



**T.C.
GEBZE TEKNİK ÜNİVERSİTESİ**

Bilgisayar Mühendisliği Bölümü

BİL 495

**AÇIK UÇLU SINAV
SORULARININ OTOMATİK
OLARAK
DEĞERLENDİRİLMESİ**

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

**Aralık, 2017
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	Doç. Dr. Erchan Aptoula	
Üniversite	Gebze Teknik Üniversitesi	
Fakülte	Bilgisayar Mühendisliği	

ÖNSÖZ

Proje raporumun son halini almasında yol gösterici olan Sayın Prof. Dr. Yusuf Sinan Akgül hocama ve bu çalışmayı destekleyen Gebze Teknik Üniversitesi'ne içten teşekkürlerimi sunarım.

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

Aralık,2017

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	4
ŞEKİL 2.3 İkinci Yöntem Genel Sistem Tasarımı	6
ŞEKİL 2.4 Kosinüs Benzerliği	7

KISALTMA LİSTESİ

GTÜ	: Gebze Teknik Üniversitesi
AES	: Otomatik Metin Puanlandırma(Automated Essay Scoring)
ÇDR	: Çoklu Doğrusal Regresyon

TABLO LİSTESİ

TABLO 4.1 Sistem Başarısı	9
TABLO 4.2 Birinci Yöntem Word2Vec Kelime Vektörleri ile Örnek Çıktı	9
TABLO 4.3 Birinci Yöntem FastText Kelime Vektörleri ile Örnek Çıktı	10
TABLO 4.4 İkinci Yöntem Word2Vec Kelime Vektörleri ile Örnek Çıktı	10
TABLO 4.5 İkinci Yöntem FastText Kelime Vektörleri 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-495 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 Makine Öğrenmesi yöntemlerinden Çoklu Doğrusal Regresyon yöntemi kullanılarak iki adet eğitmen 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 two trainers using Multiple Linear Regression method from Machine 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 aynı soruya verilmiş 92 adet cevap bulunmaktadır. Bu cevaplar iki ayrı eğitimci 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 Puanlandırıcı (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 puan-landırılmaktadır.

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 olmak-tan 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 ve FastText derin öğrenme yazılımları kullanarak iki ayrı sözcük vektörü oluşturulmuştur. Aynı soruya ait eğiticiler tarafından puanlandırılmış

cevapların elde edildikten sonra bu cevap cümlelerinin vektör olarak ifade edilebilmesi için derin öğrenmeden elde ettiğimiz sözcük vektörlerini kullanarak cümle vektörü haline getirilmiştir. Bu cümle vektörleri aynı şekilde sistem tarafından puanlandırılması istenen yeni cevabın için de oluşturulacaktır. Ardından çoklu doğrusal regresyon yöntemi kullanılarak puanlandırma yapılmaktadır. Burada iki ayrı yöntem kullanılmıştır bunlar; Cümle Vektörlerinin Çoklu Doğrusal Regresyonu yöntemi ve ilk yöntemden elde edilen puanlara göre en iyi cevapların elde edilmesinden sonra girilen cevabın her puanın en başarılı cevabına olan Cosine benzerliği değerleri ile tekrar çoklu doğrusal regresyon işlemine tabi tutulması ile elde edilen ikinci yöntemdir. Ve bu yöntemin ismi En Güzel Cevaplara Benzerliğin Çoklu Doğrusal Regresyonu'dur.

