

Kodlama Dillerinin Popülerliklerinin Analizi

Ahmet Okan Karasu*

Özet

Çeşitli kodlama dillerinin, internetteki popülerliğine etki eden parametreler belirlenmesi

1 Giriş

Popülerlik öncelikle tanınma gerektiren bir kavramdır. Bu tanınma durumuna etki eden çeşitli faktörler bulunmaktadır. Bu faktörlere örnek olarak, bilginin aktarılması, ihtiyaç hissedilmesi gibi bazı süreçlerdir. Yazılım sektöründe de değişen zamana bağlı olarak kullanılan yazılım dillerinde de değişimler görünmektedir. Yazılım dillerindeki bu popülerliğe etki eden, Wikipedia ve Github sitelerinden gelen bilgiler dahil olmak üzere inceleme yapılmıştır. Bu çalışmada kullanılan 50 dil için kullanıcı sayıları ve sektördeki iş sayılarına etki eden faktörlerin analizleri yapıp yorumlanmıştır.

1.1 Çalışmanın Amacı

Bu çalışmada kapsamında 50 farklı programlama dili ve bunlara bağlı 10 ölçütün birbirleriyle olan ilişkilerinin tespiti için çeşitli veri analizi yöntemlerinin kullanılması amaçlanmıştır. Bu doğrultuda, github ve wikipedia gibi bilgi kaynaklarının da yazılım alanındaki etkilerinin hangi yönde ve güçte olduğunu tespit edilmesi planlanmaktadır.

1.2 Literatür

Dil çeşitli semboller aracılığıyla ifade edilen ve bir bilginin aktarılması haricinde, bir komutun birseye (örneğin bir işçi veya bilgisayar gibi) aktarılmasında sağlayan bir kavramdır. Bu kavram sadece konuşma dili sınırlı olmayan ve kendi içinde kuralları, standartları, kavramlar, isaretler vb. bulunmaktadır. Örneğin; matematik doğayı anlamakta kullandığımız bir dildir. Yazılım dünyasında bu türde bilinen veya çoğunlukla çok az kişinin bildiği diller de bulunmaktadır. Bunlar, kullanım alanlarına (web geliştirme, istatistiksel analiz vb.) veya

*19080760, [Github Repo](#)

tiplerine göre (programlama dili, sorgulama dili veya veri transferi vb.) çok farklı kategorilerde olabilmektedir. Örneğin , bu çalışma kapsamında da kullanılan R programlama dili istatistik alanı için geliştirilmiş açık kaynak kodlu bir dildir. Kullanım alanı, kullanım kolaylığı ve gerekli kütüphanelerin fazlalığından dolayı popüler dillerden birisidir. Benzer olarak SQL sorgulama dili ise kendi alanında kullanılan başka bir popüler başka bir dildir. Bazı diller genel kullanıma uygunken (SQL, Python vb), bazıları da ise sınırlı kullanım alanlarında dolayı daha az tercih edilmektedir. Bunun nedenleri, çok eski ve kullanım alanlarının olmaması, firmaların kendi ürünlerinde geliştirmekte kullanmak için ürettikleri diller olmasından kaynaklanmaktadır. Ayvaz vd. (2016) , Çobanoğlu (2020), Köseoğlu (2008), Şahin (2007).

