

Büyük Veri İşlemesi Final Projesi

Amaç

1. Kullanıcı checkin sayılarının bulunması (1234 id'li kullanıcı toplam 1500 checkin yapmış) - 30p
2. Yer checkin sayılarının bulunması (1234 id'li yerde toplam 350 checkin yapılmış.) - 30p
3. Sonraki Kişi bilgisi oluşturulması (Bir kullanıcının checkin yaptığı yerlerde en çok checkin yapan kullanıcıların bulunması) - 40p

BONUS (NOT YÜKSELTMEDE KULLANILMAK ÜZERE - 30P)

1. Arkadaşlık ağ bilgisi kullanılarak Collaborative Filtering önerme yapma (bir kişiye arkadaşlarının checkin yaptığı yerlere göre yer önermesi) *ipucu: TF-IDF kullanılabilir.*

Araçlar

Tercih sırasıyla:

1. Apache Spark
2. MongoDB

ile geliştirme yapılabilir. İstenilen programlama dili kullanılabilir (seçilen platformun izin verdiği)

Veriseti

[Stanford SNAP Dataset](#) kullanılacaktır. Gerekli data ~105MB boyutlarında olup [bu](#) ve [bu](#) linkten indirilebilir.

Çıktılar

1. Kişisel olarak hazırlanmış rapor (İçerik: Ne yaptım, nasıl yaptım) -> 1 sayfa A4 Çıktı olacak, ad ve numara yazılmış olmalı
2. Top 100 checkin yapan kullanıcı sıralı listesi (Listedeki her doğru kullanıcı için 1p) -> Elektronik ortamda .txt olarak gönderilecek
3. Top 100 checkin yapılan yer sıralı listesi (Listedeki her doğru konum için 1p) -> Elektronik ortamda .txt olarak gönderilecek
4. Top 100 checkin yapılan yer için hazırlanacak sıralı liste (Listedeki her doğru konum için 1p) -> Elektronik ortamda .txt olarak gönderilecek

Bonus görevini yapan öğrenciler Bonus Görev Raporu gönderecektir (PDF olarak). Rapor, ilgili görevin nasıl yapıldığını ve örnek 3 user için önermeleri gösterecektir. Bonus raporun kodu da iletilecektir.

Önemli Not:

1 2-3 Kişilik gruplar halinde çalışılabilir, ancak her öğrenci raporu tekil olarak hazırlanacaktır.

2 Final Sınavı notu olarak değerlendirileceğinden, öğrencilerin raporlarını sekreterliğe teslim etmeleri ve **Final Sınavı Yoklama Listesine imza atmaları ÖNEMLİDİR** Aksi takdirde, DZ/FF olarak değerlendirilecektir.