



**T.C.
GEBZE TEKNİK ÜNİVERSİTESİ**

Bilgisayar Mühendisliği Bölümü

BİL 496

**AÇIK UÇLU SINAV
SORULARININ OTOMATİK
OLARAK
DEĞERLENDİRİLMESİ**

**Süleyman Balaban
Danışman
Prof. Dr. Yusuf Sinan Akgül**

**Mayıs, 2018
Gebze,
KOCAELİ**

Bu çalışma .../.../... tarihinde aşağıdaki jüri tarafından Bilgisayar Mühendisliği Bölümünde Lisans Bitirme Projesi olarak kabul edilmiştir.

Bitirme Projesi Jürisi

Danışman Adı	Prof. Dr. Yusuf Sinan Akgül	
Üniversite	Gebze Teknik Üniversitesi	
Fakülte	Bilgisayar Mühendisliği	

Jüri Adı	Yrd. Doç. Dr. Burcu Yılmaz	
Üniversite	Gebze Teknik Üniversitesi	
Fakülte	Bilgisayar Mühendisliği	

ÖNSÖZ

Proje raporumun son halini almasında yol gösterici olan Sayın Prof. Dr. Yusuf Sinan Akgül hocama, bu çalışmayı destekleyen Gebze Teknik Üniversitesi'ne ve proje için gerekli verileri sağlayan Mia Teknoloji A.Ş.'ye içten teşekkürlerimi sunarım.

Ayrıca eğitimim süresince bana her konuda tam destek veren aileme ve bana hayatlarıyla örnek olan tüm hocalarıma saygı ve sevgilerimi sunarım.

Mayıs,2018

Süleyman Balaban

İÇİNDEKİLER

İÇİNDEKİLER.....	IV
ŞEKİL LİSTESİ	V
KISALTMA LİSTESİ.....	VI
ÖZET.	VII
1.GİRİŞ	2
1.1 PROJE TANIMI	2
1.2 PROJENİN NEDEN ve AMAÇLARI.....	3
2. GEREKSİNİMLER VE YÖNTEMLER.....	3
2.1 PROJE GEREKSİNİMLERİ	3
2.2 KULLANILAN YÖNTEMLER	5
3.BULGULAR.....	8
3.1 Yöntemlerin Test Edilmesi.....	9
4.TARTIŞMA VE SONUÇLAR.....	11
5.KAYNAKLAR	12

ŞEKİL LİSTESİ

ŞEKİL 2.2 İlk Yöntem Genel Sistem Tasarımı	5
ŞEKİL 2.3 Kosinüs Benzerliği	6
ŞEKİL 2.4 LSTM Kullanımı	7
ŞEKİL 2.4 İkinci Yöntem Genel Sistem Tasarımı	7

KISALTMA LİSTESİ

GTÜ	: Gebze Teknik Üniversitesi
AES	: Otomatik Metin Puanlandırma(Automated Essay Scoring)
LSTM	: Long Short Term Memory
RNN	: Recurrent Neural Network
KNN	: K Nearest Neighborhood

TABLO LİSTESİ

TABLO 3.1 Sistem Başarısı	9
TABLO 3.2 Birinci Yöntem ve Birinci Soru ile Örnek Çıktı.....	9
TABLO 3.3 Birinci Yöntem ve İkinci Soru ile Örnek Çıktı	10
TABLO 3.4 İkinci Yöntem ve Birinci Soru ile Örnek Çıktı	10
TABLO 3.5 İkinci Yöntem ve İkinci Soru ile Örnek Çıktı	10

ÖZET

Bu raporda G.T.Ü. Bilgisayar Mühendisliği Bölümü son sınıfında hazırladığım Lisans Bitirme Projesinin içeriğini açıklayan bilgiler ve tez teslimi sırasında gerekli belgelerin listesi verilmektedir.