2 Veri

Bu çalışmada kullanılan veri Github'daki [tidytuesday reposunda](#) bulunan programlama dillerine ait bir veriseti kullanılmıştır. Verinin orijinal hali [Programming Language DataBase](#)'den paylaşılmıştır. Ayrıca 4303 dile ait 49 değişkene ait bilgileri içermektedir. Verinin çalışmaya uygun hale getirilmesi için 15 değişken (2 kategorik ve 8 numerik) seçildi ve **language_rank** değişkenin olarak ilk 50'li olacak şekilde fitreleme yapılmıştır. 8 numerik değişkene ait ortalaması, standart sapması, max, min ve ortanca değerini gösteren (Tablo1) oluşturulmuştur.

```
## \begin{table}[ht]
## \centering
## \caption{Özet İstatistikler}
## \label{tab:ozet}
## \begin{tabular}{lcccc}
## \toprule
## & Ortalama & Std.Sap & Min & Medyan & Mak \\
## \midrule
## book\_count & 107.20 & 104.02 & 0.00 & 72.00 & 401.00 \\
## github\_language\_repos & 1827455.49 & 3665709.44 & 0.00 & 219084.00 & 16046489.00 \\
## language\_rank & 25.00 & 14.87 & 0.00 & 25.00 & 50.00 \\
## number\_of\_jobs & 34238.63 & 111362.17 & 53.00 & 4682.00 & 771996.00 \\
## number\_of\_users & 1073710.90 & 1811402.91 & 3796.00 & 208708.00 & 7179119.00 \\
## semantic\_scholar & 19.47 & 14.43 & 0.00 & 19.00 & 52.00 \\
## wikipedia\_daily\_page\_views & 2295.00 & 1999.81 & -1.00 & 1574.00 & 9882.00 \\
## wikipedia\_revision\_count & 2693.73 & 2182.54 & 2.00 & 2219.00 & 10104.00 \\
## \bottomrule
## \end{tabular}
## \end{table}
```

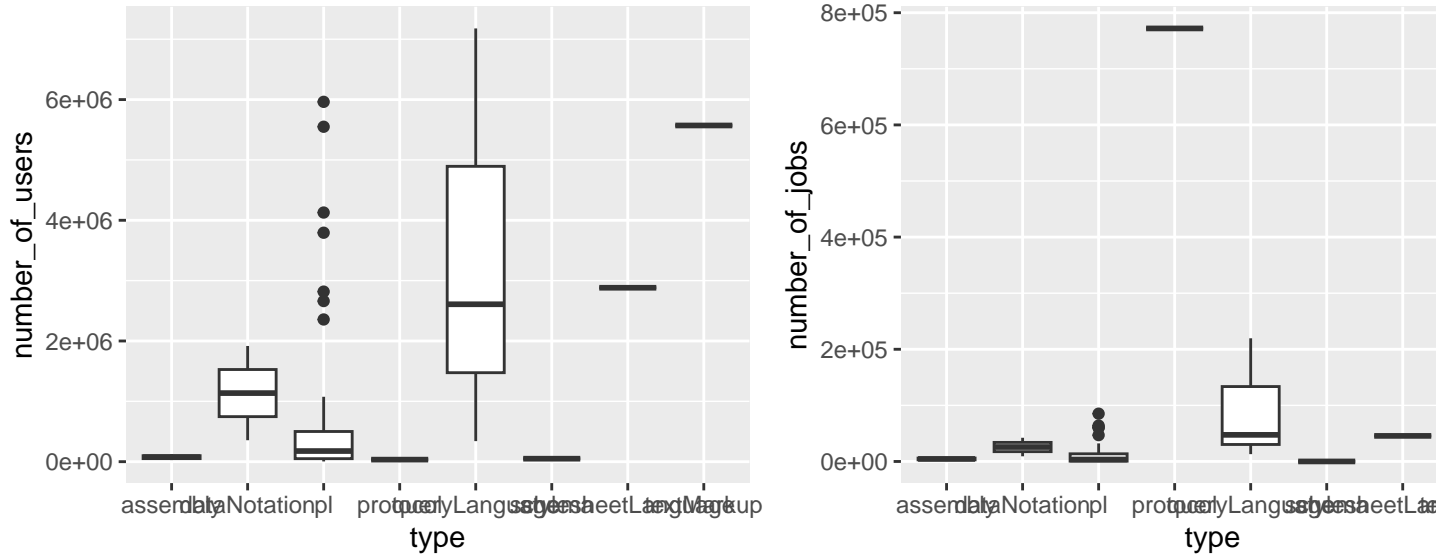
type is_open_source

pl :40 Mode :logical
queryLanguage: 3 FALSE:1

assembly : 2 TRUE :19
dataNotation : 2 NA's :31
protocol : 1
schema : 1
(Other) : 2

3 Yöntem ve Veri Analizi

Yazılım alanında kullanılan diller hakkında yorum yapabilmek için öncelikle farklı dil tiplerinin kullanıcı sayıları ve bu alandaki iş değerlerine bakmamız daha uygundur.

