

T.C.
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ



MUONLARIN ÇEVRESİNDEKİ JETLERİN ENERJİ VE KÜTLE
HESAPLAMALARI İÇİN MAKİNE ÖĞRENİMİ YAKLAŞIMI

BİTİRME PROJESİ

BEYAZIT UYSAL

BOLU,2023

ÖZET

Bu bitirme projesi, CERN ROOT dosyasından muon ve jetlerin özelliklerini içeren verileri okuyarak fiziksel bir analiz gerçekleştirmeyi hedeflemektedir. Projenin ana amacı, veritabanındaki muonun çevresinde bulunan jetlerin toplam enerjisini ve çevresel kütlesini hesaplamaktır.

Proje, öncelikle uproot kütüphanesini kullanarak ROOT dosyasını açar ve gerekli dal adlarını kullanarak verileri okur. Ardından, muon ve jetlerin dört momentumunu hesaplayan ve çevresel kütleyi hesaplayan fonksiyonlar tanımlanır.

Veri seti, muonun özellikleri (PT, Eta, Phi) ve jetlerin özellikleri (PT, Eta, Phi, Kütlev) gibi özelliklerden oluşur. Veri seti, önceden tanımlanan fonksiyonlar aracılığıyla ön işleme adımlarından geçirilir ve çevresel kütleyi hesaplamak için kullanılan fonksiyonlarla birleştirilir.

Daha sonra, veri seti bağımsız değişkenler (muon özellikleri) ve hedef değişken (çevresel kütlev) olarak ayrılır. Veri seti, eğitim ve test setlerine bölünerek makine öğrenimi modellerinin performansını değerlendirmek için kullanılır.

Projede kullanılan modeller arasında Random Forest, Linear Regression ve SVR gibi farklı makine öğrenimi yöntemleri yer alır. Bu modellerin özellik önem derecelerini hesaplamak için permütasyon önem değerlendirmesi kullanılır ve elde edilen sonuçlar özellik önem sıralamasına göre yazdırılır.

Son olarak, farklı makine öğrenimi modelleri (DummyRegressor, KNeighborsRegressor, LinearRegression, vb.) tanımlanır ve bu modellerin performansını ölçmek için farklı metrikler kullanılır. Ortalama mutlak hata, medyan mutlak hata, R-kare skoru ve açıklanan varyans skoru gibi metrikler değerlendirme sürecinde kullanılır. Elde edilen metrik sonuçları bir veri çerçevesinde toplanır ve yazdırılır.

ABSTRACT

This code demonstrates an example of a physical analysis used in a graduation project by reading data from a CERN ROOT file. The processed data contains properties of muons and jets. The main objective of the project is to calculate the total energy and invariant mass of jets surrounding a muon in the database.

The code first opens the ROOT file using the uproot library and reads the data using the required branch names. Then, it defines functions to calculate the four-momentum of muons and jets and to calculate the invariant mass.

The dataset consists of properties such as PT, Eta, and Phi for muons, and PT, Eta, Phi, and mass for jets. The dataset undergoes preprocessing steps and is combined with the functions used to calculate the invariant mass.

In the next step, the dataset is separated into independent variables (muon properties) and the target variable (invariant mass). Then, the dataset is split into training and test sets. Next, different machine learning models (Random Forest, Linear Regression, SVR) are defined, and permutation importance evaluation is used to calculate the feature importance ranks. The results obtained are printed based on the feature importance rankings.

Finally, various machine learning models (DummyRegressor, KNeighborsRegressor, LinearRegression, etc.) are defined, and they are evaluated using different performance metrics (mean absolute error, median absolute error, R-squared score, explained variance score). The metric results obtained are collected in a DataFrame and printed.

İÇİNDEKİLER

Özet	1
Abstract	2
İçindekiler	3
Giriş	4
Ana Bölüm Başlıkları	5
Sonuç ve Öneriler	9
Kaynaklar	11

GİRİŞ

Parçacık fiziği, temel parçacıkların doğasını, etkileşimlerini ve evrenin yapı taşlarını anlamamızı sağlayan heyecan verici bir bilim dalıdır. Büyük Hadron Çarpıştırıcısı (LHC) gibi yüksek enerjili parçacık hızlandırıcıları ve dedektörleri kullanılarak elde edilen veriler, parçacık fiziği araştırmalarında büyük bir rol oynamaktadır.

