



YAZILIM GEREKSİNİM BELİRTİM DOKÜMANI (SRS)

Derin Öğrenme İle Twitter Platformunda

Yalan Haber Tespiti

Cansu AYTEN – 171180010

Gamze AKSU – 171180005

BM495 BİLGİSAYAR PROJESİ I

ARALIK 2021

İÇİNDEKİLER

Sayfa

İÇİNDEKİLER	i
KISALTMALAR	iii
1. KAPSAM.....	1
1.1 Tanım	1
1.2. Sisteme Genel Bakış	1
1.3. Dokümana Genel Bakış	2
2. İLGİLİ DOKÜMANLAR.....	2
3. GEREKSİNİMLER.....	3
3.1. Gerekli Durum ve Modlar	3
3.2. YKE (Yazılım Konfigürasyon Elemanı) Fonksiyonel Gereksinimleri.....	3
3.2.1. Haberin Doğruluğunun Tespiti.....	3
3.3. YKE Dış Ara Yüz Gereksinimleri.....	3
3.4. YKE Dahili Ara Yüz Gereksinimleri.....	3
3.5. YKE Dahili Veri Gereksinimleri.....	4
3.6. Uyarılama Gereksinimleri	4
3.7.Emniyet Gereksinimleri	4
3.8. Güvenlik ve Gizlilik Gereksinimleri	4
3.9. YKE Ortam Gereksinimleri	4
3.10. Bilgisayar Kaynak Gereksinimleri	4
3.2.1. Bilgisayar Donanım Gereksinimleri	4
3.2.2. Bilgisayar Donanımı Kaynak Kullanımı Gereksinimleri.....	5
3.2.3. Bilgisayar Yazılım Gereksinimleri	5
3.2.4. Bilgisayar İletişim Gereksinimleri.....	5
3.11. Yazılım Kalite Faktörleri	5
3.12. Tasarım ve Uygulama Kısıtlamaları	6
3.13. Personelle İlgili Gereksinimler	6
3.14. Eğitimle İlgili Gereksinimler.....	6

3.15. Lojistikle İlgili Gereksinimler	7
3.16. Diğer Gereksinimler	7
3.17. Ambalajlama Gereksinimleri	7
3.18. Gereksinimlerin Önceliği ve Kritikliği	7
4. VASIFLANDIRMA YÖNTEMLERİ	7
4.1. Vasıflandırma Yöntemleri.....	7
4.2. Vasıflandırma Tablosu	8

KISALTMALAR

Kısaltmalar	Açıklamalar
CBOW	The Continuous Bag Of Words
GUI	Graphical User İnterface
IDE	Integrated Development Environment
LSTM	Long Short-Term Memory
MB	Megabayt
NLTK	Natural Language Toolkit
SDD	Software Design Description
SPMP	Software Project Management Plan
SRS	Software Requirement Specifications
STD	Software Test Description
TPU	Tensor Processing Unit
YKE	Yazılım Konfigürasyon Elemanı

1. KAPSAM

1.1. Tanım

Twitter’da Gerçek-Sahte Haber Tespit Projesi, sürüm 1.0 (beta)

