Bu metin, Amerika'daki işletmelerin doğru yetenekleri bulma ve çekme konusunda karşılaştıkları zorluklardan ve bu konuda nasıl bir çözüm arandığından bahsetmektedir.

Öncelikle, metin, Amerika'daki işletmelerin, rekabet gücünü koruyabilmek için sıkı çalışan, yetenekli ve nitelikli bireyleri bulma ihtiyaçlarına dikkat çekiyor. Bu yetenekler yerel olabileceği gibi, yabancı işçiler de olabilir.

Amerika'daki İmmigrasyon ve Vatandaşlık Yasası (INA), yabancı işçilerin Amerika'ya geçici ya da kalıcı olarak çalışmaya gelmelerine izin veriyor. Ancak aynı yasa, Amerikalı işçilerin maaşlarının veya çalışma koşullarının olumsiye etkilenmemesi için Amerikalı işverenlerin yasal gerekliliklere uymasını da sağlıyor.

Bu işlem, Dış İşçi Sertifikasyon Ofisi (OFLC) tarafından yönetiliyor. OFLC, yabancı işçileri Amerika'ya getirmek isteyen işverenlerin iş sertifikasyonu başvurularını işler ve işverenlerin, belirlenen ücretle işi yapacak yeterli Amerikalı işçi bulunmadığını kanıtlayabildikleri durumlarda sertifikalar verir.

Son yıllarda, bu başvuruların sayısında artış oldu ve her bir dosyanın incelenmesi zorlaştı. 2016 yılında, OFLC, geçici ve sürekli çalışma sertifikaları için 1.699.957 pozisyon için 775.979 işveren başvurusunu işlemiştir. Bu, bir önceki yıla göre toplam başvuru sayısında %9'luk bir artış anlamına gelmektedir.

Bu artan başvuru sayısı, Vize onay sürecini kolaylaştırabilecek ve vizesi onaylanma şansı yüksek adayları önseçebilecek bir Makine Öğrenmesi çözümünü gerektiriyor. OFLC, bu çözüm için sizin şirketinizi, TopVistos'u, görevlendirdi. Veri Bilimcisi olarak, sağlanan verileri analiz etmeli ve bir sınıflandırma modeli yardımıyla aşağıdaki iki şeyi yapmanız gerekiyor:

1. Vize onay sürecini kolaylaştırmak.
2. Vizenin sertifikalandırılması veya reddedilmesi gereken adaylar için uygun bir profil önermek. Bu profil, bir durumun durumunu önemli ölçüde etkileyen faktörlere dayanmalıdır.

Bu metin, bir veri setinin içeriği ve kullanımı hakkında bilgi veriyor. Veri seti, bir vize başvurusunda bulunan kişi ve başvurunun yapıldığı işveren hakkında farklı bilgileri içeriyor. Bu veri seti, bir makine öğrenmesi modeli eğitmek için kullanılacak. Modelin amacı, bir vize başvurusunun onaylanıp onaylanmayacağını tahmin etmek.

Aşağıda, veri setinde bulunan bilgilerin bir listesi bulunmaktadır:

* **id\_do\_caso**: Her vize başvurusunun ID'si.
* **continente**: Çalışanın kıtası.
* **educação\_do\_empregado**: Çalışanın eğitim bilgisi.
* **tem\_experiência\_de\_trabalho**: Çalışanın iş deneyimi olup olmadığı. S = Evet; N = Hayır.
* **requer\_treinamento\_de\_trabalho**: Çalışanın iş eğitimine ihtiyacı olup olmadığı. S = Evet; N = Hayır.
* **num\_de\_empregados**: İşverenin şirketindeki çalışan sayısı.
* **ano\_de\_estabelecimento**: İşverenin şirketinin kuruluş yılı.
* **região\_de\_emprego**: Yabancı işçinin Amerika'da çalışmak istediği bölgenin bilgisi.
* **salário\_prevalecente**: Yabancı işçinin çalışmayı planladığı bölgede benzer işlerde çalışan kişilere ödenen ortalama maaş. Prevalecent maaşın amacı, yabancı işçinin aynı hizmeti veren veya benzer hizmet veren diğer işçilere kıyasla daha az ücret almasını engellemektir.
* **unidade\_de\_salário**: Prevalecent maaşın birimi. Saatlik, Haftalık, Aylık ve Yıllık gibi değerler olabilir.
* **posição\_em\_tempo\_integral**: İş pozisyonunun tam zamanlı olup olmadığı. S = Tam Zamanlı Pozisyon; N = Yarı Zamanlı Pozisyon.
* **status\_do\_caso**: Vizenin onaylanıp onaylanmadığını gösteren gösterge.