2.1 PROJE GEREKSİNİMLERİ

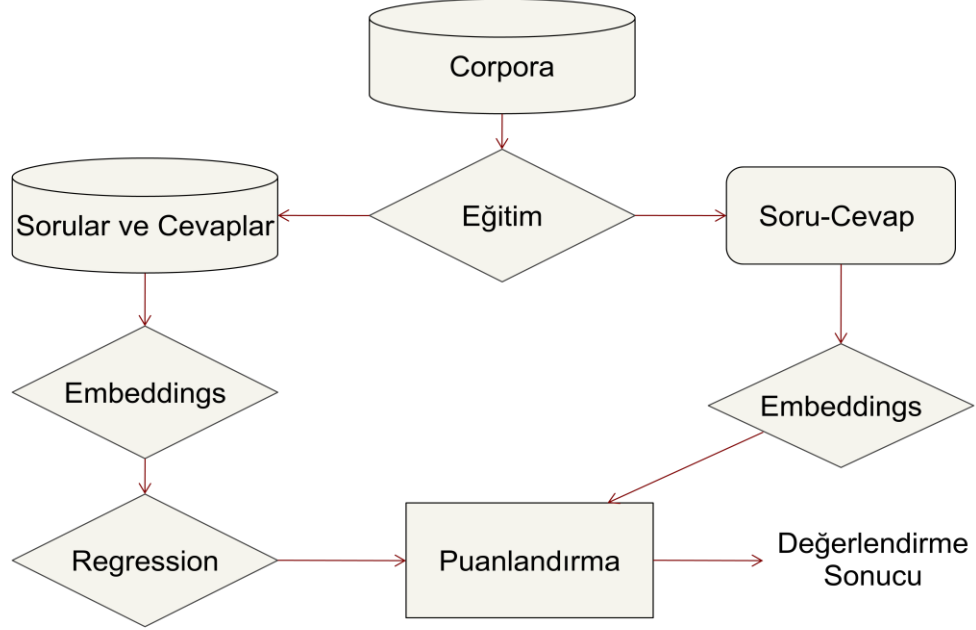
Bunların sağlanması için gerekli ihtiyaçlar:

- Windows 8.1 işletim sistemli bir bilgisayar.
- Eclipse IDE
- Java Programlama dili kullanımı.
- Wikipedia'daki tüm Türkçe metinler ve Türkçe ders kitaplarındaki metinlerden oluşan metin kümesi.
- Word2Vec ve FastText derin öğrenme algoritmaları.
- Derin öğrenme algoritmalarından 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. CÜMLE VEKTÖRLERİNİN ÇOKLU DOĞRUSAL REGRESYONU



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 ve FastText derin öğrenme algoritmaları ile eğitilmesi ile kelime vektörleri elde edilir. Daha sonra önceden puanlandırılmış soru-cevaplar ile kullanıcının sisteme girdiği soru-cevapların cümle vektörleri oluşturulur ve ardından ÇDR işlemine tabi tutulur ve son olarak puanlama işlemi yapılır. Bu yöntem sistem başarısı hesabı yaparken veri kümesindeki cevaplardan birinin dışarıda bırakılarak geriye kalanlarının eğitilmesi ile dışarıda bırakılan bir cümlenin tahmini ve hata payı hesaplanır. Bu işlem eğiticilerin 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.

Cümle Vektörleri İle Çoklu Doğrusal Regresyon

Çoklu Doğrusal Regresyon yöntemi önceden verilmiş bağımsız değişken sayısı ikiden fazla ve bağımlı değişkeni belli olan denklemlerin eğitilmesi ise bağımsız değişkenlerin önüne katsayıların elde edilmesi ile gerçekleştirilir [5]. Formülize eder isek;

$$Y = (\beta_1 * X_1) + (\beta_2 * X_2) + (\beta_3 * X_3) + \dots + (\beta_{85} * X_{85}) + \varepsilon$$

Y : Cevabın eğitici tarafından verilmiş puanıdır.

X : Cümle vektörünün değerleridir.

β : İşlem sonunda elde edeceğimiz katsayı değerleridir.

ε : Eğitici puanı ile sistemin tahmin ettiği puan arasındaki farktır.

