ÖRNEKTİR

#### PATENT / F.MODEL BAŞVURULARI İÇİN

#### ÖRNEK ÖN HAZIRLIK KILAVUZU

Bu kılavuza uygun olarak yapacağınız çalışma,

* Buluşunuzun tarafımızdan en iyi şekilde tanımlanabilmesi ve
* İleride hukuk önünde istenen korumanın elde edebilmesi için temel teşkil edecektir.
1. **Buluş başlığı**

*ÖĞRENEBİLME YETENEĞİNE SAHİP HASTALIK TEŞHİS YAZILIMI*

1. **Buluşunuz hangi alanda kullanılacaktır?**

*Buluş, sağlık alanında, hastalıkların teşhisinde kullanılacak hastalıkların teşhis/tedavisinde kullanılacak bir öngörücü ve tanımlayıcı analiz yazılımı ile ilgilidir. Buluş özellikle, hasta şikâyet formunda veri madenciliği ile analiz yapan, bu analizlerle hekimlerin hastalık teşhis ve tedavi sürecini kolaylaştıran, hastaların şikayetlerini Makine Öğrenmesi algoritmaları ve Doğal Dil İşleme ile hastalıkların teşhis/tedavisinde kullanan bir öngörücü ve tanımlayıcı analiz robotu ile ilgilidir.*

1. **Mevcut uygulamalar hakkında detaylı bilgi veriniz. Bahsedilen uygulamaların eksikliklerini ve bu uygulamalar neticesinde ortaya çıkan en az bir teknik problemi belirtiniz.**

**(Varsa mevcut tekniğe ait resim ya da fotoğrafları da forma ilave ediniz.)**

*Mevcut teknikte makine öğrenmesi algoritmalarının hisse senedi tahmini, banka kredisi için müşteri risk analizi, deprem tahmini, metin sınıflandırma, hastalık teşhisi, genetik çalışmalar gibi çok çeşitli alanlarda kullanıldığı bilinmektedir. Ancak sağlık sektöründe 15 makine öğrenmesi ile doğal dil işleme tekniklerinin geliştirilip, hekimlerin karar destek süreçlerine katkı sağlayacak spesifik olarak uygulama bulunmamaktadır.*

*Sonuç olarak, yukarıda anlatılan olumsuzluklardan dolayı ve mevcut çözümlerin konu hakkındaki yetersizliği nedeniyle ilgili teknik alanda bir geliştirme yapılması gerekli kılınmıştır.*

1. **Buluşunuzun 3. maddede belirtilen teknik problemleri ortadan kaldırmak üzere sunduğu çözüm nedir? Bu çözümü sağlamak üzere hangi unsur ya da özelliklerden faydalanılmaktadır.**

*Buluş, mevcut durumlardan esinlenerek oluşturulup yukarıda belirtilen olumsuzlukları çözmeyi amaçlamaktadır. Buluşun ana amacı, geliştirilen yapay zekâ (Makine Öğrenmesi ve Doğal Dil İşleme) algoritmaları ile hasta şikayetlerinden yola çıkarak hastanın muayene olması gereken uzmanlık alanı belirlenmesi ve hastalık teşhisi yapılabilmesidir. Buluşun bir amacı, Tokenize” işlemi ile Veri İşleme süresince metin dokümanını analiz edilmesini sağlamaktır. Buluşun diğer bir amacı, filtreleme işlemi ile metin içerisindeki tüm ortak ayraç̧, işaret, noktalama işaretleri ve yazdırılmayan karakterlerin kaldırılması ve sık kullanılan gereksiz kelimelerin (bağlaçlar ve edatlar) filtrelenmesinin sağlanmasıdır.*

1. **Buluşunuzda yer alan unsurları çizim üzerinde işaretleyiniz ve bir referans tablosu oluşturunuz.**

*Şekil 1, buluşa konu olan sistem mimarisidir*

*Şekil 2, buluşa konu olan uygulama mimarisidir*

*Şekil 3, buluşa konu olan sistemin kullanıcı senaryo diyagramıdır*

*Şekil 4, buluşa konu olan sistemin sıralı olay diyagramıdır*

|  |  |
| --- | --- |
| **1.**  | Doğal Dil İşleme |
| **2.**  | Makine öğrenmesi  |
| **10.**  | Uygulama katmanı  |
| **11.**  | Http api servisleri  |
| **13.**  | Veri tabanı |
| **14.**  | Model |
| **14.1** | Eğitilmiş model |
| **15.**  | Temsili Durum Transferi (Rest) |
| **16.**  | Ağ Hizmetleri |
| **17.**  | Arayüz |
| **18.**  | Kontrolcü |
| **19.**  | Spring servisleri |
| **20.**  | Spring JDBC |
| **21.**  | Spring data/JPA |
| **..** |  |