BIL-496 dersi kapsamında gerçekleşecek olan proje açık uçlu sınav sorularının yazılım tarafından otomatik olarak değerlendirilmesini kapsamaktadır. Açık uçlu sınav soruları öğrenme çıktılarını değerlendirmek için kullanılan en önemli araçtır. Bu projede bu durum göz önünde bulundurulur. Güvenilirlik ve zaman tasarrufu en önemli kazancımızdır.

Bu projede, açık uçlu sınav sorularının değerlendirilebilmesi için Doğal Dil İşleme yöntemleri kullanılmaktadır. Kelime vektörleri derin öğrenme yöntemi ile elde edilir ve Derin Öğrenme yöntemlerinden LSTM yöntemi kullanılarak eğitmenler tarafından alınan değerlendirme sonuçlarına göre eğitilmektedir. Yapılan deneyler sonucunda otomatik olarak değerlendirilen sınav sonuçlarının eğitmenlerin verdikleri sonuçlarla uyumlu oldukları gözlemlenmiştir.

SUMMARY

This report provides information on the content of the undergraduate completion project. I have prepared in the last year of Computer Engineering Department of Gebze Technical University and a list of the documents required during thesis sub-mission.

The project, which will take place within the scope of BIL-495, involves the automatic evaluation of open-ended exam questions by the software. Open-ended exam questions are the most important tool used to evaluate learning outcomes. This situation is taken into account in this project. Reliability and time saving are our most important gains.

In this project, Natural Language Processing methods are used to evaluate open-ended exam questions. Word vectors are obtained by deep learning method and they are trained according to the evaluation results obtained by trainers using LSTM method from Deep Learning methods. It was observed that the test results, which were evaluated automatically at the end of the experiments, corresponded to the results of the instructors.

1. GİRİŞ

Bu doküman G.T.Ü. Bilgisayar Mühendisliği Bölümü Lisans Bitirme Projesi dersi için yapılacak olan projenin amacını, yapılma nedenlerini, içeriğini ve tasarım süreçlerini açıklamak için düzenlenmiştir.

Eğitimde öğrenme çıktılarını değerlendirmek çok önemlidir. Bu değerlendirmeler ülkemizde eğitimciler tarafından değerlendirilmektedir. Bu değerlendirmeler çok fazla zaman almaktadır. Hele ki binlerce öğrencinin katıldığı merkezi sınavlarda eğitimciler tarafından değerlendirilmesi mümkün olmayacaktır. Ayrıca değerlendirme sonuçlarına güvenilirlik önemli bir sorun teşkil etmektedir. Bu proje bu iki ana sorunu ortadan kaldırmaktadır.

Proje kapsamında iki farklı soruya ait toplam 340 adet cevap bulunmaktadır. Bu cevapların 270'i bir soruya ait diğer 170'i ise diğer soruya aittir. Cevaplar bir eğitici tarafından puanlandırılmıştır. Puanlandırma yazılım tarafından yapıp eğitimcilerin verdikleri puanlar ile karşılaştırılması yapılmıştır. Projeye ilgili yapılan ön çalışmada şu makale ve yayınlardan faydalanılmıştır: "An overview of automated scoring of essays." [1], "Efficient estimation of word representations in vector space." [3], " Essay assessment with latent semantic analysis." [12]. Sorunun doğru cevabı eğitimciler tarafından önceden belirlenerek doğru cevabın ve öğrencilerin cevaplarının özellik vektörü çıkarılan ve ardından SVM benzerlik yöntemi ile puanlandırma yapan bir çalışma olan "Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments." [2] makalesi bu projeye en benzer yapıda yapılan bir fikri barındırıyor.