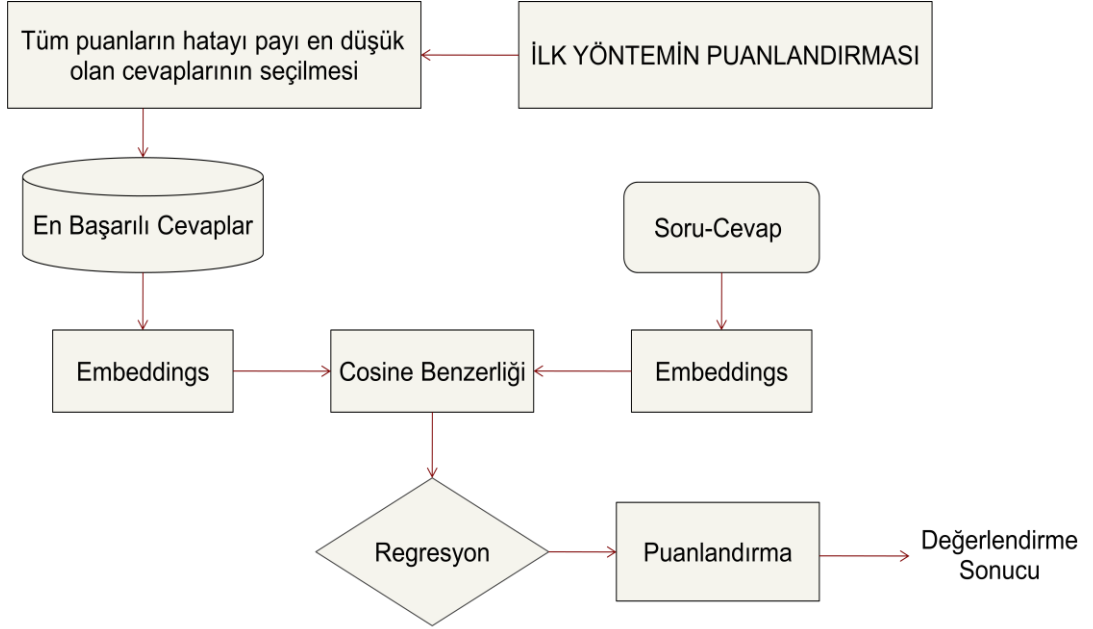
Her cevap için bu denklemi oluşturulmuştur ve ÇDR işlemine sokulmuştur. β ve X değerlerinin 200'e kadar değil de 85'e kadar gitmesinin sebebi; ÇDR işleminde bağımsız değişken sayısından az sayıda denklem var ise ÇDR işlemi doğru sonuçlar üretmez. Bizim eğiticiler tarafından puanlandırılmış soru-cevap veri kümemizde 92 adet soru-cevap yani denklem bulunmaktadır. ÇDR işleminin en basit gösterimiyle formülü şu şekildedir;

$$\beta = \text{multiply}(\text{multiply}(\text{inverse}(\text{multiply}(\text{transpose}(x), x)), \text{transpose}(x)), y);$$

Puanlandırma

Çoklu Doğrusal Regresyon yöntemiyle elde ettiğimiz β yani bağımsız değişken katsayıları kullanılarak iki farklı işlem yapılabilmektedir. Öncelikle bu katsayılar ile eğiticiler tarafından verilen bir cevabın puanı tekrardan hesaplanabilir ve böylelikle sistem başarısını test edebiliriz. Bu konuya daha sonra

B. BAŞARILI CEVAPLARA BENZERLİĞİN ÇOKLU DOĞRUSAL REGRESYONU



Şekil 2.3 İkinci Yöntem Genel Sistem Tasarımı

Bu yöntemde ilk yöntemden ile elde ettiğimiz ÇDR katsayılarını kullanarak eğiticiler tarafından puanlandırılan veri kümemizdeki cevaplar sistem tarafından puanlandırılmıştır. Ardından bu puanlandırmaya göre her cevap için elde edilen ε değeri yani hata payları her puan dilimi için ayrı ayrı karşılaştırılmıştır. Hata payı en düşük olan her puan için cevaplar belirlenmiştir. Ardından kullanıcı tarafından girilen veya sistem başarısı ölçümü sistemden ayrılan cevabın bu en başarılı cevaplara ayrı ayrı kosinüs benzerliği hesaplanmıştır.

Kosinüs Benzerliği

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

A ve B iki ayrı cümleyi temsil etmektedir. Bu formül iki cümlenin vektör değerleri üzerinden gerçekleştirilmiştir. kosinüs benzerliği kullanılarak her puana ait en güzel cevaplara karşı benzerlik hesaplanır. Ardından bu benzerlik değerleri tekrardan Çoklu Doğrusal Regresyon yöntemiyle eğitilir.

Çoklu Doğrusal Regresyon ve Puanlandırma

İlk yöntemde ÇDR işlemine cümle vektörlerini bağımsız değişkenler olarak ve eğiticilerin puanlarını da bağımlı değişkenler olarak veriyorduk. Bu yöntemde ise bağımsız değişkenlerin yerini puanlandırılmak istenen cevabın en başarılı cevaplara göre kosinüs benzerliği değerleri girilmektedir.

Örnek verecek olursak; bir sorunun maksimum puanı 5 ise 0 ile 5 arasındaki altı adet puan dilimindeki en güzel yani hata payı en düşük olan cevaplara göre kosinüs benzerliği hesaplanır. Ardından bu değerlere ek olarak cümlenin kaç sözcükten oluştuğu bilgisi ile birlikte ÇDR işlemine tabi tutulur. ÇDR işlemine verilen bağımlı değişkenler ilk yöntemde olduğu gibi eğiticilerin verdikleri puanlardır.

