

GAZİ ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ

BM495 BİLGİSAYAR PROJESİ I
SDD Belgesi

Abdullah Akalın
Karim El Guermai
Muhammed Emre Emrah

11.12.2017

İçindekiler

1 Genel Bakış	1
1.1 Kapsam	1
1.2 Amaç	1
1.3 Hedef Kitlesi	1
2 Tanımlar	1
3 Veri Bilimi Projelerinde Tasarım Sürecine Dair	1
4 Tasarım Bilgi İçeriği	2
4.1 Belgenin Organizasyonu	2
4.2 Projenin Beklentileri	2
4.3 Tasarım Katmanları	2
4.4 Tasarım Gerekçeleri	2
5 Proje Tasarımında İzlenen Yollar	2
5.1 İlginç Bir Konunun Belirlenmesi	2
5.2 Verinin Temin Edilmesi	4
5.3 Verinin Keşfedilmesi	4
5.4 Verinin Modellenmesi	4
5.5 Sonucun Tartışılması ve Görselleştirilmesi	4
6 Sonuç	4
Kaynaklar	5

1 Genel Bakış

1.1 Kapsam

Bu belgenin amacı grubumuzun üzerinde çalıştığı yapay öğrenme projesinin tasarım sürecini ele almak, takip edilen yaklaşımları açıklamak ve beklentileri değerlendirmektir.

1.2 Amaç

Bu tasarım belgesi, projenin yapısal modelini belirtir, projenin hayata geçirilmesi için göz önüne alınacak tasarım unsurlarını açıklar ve gerekli görülen yerlerde bu unsurların birbirleriyle olan ilişkisini gösterir.

1.3 Hedef Kitle

Belgenin hedef kitle, projenin geliştiricileri, gözetmeni ve bu projeyle ilgilenen herkeştir.

2 Tanımlar

Jupyter Notebook: İçinde düz yazıların, kodların ve kod çıktılarının, hatta grafiklerin ve görsellerin bir arada bulunabildiği açık kaynaklı bir web uygulamasıdır. [1]

3 Veri Bilimi Projelerinde Tasarım Sürecine Dair

Grubumuzun geliştirmeyi hedeflediği proje, bir veri bilimi araştırma projesidir. Veri bilimi usülleri, yazılım geliştirme usüllerinden biraz farklı bulunmaktadır. Zira yazılım geliştirme sürecinin sonucu bir bilgisayar kodudur, öte yandan bir veri bilimi projesi sonuç olarak bir konu hakkında çeşitli faydalı bilgilere vakıf olunmasını sağlar.

Veri bilimi, yazılım mühendisliğine nispetle yeni bir alandır. Bu durumun doğal bir sonucu olarak bu alanda proje geliştirmenin gelenekleşmiş bir yolu henüz tam olarak bulunmamaktadır. Aynı zamanda bir veri bilimi projesinde, projenin nasıl bir çıktı üreteceğinin önceden bilinmesi olanak dışıdır.

Bu nedenlerle yazılım mühendisliğine dayalı belgelendirmenin bir numunesi olan bu belgenin, veri bilimine has yukarıda bahsi geçen durumlar göz önüne alınarak değerlendirilmesi rica olunur.[2]

4 Tasarım Bilgi İçeriği

4.1 Belgenin Organizasyonu

İlk bölümde belgeye genel bakış yapılmıştır. Belgenin kapsam, amaç ve hedef kitlesi belirtilmiştir. Daha sonra bu belgenin hazırlanmasında etkili olan kriterlere değinilmiştir. Ardından projenin tasarımında izlenen yollar açıklanıp belge sonuçlandırılmıştır.

4.2 Projenin Beklentileri

Bu projenin beklentisi, bir konu hakkında bir problemin, yapay öğrenme yöntemleriyle çözülmesiyle, o konuya dair çıkarımlarda bulunulmasıdır. Her ne kadar proje başlangıcında görüntü verilerine dair bir problem çözümü düşünülmüş olsa da, proje danışmanının yönlendirmesiyle sağlık alanında bit problem incelenmektedir.

