

LGS TÜRKÇE SORU TAHMİN PROJESİ SWOT ANALİZİ

Bu doküman, LGS Soru Tahmin Projesi için proje ekibinin başlangıçta ihtiyaç duyacağı tüm stratejik çerçeveyi sunmak amacıyla hazırlanmış kapsamlı bir SWOT analizini içermektedir.

1. Proje Tanımı ve Amaç

Projenin temel amacı, 2018–2025 yılları arasında yayımlanmış LGS sorularını —başlangıçta Türkçe olmak üzere ilerleyen aşamalarda diğer dersleri de kapsayacak şekilde— toplayıp yapay zekâ ve doğal dil işleme (NLP) teknikleriyle analiz etmektir. Bu analizler doğrultusunda gelecekteki sınavlarda çıkma olasılığı yüksek soru türlerini, konu dağılımlarını, paragraf yapılarını ve bilişsel zorluk seviyelerini tahmin etmek hedeflenmektedir.

Bu sayede öğrencilerin daha odaklı çalışması, öğretmenlerin soru yazma ve deneme oluşturma süreçlerinin desteklenmesi ve eğitim kurumlarının soru bankalarını stratejik biçimde yönetebilmesi amaçlanmaktadır. Ekip olarak yalnızca bir yazılım geliştirmek değil, öğrencilere doğrudan katkı sağlayacak öngörü tabanlı bir eğitim aracı oluşturmak istiyoruz. Uzun vadede ise bu tahmin altyapısının yalnızca LGS Türkçe ile sınırlı kalmayıp diğer derslere de uyarlanabilir bir yapıya kavuşturulması hedeflenmektedir.

2. Projenin Kapsamı ve Varsayımlar

Projenin başlangıç kapsamı aşağıdaki şekilde tanımlanabilir:

- 2018–2025 yılları arasında yayımlanmış resmi LGS sorularının PDF formatında indirilmesi ve işlenmesi.
- Öncelikli odak: Türkçe dersi; sonraki fazlarda Matematik, Fen Bilimleri, İnkılap Tarihi, Din Kültürü ve İngilizce derslerine genişleme.
- OCR (Optik Karakter Tanıma) ve görsel işleme teknikleriyle soruların metin ve gerekiyorsa yapılandırılmış veri hâline dönüştürülmesi.
- Makine Öğrenimi, Derin Öğrenme ve LLM tabanlı modellerle soru sınıflandırma, konu etiketleme, zorluk düzeyi ve çıkma olasılığı tahmini yapılması.
- Elde edilen tahminlerin öğrenci, veli ve öğretmenlerin anlayabileceği görselleştirmeler ve raporlar hâlinde sunulması.

Proje, ařağıdaki temel varsayımlar zerine kuruludur:

- MEB tarafından yayımlanan gemiř LGS sorularının eęitim ve arařtırma amalı analizinde hukuki bir engel bulunmadığı varsayılmaktadır.
- 2018–2025 arası soru formatı, belirli bir oranda tutarlılık gösterdiği iin istatistiksel ve yapay zekâ tabanlı tahmine imkân tanımaktadır.
- Proje ekibi, zaman ierisinde NLP, derin ęrenme ve veri mhendislięi konularında kendini geliřtirmeye istekli ve ęrenmeye aıktır.

3. SWOT Analizi

3.1.Gl Ynler (Strengths)

3.1.1. Veri Odaklı Gl Ynler

- Dzenli ve eriřilebilir LGS veri kaynağı: 2018'den bu yana MEB tarafından her yıl yayımlanan LGS sorularının PDF formatında dzenli biimde arřıvlenmiř olması, veri toplama srecini olduka kolaylařtırır. Bu yapı, soruların toplu indirilmesi ve OCR iřlemlerine hızlıca alınmasını mmkn kılar.
- Geniř ve grece homojen veri seti (2018–2025): Sekiz yıllık tutarlı soru formatı; konu daęılımlarını, yeni nesil soru rntlerini ve yıllar arasındaki deęiřimleri analiz etmek iin yeterli bir zaman aralıęı saęlar. Bu da hem trend analizi hem de geleceęe ynelik tahmin modelleri geliřtirmeyi mmkn kılar.
- Veri setinin geniřletilebilirlięi: Resm LGS sorularının yanı sıra MEB rnek soruları ve gvenilir yayınevlerinin deneme sınavları eklenerek daha kapsamlı bir veri seti oluřturulabilir. Bylece model, grmedięi ama benzer sluba sahip sorular zerinden genelleme yeteneęini artırır.

