli

Bu çalışmada, şiir metinlerinin konu başlıklarına göre sınıflandırılması amacıyla doğal dil işleme (NLP) alanında yaygın olarak kullanılan ve önceden eğitilmiş bir transformer

modeli olan **BERT** tercih edilmiş; model, çok sınıflı metin sınıflandırma görevine uygun şekilde yapılandırılarak eğitilmiştir

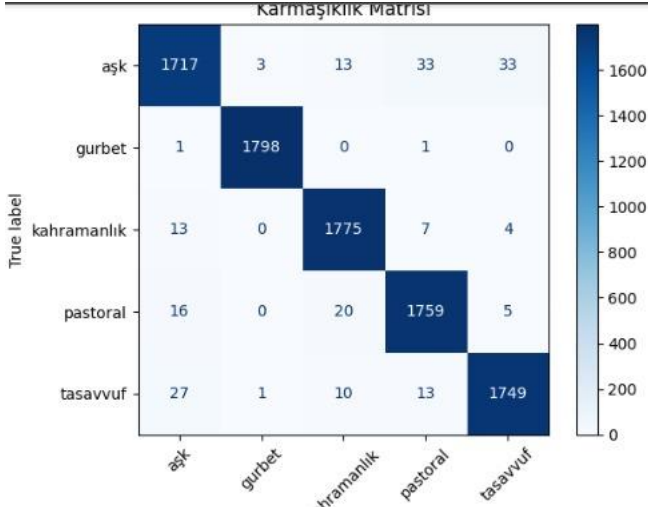
- **Model İçin Kullanılan Parametreler**

Hiperparametre	Değer
Model	savasy/bert-base-turkish-sentiment-cased
Epoch Sayısı	50
Erken Durdurma	3 epoch
Batch Size (Train)	16
Batch Size (Eval)	16
Learning Rate	5e-6
Weight Decay	0.01
Label Smoothing	0.05
Warmup Ratio	0.1
Dropout Oranı	0.1
Metric for Best Model	eval_loss
Seed	42
Maksimum Token Uzunluğu	256

Yukarıdaki tabloda,savasy/bert-base-turkish-sentiment-cased adlı önceden eğitilmiş Türkçe BERT modeliyle gerçekleştirilen çok sınıflı metin sınıflandırma görevi için kullanılan temel hiperparametreler sunulmaktadır.Modelin eğitimi maksimum 50 epoch boyunca planlanmış olup,dogrulama kaybının belirli bir sürede iyileşmemesi durumunda eğitimi durdurmak amacıyla 3 epoch'luk erken durdurma (early stopping) mekanizması etkinleştirilmiştir.Bu mekanizma,aşırı öğrenmenin önüne geçilmesi ve modelin genelleme kabiliyetinin korunması açısından kritik rol oynamaktadır.Eğitimde kullanılan batch size değeri hem eğitim hem de doğrulama seti için 16 olarak belirlenmiş,bu da modelin belleği verimli kullanmasını ve istikrarlı bir öğrenme süreci yürütmesini sağlamıştır.Öğrenme oranı 5e-6 gibi oldukça düşük bir değerle belirlenerek modelin parametre güncellemelerinde hassas ilerlemesi hedeflenmiştir.Overfitting'i önlemek amacıyla weight decay oranı 0.01 olarak ayarlanmış,ayrıca etiket yumuşatma (label smoothing) yöntemiyle %5 oranında etiket belirsizliği tanımlanarak modelin daha sağlam kararlar verebilmesi desteklenmiştir.Modelin eğitiminin başlangıcında öğrenme oranının ani artışı önlemek adına warmup ratio değeri %10 olarak tanımlanmış,böylece modelin erken dönem öğrenme davranışı daha kararlı hale getirilmiştir.Transformer mimarisine özgü olan dropout oranı hem gizli katmanlarda hem de dikkat (attention) mekanizmalarında %10 olarak uygulanmıştır.Bu oran,eğitim sürecinde ağırlıkların rastgele kapatılarak modelin daha esnek hale getirilmesine katkı sağlamıştır.En iyi modelin belirlenmesinde esas alınan performans ölçütü eval_loss olarak seçilmiş,bu da doğrulama kaybı en düşük olan modelin eğitim sonunda otomatik olarak kaydedilmesini sağlamıştır.Ayrıca,rastgelelikten kaynaklı farkları azaltmak için eğitim süreci seed=42 olarak sabitlenmiş ve tüm metin girdileri max_length=256 olacak şekilde

