

Veri Düzenleme Kontrol Listesi

Faaliyetlerinizi yürütme şeklinizin makine öğrenimi ve derin öğrenim gibi veri yoğunluklu teknolojilere dayandığı yapa zeka (YZ) çağına hoş geldiniz. Bu yeni YZ araçlarından yararlanabilmek için kuruluşunuzdaki veri «evinin» düzenli durumda bulunduğundan emin olmalısınız.

Bir temiz veri evine doğru uzanan yolda harekete geçmenizi sağlayacak bir kontrol listesi, düzenlemenin iki önemli safhasına, yani eğitim ve müdahale kısımlarına ayrılmış olarak burada sunuluyor.

Bir YZ ustası haline gelmenize yardımcı olacak bu adımları izleyin. YZ'yi bir kavram kanıtlama modelinden tüm üretim ve ölçeğe nasıl taşıyacağınız hakkında daha kapsamlı bir açıklama için bu IDC raporuna göz atın.

[YZ için Optimize Edilmiş Altyapı Kullanarak YZ Uygulamalarını Hızlandırın ve İşlevselleştirin.](#)

Eğitim

YZ'ye hazırlanmanın eğitim kısmında bir veri setini anlayacak algoritmalar geliştireceksiniz. Ana ilgi alanınız mevcut verileri toplamak ve yeni bir yeteneği öğrenmesi için YZ'yi kullanmak olacaktır.

- YZ'yi kullanarak çözmek isteyeceğiniz özel iş sorununu tespit edin (öğrenmenize yardımcı olacak daha küçük projelerle başlayın)
- İlgili kaynaklarda bu sorunu çözebilecek verilerin yerini tespit edin (Muhtemelen hepsi tek bir yerde bulunmayacaktır)
- Uygun verileri bulmak için gerekli süreyi önemli oranda azaltmak amacıyla verilerinizi üst veri etiketleriyle hazırlayın
- Verilerinizin düzgün şekilde eşitlendiğinden ve kullanacağınız tüm veri kümeleri genelinde bağlantılandırıldığından emin olun (zaman eşitleme dahil)
- Müşteriler açısından hassas olan her türlü veriyi ve diğer özel verileri, kesinlikle güvenli ve tüm ilgili idari ve yasal yönetmeliklere uygun şekilde tuttuğunuzdan emin olmanızı sağlayacak biçimde işaretleyin (bu konuda üst veri etiketleme işlemi yardımcı olabilir)
- Kullandığınız veri türü ve biçimlendirilme şekli için doğru geliştirme ortamını seçin (yani resimler, video, serbest biçimli metin ve ses verilerinin her biri tipik olarak tek türde ortama sahip olmalıdır)
- Veri havuzundan veri kümelerini çekin ve bunları kendi geliştirme ortamınıza aktarın
- Model geliştirme sürecinize yardımcı olması için verilerinizi iki gruba ayırın («eğitim» adlı bir klasörde bir set ve «test» adlı klasörde başka bir set tutun)
- Verilerinizin nereden/hangi kaynaktan geldiğini takip ederek verilerin izlenebilirliğini sağlayın (süreci otomatikleştirmeye yardım edecek araçlar kullanmayı düşünün)
- Verileri bir model oluşturmaya hazırlamak için temel veri hijyeni işlemlerini gerçekleştirin (örneğin eksik veri girişlerinin doldurulması ve geçersiz girişlerin silinmesi işlemlerine yer verin)
- Tahmin işlemi için cevabını zaten bildiğiniz bir veri altkümü örneği kullanın (buna «eğitim seti» denilmektedir) ve verileri bir tahmin yapmaya hazırlamak için gerekli tüm süreç öncesi adımları tespit edin
- Bu eğitim setinden elde ettiğiniz bilgileri tutarlılık skorlarını hesaplamakta kullanın; bu skorlar, aynı modeli daha önce hiç açıkça denenmediği yeni veriler için uygulamanız için size güven verebilecektir

Çıkarım

İş sorununuzu çözmek için işe yarayan bir model geliştirdiğinizde eğitimden müdahaleye geçeceksiniz. Bu safhada, başarılı olan modeli alıp, bir nebze sürekliliği veri düzenlemeyi gerektiren başka, yeni verilere uygularsınız.

- Gecikmeleri azaltmak, bant genişliği ihtiyaçlarını azaltmak ve genel model performansını iyileştirmek için YZ modelinizi verilerinizin yakınında konumlandırın
- Verimli bir veri hattı süreci geliştirin ve verileriniz geldikçe onlara üst veri etiketlerini uygulayın, böylece yeni veriler biriktirilerek, modeli ileriyeye doğru geliştirmek için kullanılabilir
- Verileri, birbiriyile bağlantılı ve zaman eşlemeli bir şekilde etiketleyin (örneğin veriler zamana göre sıralıysa veri kümeleri arasında zaman eşlemesi uygulayabilir ya da — bir müşterinin adı — gibi gelen tüm veriler arasında tek bir alanı seçerek bağlantılandırma yapabilirsiniz)
- Veri hacmini ve hızını, veriler geldikçe ve siz arşivledikçe nasıl yöneteceğiniz hakkında uzun vadeli bir veri yaşam döngüsü depolama planı geliştirin
- Gelecekteki YZ, derin öğrenim ve diğer veri güdümlü projeler için kuruluşunuzun veri yönetimini sürdürmek amacıyla bir Baş Veri Yetkilisi görevlendirmeyi düşünün