Bu yöntem baştan sona çapraz doğrulama ile yapılmaktadır. İlk yöntem de her cevap eğitim kümesinden ayrılır ardından ayrılmış olan eğitim kümesine ikinci yöntem uygulanır. Böylelikle sistem başarısı bu yöntemde daha başarılı olmuştur.

3. BULGULAR

Sistem başarısı ÇDR işlemlerinin sonucunda tahmin edilen puan ile eğiticilerin verdikleri puan arasındaki farkın eğiticilerin verdikleri puana göre oranlarının ortalaması ile bulunmaktadır. Sistem kelime vektörlerinin hangi derin öğrenme

algoritması ile elde edildiği, hangi eğiticinin veri kümesindeki cevapları değerlendirdiği ve son olarak hangi yöntemi kullanarak puanlandırma yapıldığı gibi kriterlere bağlıdır. Aşağıda belirtilen özelliklere göre çalıştırılan sistemlerin ortalama korelasyonu gösterilmiştir.

Tablo 3.1 Sistem Başarısı

	Word2Vec		FastText	
Yöntem	Eğitmen 1	Eğitmen 2	Eğitmen 1	Eğitmen 2
I	% 31	% 42	% 13	% 20
II	% 81	% 86	% 81	% 83

Eğitmcilerin değerlendirmesi sistemin çalışmasını etkilemektedir. Kelime vektörlerinin elde edilmesi için kullanılan Word2Vec ve FastText derin öğrenme yazılımlarının etkisi sistem başarısına çok fazla etki etmemektedir. Fakat kullanılan yöntemler sistem başarısına etkisi çok yüksektir. Eğiticilerin değerlendirmesinin ortalama korelasyonu %80.58'dir. Bu değerin yüksek olması eğiticilerin değerlendirmedeki tutarlılığını gösterir.

3.1 Yöntemlerin Test Edilmesi

İki farklı yöntem mevcuttur ve bu yöntemler kullanılarak aynı sorulara verilmiş yanıtların puanlandırılması yapılmıştır. Eğiticilerin verdiği puanlar ve sistemin verdiği puanlar tablolarda gösterilmiştir. Soru olarak şu soruya verilen cevaplar değerlendirilmiştir:

"Midesi alınan erişkin bir insanın sindirimle ilgili yaşayacağı sorunları yazınız?"

3.1.1 Cümle Vektörlerinin ÇDR' si Yöntemi Testi

Aşağıda Cümle Vektörlerinin ÇDR' si yöntemi kullanılarak ve Word2Vec'den elde ettiğimiz kelime vektörleri kullanılarak test edilen sistemin örnek çıktılarından belirli bir kısmı aşağıdaki tabloda gösterilmiştir.

Tablo 3.2 Birinci Yöntem Word2Vec Kelime Vektörleri ile Örnek Çıktı

Cevap	Eğitmen 1	Sistem 1	Eğitmen 2	Sistem 2
Proteinlerin mekanik sindirimi yapılamaz.	2	-0.83	2	4.54
Proteinlerin kimyasal sindirimi aksar. Fazladan besin depolayamaz.	5	2.83	3	4.84
Besinler yeterince fiziksel olarak parçalanamaz. Besinler depo edilemez. Sindirim süresi uzar.	5	4.92	3	8.55
Protein sindiriminde zorlanır.	2	2.19	1	-0.78

Yukarıdaki tabloda Word2Vec algoritması ile elde ettiğimiz kelime vektörleri kullanılmıştır. Aşağıdaki tabloda ise Facebook'un geliştirdiği FastText algoritmasından elde edilen kelime vektörleri kullanılmıştır.