Teknolojinin gelişmesinden dolayı artan internet ve akıllı cihaz kullanımı ile birlikte insanların sosyal medya kullanım oranı artmıştır. İnsanlar kolaylıkla ulaşabildikleri bu teknoloji sayesinde gerçek zamanlı olarak sosyal medya üzerinden yaşadıkları olayları paylaşabilmektedirler. Aynı şekilde haber ajanslarının da sosyal medya üzerinde hesap sahibi olmaya başlamasıyla birlikte bu ajanslar da haberlerini sosyal medya üzerinden paylaşmaya başlamışlardır. Haber ajansları veya insanlar tarafından paylaşılan bu bilgiler kimi zaman abartılıp anlatılırken kimi zaman doğru bir şekilde anlatılmaktadır. Bazen ise doğru paylaşılan bu bilgiler kulaktan kulağa dolaşırken veya gündem değiştirmek amacıyla çarpıtılır. Sosyal medya ile çoğu zaman doğruluğu kanıtlanamamış bu bilgilerin hızlı bir şekilde yayılması bilgi kirliliği yaratabilmektedir. Dolayısıyla bu durum sosyal medya aracılığıyla elde edilen bilgilerin doğruluğunu tartışmalı hale getirmektedir. Doğruluğu kanıtlanmamış bu bilgilerin hızlı bir şekilde yayılması insanları, ülkelerin ekonomisini, siyaseti vb. konuları etkileyebilmekle birlikte toplum içerisinde kargaşa ortamı yaratabilmektedir. Bu bilgilerin en hızlı yayıldığı sosyal medya ortamlarından biri olan Twitter’da herkes kendi hesabında anlık olarak en fazla 280 karakter uzunluğunda yazılar paylaşabilmektedir. Eğer bir konu bir ülkede veya dünyada çok fazla konuşuluyor ise o ülke veya dünya gündeminde yer alır. Bu gündemi takip eden kişiler o konuya ait tweetleri görüntüleyebilir. Böylelikle doğru ya da sahte bir haberin yayılımı kolaylaşabilir. Çünkü çoğu insan okuduklarının doğruluğunu araştırmadan inanma eğilimindedir. Bunun önlemek amacı ile sosyal medya platformlarında gerçek ya da sahte bilgileri ayırt edebilecek bir sisteme ihtiyaç duyulmaktadır. Yazılım Gereksinim Belirtim Dökümanı içerisinde bu soruna alternatif bir çözüm olarak geliştirilecek olan derin öğrenme ile Twitter platformunda yalan haber tespiti konulu projenin gereksinimleri belgelenecektir.

1.2 Sisteme Genel Bakış

Derin öğrenme ile Twitter platformunda yalan haber tespiti konulu proje basit bir ara yüz aracılığıyla girilen haber içerikli bir tweetin içeriğinin gerçek ya da sahte olmasını doğru bir şekilde tahmin etmeye çalışmaktadır. Bu proje kapsamında derin öğrenme ve doğal dil işleme teknikleri birlikte kullanılarak bir Python uygulaması geliştirilecektir. Uygulamanın kullanıcı kitlesi Twitter üzerinde elde edilen bir haberin doğru olup olmadığını öğrenmek isteyen

kullanıcılardan oluşmaktadır. Proje Twitter'dan elde edilen en fazla 280 karakter uzunluğundaki bilgiler için uygundur.

Derin öğrenme ile Twitter platformunda yalan haber tespiti konulu proje Gazi Üniversitesi Mühendislik Fakültesi'nin Bilgisayar Mühendisliği bölümünde 4. sınıfta öğrenim gören iki kişilik bir öğrenci grubu tarafından gerçekleştirilecektir. Projenin dokümantasyon takvimi şu şekildedir:

1. Yazılım Proje Yönetim Planı (SPMP): **19 Kasım 2021**
2. Yazılım Gereksinim Belirtilim Dokümanı (SRS): **10 Aralık 2021**
3. Yazılım Tasarım Tanımlama Dokümanı (SDD): **14 Ocak 2022**
4. Yazılım Test Tanımlama Dokümanı (STD): **21 Ocak 2022**
5. Proje Rapor Teslimi ve Sunumu: **28 Ocak 2022**

1.3 Dokümana Genel Bakış

SRS (Software Requirements Specifications) dokümanı; kullanıcı ve sistem gereksinimlerini içeren, sistem geliştiricilerinden neler beklendiğini gösteren bir dokümandır. Bu dokümanda sistemin ne yapacağı açıklanmaya çalışılacaktır. Doküman, yazılım gereksinimlerinin detaylı bir şekilde incelenmesini ve her gereksinim için seçilen vasıflandırma yöntemini içerir.

2. İLGİLİ DOKÜMANLAR

SRS raporu hazırlanırken rapor düzeni ve içeriği anlamında aşağıdaki kaynaklardan yararlanılmıştır.