3.1.2. Teknik ve Yapay Zekâ Gl Ynleri

- NLP/LLM modellerinin yksek bařarısı: Trke metin iřleme alanında geliřen yapay zekâ modelleri; sınıflandırma, soru tr belirleme, paragraf analizi ve soru retme gibi grevlerde bařarılıdır. Bu da projenin teknik uygulanabilirlięini glendirir.
- Yksek tahmin gc: Makine ęrenimi modelleri; soru tarzlarındaki ince deęiřiklikleri, konu eęilimlerini ve biliřsel zorluk seviyelerini insan gznn fark edemeyeceęi kadar derin zelliklerle analiz edebilir. Bylece konu, zorluk ve soru tipi bazında gl tahminler retilir.

- Bilişsel yük analizi yapabilme: Metin uzunluğu, çeldirici sayısı, görsel karmaşıklığı ve dilsel yoğunluk gibi faktörleri dikkate alan bilişsel yük modeli sayesinde yalnızca “konu tahmini” değil, “zorluk ve yük tahmini” de yapılabilir. Bu, projeyi sıradan bir analiz sisteminden ileri düzey bir öğrenme aracı hâline getirir.

3.1.3. Pedagojik ve Ürün Odaklı Güçlü Yönler

- Hem akademik hem pratik kullanım değeri: Oluşturulacak temiz ve etiketlenmiş LGS soru veri seti; akademik araştırmalarda, okul projelerinde, lisansüstü çalışmalarında ve özel eğitim kurumlarında kullanılabilir niteliktedir.
- Hedef kitleden yüksek talep: LGS öğrencileri, veliler ve öğretmenler sınav kaygısını azaltacak ve çalışma sürecini planlı hâle getirecek araçlara büyük ilgi göstermektedir. Proje bu ihtiyaca doğrudan karşılık verir.
- Yenilikçi ve ilk olma potansiyeli: Türkiye’de bu ölçekte, veri odaklı ve yapay zekâ destekli bir LGS tahmin sistemi oldukça sınırlıdır. Bu durum projeye pazar liderliği ve erken benimsenme avantajı sunar.
- Karmaşık çıktıları pedagojik içgörüyü dönüştürme: Yapay zekâ tarafından üretilen teknik verilerin, öğretmen ve öğrencilerin kolayca anlayabileceği “Çalışma Önerileri”, “Öncelikli Konu Haritaları” ve “Soru Tipi Uyarıları”na dönüştürülebilmesi önemli bir kullanıcı deneyimi avantajıdır.

3.1.4. Ekip ve Yönetim Güçlü Yönleri

- 12 kişilik geniş ve uzmanlaşabilir ekip: OCR, veri temizleme, etiketleme, model geliştirme, test, raporlama ve dokümantasyon gibi süreçlerin paralel yürütülmesini sağlar. Bu da projenin verimliliğini ve teslim sürelerini artırır.
- Güçlü iletişim ve sistematik planlama: Net görev dağılımı, düzenli toplantılar ve Trello üzerinden yürütülen sprint yönetimi sayesinde ekip içi koordinasyon yüksektir. Bu yapı, hataları azaltır ve projenin ilerleyişini hızlandırır.
- Motivasyon ve toplumsal katkı bilinci: Ekip üyeleri projeyi sadece teknik bir görev değil, öğrenciler için anlamlı bir fayda üretme imkânı olarak görmektedir. Bu motivasyon, proje kalitesini doğrudan yükseltir.