Tablo 3.3 Birinci Yöntem FastText Kelime Vektörleri ile Örnek Çıktı

Cevap	Eğitmen 1	Sistem 1	Eğitmen 2	Sistem 2
Proteinlerin mekanik sindirimi yapılamaz.	2	3.88	2	1.70
Proteinlerin kimyasal sindirimi aksar. Fazladan besin depolayamaz.	5	-5.33	3	-6.45
Besinler yeterince fiziksel olarak parçalanamaz. Besinler depo edilemez. Sindirim süresi uzar.	5	8.56	3	10.58
Protein sindiriminde zorlanır.	2	-0.58	1	-1.21

3.1.2 Başarılı Cevaplara Benzerliğin ÇDR' si Yöntemi Testi

Aşağıda kullanmamız gereken en iyi yöntem ve kelime vektörleri ile hazırlanmış örnek çıktılar tabloda gösterilmiştir. Yöntem olarak birinci yöntemin üzerine inşa edilen ikinci yöntem ve Word2Vec'den elde ettiğimiz kelime vektörleri kullanılmıştır. Sistem test edilmiş ve örnek çıktılarında belirli bir kısmı aşağıdaki tabloda gösterilmiştir.

Tablo 3.4 İkinci Yöntem Word2Vec Kelime Vektörleri ile Örnek Çıktı

Cevap	Eğitmen 1	Sistem 1	Eğitmen 2	Sistem 2
Proteinlerin mekanik sindirimi yapılamaz.	2	3.49	2	2.22
Proteinlerin kimyasal sindirimi aksar. Fazladan besin depolayamaz.	5	3.09	3	2.30
Besinler yeterince fiziksel olarak parçalanamaz. Besinler depo edilemez. Sindirim süresi uzar.	5	3.71	3	3.49
Protein sindiriminde zorlanır.	2	1.77	1	1.15

Aşağıda FastText algoritmasından elde ettiğimiz kelime vektörleri ile ikinci yöntem kullanarak sistem eğitimi yapıldığında oluşan örnek çıktılarından bir kısmı bulunmaktadır.

Tablo 3.5 İkinci Yöntem FastText Kelime Vektörleri ile Örnek Çıktı

Cevap	Eğitmen 1	Sistem 1	Eğitmen 2	Sistem 2
Proteinlerin mekanik sindirimi yapılamaz.	2	3.36	2	2.13
Proteinlerin kimyasal sindirimi aksar. Fazladan besin depolayamaz.	5	3.18	3	2.56
Besinler yeterince fiziksel olarak parçalanamaz. Besinler depo edilemez. Sindirim süresi uzar.	5	3.20	3	3.26
Protein sindiriminde zorlanır.	2	2.01	1	2.26

4. TARTIŞMA VE SONUÇLAR

Algoritmanın sonuçları incelendiğinde eğitmenlerle sistemin ürettiği puanlar arasında belirli bir uyumun olduğu söylenebilir. İlk kullanılan yöntemin üzerine geliştirilen Başarılı Cevaplara Benzerliğin ÇDR' si 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. Buna kanıt olarak yöntem test aşamasından da anlaşılacağı üzere ikinci eğiticinin değerlendirdiği veri kümesini kullanarak sistemi eğittiğimizde sistem başarı yüzdesinin daha yüksek çıkmasıdır.

Regresyon bir istatistiksel tekniktir ve her zaman veri miktarı ile doğru orantılı olarak çalışır. ÇDR algoritması daha büyük bir veri ile çalıştırılırsa katsayıların doğruluğu

artacaktır. Ayrıca apraz doęrulama tekniklerinin doęruluk oranı da veri miktarına baęlı olarak deęişmektedir. Soru-cevap veri kümemizdeki deęerlendirilmiş cevap sayısı arttırılırsa DR algoritmasına cümle vektörlerinin 85 adet deęerini deęil de daha fazla vektör deęeri ile alışılabilecektir. ünkü; daha önce de belirtildięi gibi baęımsız deęişken katsayısından az sayıda denklem yani cümlemiz mevcut olursa DR işleminin kötü sonuçlar verecektir. Cümle vektörlerinin 200 tane deęeri kullanılır ise daha başarılı sonuçlar elde edilecektir.

Bu sistemde sözcüklerin arasında dil bilgisi açısından analiz yapılmamaktadır. Sözcüklerin birbirleriyle olan ilişkilerini incelenmeli, cümlelerin dil bilgisi yapısı dikkate alınmalıdır.

Yabancı ülkeler AES üzerine bir ok alışmalar yapılmaktadır. Fakat ülkemizde bu konu üzerine pek alışma yapılmamaktadır. ABD'de GMAT adı verilen merkezi sınavda AES kullanılmaktadır. Yapılan alışmanın bir sonraki aşamasında dil bilgisi kuralları da sistem eęitimine 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

