

# AÇIK UÇLU SINAV SORULARININ OTOMATİK OLARAK DEĞERLENDİRİLMESİ

Prof. Dr. Yusuf Sinan AKGÜL, Süleyman BALABAN  
Gebze Teknik Üniversitesi, Bilgisayar Mühendisliği  
akgul@gtu.edu.tr  
suleymanbalaban3@gmail.com

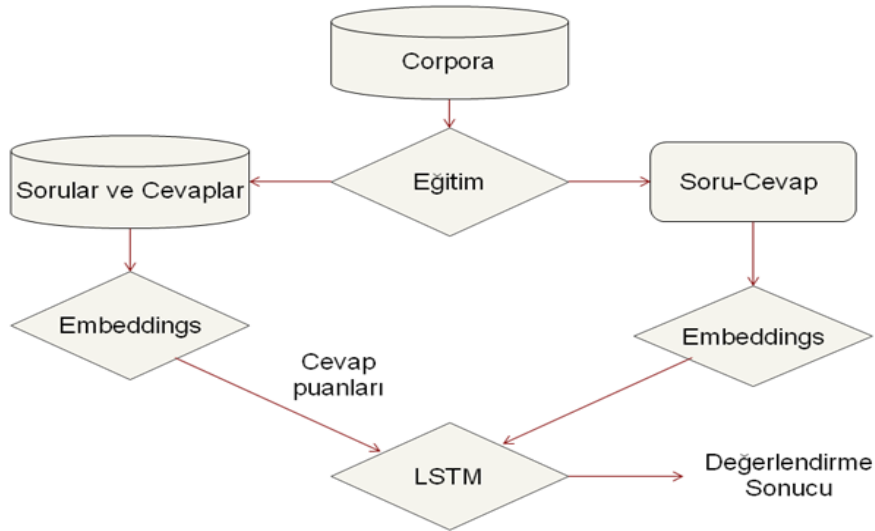
**Anahtar Kelimeler:** Doğal Dil İşleme, Derin Öğrenme, Otomatik Sınav Puanlandırma

## 1. GİRİŞ (AMAÇ)

Günümüzde öğrenci sayısı gittikçe artmakta ve öğrencilerin eğitim başarılarını en iyi değerlendirme yöntemi açık uçlu sorulardır. Merkezi sınavlara milyonlarca öğrenci katılmaktadır. Bu sınavların eğiticiler tarafından değerlendirilmesi mümkün değildir. Ayrıca insana dayalı değerlendirmede güvenilirlik hiç bir zaman tartışma konusu olmaktan çıkmayacaktır. Bu proje sayesinde zamandan tasarruf edilecektir, çoktan seçmeli sorular yerine daha seçici olan açık uçlu sınav sorularının kullanımı merkezi sınavlarda da yaygınlaşacaktır, en önemlisi ise tarafsız bir sistem tarafından değerlendirilmesi güvenilirliği de beraberinde getirecektir. Bu proje Mia Teknoloji A.Ş.<sup>1</sup>'nin katkılarıyla gerçekleştirilmiştir.

## 2. YÖNTEM

### 2.1. Kelime Vektörleri ile Derin Öğrenme



<sup>1</sup> Mia Teknoloji A.Ş., kamu ve özel sektör için bilişim yazımları gerçekleştiren bir firma.

Yöntem genel olarak yukarıdaki tasarımda gösterilmiştir. Corpora<sup>2</sup>, büyük Türkçe metinlerinden ve Türkçe ders kitaplarından oluşmaktadır. Corpora'nın Word2Vec derin öğrenme algoritması ile eğitilmesi ile kelime vektörleri elde edilir. Daha sonra önceden puanlandırılmış soru-cevaplar ile kullanıcının sisteme girdiği soru-cevapların sözcük vektörleri oluşturulur ve ardından geçen LSTM<sup>3</sup> derin öğrenme algoritmasına tabi tutulur ve son olarak puanlama işlemi yapılır.

### 3. BULGULAR ve YORUMLAR

Sistem başarısı puanlandırma işlemlerinin sonucunda tahmin edilen puan ile eğiticilerin verdikleri puan arasındaki farkın eğiticilerin verdikleri puana göre oranlarının ortalaması ile bulunmaktadır. Eğiticinin veri kümesindeki cevapları değerlendirirken adil davranıp davranmadığı, bir cevaba ait önceden değerlendirilmiş cevapların çokluğuna, her puan dilimine ait cevapların sayılarının arasında çok fark olmamasına ve son olarak hangi yöntemi kullanarak puanlandırma yapıldığı gibi kriterlere bağlıdır Aşağıda Sistemlerin ortalama korelasyonu gösterilmiştir.

Soru	I	II
Sistem Başarısı	%82	%95

Biyoloji dersi ile alakalı bir soru ile test edilen sistemin örnek çıktılarının belirli bir kısmı gösterilmiştir. Sonuçlar şu şekildedir;

Midesi alman erişkin bir insanın sindirimle ilgili yaşayacağı sorunları yazınız? (5 puan)		
Cevap	Eğitmen	Sistem
Proteinlerin mekanik sindirimi yapılamaz.	2	2
Proteinlerin kimyasal sindirimi aksar. Fazladan besin depolayamaz.	3	3
Besinler yeterince fiziksel olarak parçalanamaz. Besinler depo edilemez. Sindirim süresi uzar.	3	3
Protein sindiriminde zorlanır.	1	1

<sup>2</sup> Büyük metin kümesidir.

<sup>3</sup> Derin öğrenme algoritmasıdır. Açılımı Long-Short Term Memory'dir.

#### 4. SONUÇLAR

Algoritmanın sonuçları incelendiğinde eğitimcilerle sistemin ürettiği puanlar arasında belirli bir uyumun olduğu söylenebilir. Yöntem başarılı sonuçlar vermektedir. Eğitimcilerin sistem çalışmasının öncesinde değerlendirdiği cevapları daha tutarlı ve başarılı bir şekilde yaptığı taktirde sistem başarısı da artacaktır. Cevap veri seti ne kadar fazla ve her puan dilimine ait cevapların sayılarının arasındaki farklar ne kadar az ise sistem daha da başarılı sonuçlar verecektir. Yabancı ülkeler AES<sup>4</sup> üzerine bir çok çalışmalar yapılmaktadır. ABD'de GMAT adı verilen merkezi sınavda AES kullanılmaktadır. Ülkemizde de merkezi sınavlar olmak üzere bir çok sınavda kullanılabilir.

---

<sup>4</sup> Automated Essay Scoring.

## KAYNAKÇA

Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv: 1301.3781* (2013).

Markoff, John. "Essay-grading software offers professors a break." *New York Times* 4 (2013).

LSTM, <https://medium.com/@ishakdolek/lstm-d2c281b92aac>. Accessed 29 May 2018.