4.3 Tasarım Katmanları

Söz konusu bir veri bilimi araştırma projesi olduğu için, klasik yazılım projelerine has veritabanı, önyüz, sunucu vb. katmanlar projemizde bulunmamaktadır. Ancak problemin tespiti, verinin elde edilmesi, analizi ve modellenmesi aşamaları ayrı katmanlar olarak düşünülebilir. Bu adımlar 5 numaralı bölümde ayrıntılı olarak ele alınmıştır.

4.4 Tasarım Gerekçeleri

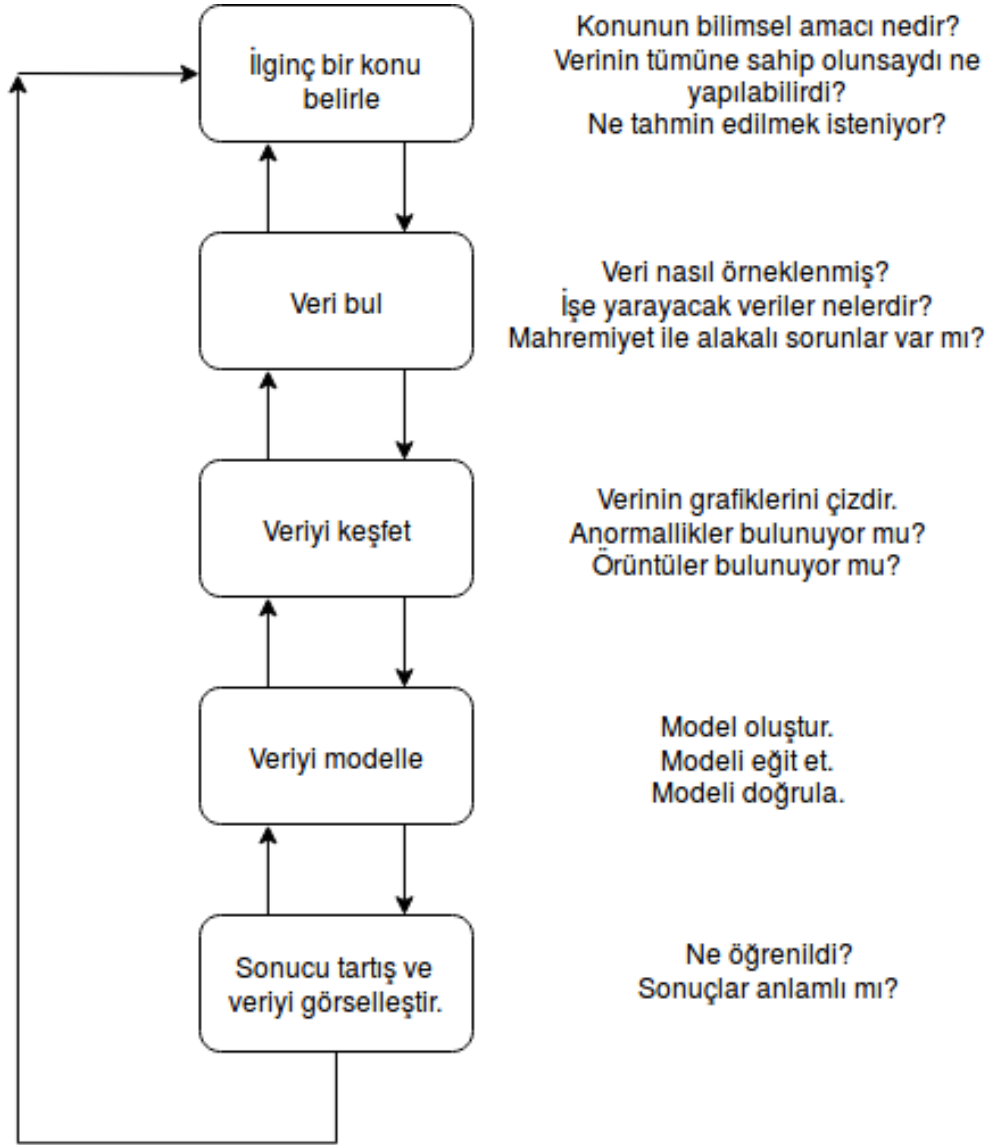
Projenin tekrar kullanılabilmesi ve değerlendirilebilmesi, atılan tüm adımların kaydedilmesine bağlıdır. Veri üzerinde yapılan tüm değişiklikler, öznel çıkarımları, veri üzerinde uygulanan fonksiyonlar aynı sırada ve aynı koşullar altında uygulanmalıdır. Bu yüzden projenin tasarım aşamalarında *Jupyter Notebook* kullanılmıştır.