3.2. Weaknesses (Zayıf Yönler)

3.2.1. Veri Kalitesi ve Etiketleme Zayıflıkları

- Heterojen PDF kalitesi ve OCR hataları: LGS PDF'leri yıllara göre farklı kaliteye sahiptir; düşük çözünürlüklü veya tarama yoluyla oluşturulmuş dosyalarda OCR doğruluğu düşer. Özellikle uzun paragraf içeren sorularda metin bozulmaları veri güvenilirliğini azaltır.
- Görsel tabanlı soruların işlenmesinde zorluk: Grafik, tablo, şekil veya görsel içeren sorularda OCR yeterli değildir. Bu soruların metne dönüştürülmesi için ek görsel işleme teknikleri veya manuel etiketleme gerekir. Bu durum iş yükünü ve zaman maliyetini artırır.
- Tutarsız etiketleme riski: Zorluk seviyesi, bilişsel yük, alt kazanım ve tema gibi etiketler insan yorumuna bağlı olduğundan, ekipler arasında etiketleme tutarsızlıkları oluşabilir. Bu tutarsızlıklar, modelin yanlış öğrenmesine ve tahmin isabet oranının düşmesine neden olabilir.
- Sınırlı veri seti büyüklüğü: LGS yalnızca sekiz yıldır uygulandığı için Türkçe branşında yaklaşık 160 soru bulunmaktadır. Bu da makine öğrenimi açısından sınırlı bir örneklem büyüklüğü anlamına gelir.
- Ek veri kaynaklarının uyumsuzluğu: MEB örnek soruları ve özel yayınevlerinin denemeleri resmî sorularla tam olarak aynı format ve dil yapısına sahip değildir. Bu uyumsuzluk veri setinin homojenliğini bozabilir.

3.2.2. Teknik ve Uzmanlıkla İlgili Zayıflıklar

- Sınırlı NLP ve derin öğrenme uzmanlığı: Gelişmiş dil modelleriyle çalışmak; veri ön işleme, model seçimi, hiperparametre ayarlama ve değerlendirme metriklerinin analizinde deneyim gerektirir. Ekipte bu uzmanlığın sınırlı olması geliştirme sürecini uzatabilir.
- Türkçe NLP'de yüksek lokalizasyon maliyeti: Birçok hazır NLP aracı İngilizce odaklıdır. Türkçe için tokenizer, stop-word listesi, embedding yapıları ve modellerin özelleştirilmesi gerekir. Bu durum teknik maliyeti ve zaman gereksinimini artırır.
- Modelin black-box (kara kutu) yapısı: Derin öğrenme modelleri yüksek doğruluk sağlasa bile karar mekanizmasını açıklamak zordur. "Neden bu konuya ağırlık verildi?" gibi sorulara net yanıt verilememesi, kullanıcı güvenini

etkileyebilir.

- Modelin yıllık soru tarzı değişimlerine duyarlılığı: LGS’de soru tarzının her yıl kısmen değişmesi, modelin yanlış örüntü öğrenme veya eski trendlere göre tahmin yapma riskini artırır.
- MEB format değişikliklerine uyum zorluğu: Sınav yapısında yapılacak ani müfredat veya soru tipi değişikliklerinin modele hızlı uygulanması zor olabilir.

3.2.3. Süreç ve Organizasyon Zayıflıkları

- Temiz veri seti oluşturmaının uzun ve zahmetli olması: Her sorunun OCR çıktısının elle doğrulanması, karakter hatalarının düzeltilmesi ve görsel soruların yeniden yazılması yoğun emek gerektirir. Küçük hatalar bile model performansını ciddi şekilde etkileyebilir.
- Ekip organizasyonunun başlangıçta geç şekillenmesi: Ekip geç kurulduğu için görev dağılımı ve planlama sürecinde başlangıçta zaman kaybı yaşanmıştır.
- Eğitimcilerin güvenini kazanma zorluğu: Eğitim camiası “tahmine dayalı” yapay zekâ araçlarına temkinli yaklaşabilir. Modelin metodolojisinin şeffaf şekilde sunulması ve pilot çalışmalar yapılması gerekebilir.
- Zaman kısıtı: Proje akademik takvime paralel ilerlediğinden teslim tarihleri baskı yaratır ve süreç optimizasyonu kritik hâle gelir.

3.3.Fırsatlar (Opportunities)

3.3.1. Eğitim Ekosistemi ve Toplumsal Fırsatlar

- Türkiye’de bu alanda az sayıda yapay zekâ projesi bulunması: LGS özelinde kapsamlı veri analizi, yeni nesil soru çözümleme ve tahmin yapan araçların yok denecek kadar az olması, projeyi yenilikçi bir referans proje hâline getirebilir.
- Eğitimde dijitalleşme trendi: COVID-19 sonrası hızlanan çevrim içi eğitim, dijital soru bankaları ve yapay zekâ destekli öğrenme araçları, bu tür analiz araçlarına olan ilgiyi artırmaktadır.
- Toplumsal fayda ve geniş ilgi potansiyeli: Doğru tahminler üreten bir sistem; öğrenciler, veliler ve öğretmenler için yüksek değer sunarak geniş kitlelere

ulařabilir.