1. Karacan,H. (2021).*Yazılım Mühendisliği* [Power Point Slayt]
2. Wright,S.(1999). SOFTWARE REQUIREMENTS SPECIFICATION (SRS).
http://simonjwright.users.sourceforge.net/pushface.org/mil_498/srs-did.htm
3. Johnson,B.J. TY. *Requirements Specification Document*. [Word Belgesi]
4. Ericson, L. Shine, S. (2014). Minutia Deviation Tool: Software Requirements Specification (SRS), (Version 1.1)
<https://www.ojp.gov/pdffiles1/nij/grants/249556.pdf>
5. Ayten, C. Aksu, G. (2021) YAZILIM PROJE YÖNETİM PLANI (SPMP) Derin Öğrenme İle Twitter Platformunda Yalan Haber Tespiti [PDF Belgesi]

3. GEREKSİNİMLER

3.1 Gerekli Durum ve Modlar

Herhangi bir gerekli durum ve mod bulunmamaktadır.

3.2 YKE (Yazılım Konfigürasyon Elemanı) Fonksiyonel Gereksinimleri

3.2.1 Haberin Doğruluğunun Tespiti

- i. Haber içerikli tweet metin kutusu ile kullanıcıdan alınmalıdır.
- ii. Alınan haber içerikli tweetin uzunluğu en fazla 280 karakter olmalıdır.
- iii. Kullanıcı Tahmin butonuna tıkladıktan sonra girilen tweetin gerçek veya sahte olduğu eğitilmiş model kullanılarak tespit edilmelidir.
- iv. Yapılan çıkarım kullanıcıya en fazla 1 saniye içerisinde gösterilmelidir.
- v. Yapılan çıkarımın doğruluğu en az %85 orana sahip olmalıdır.
- vi. Kullanıcı yeni bir tweet girmek istediğinde Sayfayı Temizle butonuna tıklaması sonucu sayfa ilk haline dönmelidir.

3.3 YKE Dış Ara Yüz Gereksinimleri

Verilerin yüklenmesi, tanınması, verilere ön işlemlerin uygulanması, model oluşturulması, model eğitimi, oluşturulan modellerin testi ve testlerin raporlanması Google Colab üzerinden gerçekleştirilecektir. Daha sonrası bu model .h5 dosyası olarak kaydedilecek ve .h5 dosyasını da kullanarak Spyder IDE'si üzerinden ara yüz oluşturulacaktır. Oluşturulan bu basit ara yüzde:

- 3.2.1 Haberin Doğruluğunun Tespiti başlığında belirtildiği gibi kullanıcı tweet için belirtilen alana 280 karakter uzunluğunda haber içerikli bir tweeti yazabilmelidir.
- 3.2.1 Haberin Doğruluğunun Tespiti başlığında belirtildiği gibi tweet yazıldıktan sonra Tahmin butonuna tıklandığında tahmin sonucu ekrana gelmelidir.
- 3.2.1 Haberin Doğruluğunun Tespiti başlığında belirtildiği gibi kullanıcı bir tweete dair sonucu öğrendikten sonra yeni bir tweet yazabilmek için Sayfayı Temizle butonuna tıkladığında sayfa temizlenmelidir.

3.4 YKE Dâhili Ara Yüz Gereksinimleri

Herhangi bir dahili ara yüz gereksinimi bulunmamaktadır.

3.5 YKE Dâhili Veri Gereksinimleri

Derin öğrenme ile Twitter platformunda yalan haber tespiti konulu projede bir tweetin sahte ya da gerçek olduğunun tespit edilebilmesi beklenmektedir. Bu nedenle projedeki modelin önceden sahte ve gerçek tweetlerin bulunduğu bir veri seti ile eğitilmesi gerekmektedir. Başka bir deyişle model bir veri seti kullanılarak eğitilmeden herhangi bir tweetin sahte ya da gerçek olduğunun çıkarımını yapamaz. Modelin çıkarımı doğru bir şekilde yapabilmesi için eğitilmesine ihtiyaç vardır ve dolayısıyla eğitim için de bir veri setine ihtiyaç duyulmaktadır.

3.6 Uyarılama Gereksinimleri

Herhangi bir uyarılama gereksinimi bulunmamaktadır.