5 Proje Tasarımında İzlenen Yollar

Projenin tasarımında "*Harvard data science*" dersinde kullanılan aşamalar esas alınmıştır (bkz. Şekil 1). Bu aşamalar müteakip bölümlerde açıklanmıştır. Belirtilmesi gereken bir husus şudur ki, bu aşamalar sadece yukarıdan aşağı ilerlemeyip, gerektiğinde aşağıdan yukarı geçişlere müsaade etmektedir.

5.1 İlginç Bir Konunun Belirlenmesi

Bu aşama, projenin ana gayesini belirleyen aşamadır. Belirlenen konuya dair bilgi edinilmesi de gerekebilmekte olup başarı kriterleri de yine konuyu bilenler ile birlikte



Şekil 1: Projede tasarımında kullanılan aşamalar. ([2] kaynağından uyarlanmıştır.)

değerlendirilmelidir.

5.2 Verinin Temin Edilmesi

Bu projede veri tarafımıza danışmanımızca sağlanmıştır. Ancak hazır verinin bulunmadığı taktirde verinin ölçümler sonucunda elde edilmesi gerekmektedir.

5.3 Verinin Keşfedilmesi

Bu aşama, eldeki verinin didiklenmesidir. Bu aşamada grafikler ve görseller kullanılarak verinin mahiyeti gözleme alınır. Bir sonraki aşama için gerekli altyapının oluşmasını sağlar.

5.4 Verinin Modellenmesi

İrdelenen veri, uygun görülen yöntemler kullanılarak modellenir. Hiçbir zaman bir model en iyisi değildir, bu yüzden birden çok modelin oluşturulması muhtemeldir. Modellerin sonuçları bir sonraki aşamaya aktarılır.

5.5 Sonucun Tartışılması ve Görselleştirilmesi

Modellerin sonuçları, belirlenen başarı kriterlerine göre belirlenir. Bir konuda, başarı gösteremeyen bir model, üstün başarı gösteren bir modelin ifade ettiğinden daha az değere sahip değildir. Burada ölçüt hedeflenen ve ulaşılmak istenen bilgi ve kavrayıştır.

6 Sonuç

Bu belgede yapılmakta olan projenin tasarım aşamalarına ve unsurlarına yer verilmiştir. Veri bilimi projelerinin doğası gereği şimdi göz önünde bulundurulan bir hususun önemsiz olduğunun daha sonra anlaşılması yahut tam tersi durumların ortaya çıkması muhtemel olduğundan genel bir çerçeve çizilmeye çalışılmıştır. Nihai sonuç ancak proje tamamlandıktan sonra anlaşılabilir.

Kaynaklar

- [1] “Project jupyter.” <http://jupyter.org/index.html>. Erişim tarihi: 11.12.2017.
- [2] C. Byrne, *Development Workflows for Data Scientists*. O'Reilly, first ed., 2017.