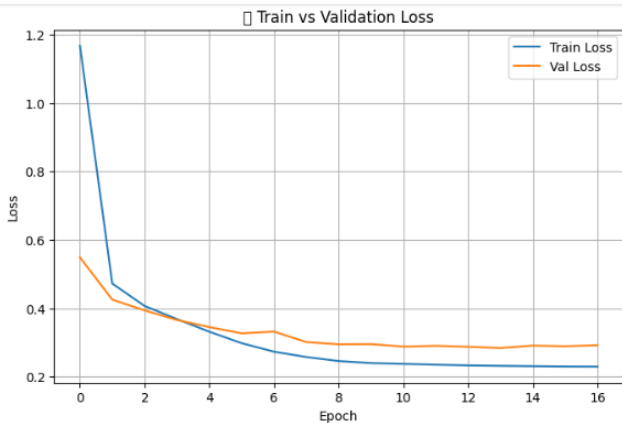
sınırlandırılmıştır.Bu hiperparametre seçimleri,modelin dengeli,doğru ve kararlı bir sınıflandırma performansı göstermesi amacıyla dikkatle optimize edilmiştir.

- Karmaşıklık Matrisi Sonuçları**

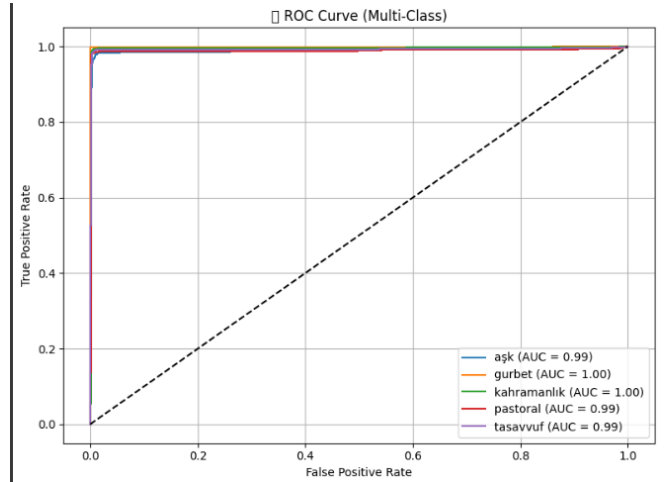


Yukarıdaki karmaşıklık matrisi,geliştirilen BERT tabanlı çok sınıflı sınıflandırma modelinin doğrulama veri kümesi üzerindeki performansını görselleştirmektedir.Matrisin satırları gerçek sınıfları,sütunları ise modelin tahmin ettiği sınıfları temsil etmektedir.Diyagonal üzerindeki yüksek değerler,modelin ilgili sınıfları doğru bir şekilde tahmin ettiğini göstermektedir.Model,aşk sınıfında 1717 doğru tahmin gerçekleştirmiş olup,sadece 82 örneği diğer sınıflarla karıştırmıştır.Gurbet sınıfında model neredeyse mükemmel performans göstermiş,1798 örneğin yalnızca biri yanlış sınıflandırılmıştır.Kahramanlık sınıfında da modelin başarısı yüksektir;1775 doğru tahmine karşılık yalnızca sınırlı sayıda karışıklık gözlemlenmiştir.Pastoral sınıfında 1759 doğru tahmin yapılmış,toplam hata oranı oldukça düşük kalmıştır.Tasavvuf sınıfında ise 1749 örnek doğru tahminlenmiş ve diğer sınıflarla karışıklık minimal seviyede kalmıştır.Genel olarak,diyagonal dışındaki değerlerin düşük olması,modelin sınıflar arası ayrımı yüksek doğrulukla gerçekleştirebildiğini göstermektedir.Özellikle gurbet,kahramanlık ve pastoral sınıflarındaki yüksek netlik,modelin semantik olarak ayırt edici temaları başarıyla öğrendiğini ortaya koymaktadır.Bu karmaşıklık matrisi,modelin sınıf bazında tutarlı ve güçlü bir sınıflandırma performansı sergilediğini desteklemektedir.