3.7 Emniyet Gereksinimleri

Herhangi bir emniyet gereksinimi bulunmamaktadır.

3.8 Güvenlik Ve Gizlilik Gereksinimleri

Herhangi bir güvenlik ve gizlilik gereksinimi bulunmamaktadır.

3.9 YKE Ortam Gereksinimleri

Derin öğrenme ile Twitter platformunda yalan haber tespiti konulu proje bir masaüstü uygulaması olarak geliştirilecektir. Proje .py dosyası olarak oluşturulduktan sonra .exe dosyasına dönüştürüleceği için .exe dosyasını açabilen uygulamalarda ve sistemlerde çalışabilecektir. Mobil ortama uyum sağlanması için yeniden ara yüz geliştirilmesi gerekeceğinden mobil ortamlara uygun değildir.

3.10 Bilgisayar Kaynak Gereksinimleri

Bilgisayar kaynak gereksinimleri; bilgisayar donanım gereksinimleri, bilgisayar donanım kaynak kullanımı gereksinimleri, bilgisayar yazılım gereksinimleri ve bilgisayar iletişim gereksinimleri başlıkları altında açıklanmıştır.

3.10.1 Bilgisayar Donanım Gereksinimleri

3.9. YKE Ortam Gereksinimleri başlığında belirtildiği gibi bu proje bir masaüstünde çalışan bir uygulama olarak geliştirilecektir.

3.10.2 Bilgisayar Donanımı Kaynak Kullanımı Gereksinimleri

- i. Uygulama en fazla 2 Megabayt (MB) dâhili depolama alanı gerektirmektedir.

3.10.3 Bilgisayar Yazılım Gereksinimleri

- i. Verilere dair işlemler Google Colab üzerinden gerçekleştirilmelidir.
- ii. Ara yüz gerçekleştirimi Spyder IDE'si üzerinden yapılmalıdır.
- iii. Verilere dair işlemler yapılırken Pandas, Numpy, Matplotlib, Seaborn, Wordcloud, Tensorflow Keras, NLTK ve Re gibi kütüphaneler kullanılacaktır.
- iv. Ara yüz gerçekleştirilirken üstteki bazı kütüphanelerle birlikte Tkinter kütüphanesi kullanılmalıdır.

3.10.4 Bilgisayar İletişim Gereksinimleri

Herhangi bir bilgisayar iletişim gereksinimi bulunmamaktadır.

3.11 Yazılım Kalite Faktörleri

Güvenilirlik: Güvenilirlik faktörü modelin eğitimi bittikten sonra modelin doğruluk oranına bağlıdır. Modelin doğruluk oranını etkileyen birçok kısıt vardır. Bunların başında verilerin doğruluğu gelir. Modelin eğitimi için kullanılan verinin içerisinde hatalar olması örneğin gerçek bir haberin sahte olarak etiketlenmesi ya da veriler içerisinde haberdan başka bilgilerin bulunması gibi veri gürültüleri modelin doğruluk oranını etkilemektedir. Bir diğer kısıt ise uygulama geliştirilirken dâhil edilen verilerin yeterli sayıda olmaması ya da veri temizliğinden sonra yetersiz sayıda veri kalmasıdır. Ek olarak verilerin içerisindeki sahte ve gerçek haber sayısının modelin doğru çıkarım yapması için uygun bir şekilde dağılmış olmaması da modelin doğruluk oranını etkilemektedir. Verilere yapılan ön işlemlerin eksik bir şekilde tamamlanmış olması sonrasında eksik ön işlemlerle model oluşturulması/modelin yanlış kurulması, modele uygun hiperparametrelerin seçilmemiş olması, hiperparametrelerin optimize edilememesi modelin doğruluk oranını olumsuz yönde etkilemektedir. Dolayısıyla uygulamanın haber içerikli bir tweetin sahte ya da gerçek olduğunun çıkarımını en doğru şekilde yapabilmesi için bu kısıtlamaları eksiksiz ve doğru bir şekilde yerine getirilmesi gerekmektedir.