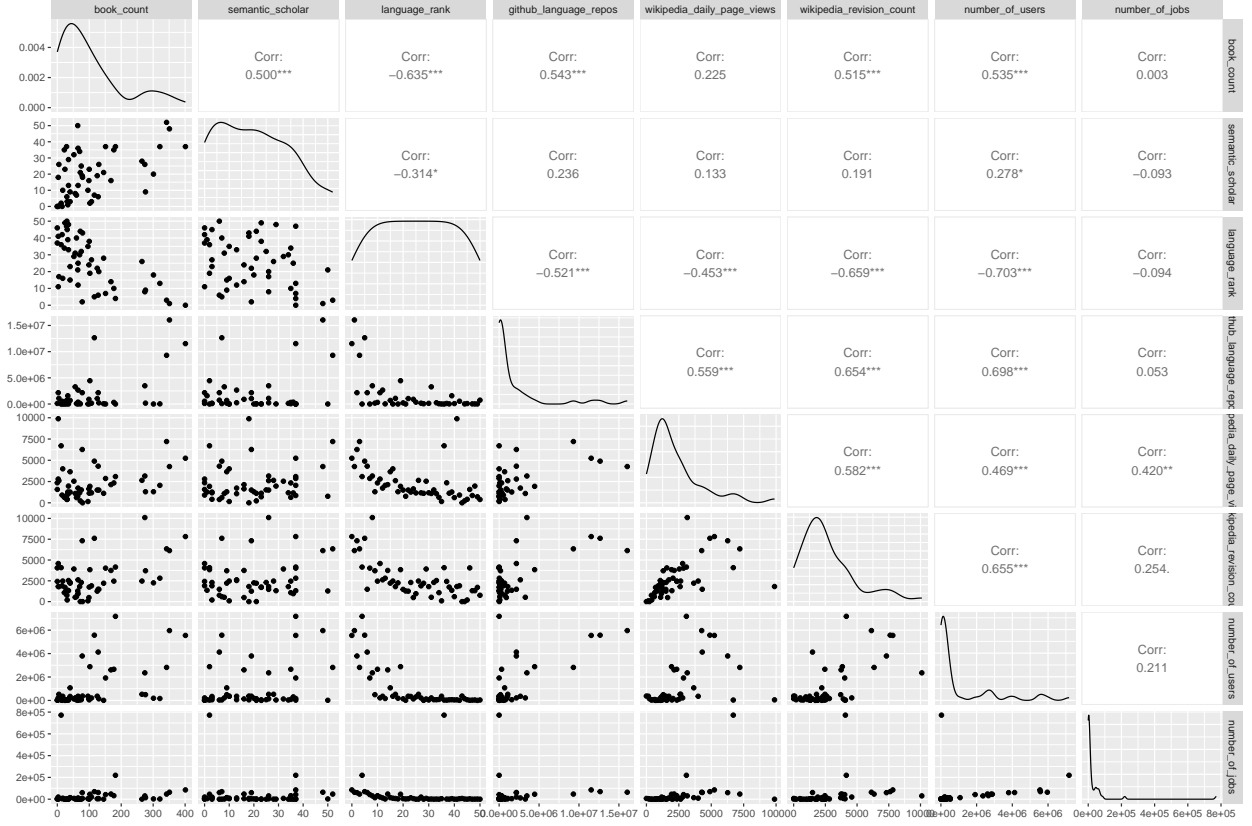


Şekil 1 ve 2 incelendiğinde özellikle sorgulama dillerinin (query language) median değer olarak diğerlerinden daha fazla tercih edildiği görülmektedir. Bunun nedeni çoğu sektörde verilerin depolanması veya sorgulanması aşamasında gerekmektedir. Bu nedenle de veri içerikli sektörlerde özellikle aranan kriterler arasındadır. Ayrıca web geliştirmede kullanılan HTML ve CSS gibi dilleri (ek olarak sorgu dilleri) alanlarında alternatifi olmadığı için özellikle tercih edilmektedir. PL ile gösterilen programlama dillerinde ise hem kullanıcı sayısı hemde iş sayısı sorgulama dillerine göre farklı bir dağılım göstermektedir. Bunun nedeni ise 40 farklı programlama dilinden bazıları aşırı derece kullanılması ve temel gereksinimlerinin daha fazla olmasındandır. Örneğin algoritma geliştirme gibi kavramlar sorgulama dillerinde pek kullanılan ifadeler değildir.