Bu projeye benzer çalışmalar İngilizce üzerine çokça yapılmaktadır. AES sistemlerinin tarihi eskilere dayanmaktadır. İlk olarak 1966'da, Page bir Otomatik Yazısal İletim Puanlandırma (Automated Essay Scoring, AES) sistemi geliştirdi ve otomatik bir değerlendirme yazılımının insanın değerlendirmesinden ayırt edilemez olduğunu gösterdi (Page, 1966). Son yıllarda daha fazla sistemler geliştirilmiştir ki bunların en önemlileri; Intelligent Essay Assessor (Landauer, Laham, & Foltz, 2003), Intellimetric (Elliot, 2001), Project Essay Grade'in yeni

bir versiyonu (PEG, Page, 1994), e-rater (Burstein, Kukich, Wolff, Lu, & Chodrow, 1998) ve e-rater v2 (Attali & Burstein, 2006) [4], c-rater (Sukkarieh, Jana Zuheir, ve John Blackmore, 2009). Bu sistemlerin yanında AES üzerine yeni çalışmalarda yapılmıştır. Daha önce kullanılan yöntemlerin yerine AES bir sıralama problemi olarak görülmüş ve başarılı sonuçlara ulaşılmıştır (Chen, Hongbo, et al., 2012).

1.1 PROJE TANIMI

“Açık Uçlu Sınav Sorularının Otomatik Olarak Değerlendirilmesi” projesinde veri kümesinde bulunan soru ve cevapları bulunmaktadır. Sisteme bu veri kümesinde bulunan her hangi bir soruya ait cevaplar verilmektedir ve bu cevaplar otomatik olarak puanlandırılmaktadır.

1.2 PROJENİN NEDEN ve AMAÇLARI

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.

2. GEREKSİNİMLER VE YÖNTEMLER

Wikipedia'da geçen tüm Türkçe metinler ve Türkçe ders kitaplarında geçen metinler birleştirilip bir Türkçe metin kümesi oluşturulmuştur. Bu metin kümesinden Word2Vec derin öğrenme yazılımı kullanarak sözcük vektörü oluşturulmuştur. Aynı soruya ait eğiticiler tarafından puanlandırılmış cevapları elde ettikten sonra bu

cevap cümlelerinin vektör olarak ifade edilebilmesi için derin öğrenmeden elde ettiğimiz sözcük vektörlerini kullanılmıştır. Cevapların içersindeki sözcüklerin vektörel değerleri ve cevabın eğitici tarafından verilen puanları LSTM derin öğrenme algoritması ile eğitilerek sınıflandırma yapılmaktadır. Burada LSTM algoritmasına iki farklı şekilde bilgi verilmektedir. Bunlar Kelime Vektörleri ile LSTM ve Kelime & Tf-Idf Vektörleri LSTM'dir. Tf-Idf vektörleri sistemin daha iyi çalışmasını sağlamak için geliştirilmiştir. Her iki yöntemin daha iyi çalışması ve LSTM'in eksiklerini gidermek için KNN sınıflandırma algoritması kullanılarak geliştirilen bir yöntem ile girilen cevabın LSTM tarafından puanlandırılması iyi değilse bu yöntem devreye girer ve puanlamayı istenen başarıya ulaştırır. Eğer girilen cevabın cümle vektörü eğiticiler tarafından önceden değerlendirilen cevapların cümle vektörlerine Kosinüs Benzerliği en yüksek olan 3 cevabın yani K-3 sınıflandırma yöntemi ile elde edilen puanın en benzer üç cevap içerisinde ortalama Kosinüs Benzerliği değeri 0.8'den büyük ise bu algoritmanın ürettiği puan geçerli kabul edilmektedir.