- Eğitim Doğrulama Grafikleri ve Roc Eğrisi**



Yukarıdaki grafik,modelin eğitim sürecinde her bir epoch boyunca elde edilen eğitim kaybı (train loss) ve doğrulama kaybı (validation loss) değerlerinin karşılaştırmalı olarak sunulduğu performans izleme çıktısını göstermektedir.Eğitim sürecinin ilk epochlarında her iki kayıpta da hızlı bir düşüş gözlemlenmiş,bu da modelin veriyi hızla öğrenmeye başladığını göstermektedir.İlk epochta 1.2 seviyesinden başlayan eğitim kaybı birkaç epoch içinde 0.3 seviyelerine kadar gerilemiş,aynı şekilde doğrulama kaybı da ilk etapta 0.6 seviyesinden başlayarak kısa sürede istikrarlı şekilde düşüş göstermiştir.Epoch 6'dan itibaren doğrulama kaybındaki azalış yavaşlamış ve yaklaşık 0.3 seviyelerinde sabitlenmiştir.Bu noktadan sonra eğitim kaybı düşmeye devam ederken doğrulama kaybı nispeten yatay seyretmiştir.Bu durum,modelin eğitimi sürdürdükçe eğitim verisine daha iyi uyum sağladığını ancak doğrulama verisi üzerindeki iyileşmenin sınırlı kaldığını göstermektedir.Grafikte eğitim kaybı ile doğrulama kaybı arasındaki farkın artmaması ve paralel şekilde seyretmesi,modelin aşırı öğrenme (overfitting) eğilimi göstermediğine işaret etmektedir.Bu da kullanılan düzenleme tekniklerinin (örneğin dropout,weight decay,label smoothing) etkili biçimde uygulandığını desteklemektedir.Sonuç olarak,hem eğitim hem de doğrulama kaybının belirgin şekilde azalarak düşük seviyelerde sabitlenmesi,modelin istikrarlı ve genellenebilir bir öğrenme süreci geçirdiğini ortaya koymaktadır.



Yukarıdaki grafik,geliştirilen çok sınıflı metin sınıflandırma modelinin ROC (Receiver Operating Characteristic) eğrilerini ve her sınıfa ait AUC (Area Under Curve) değerlerini göstermektedir.ROC eğrileri,modelin farklı eşik değerlerinde doğru pozitif oranı (True Positive Rate) ile yanlış pozitif oranı (False Positive Rate) arasındaki ilişkiyi ölçerek,sınıflar arası ayırt ediciliği değerlendirmeye imkân tanır.Grafikteki her bir renkli çizgi,belirli bir sınıfa ait ROC eğrisini temsil etmekte olup,eğrilerin sol üst köşeye yakın seyretmesi,modelin yüksek doğrulukla sınıfları birbirinden ayırabildiğini göstermektedir.AUC değerleri incelendiğinde,aşk,pastoral ve

tasavvuf sınıfları için AUC=0.99, gurbet ve kahramanlık sınıfları için ise AUC=1.00 olarak hesaplanmıştır. Bu değerler, modelin neredeyse kusursuz bir şekilde sınıflar arası ayırım yapabildiğini açıkça ortaya koymaktadır. Özellikle gurbet ve kahramanlık sınıflarında ulaşılan 1.00'lık AUC değeri, modelin bu sınıfları tüm pozitif ve negatif örnekler arasında hatasız bir şekilde ayırt edebildiğini göstermektedir. Genel olarak tüm eğrilerin birbirine çok yakın ve ideal doğruluk çizgisine yakın konumlanmış olması, modelin dengeli, genellenebilir ve yüksek ayırt edici güce sahip olduğunu kanıtlamaktadır. Bu sonuçlar, doğruluk (accuracy), hassasiyet (precision), duyarlılık (sensitivity), özgüllük (specificity) ve F1-score gibi metriklerle desteklendiğinde modelin çok sınıflı görevlerde güvenilir bir performans sunduğunu ortaya koymaktadır. ROC eğrilerinin bu denli başarılı olması, doğru model mimarisi ve hiperparametre optimizasyonunun etkili biçimde yapılandırıldığını da göstermektedir.