 ** **

1. **Buluşunuzda yer alan unsurlar nelerdir?**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Unsur,****No** | **Unsur Adı** | **Yeni** | **Önceki****Tekniğe Ait** | **Buluşum için çok önemli** | **Unsurun işlevini ve bu işlevi gerçekleştirmesini sağlayan teknik özelliğini belirtiniz.** |
| **1.**  | Doğal Dil İşleme | [x]  | [ ]  | [x]  | Aldığı komutlara göre ürünlerin dağıtımını gerçekleştiren ve uygulayan yazılımdır. |
| **2.**  | Makine öğrenmesi  | [x]  | [ ]  | [x]  | Dağıtma makinesinin (10) kontrol edilmesi için kullanılmaktadır. |
| **10.**  | Uygulama katmanı  | [ ]  | [x]  | [ ]  | Yazılımın katman içeriğini ifade etmektedir. |
| **11.**  | Http api servisleri  | [ ]  | [x]  | [x]  | Ağ servis sağlayıcısını ifade etmektedir. |
| **13.**  | Veri tabanı | [ ]  | [x]  | [x]  |  |
| … | … | [x]  | [ ]  | [x]  |  |
|  |  | [x]  | [ ]  | [x]  |  |
|  |  | [x]  | [ ]  | [x]  |  |

**Tablo 1**

1. **Buluşun unsurlarına atıfta bulunarak çalışma mantığını ya da prensibini özetleyeniz.**

**(Tabloda belirtilen tüm unsurlara atıfta bulunulması gerekmektedir. Anlatımda belirlenen unsur adları ve referans numaralarını kullanmaya özen gösteriniz. )**

*Bu detaylı açıklamada, buluşa konu olan hastalıkların teşhis/tedavisinde kullanılacak bir öngörücü ve tanımlayıcı analiz robotunun tercih edilen yapılanmaları, sadece konunun daha iyi anlaşılmasına yönelik olarak açıklanmaktadır. Buluşa konu olan sistem mimarisi şekil 1’ de verilmektedir.*

*Şekil 3 ve şekil 4 sırasıyla kullanıcı senaryo diyagramı ve sıralı olay diyagramı verilen buluş konusu geliştirilen yapay zekâ (Makine Öğrenmesi ve Doğal Dil İşleme) algoritmaları ile hasta şikayetlerinden yola çıkarak hastanın muayene olması gereken uzmanlık alanı belirlenmesi ve hastalık teşhisi yapılabilmesi amaçlanmaktadır. Makine Öğrenmesi (2) ve Doğal Dil İşleme (1) algoritmaları geliştirilirken Google Jigsaw Toxic Comment Multi Label Classification veri seti kullanılmıştır. Google çatısı altında yer alan Jigsaw çalışma grubu internette yer alan oransız, rahatsız edici aşağılayıcı yorumları tespit edebilmek için deneysel makine öğrenmesi çalışmalarında kullanılmak üzere çoklu etikete sahip veri seti 6 adet kategori içermekte ve birden fazla etikete sahip olabilmenin yanı sıra veri seti 32.000 satırdan oluşmaktadır. Her satırda ‘hakaret’, ‘ağır hakaret’, ‘aşağılama’, ‘tehdit’, ‘müstehcen’, ‘ırkçı’ başlıkları altında kategorize edilmiş̧ ifadeler bulunmaktadır.*