2.1 PROJE GEREKSİNİMLERİ

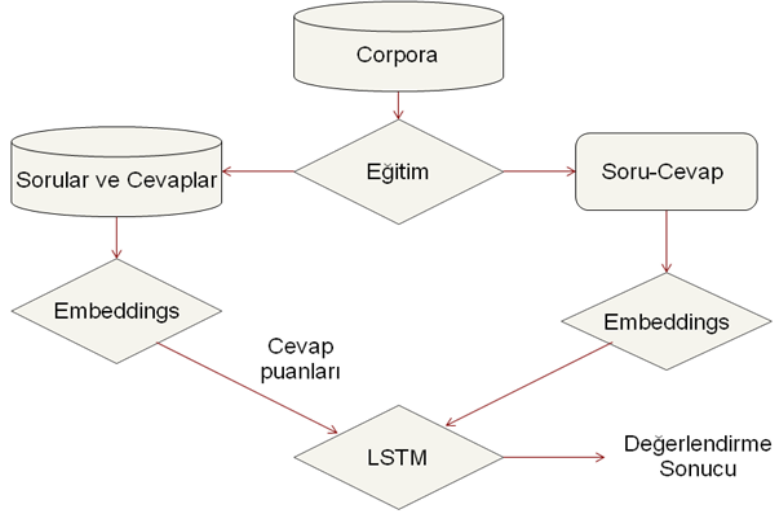
Bunların sağlanması için gerekli ihtiyaçlar:

- Ubuntu 14.04 işletim sistemli bir bilgisayar.
- Python Programlama dili kullanımı.
- Keras, Gensim, Numpy, Scikit-Learn, Scipy kütüphanelerinin kurulumu.
- Wikipedia'daki tüm Türkçe metinler ve Türkçe ders kitaplarındaki metinlerden oluşan metin kümesi.
- Word2Vec derin öğrenme algoritmaları.
- Derin öğrenme algoritmasından kelime vektörlerinin elde edilmesi ve metin dosyası halinde tutulması.
- Aynı soruya verilmiş cevapların daha önceden eğiticiler tarafından puanlandırılmış halde elde edilip metin dosyası halinde tutulması.

Kullanıcı sisteme girdi olarak hangi soruların cevapları eğitilmiş ise o soruya göre cevaplar girip puanını öğrenebilir veya kullanıcı girdi olarak bir şey girmeden sistemin başarısını ölçebilir. Sistem başarısını etkileyen birden çok faktör bulunmaktadır.

2.2 KULLANILAN YÖNTEMLER

A. KELİME VEKTÖRLERİ İLE LSTM



Şekil 2.2 İlk Yöntem Genel Sistem Tasarımı

Yöntem genel olarak yukarıdaki tasarımda gösterilmiştir. Corpora, Wikipedia'daki tüm Türkçe metinler 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 LSTM derin öğrenme algoritmasına tabi tutulur ve son olarak puanlama işlemi yapılır. Bu yöntem sistem başarısı hesabı yaparken veri kümesindeki cevapların %7'sinin dışarıda bırakılarak geriye kalanlarının eğitilmesi ile dışarıda bırakılan bir cümle tahmini ve hata payı hesaplanır. Bu işlem eğitimcilerin puanlandığı cevapların yer aldığı veri kümemizdeki her cevap için ayrı ayrı uygulanır. Bu işleme çapraz doğrulama denir.

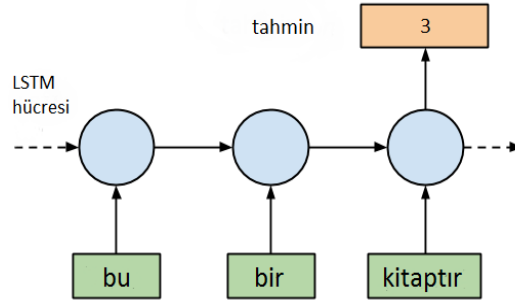
Kelime Vektörleri

Her kelimenin bir vektörel değeri bulunmaktadır. Bu değerler her bir sözcük için sözlüğümüzde ne kadar sözcük var ise o kadar sayı içermektedir. Bu değerler kelimelerin birbirlerine benzerliğini içermektedir [6][7]. Fakat bütün vektörü kullanmak yerine ilk dense vektör denen en önemli 200 vektör değerini sisteme yüklenmiştir. Cümledeki bütün kelimelerin vektörlerinin karşılıklı olarak toplanıp ardından her koordinattaki vektör değerinin kelime sayısına bölünmesiyle cümle

vektörü elde edilir.