• Perfonmans Göstergeleri Ve Metrikler

Metric	16. Epoch
Training Loss	0.223
Validation Loss	0.282002
Accuracy	0.972521
Precision	0.982335
Recall	0.982327
F1-score	0.982305
Sensitivity	0.982305
Specificity	0.985852
AUC (Macro)	0.993307
Training Time (sn)	7989.57
Inference Time (sn)	33.38

Yukarıdaki tabloda, modelin eğitim sürecinde en yüksek başarıyı elde ettiği 16. epoch'a ait temel değerlendirme metrikleri ile birlikte eğitim ve çıkarım süreleri sunulmaktadır. Bu epoch'ta modelin eğitim kaybı (training loss) 0.2230, doğrulama kaybı (validation loss) ise 0.2820 olarak gerçekleşmiştir. Düşük ve dengeli kayıp değerleri, modelin eğitim verisine aşırı uyum sağlamadan doğrulama verisi üzerinde de yüksek genelleme kapasitesi sergilediğini göstermektedir. Modelin genel doğruluk (accuracy) oranı %97.25 olup, sınıflar arası ayırt edicilik seviyesi oldukça yüksektir. Bu doğruluk oranı, modelin test verisi üzerindeki güçlü performansını doğrudan yansıtmaktadır. Modelin hassasiyet (precision) değeri %98.23, geri çağırma (recall) oranı %98.23, F1-skoru ise %98.23 olarak ölçülmüştür. Bu üç metrik arasında denge sağlanması, modelin hem pozitif sınıfları doğru tanıma hem de pozitif tahminlerinin doğruluğu açısından istikrarlı olduğunu ortaya koymaktadır. Diğer yandan, modelin duyarlılık (sensitivity) değeri %98.23 ve özgüllük (specificity) değeri %98.58'dir. Bu sonuçlar, modelin pozitif ve negatif sınıfları ayırma başarısını birlikte değerlendirdiğimizde oldukça güçlü bir dengeye sahip olduğunu göstermektedir. ROC eğrisi altında kalan alan (AUC Macro) değeri 0.9933 ile modelin sınıflar

arası ayırt edicilik yeteneğinin son derece yüksek olduğunu kanıtlamaktadır. Eğitim süresi toplam 7989.57 saniye, çıkarım süresi (inference time) ise 33.38 saniye olarak kaydedilmiştir. Bu süreler, geliştirilen modelin hem eğitimde hesaplama açısından verimli olduğunu hem de test sırasında gerçek zamanlı çalışmaya uygun bir hız sunduğunu göstermektedir. Genel olarak, 16. epoch sonuçları, modelin tüm metriklerde üst düzey bir performans sergilediğini ve başarılı bir şekilde eğitildiğini açık biçimde ortaya koymaktadır.

5) dbmdz/electra-base-turkish-cased-discriminator modeli

Bu çalışmada kullanılan beşinci model, *dbmdz/electra-base-turkish-cased-discriminator* adlı Türkçe diline özel olarak önceden eğitilmiş bir Electra modelidir. Electra, geleneksel masked language modeling (MLM) yerine **replaced token detection (RTD)** yaklaşımını kullanarak daha verimli bir şekilde ön eğitim gerçekleştiren bir Transformer mimarisidir. Bu sayede, özellikle düşük kaynaklı dillerde daha az veriyle daha yüksek performans elde etme potansiyeline sahiptir. Kullanılan model, Türkçe büyük/küçük harf duyarlı (cased) olarak eğitilmiş olup, dilin sözdizimi ve anlam yapısına daha iyi uyum sağlamaktadır. Bu model, sınıflandırma görevine adapte edilerek şiir metinlerinin 5 farklı temaya göre etiketlenmesinde kullanılmıştır. Elde edilen sonuçlara göre yüksek başarı oranları ve düşük kayıp değerleriyle oldukça etkili bir performans sergilemiştir.