Ayrıca, modelin doğru bir şekilde değerlendirilmesi için, verilerin **sample\_submission.csv** formatında gönderilmesi gerekiyor. **train.csv** ve **test.csv** dosyaları modeli eğitmek ve test etmek için kullanılacak.

Contexto:

As comunidades empresariais nos Estados Unidos enfrentam uma alta demanda por recursos humanos, mas um dos desafios constantes é identificar e atrair o talento certo, que é talvez o elemento mais importante para se manter competitivo. Empresas nos Estados Unidos procuram indivíduos trabalhadores, talentosos e qualificados tanto localmente quanto no exterior.

A Lei de Imigração e Nacionalidade (INA) dos EUA permite que trabalhadores estrangeiros venham trabalhar nos Estados Unidos temporária ou permanentemente. A lei também protege os trabalhadores americanos contra impactos adversos em seus salários ou condições de trabalho, garantindo que os empregadores americanos cumpram os requisitos legais ao contratar trabalhadores estrangeiros para suprir a escassez de mão de obra. Os programas de imigração são administrados pelo Escritório de Certificação de Trabalho Estrangeiro (OFLC).

O OFLC processa pedidos de certificação de emprego para empregadores que buscam trazer trabalhadores estrangeiros para os Estados Unidos e concede certificações nos casos em que os empregadores podem demonstrar que não há trabalhadores americanos suficientes disponíveis para realizar o trabalho com salários que atendam ou excedam o salário pago para a ocupação na área de emprego pretendida.

Objetivo:

No ano fiscal de 2016, o OFLC processou 775.979 pedidos de empregadores para 1.699.957 posições de certificações de trabalho temporárias e permanentes. Isso representou um aumento de nove por cento no número total de pedidos processados em relação ao ano anterior. O processo de revisar cada caso está se tornando uma tarefa tediosa à medida que o número de candidatos aumenta a cada ano.

O aumento do número de candidatos a cada ano demanda uma solução baseada em Aprendizado de Máquina que possa ajudar na pré-seleção dos candidatos com maiores chances de aprovação de VISTO. O OFLC contratou sua empresa, TopVistos, para soluções baseadas em dados. Como Cientista de Dados, você deve analisar os dados fornecidos e, com a ajuda de um modelo de classificação:

Facilitar o processo de aprovação de vistos.

Recomendar um perfil adequado para os candidatos para os quais o visto deve ser certificado ou negado, com base nos fatores que influenciam significativamente o status do caso.

Dataset Description

Os dados contêm diferentes atributos do empregado e do empregador. O dicionário de dados detalhado é fornecido abaixo.

Temos alguns arquivos disponibilizados para o treino dos modelos. Note que para seu modelo ser corretamente avaliado, você precisa enviar os seus dados no formato do arquivo sample\_submission.csv.

Arquivos

train.csv - dataset de treino

test.csv - dataset de testes

sample\_submission.csv - um exemplo de como o arquivo enviado deve ser composto

Dicionário dos Dados

id\_do\_caso: ID de cada solicitação de visto

continente: Informação do continente do empregado

educação\_do\_empregado: Informação sobre a educação do empregado

tem\_experiência\_de\_trabalho: O empregado possui alguma experiência de trabalho? S = Sim; N = Não

requer\_treinamento\_de\_trabalho: O empregado requer algum treinamento de trabalho? S = Sim; N = Não

num\_de\_empregados: Número de funcionários na empresa do empregador

ano\_de\_estabelecimento: Ano em que a empresa do empregador foi estabelecida

região\_de\_emprego: Informação da região de emprego pretendida pelo trabalhador estrangeiro nos EUA.

salário\_prevalecente: Salário médio pago a trabalhadores em ocupações semelhantes na área de emprego pretendida. O objetivo do salário prevalecente é garantir que o trabalhador estrangeiro não seja subremunerado em comparação com outros trabalhadores que oferecem o mesmo serviço ou serviço similar na mesma área de emprego.

unidade\_de\_salário: Unidade do salário prevalecente. Os valores incluem Por Hora, Por Semana, Por Mês e Por Ano.

posição\_em\_tempo\_integral: A posição de trabalho é em tempo integral? S = Posição em Tempo Integral; N = Posição em Meio Período

status\_do\_caso: Indicador se o visto foi certificado ou negado