- Velilerin kaygı odaklı yüksek talebi: LGS'nin rekabetçi yapısı, velileri "daha kesin ve güvenilir yönlendirme sunan" araçlara yöneltmektedir. Proje, doğru konumlandırıldığında sınav kaygısını azaltan bir rehber olarak hızla benimsenebilir.

3.3.2. Kurumsal ve Finansal Fırsatlar

- Bakanlık ve devlet destekleri: MEB, TÜBİTAK, Teknofest ve bölgesel kalkınma ajansları; yapay zekâ ve eğitim teknolojileri projelerine fon, tanıtım ve altyapı desteęi sağlayabilmektedir.
- Kurumsal işbirlikleri ve API modeli: Deneme yayıncıları, çevrim içi özel ders platformları ve kurs merkezleri; analizlerini güçlendirmek için bu sistemin API'sini kullanmak isteyebilir.
- B2C ve B2B iş modeli açılımı:
Proje zamanla řu alanlara dönüşebilir:
 - öğretmenler için otomatik soru üretim ve analiz aracı
 - öğrenciler için kişiselleştirilmiş çalışma seti
 - okullar ve yayınevleri için soru bankası analiz sistemi

Böylece hem geniş kitleye (B2C) hem de kurumlara (B2B) hitap eden sürdürülebilir bir gelir modeli oluşabilir.

- Ücretsiz temel sürüm ile hızlı yayılım: Uygulamanın ücretsiz sunulan kısmı geniş kullanıcı tabanı oluşturur, kurumsal sürümler için farkındalığı artırır.

3.3.3. Teknolojik ve Ölçeklenebilirlik Fırsatları

- LLM tabanlı Türkçe yapay zekâ modellerinin hızlı gelişimi: Her yıl ortaya çıkan daha başarılı ve ekonomik modeller sayesinde sistemin doğruluk oranı, hız ve kalite sürekli artabilir.
- Geniş tarihsel veri sayesinde trend analizi: Yıllara göre konu ağırlıkları, soru tipleri, yeni nesil soru yapıları ve değişen ölçme yaklaşımları istatistiksel olarak tespit edilebilir.
- Dikey büyüme potansiyeli: LGS'de başarı elde edildiğinde proje kolaylıkla:

- YKS
- ALES
- KPSS

gibi çok daha büyük pazarları da kapsayacak şekilde genişletilebilir.

- Farklı derslere uyarlanabilirlik: Türkçe modeli başarılı olduğunda Matematik, Fen, İngilizce gibi diğer derslere tahmin ve soru üretim sistemi olarak kolayca adapte edilebilir.

3.3.4. Stratejik Fırsatlar

- Ebeveynlerin Kaygı Odaklı Yüksek Talep Piyasası:

LGS yüksek rekabet ve stres yaratan bir sınavdır. Veliler, risk azaltma ve doğru çalışma yönlendirmesi sunan sistemlere güçlü bir talep göstermektedir.

Stratejik Önemi: Proje, “kaygıyı azaltan ve odaklanmayı optimize eden analiz aracı” olarak konumlandırılırsa, geleneksel eğitim uygulamalarına göre premium fiyatlandırma avantajı oluşur.

- Yeni Nesil Soruların Rakiplerde Yarattığı Boşluk:

LGS Türkçe sorularında ezberden uzak, derin anlamsal analiz gerektiren yeni nesil yapılar rakiplerin klasik yöntemlerle çözemediği bir alan oluşturmuştur.

Stratejik Önemi: Yapay zekâ modeli bu boşluğu doldurarak “Mavi Okyanus” alanı yaratır; rekabet düşük, talep yüksek olur. Projenin hızlıca pazar lideri olma ihtimali artar.

- Büyük Veri Ortaklıkları ve API Entegrasyonu:

Deneme yayıncıları, çevrim içi eğitim platformları ve özel ders uygulamaları; soru analizlerini geliştirmek için bu sistemin tahmin API’sine ihtiyaç duyabilir.

Stratejik Önemi: Proje yalnızca bireysel öğrencilere değil, kurumlara da lisanslanabilir hâle gelir. Bu da ölçeklenebilir, sürdürülebilir ve yüksek getirili bir iş modeline dönüşmesini sağlar.