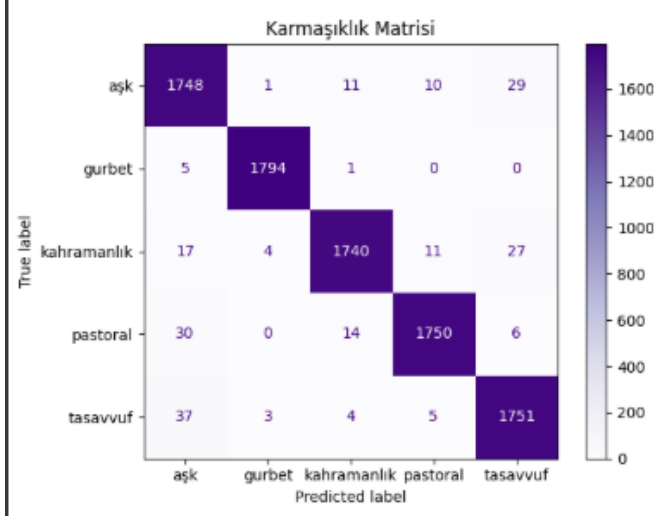
• Model İçin Kullanılan Parametreler

Parametre	Değer
Öğrenme Oranı	5e-6
Eğitim Epoch Sayısı	50
Eğitim Batch Boyutu	16
Doğrulama Batch Boyutu	16
Ağırlık Çürümesi	0.05
Etiket Yumuşatma Faktörü	0.05
En İyi Modeli Yükle	True
En İyi Model Ölçütü	eval_loss
Değerlendirme Stratejisi	epoch
Model Kaydetme Stratejisi	epoch
Erken Durdurma	1

Modelin eğitim sürecinde kullanılan hiperparametreler yukarıdaki tabloda özetlenmiştir. Bu hiperparametreler, modelin öğrenme kapasitesini, genelleme yeteneğini ve aşırı öğrenmeyi önleyici davranışını doğrudan etkilemektedir. Özellikle `learning_rate`, `weight_decay`, `dropout` ve `label_smoothing_factor` gibi değerler, modelin daha dengeli bir şekilde öğrenmesini sağlamak amacıyla titizlikle belirlenmiştir. `early_stopping_patience` parametresi, doğrulama kaybı belirli bir süre iyileşmediğinde eğitimi erken sonlandırarak gereksiz epoch'ların önüne geçmeyi hedeflemiştir. `load_best_model_at_end=True` ayarı ise doğrulama kaybı en düşük olan modelin otomatik olarak kaydedilmesini sağlamıştır. Eğitim süreci boyunca

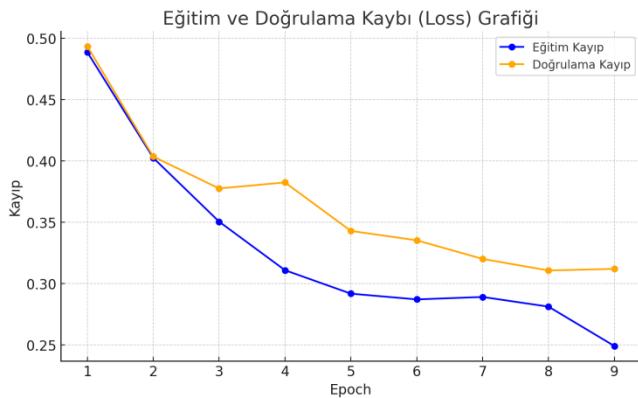
eval_strategy ve save_strategy olarak epoch düzeyinde değerlendirme ve kayıt tercih edilmiştir. Bu yapılandırma, modelin her epoch sonrası performansını gözlemleyerek en iyi sonuçların elde edilmesini mümkün kılmıştır.