Erişilebilirlik: Proje bir masaüstü uygulaması olarak geliştirilmektedir, mobil cihazlar üzerinde çalıştırılmaz. Mobil cihazlar üzerinde çalışabilmesi için farklı uygulamalar ya da mobil cihazlara uygun ara yüz oluşturulmasını gerektirir.

Esneklik ve İdame Edilebilirlik: Proje şelale yazılım tasarım süreci kullanılarak geliştirileceği için bir hata olması sonucunda projedeki bu hatanın kolayca düzeltilebilir olma yeteneği düşüktür. Aynı şekilde düzeltme işlemi ile değişen gereksinimlere kolayca adapte olma yeteneği de düşüktür. Çünkü şelale yazılım tasarım sürecinde bir hatanın düzeltilebilmesi için

önceki aşamaları dönmek, birçok aşamayı tekrarlamak gerekmektedir. Bu işlemler maliyeti artıracığından gereksinimlerin iyi belirlenmiş olması, proje aşama sıralaması ve model tasarımının doğru yapılmış olması sonrasında çıkan herhangi bir hata sonucu önceki aşamalara dönme maliyetini ortadan kaldıracaktır.

Taşınabilirlik: Uygulama .exe dosyalarını çalıştırabilen her ortamda kullanılabilir. Yeni bir ortama geçiş yapılacağı zaman proje kolayca yeni ortama taşınabilir. Örneğin mobil ortama geçiş yapılacağı zaman oluşturulan basit bir ara yüze proje entegre edilerek taşınabilmektedir.

Yeniden Kullanılabilirlik: Modelin .h5 olarak kaydedilmesi sonucu bu model bir çok farklı uygulama içinde kullanılabilir hale gelir.

Kullanılabilirlik: Proje için basit, sade ve anlaşılabilir bir ara yüz geliştirilecektir. Kısacası ara yüz kullanıcı dostu olacaktır. Bu nedenle kullanımı son kullanıcı tarafından kolaylıkla öğrenilebilir.

3.12 Tasarım Ve Uygulama Kısıtlamaları

Uygulama Python programlama dili kullanılarak geliştirilecektir. 3.10.3 Bilgisayar Yazılım Gereksinimleri başlığında belirtildiği üzere geliştirme ortamı olarak Google Colab ile birlikte Spyder IDE'si kullanılacaktır.

3.10.3 Bilgisayar Yazılım Gereksinimleri başlığında belirtildiği üzere verileri işlemek, görselleştirmek, model oluşturmak, eğitmek ve GUI oluşturmak amacıyla için Pandas, Numpy, Matplotlib, Seaborn, Wordcloud, Tensorflow Keras, Tkinter, NLTK ve Re gibi kütüphaneler kullanılacaktır.

- i. Google Colab ortamı içerisindeki TPU desteği ile projedeki veri sayısının artması sonucu model eğitim süresinin kısaltılması sağlanacaktır.

3.13 Personelle İlgili Gereksinimler

Personelle ilgili herhangi bir gereksinim bulunmamaktadır.

3.14 Eğitimle İlgili Gereksinimler

Proje içerisinde oluşturulan modelin haberler hakkında doğru çıkarımı yapabilmesi için oluşturulan modelin veriler ile eğitilmesi gerekmektedir. Veriler ham halde modele verilemez bu yüzden verilere ön işlemlerin yapılması gerekmektedir. Bu ön işlemler:

- i. Verilerdeki büyük harflerin küçük harfe dönüştürülmesi

- ii. Verilerdeki noktalama işaretlerinin kaldırılması
- iii. Verilerdeki kelimelerin kök haline getirilmesi
- iv. Verilerdeki etkisiz kelimelerin kaldırılması
- v. Tokenize işlemi

Ayrıca verilerin vektöre dönüştürülmesi için Word2Vec yapısı ve Word2Vec yapısının kullandığı CBOW (Continuous Bag of Words) ve Skip-Gram Model'den faydalanılacaktır.

3.15 Lojistikle İlgili Gereksinimler

Lojistikle ilgili herhangi bir gereksinim bulunmamaktadır.