3.1 Dağılım matrisi ve değişkenlerin birbirleri ile olan ilişkisi

Şekil 3'deki corelasyon-dağılım matrisine göre kodlama dillerinin kullanıcı sayılarının, kitap sayısı, wikipedia görüntülenmesi, sayfa düzenlemesi ve github'taki repo sayılarıyla yüksek düzeyde pozitif korelasyona sahip olduğu görünmektedir. Özellikle github korelasyon değeri

0.698 değerine sahiptir. Buna rağmen dil ile alakalı yayınların daha düşük korelasyon değerine (0.278) sahip olduğu görülmektedir. Aynı grafikte dil ile alakalı iş sayısının ise günlük wikipedia görüntülenmesi ile anlamlı bir korelasyon değerine (0.420) sahiptir. Dili kullananlar ile dile ait iş sayısının ise düşük bir korelasyon değerine sahiptir.



Şekil 1: Kodlama dillerinin kullanıcı ve iş sayılarının diğer parametrelerle ilişki matrisi

4 Sonuç

Yazılım alanında kullanılan dillerin, kullanıcı popülarlığı ile iş alanındaki popülarlıkları arasında çeşitli farklar görölmektedir. Bunların nedenleri olarak, iş sektöründe daha iş amaçlı diller tercih edilirken, ilgili dili öğrenenler eğitim hobi veya kendi işlerini kolaylaştırmak amacıyla belirli düzeyde öğrenmiş olabilir. Ayrıca zamanın değişimine bağlı olarak sektörel gereksinimlerde de değişimler gerçekleşebilir. Bu nedenle popülarlık kavramını kullanıcı ve sektörel olarak ayrı ayrı düşünmek gerekmektedir. Eğer iş odaklı yazılım öğrenmek gerekiyorsa sektör ihtiyaçlarına göre hareket edilmesi daha uygundur. Yazılım dillerinde ayrıca dillere ait bilgi kaynaklara ulaşımında önemli olduğu yapılan analiz sonucunda çıkan önemli sonuçlardan birisidir. Wikipedia gibi hızlı rahat erişilebilir kaynakların dilin kullanıcı sayılarına etkisi dile ait kitap sayısı kadar fazladır. Ayrıca Github gibi hem bilgi hemde farklı uygulamaları görölebileceğimiz platformlar kullanıcı sayılarında önemli etkisi bulunmaktadır. Yapılan analizlerin bazı noktalarda yetersiz olabileceği durumlarda bulunmaktadır. Örneğin, R dili yüksek oranda istatistiksel analizlerde kullanılan bir dil olmasına rağmen Java gibi bazı diller birden fazla alanda kullanılabilir. Dilin toplam kullanıcı sayısı yerine o dilin hangi amaçla kullanıldığı bilgisinin olduğu bir veriseti daha kapsamlı sonuçlar elde etmemizi sağlayabilir. Verisetine ek olarak kullanılan analiz yöntemi de daha iyi sonuçlar elde etmemizi sağlayabilir. Örneğin doğrusal regresyon analizi bu veri setinde kullanarak daha kapsamlı sonuçlar elde edilebilir.

5 Kaynakça

- Ayvaz, U., Çoban, A., Gürüler, H. ve Peker, M. (2016). Python Dilinin Öznitelikleri, Programlama Eğitiminde ve Yazılım Dünyasındaki Yeri.
- Çobanoğlu, B. (2020). *Herkes için Python*. Pusula.
- Köseoğlu, K. (2008). *Veritabanı mantığı*. Pusula.
- Şahin, M. (2007). Java, python ve ruby dillerinin performans karşılaştırması. *Akademik Bilişim*, 529-532.