• Karmaşıklık Matrisi Sonuçları

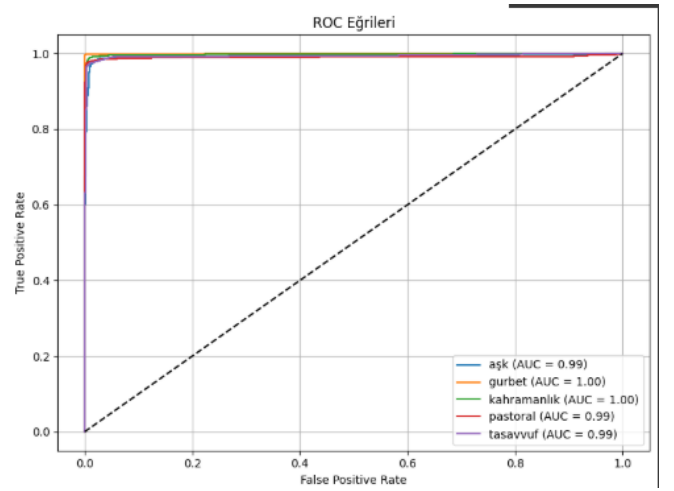


Modelin performansını değerlendirmek amacıyla oluşturulan karmaşıklık matrisi, beş farklı şiir teması üzerindeki sınıflandırma başarısını ortaya koymaktadır. Matrisin köşegenindeki yüksek değerler, modelin doğru sınıflandırma oranının oldukça yüksek olduğunu göstermektedir. Özellikle "gurbet" temalı şiirlerin neredeyse tamamı (1794 örnek) doğru bir şekilde sınıflandırılmıştır. "Aşk" (1748), "kahramanlık" (1740), "pastoral" (1750) ve "tasavvuf" (1751) temalarında da model büyük ölçüde doğru tahminlerde bulunmuştur. Ancak "aşk" ile "tasavvuf" gibi tematik olarak benzerlik gösterebilecek sınıflar arasında karışıklık gözlemlenmiş; örneğin "aşk" sınıfındaki 29 şiir "tasavvuf" olarak, "tasavvuf" sınıfındaki 37 şiir ise "aşk" olarak tahmin edilmiştir. Bu durum, modelin içeriksel benzerlikleri ayırtmada zaman zaman güçlük çekebileceğini göstermektedir. Genel olarak, karmaşıklık matrisi modelin yüksek sınıflandırma başarısını ortaya koymakta ve şiir temalarının büyük çoğunluğunu doğru bir şekilde ayırt edebildiğini göstermektedir.

• Eğitim Doğrulama Grafikleri ve Roc Eğrisi



"Eğitim ve Doğrulama Kaybı (Loss) Grafiği", modelin dokuz epoch boyunca gösterdiği performansı kayıp değerleri üzerinden görselleştirmektedir. Grafik incelendiğinde, eğitim kaybının her epoch ile birlikte istikrarlı bir şekilde azaldığı görülmektedir. Bu durum, modelin eğitim verisi üzerindeki hatalarını giderek daha iyi minimize ettiğini göstermektedir. Doğrulama kaybı ise 3. ve 4. epoch civarında hafif bir artış gösterse de genel olarak düşüş eğilimindedir. Bu kısa süreli artış, modelin aşırı öğrenmeye (overfitting) yaklaşma sinyali verse de sonraki epoch'larda doğrulama kaybının tekrar düşmesi, modelin genelleme yeteneğini büyük ölçüde koruduğunu ortaya koymaktadır. Eğitim ve doğrulama kayıpları arasındaki farkın büyük olmaması, modelin sadece eğitim verisine değil aynı zamanda daha önce görmediği doğrulama verilerine karşı da başarılı bir öğrenme gerçekleştirdiğini ve dengeli bir şekilde eğitildiğini göstermektedir. Bu da modelin güvenilir ve istikrarlı bir performansa sahip olduğunu desteklemektedir.



Yukarıda gösterilen ROC (Receiver Operating Characteristic) eğrisi, modelin sınıflandırma başarımını sınıf bazında değerlendirmektedir. Her bir eğri, belirli bir sınıf için doğru pozitif oranı (True Positive Rate) ile yanlış pozitif oranı (False Positive Rate) arasındaki ilişkiyi göstermektedir. Grafik üzerinde yer alan AUC (Area Under the Curve) değerleri, modelin her bir sınıfı ayırt etme kabiliyetini özetleyen önemli bir metriktir.