LSTM

LSTM bir derin öğrenme algoritmasıdır yani bir yapay sinir ağıdır. RNN özelliği taşır yani tekrarlayan bir yapısı ve zaman serisi vardır. Bu sayede hafızaya sahiptir bu algoritma. Cevap olarak girilen cümlelerin özelliklerini sayısal olarak bu algoritmaya verilmektedir[16]. Ayrıca daha önceden eğiticiler tarafından değerlendirilen cevapların puanları da bu algoritmaya sınıflandırmak için verilmektedir. Bu eğitim algoritması sayesinde girilen bir cevabı sınıflandırabiliyor yani puanlandırabiliyoruz. Daha önceki çalışmamızda lineer regresyon yöntemi ile bu projeyi yaptığımızda cümle vektörlerini eğitim için kullandığımız için sözcüklerin cümle içeri-sindeki yerlerinin önemi yoktu. Bu büyük bir sorun teşkil ediyordu. Fakat LSTM'de bu sorunu aşmış olduk.



Şekil 2.3 LSTM Kullanımı

Kosinüs Benzerliği

Bu metot Doğal Dil İşleme ve Makine Öğrenmesi yöntemlerinde sıkça kullanılan bir benzerlik hesabı bulan bir metottur. Cümlelerin vektörleri kullanılmaktadır. Böylece iki cümlelerin benzerliği bulunabilmektedir.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

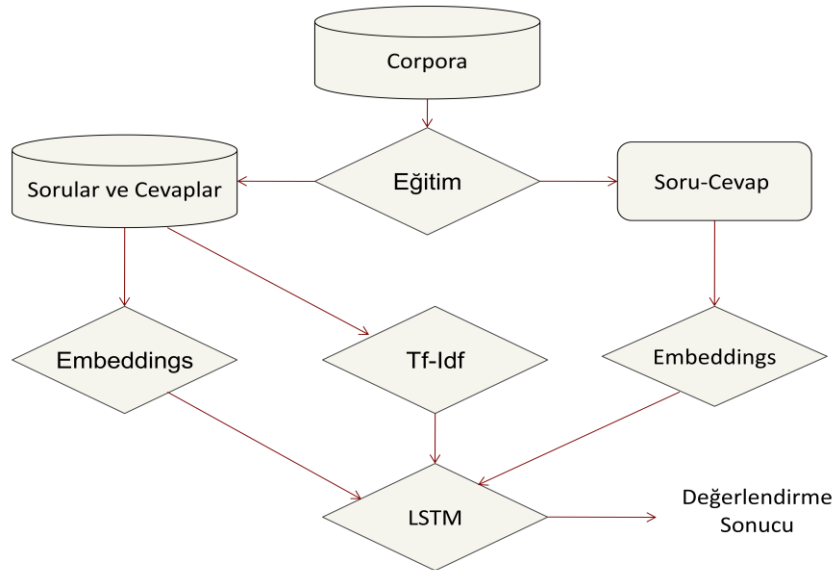
Şekil 2.4 Kosinüs Benzerliği

KNN

Sınıflandırmada kullanılan bir Makine Öğrenmesi algoritmasıdır. Sınıflandırma sırasında çıkarılan özelliklerden, sınıflandırılmak istenen yeni cümle için daha önceki cümlelerden kendisine en yakın k tanesine yakınlığına bakılmasıdır[17]. Yakınlık ise cümlelerin kosinüs benzerliğinden elde ettiğimiz değerdir.

Bu k tane cümle içerisinde hangi puan dilimine ait cevap daha çok var ise o puan diliminde ait cevapların kosinüs benzerliği değerlerinin ortalaması alınır. Eğer ki bu değer 0.75'den büyük ise ve en benzer k adet cümle içerisinde LSTM'in ürettiği sonuçla aynı puana sahip cümle yoksa LSTM'in sonucuna bakılmaksızın bu puan dilimi sistem puanı olarak kabul edilir. Böylelikle LSTM'in eğitiminde kullanılan veri setinin azlığından dolayı nadiren oluşan hatalarının önüne geçilmektedir.