3.4.Tehditler (Threats)

3.4.1. Düzenleyici ve Hukuki Tehditler

- MEB'in soru tarzını radikal şekilde değiştirmesi:
Soru formatı, konu dağılımı veya ölçme yaklaşımı ani değişikliklere uğrayabilir. Bu tür bir değişiklik, modelin geçmiş veriden öğrendiği örüntüleri tamamen geçersiz hâle getirerek tahmin gücünü bir gecede sıfırlayabilir.
Bu durum en kritik dış risklerden biridir.
- Sınav güvenliği ve etik kısıtlamalar:
MEB veya eğitim kurumları, "soruları tahmin eden sistem" algısı nedeniyle projeyi sınav güvenliği açısından riskli görerek resmî olarak sınırlandırabilir.
- Fikri mülkiyet ve telif riskleri:
LGS soruları telif kapsamındadır. Soruların AR-GE, eğitim veya ticari üründe hangi kapsamda kullanılacağı netleştirilmezse hukuki sorunlar ortaya çıkabilir.
- KVKK ve veri güvenliği riskleri:
Öğrenci performans verilerinin saklanması veri ihlali, siber saldırı veya anonimleştirme eksikliği durumunda ciddi itibar kaybı ve hukuki yaptırımlar doğurabilir.
Bu risk, projenin itibarı için en ölümcül tehditlerden biridir.

3.4.2. Teknik Tehditler

- OCR doğruluk sorunları:
Türkçe'de uzun paragraf, tablo, grafik veya şiirsel metinlerde OCR hataları artabilir.
Bu hatalar → veri setinin bozulması → modelin tahmin doğruluğunun düşmesi → güven kaybı şeklinde zincirleme etki oluşturur.
- Model eğitim maliyetleri:
Dil modeli eğitimi veya fine-tuning aşamaları GPU, bulut ve API maliyetleri gerektirir. Öğrenci projesi ölçeğinde bu maliyetler sürdürülemez olabilir.
- Türkçe'nin dilsel karmaşıklığı:
Anlam, bağlam, mecaz, soru kökü-karşılık ilişkileri gibi yapılar bazı modellerde tam olarak yakalanamayabilir. Bu da doğruluk sınırlaması yaratır.
- Veri güvenilirliği riski:
Yanlış OCR çıktıları, eksik veri veya hatalı etiketleme modelin öğrenmesini

bozarak tahmin performansını geriletir.

3.4.3. Pazar ve Rekabet Tehditleri

- Büyük EdTech firmalarının ücretsiz araç çıkarması:
Büyük kurumlar pazarlama amaçlı ücretsiz tahmin veya analiz araçları sunabilir. Kullanıcılar kısa vadede ücretsiz ürünlere yönelerek projenin gelir modelini tehdit edebilir.
- Rakiplerin hızlı şekilde ortaya çıkması:
Proje ilgi gördüğünde, benzer yapay zekâ tahmin sistemleri hızla geliştirilebilir. Bu durum rekabetçi baskıyı artırır.
- Eğitim kurumlarının yapay zekâ tabanlı sınav tahmin araçlarına temkinli yaklaşması:
Bazı kurumlar etik, güvenlik veya pedagojik gerekçelerle böyle bir sistemi kullanmak istemeyebilir. Bu da kurumsal işbirliklerini yavaşlatabilir.
- Kamu algısı riski:
“Soruları önceden tahmin eden sistem” algısı bazı çevrelerde etik tartışmalara, hatta kamu tepkisine yol açabilir.

3.4.4. Stratejik ve Operasyonel Tehditler

- Modelin öğrendiği örüntülerin bir gecede geçersizleşmesi:
MEB’in örnek soru politikasını değiştirerek eski kalıplarla ilişkisi olmayan yeni bir format uygulaması, modelin yeniden eğitilmesini ve güvenilirliğin sıfırdan inşa edilmesini gerektirebilir.
- Veri toplama sürecinin yavaşlaması:
Yeni soru formatı veya ek güvenlik önlemleri veriye erişimi zorlaştırabilir.
Bu durum tahmin modelinin güncellenme hızını düşürerek pazardaki rekabet avantajını kaybettirebilir.
- İtibar kaybı riski:
Veri ihlali, yanlış tahminler veya etik tartışmalar, projenin güvenilirliğini kalıcı olarak zedeleyebilir. Eğitim alanında itibar kaybı kritik bir tehdit olarak öne çıkar.