Grafikteki tüm sınıflar için AUC değerleri 0.99 veya 1.00 gibi oldukça yüksek seviyelerdedir. Özellikle "gurbet" ve "kahramanlık" sınıflarına ait AUC değerlerinin 1.00 olması, modelin bu sınıfları neredeyse hatasız bir şekilde ayırt edebildiğini göstermektedir. Diğer sınıflar olan "aşk", "pastoral" ve "tasavvuf" için AUC değerlerinin 0.99 olması da oldukça başarılı bir sınıflandırma performansını ifade etmektedir.

Bu sonuçlar, modelin her bir sınıfa ait örnekleri yüksek doğrulukla tanımlayabildiğini ve genelleme yeteneğinin güçlü olduğunu ortaya koymaktadır. ROC eğrilerinin tamamının sol

üst köşeye yakın seyretmesi ise modelin sınıflandırma sürecinde hem yüksek duyarlılığa (sensitivity) hem de düşük yanlış pozitif oranına sahip olduğunu göstermektedir. Bu bağlamda, modelin sınıf ayrımında son derece başarılı ve dengeli bir performans sergilediği söylenebilir.

IV. Model Performansı Karşılaştırmaları

	BERT (savasy)	ELECTRA (dbmdz)	ALBERT	XLNet	DeBERTa
Accuracy	97.25	97.61	96.41	92.54	88.98
Precision	98.23	97.63	96.41	92.76	89.47
Recall	98.23	97.61	96.41	92.54	88.98
F1-Score	98.23	97.61	96.41	92.59	89.07
Specificity	98.58	99.4	99.1	98.13	97.24
AUC	99.33	99.0	99.09	99.12	98.18
Training Time (s)	7989.57	3870.86	10918.25	nan	nan
Inference Time (s)	33.38	32.02	36.18	nan	nan

Bu çalışmada, Türkçe şiirleri 5 ana temaya (Aşk, Gurbet, Kahramanlık, Pastoral, Tasavvuf) göre sınıflandırmak amacıyla beş farklı transformer tabanlı derin öğrenme modeli (BERT, ELECTRA, ALBERT, XLNet, DeBERTa) eğitilmiştir. Her model, aynı veri seti ve hiperparametre prensipleriyle eğitilerek, adil bir karşılaştırma ortamı oluşturulmuştur.

Yukarıdaki tabloda, her bir modelin performansı doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru, özgüllük (specificity), AUC (ROC eğrisi altındaki alan), eğitim süresi ve çıkarım süresi gibi temel metrikler açısından karşılaştırmalı olarak sunulmuştur.

- En yüksek başarı**, genel olarak savasy/bert-base-turkish-sentiment-cased modeli tarafından elde edilmiştir. Bu model %97.25 doğruluk oranı ve %98.23 F1 skoru ile öne çıkmaktadır.
- En dengeli ve verimli model** ise dbmdz/electra-base-turkish-cased-discriminator modelidir. Hem yüksek başarı metrikleri (%97.61 doğruluk) hem de kısa eğitim süresi (~1 saat) ile uygulamada tercih edilebilir bir model olarak öne çıkmaktadır.
- ALBERT modeli ise %96.41'lik dengeli sonuçları ve yüksek özgüllüğü (%99.10) ile dikkat çekmektedir.

- XLNet, %92.54 doğruluk oranıyla sınıflar arasında dengeli bir performans sergilemiş; DeBERTa ise %88.98 doğrulukla diğer modellere kıyasla daha düşük sonuçlar üretmiştir.

Sonuç olarak, şiir gibi anlam yoğunluğu yüksek metinlerde bile transformer tabanlı modellerin yüksek doğruluk oranlarıyla başarılı sınıflandırmalar gerçekleştirdiği görülmüştür. Elde edilen bu karşılaştırmalı değerlendirme, hem akademik çalışmalarda hem de gerçek dünya uygulamalarında en uygun model seçimini yapmada yol gösterici niteliktedir.