B. KELİME VE TF-IDF VEKTÖRLERİ İLE LSTM



Şekil 2.5 İkinci Yöntem Genel Sistem Tasarımı

Bu yöntemde ilk yöntemden farklı olarak LSTM'e sözcüklerin vektörlerine ek olarak sözcüklerin tf-idf vektörleri de eklenir. Bu sayede sistemin daha iyi çalışması amaçlanmıştır. Bu vektörler bir sözcüğün bir dokümanda geçme sıklığını temsil eder. Veri setimizdeki tüm sözcükler için bu vektörler oluşturulmuştur.

Tf-Idf Vektörleri

Bu vektörleri çıkarmamız için öncelikle veri setimizdeki aynı puana sahip cevapların bir araya getirilerek tek bir doküman gibi olması sağlanmalıdır. Bütün puanlar için bu işlem yapılır[18]. Ardından veri setimizdeki tüm sözcükler için aşağıdaki şu formül ile kaç doküman var ise o kadar değer üretilir. Örneğin; 5 puanlık bir soruya ait cevapların bulunduğu bir veri setinde 6 farklı puana sahip cevap yani doküman bulunmaktadır. Bu işlemleri formül ile gösterir isek;

$$tfidf_{s,d} = tf_{s,d} \times idf_s$$

s : sözcük

d : doküman

tf : sözcüğün dokümanda kaç kez geçtiği / dokümanda en çok geçen sözcüğün adeti

idf : $\log(\text{toplam doküman sayısı} / \text{kelimenin geçtiği doküman sayısı})$

LSTM'e word2vec kelime vektörlerinin yanında bu tf-idf vektörleri de verilir. İlk yöntemde her sözcük için 200 adet sayıda oluşan bir özellik vektörümüz vardı bu yöntemde ise her sözcük için 200 + sorulara verilebilecek birbirinden farklı puan sayısı kadar büyüklükte özellik vektörü eğitime verilir.

3. BULGULAR

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 belirtilen özelliklere göre çalıştırılan sistemlerin ortalama korelasyonu gösterilmiştir.

Soru-I : Midesi alınan erişkin bir insanın sindirimle ilgili yaşayacağı sorunları yazınız?

Soru-II: Atom ve molekül arasındaki fark nedir?

Tablo 3.1 Sistem Başarısı

Yöntem	Soru-I	Soru-II
I	% 82	% 95
II	% 77	% 90

Eğitimcilerin değerlendirmesi sistemin çalışmasını etkilemektedir. Fakat kullanılan yöntemler sistem başarısına etkisi çok yüksektir. Ayrıca eğitimcilerin değerlendirdiği cevapların sayısının az olması sistem başarısına doğrudan olumsuz etkilemektedir. Eğitilen cevapların puan dilimlerine ayrıldığında birbirine yakın sayıda bulunması olumlu etkilemektedir. LSTM'in her puan dilimini öğrenebilmesini sağlar. Önceden lineer regresyon kullanarak yaptığımız bu projenin birinci soru için elde ettiğimiz ortalama korelasyon değeri %80.58 olarak ölçülmüştü.

3.1 Yöntemlerin Test Edilmesi

İki farklı yöntem mevcuttur ve bu yöntemler kullanılarak aynı sorulara verilmiş yanıtların puanlandırılması yapılmıştır. Eğitimcilerin verdiği puanlar ve sistemin verdiği puanlar tablolarda gösterilmiştir. Soru olarak yukarıda göstermiş olduğumuz iki soru kullanılmıştır.