*“Veri İşleme” süresince öncelikle metin dokümanını analiz edebilmek için “Tokenize” işlemi uygulanarak kelime grupları elde edilmektedir. Sonrasında metin içerisindeki tüm ortak ayraç̧, işaret, noktalama işaretleri ve yazdırılmayan karakterlerin kaldırılması ve sık kullanılan gereksiz kelimelerin (bağlaçlar ve edatlar) filtrelenmesi için “StopWords” filtreleme işlemi uygulanmaktadır. Son olarak metin kelime hakkında dil bilgisel veya sözcüksel bilgiler sunan, son eklerin çıkartılmasıyla morfolojik kökün elde edilmesini sağlayan “Stemming” işlemi uygulanmaktadır. Tüm bu işlemler sonrasında elde edilen dokümanın içerisindeki kelimelerin terim ağırlıklarının hesaplanabilmesi için “TF-IDF” algoritması kullanılmaktadır. “Modelleme” aşamasında ve işlenmiş̧ verinin semantik sınıflandırılmasının yapılabilmesi için “Makine Öğrenmesi” algoritmaları kullanılarak modeller oluşturulmaktadır. Örnek veri kümemizde veri içerisindeki terimlerin tespit edilmesinden sonra “Random Forest”, “Linear Regression”, “Linear SVC”, “XGBoost” ve “LGBM” makine öğrenmesi algoritmaları ile sınıflandırma işlemi yapılmakta ve 92,35% oranı ile en yüksek doğruluk oranı “LGBM” algoritması ile elde edilmektedir. Diğer oranlar şu şekildedir; XGM; 92,13%, Linear SVC; 91,91%, LR; 90,74%, RF; 89,88%. Aynı zamanda f1- score metriğine göre de değerlendirilme yapılmış olup, başarı oranındaki sıralama değişmemiştir. LGBM; 0,5714, XGB; 05700, Linear SVC; 0,4794, LR; 0,3657, RF; 0,3627. Doğal dil işleme (1) ile kelimelerin köklerinin bulunması, metindeki gereksiz işaretlerin temizlenmesi, sözcük yapısının belirlenmesi, metnin özetlenmesi, konuların modellenmesi sağlanmaktadır.*

*Makine öğrenmesinde (2) belgelerdeki sınıf sayısını bulmak için, gözetimli öğrenme algoritmaları kullanılmıştır. Aktarımlı öğrenme metodolojisi kullanılarak konu bağımsız bir dil modeli geliştirilmiştir. LSTM algoritması (4) hasta şikayetlerini sınıflandırma da kullanılmıştır.*

*Buluş konusu ürün; mobil cihazlar için geliştirilen uygulaması veya kişisel bilgisayar için geliştirilen servis uygulaması üzerinden metin olarak hasta şikayetlerini almaktadır. Sonrasında text formatında Türkçe dilindeki hasta şikayetlerini içeren ham veriyi doğal dil işleme (1) algoritmaları ile işleyip sınıflandırma işlemi için uygun işlenmiş veri haline5 getirmektedir. İşlenmiş veri uygulama için geliştirilen (Long-Short Term Memory) yapay sinir ağı ile sınıflandırarak hastanın şikayetinin hangi hastalıklara ait olabileceğini ve hastanın hangi uzmanlık alanında tedavi olabileceğini tahmin ederek sonuçlarını hastaya geri döndürmektedir. Buluşa konu olan yöntemdeki iş akışı şu şekildedir;*

* *İş tanımlama,*
* *Veri anlama,*
* *Veri hazırlama,*
* *Modelleme,*
* *Geliştirme,*
* *Uygulama.*

*Buluşa konu olan yöntem ile gerçekleştirilen işlem adımları şunlardır;*

* *Mobil cihazlar için geliştirilen mobil uygulama veya kişisel bilgisayar için geliştirilen servis uygulaması üzerinden metin olarak hasta şikayetlerinin alınması,*
* *Metin/Text formatında hasta şikayetlerini içeren ham verinin doğal dil işleme (1) algoritmaları ile işlenmesi ve sınıflandırma işlemi için uygun işlenmiş veri haline getirilmesi,*
* *İşlenmiş verilerin yapay sinir ağı ile sınıflandırılarak hastanın şikayetinin hangi hastalıklara ait olabileceği ve hastanın hangi uzmanlık alanında tedavi 20 olabileceğinin makine öğrenmesi (2) ile tahmin edilmesi ve sonuçların hastaya iletilmesi.*
1. **Buluş büyük bir yapı içindeyse yapının bütünü gösteren ya da anlatan çizim ve bilgiler.**

**

*Örnektir.*

1. **Patent ön araştırmasında kullanılabilecek anahtar kelimeler**

*Makine Öğrenmesi, Hastalık Teşhisi, Öğrenme Yeteneği, Yazılım*

1. ***Benzer ürünleri üreten yurtiçi ve yurtdışı firma isimleri, varsa internet sayfaları ve markaları***

*YCP BİLGİ TEKNOLOJİLERİ , Turkcelll*

1. ***Konuyla ilgili tespit edilen patent numaraları (araştırma yapmadıysanız bu maddeyi boş bırakabilirsiniz)***
2. *KR102183320*
3. *CN212023557*
4. *2017/01582*
5. **Lütfen buluşunuzla ilgili yukarıdaki alanlara sığmayan ya da bildirmek istediğiniz ek hususları aşağıdaki ek sayfalarda belirtiniz.**