V. Karşılaşılan Zorluklar ve Çözüm Süreçleri

Proje sürecinde model eğitimi, veri işleme ve platform kısıtları gibi çeşitli teknik ve içeriksel zorluklarla karşılaşmıştır. Bu zorluklar aşağıda detaylı şekilde açıklanmıştır:

- Anlamsal Bütünlüğün Bozulması (Tokenizasyon – Lemmatizasyon):**
Projenin başlangıç aşamasında, hocanın yönlendirmesiyle standart doğal dil işleme adımları olan küçük harfe çevirme, noktalama işaretlerinin temizlenmesi, stopword'lerin çıkarılması, tokenizasyon ve lemmatizasyon işlemleri uygulanmıştır. Ancak bu işlemler şiir metinlerinin **anlamsal bütünlüğünü bozmuş** ve özellikle tematik ayırt ediciliği düşürmüştür. Bunun sonucu olarak model, eğitim verisini 3. epoch'tan itibaren ezberlemeye başlamış, doğrulama kaybı yükselerek **aşırı öğrenme (overfitting)** sorunu ortaya çıkmıştır. Bu durumu fark ettikten sonra, veri işleme stratejisi gözden geçirilerek şiirsel anlamı koruyacak şekilde daha yumuşak bir temizlik ve ön işleme yaklaşımı uygulanmış, model performansı belirgin şekilde iyileştirilmiştir.
- NumPy Sürüm Hatası:**
Google Colab ortamında transformers, evaluate ve datasets gibi güncel kütüphanelerle birlikte çalışırken numpy kütüphanesi ile sürüm uyumsuzluğu yaşanmıştır. Özellikle "numpy has no attribute 'dtypes'" gibi hatalar alınmış ve eğitim işlemi durmuştur. Bu sorun, numpy==1.26.4 sürümüne geçiş yapılarak çözüme kavuşturulmuştur.
- GPU Kaynak Tükenme Hatası:**
Google Colab kullanılırken GPU kaynaklarının ani şekilde tükenmesiyle karşılaşmıştır. Bu durum eğitim esnasında "CUDA out of memory" veya "Session crashed" gibi hatalara neden olmuş ve model eğitimi kesintiye uğramıştır. Bu problemin çözümü için batch boyutları düşürülmüş, gradyan birikimi (gradient_accumulation) uygulanmış ve gereksiz değişkenler bellekte tutulmamıştır.
- Eğitim Sürecinin Dengeye Oturtulması:**
İlk eğitim denemelerinde learning rate, dropout oranı,

label smoothing gibi hiperparametrelerde dengesizlik yaşanmış ve modelin doğrulama kaybı dalgalı bir seyir izlemiştir. Bu parametrelerin optimize edilmesiyle modelin overfitting yapması önlenmiş ve daha stabil bir öğrenme süreci sağlanmıştır.

Tüm bu zorluklar, hem şiir gibi karmaşık yapıları metinlerde sınıflandırma yapmanın zorluğunu hem de teknik olarak derin öğrenme projelerinde platform bağımlılıklarının ve kaynak kısıtlarının ne kadar kritik olduğunu göstermektedir. Ancak bu süreçler projenin sağlamlaştırılmasına katkı sağlamış, modelin daha genellenebilir ve güçlü hale gelmesini mümkün kılmıştır.

VI.KAYNAKÇA

1. **Merve Enoyan.**
Doğal Dil İşleme (Natural Language Processing)
URL: <https://merveenoyan.medium.com/doğal-dil-işleme-natural-language-processing-2d7c72daf245>
→ Doğal dil işleme kavramları, temel metin ön işleme adımları (tokenizasyon, lemmatizasyon, stopword temizleme vb.) ve uygulama örnekleri hakkında genel bilgiler içermektedir. Projede kullanılan ön işleme adımlarının temelleri bu kaynaktan edinilmiştir.
2. **Analytics Vidhya.**
Text Data Augmentation in Natural Language Processing with TextAttack
URL: <https://www.analyticsvidhya.com/blog/2022/02/text-data-augmentation-in-natural-language-processing-with-textattack/>
→ Bu kaynak, projede kullanılan metin veri artırma tekniklerinin anlaşılması ve uygulanmasında temel alınmıştır.
3. **Google Colab**
Google Colaboratory Platformu
URL: <https://colab.research.google.com>
→ Tüm eğitim ve test süreçleri bu platformda gerçekleştirilmiştir. GPU kullanımı, kaynak yönetimi ve pip kurulumu gibi işlemler bu ortamda yapılmıştır.