3.16 Diğer Gereksinimler

Önceki paragraflarda kapsanmamış bir gereksinim bulunmamaktadır.

3.17 Ambalajlama Gereksinimleri

Herhangi bir ambalajlama gereksinimi bulunmamaktadır.

3.18 Gereksinimlerin Önceliği Ve Kritikliği

Uygulamanın var olabilmesi için en önemli öncelik veridir. Veri üzerine yapılacak olan işlemler sonucu veriler tasarlanmış olan modelin eğitimi için kullanılacaktır. Eğitim sonunda model bu veriler sayesinde doğru çıkarımı yapabilir noktaya gelecektir. Verilerin doğruluğu, verilerin sayısı, kalitesi, dağılımı modelin doğruluk oranını önemli ölçüde etkilemektedir.

4. VASIFLANDIRMA YÖNTEMLERİ

4.1 Vasıflandırma Yöntemleri

YKE test edilmesi için aşağıdaki vasıflandırma yöntemlerinden yararlanılacaktır. Vasıflandırma yöntemleri olarak demo, muayene ve özel vasıflandırma yöntemleri kullanılacaktır.

Demo: Özel test ekipmanı veya ölçüm cihazları kullanımını gerektirmez. Gözleme dayalı fonksiyonel işleme ile YKE'nin test edilmesi işlemine denir.

Muayene: Kodun, belgelerin vs. görsel olarak incelenmesine denir.

Özel vasıflandırma yöntemleri: Özel araçlar, teknikler, prosedürler, tesisler ve kabul sınırları gibi her tür özel vasıflandırma yöntemlerinin uygulamayı test etmek için kullanılmasına denir.

4.2 Vasıflandırma Tablosu

3.2 YKE Fonksiyonel Gereksinimleri		
3.2.1 Haberin Doğruluğunun Tespiti		
i.	Demo	Metin kutusuna tweet girildikten sonra geçici bir labelda gösterilerek kontrol edilir.
ii.	Demo	Tweeti 280 karakterden fazla uzunlukta girip alıp almadığı kontrol edilir.
iii.	Demo	Butona basıldıktan tahmin sonucunun ekranda görünüp görünmediği kontrol edilir.
iv.	Özel	Time modülü ile geçen süre hesaplanıp kontrol edilir.
v.	Özel	Gerekli modüller ile modelin doğruluk değeri ölçülüp kontrol edilir.
vi.	Demo	Butona basıldığında ekranın ilk haline dönüp dönmediği kontrol edilir.
3.5 YKE Dâhili Veri Gereksinimleri		
i.	Muayene	Oluşturulan modelin veriler ile eğitilip eğitilmediği kontrol edilir.
3.9 YKE Ortam Gereksinimleri		
i.	Muayene	Uygulamanın masaüstü uygulaması olarak çalışıp çalışmadığı kontrol edilir.
3.10 Bilgisayar Kaynak Gereksinimleri		
3.10.2 Bilgisayar Donanımı Kaynak Kullanımı Gereksinimleri		
i.	Muayene	Uygulamanın boyutu kontrol edilir.
3.10.3 Bilgisayar Yazılım Gereksinimleri		
i.	Muayene	Kodların nerede çalıştığı kontrol edilir.
ii.	Muayene	Kodların nerede çalıştığı kontrol edilir.
iii.	Muayene	Kodlar üzerinde inceleme yapılır.
iv.	Muayene	Kodlar üzerinde inceleme yapılır.

3.12 Tasarım ve Uygulama Kısıtları

i.	Muayene	Colab üzerinden çalışma zamanı türü kontrol edilir.
-----------	---------	---

3.14 Eğitimle İlgili Gereksinimler

i.	Muayene	Kodlar üzerinde inceleme yapılır.
ii.	Muayene	Kodlar üzerinde inceleme yapılır.
iii.	Muayene	Kodlar üzerinde inceleme yapılır.
iv.	Muayene	Kodlar üzerinde inceleme yapılır.
v.	Muayene	Kodlar üzerinde inceleme yapılır.