3.1.1 Kelime Vektörleri İle LSTM Yöntemi Testi

Aşağıda Kelime Vektörleri ile LSTM yöntemi kullanılarak ve birinci soru olarak belirttiğimiz biyoloji dersi ile alakalı birinci soru ile test edilen sistemin örnek çıktılarından belirli bir kısmı aşağıdaki tabloda gösterilmiştir. Sonuçlar şu şekildedir;

Tablo 3.2 Birinci Yöntem ve Birinci Soru ile Örnek Çıktı

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

Yukarıdaki tabloda birinci sorunun birinci yöntem ile olan test sonuçları tablo halinde gösterilmiştir. Bu kısımda ise birinci yöntem kullanılarak kimya dersine ait olan ikinci soru ile sistem test edilmiştir. Sonuçlar şu şekildedir;

Tablo 3.3 Birinci Yöntem ve İkinci Soru ile Örnek Çıktı

Atom ve molekül arasındaki fark nedir?			
Cevap	Eğitmen	LSTM	Sistem
Atomlar moleküllerden oluşur.	0	0	1
Atom en küçük yapı molekül ise birleşimidir.	2	2	2
Molekül bir çok farklı maddenin birleşiminden oluşur. Atom ise homojendir.	1	1	1
Atom en küçük yapı birimdir. Moleküller atomlardan oluşur.	3	3	3

3.1.2 Kelime ve Tf-Idf Vektörleri İle LSTM Yöntemi Testi

Aşağıda Kelime ve Tf-Idf Vektörleri ile LSTM yöntemi kullanılarak iki ayrı soruya ait eğiticiler tarafından önceden puanlandırılmış cevapların belirli bir kısmı test edilmiştir. Bu yöntemi ikinci yöntem olarak belirtilmiştir. İlk olarak biyoloji dersi ile alakalı olan birinci soru test edilmiştir. İlk yöntem ikinci yönteme nazaran daha güzel sonuçlar vermektedir. Sonuçlar şu şekildedir;

Tablo 3.4 İkinci Yöntem ve Birinci Soru ile Örnek Çıktı

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

Aşağıda kimya dersine ait soruların ikinci yöntem kullanılarak elde edilen sistemin test sonuçları gösterilmektedir. Sonuçlar şu şekildedir.

Tablo 3.5 İkinci Yöntem ve İkinci Soru ile Örnek Çıktı

Atom ve molekül arasındaki fark nedir?			
Cevap	Eğitmen	LSTM	Sistem
Atomlar moleküllerden oluşur.	0	0	1
Atom en küçük yapı molekül ise birleşimidir.	2	2	3
Molekül bir çok farklı maddenin birleşiminden oluşur. Atom ise homojendir.	1	1	3
Atom en küçük yapı birimdir. Moleküller atomlardan oluşur.	3	3	3

4. TARTIŞMA VE SONUÇLAR

Algoritmanın sonuçları incelendiğinde eğitimcilerle sistemin ürettiği puanlar arasında belirli bir uyumun olduğu söylenebilir. İlk kullanılan yöntemin üzerine geliştirilen Kelime ve Tf-Idf Vektörleri ile LSTM yöntemi ilk yöntemle göre daha kötü çalışmaktadır. Bunun sebebi tf-idf vektörlerini elde ederken yeni bir cevap girdiğimizde cevap veri setinde olmayan bir sözcük var ise o sözcüğü bir kez geçmiş kabul ediyoruz buna Smoothing denir. Bu sebepten ötürü güzel sonuçlar vermemektedir. Fakat sistemin başarısını ölçerken çapraz doğrulama ile test ettiğimiz için veri setinde olmayan sözcük olması imkansız olduğu için bu yöntemin başarısı yüksektir.

Kelime Vektörleri ile LSTM yöntemi gayet başarılı sonuçlar vermektedir. Yine de sistemin tam olarak başarılı olduğu söylenemez. Bunun sebepleri şöyledir:

Eğiticilerin sistem çalışmasının öncesinde değerlendirdiği cevapları daha tutarlı ve başarılı bir şekilde yaptığı takdirde 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.

Bu sistemde sözcüklerin cümle içerisindeki yerlerinin önemi vardır. LSTM algoritmasına cümle içerisindeki sözcüklerine sayısına bakmaksızın tüm sözcüklerin vektörlerini verebildiğimiz için bunu sağlayabiliyoruz. Daha önceden yaptığımız çalışmada Çoklu Doğrusal Regresyon kullanmıştık. O çalışmada cümle vektörleri üzerinden çalıştığımız için cümle yapısını kaybediyorduk. Böylece bu sorunu halletmiş olduk.

Yabancı ülkeler AES üzerine bir çok çalışmalar yapılmaktadır. Fakat ülkemizde bu konu üzerine pek çalışma yapılmamaktadır. ABD'de GMAT adı verilen merkezi sınavda AES kullanılmaktadır. Yapılan çalışmanın bir sonraki aşamasında dilbilgisi kuralları da sistem eğitime dahil edilebilir. Veri kümesi dijital ortamdan elde edilmesi yerine kağıda yazılmış cevapların görüntü işleme yöntemleri ile tanımlanıp dijital ortama aktarımı sağlanabilir.

5. KAYNAKLAR

- [1] Dikli, Semire. "An overview of automated scoring of essays." *The Journal of Technology, Learning and Assessment* 5.1 (2006).
- [2] Michael A.G. Mohler, Razvan Bunescu, Rada Mihalcea. "Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments." *ACM Digital Library* (2011).
- [3] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv: 1301.3781* (2013).
- [4] Attali, Yigal, and Jill Burstein. "Automated essay scoring with e-rater® V. 2." *The Journal of Technology, Learning and Assessment* 4.3 (2006).
- [5] www.econstats.com/RegressionOut_3_1.htm. Accessed 15 December 2017.
- [6] An Intuitive Understanding of Word Embeddings, [www.analyticsvidhya.com /blog/2017/06/word-embeddings-count-word2vec/](http://www.analyticsvidhya.com/blog/2017/06/word-embeddings-count-word2vec/). Accessed 19 December 2017.
- [7] Text Classification & Word Representations using FastText, www.analyticsvidhya.com/blog/2017/07/word-representations-text-classification-using-fasttext-nlp-facebook/. Accessed 19 December 2017.
- [8] Dan Jurafsky, Christopher Manning. web.stanford.edu/~jurafsky/NLPCourseraSlides.html. Accessed 19 December 2017.
- [9] B. Zadrozny and C. Elkan. 2002. Transforming classifier scores into accurate multiclass probability estimates. Edmonton, Alberta.
- [10] Markoff, John. "Essay-grading software offers professors a break." *New York Times* 4 (2013).
- [11] Balfour, Stephen P. "Assessing writing in MOOCs: Automated essay scoring and calibrated peer review." *Research & Practice in Assessment* 8.1 (2013): 40-48.
- [12] Miller, Tristan. "Essay assessment with latent semantic analysis." *Journal of Educational Computing Research* 29.4 (2003): 495-512.
- [13] Breland, Hunter M. "Word frequency and word difficulty: A comparison of counts in four corpora." *Psychological Science* (1996): 96-99.

- [14] M. Mohler and R. Mihalcea. 2009. Text-to-text semantic similarity for automatic short answer grading. In *Proceedings of the European Association for Computational Linguistics (EACL 2009)*, Athens, Greece.
- [15] C. Leacock and M. Chodorow. 2003. C-rater: Automated Scoring of Short-Answer Questions. *Computers and the Humanities*, 37(4):389–405
- [16] LSTM, <https://medium.com/@ishakdolek/lstm-d2c281b92aac>. Accessed 29 May 2018.
- [17] KNN, <http://bilgisayarkavramlari.sadievrenseker.com/2008/11/17/knn-k-nearest-neighborhood-en-yakin-k-komsu/>. Accessed 29 May 2018.
- [18] Tf-idf Weighting, <https://nlp.stanford.edu/IR-book/html/htmledition/tf-idf-weighting-1.html>. Accessed 29 May 2018.