Bu bitirme projesi, LHC tarafından üretilen verilere dayanan bir fiziksel analiz örneğini sunmaktadır. Bu analiz, CERN ROOT dosyaları aracılığıyla muon ve jetlerin özelliklerini içeren verileri kullanarak gerçekleştirilmektedir. Müonlar, yüksek enerjili parçacıkların bir türüdür, ve jetler ise yüksek enerjili parçacıkların hızlandırıcıdaki çarpışmalar sonucunda oluşan gruplamaları temsil eder.

Bu projenin temel hedefi, veri tabanındaki muonun çevresinde bulunan jetlerin toplam enerjisini ve çevresel kütlelerini hesaplamaktır. Bu hesaplamalar, parçacıklar arasındaki etkileşimlerin ve enerji transferinin anlaşılmasına katkıda bulunur. Ayrıca, bu tür analizler, parçacık fiziği deneylerinin sonuçlarının doğruluğunu doğrulamak için önemlidir.

Bu proje, makine öğrenimi tekniklerinin parçacık fiziği verilerinin analizindeki rolünü göstermektedir. Veri seti, önceden belirlenmiş özelliklerden oluşur ve makine öğrenimi modelleri kullanılarak bu özelliklerin önemi değerlendirilir. Ayrıca, farklı makine öğrenimi modelleri kullanılarak, veri setinin özellikleriyle çevresel kütle tahmin edilmesi için bir regresyon analizi gerçekleştirilir.

Bu proje, parçacık fiziği araştırmalarında veri analizinin önemini ve makine öğrenimi tekniklerinin bu alandaki potansiyelini vurgulamaktadır. Elde edilen sonuçlar, parçacık fiziği topluluğuna daha iyi bir anlayış sunarak gelecekteki araştırmalara yol gösterebilir.

ANA BÖLÜM BAŞLIKLARI

I. Veri Ön İşleme ve Hazırlık

- A. ROOT Dosyalarından Veri Okuma
- B. Veri Temizleme ve Düzenleme
- C. Özelliklerin Tanımlanması ve Ölçeklendirme

II. Çevresel Kütlenin Hesaplanması

- A. Müonun Dört Momentumunun Hesaplanması
- B. Jetlerin Dört Momentumunun Hesaplanması
- C. Çevresel Kütle Hesaplama Algoritması

III. Makine Öğrenimi Modellerinin Uygulanması

- A. Veri Setinin Bölünmesi: Eğitim ve Test Setleri
- B. Model Seçimi ve Tanımlama
- C. Model Eğitimi ve Performans Değerlendirmesi

IV. Özellik Önem Analizi

- A. Permutasyon Önem Değerlendirmesi
- B. Özellik Önem Sıralamasının Belirlenmesi

V. Sonuçlar ve Tartışma

- A. Makine Öğrenimi Modellerinin Performansı
- B. Özellik Önem Analizinin Sonuçları
- C. Bulguların Yorumlanması ve Tartışılması

I. Veri Ön İşleme ve Hazırlık

A. ROOT Dosyalarından Veri Okuma

Veri analizi sürecine başlamadan önce, kullanılacak ROOT dosyalarını açmamız gerekiyor. Bu adımda uproot kütüphanesini kullanarak ROOT dosyasını açıyoruz ve içerisindeki ağaca erişiyoruz.

B. Veri Temizleme ve Düzenleme

Veri temizleme ve düzenleme adımında, DataFrame'imizi oluşturuyoruz ve gereksiz veya eksik verileri temizliyoruz. Eğer bir muonun transvers momentumu (PT) sıfır ise veya eksik ise, bu muonu veri setinden çıkarıyoruz.

C. Özelliklerin Tanımlanması ve Ölçeklendirme

Bu adımda, kullanacağımız özellikleri (PT, Eta, Phi) belirliyoruz ve DataFrame'e ekliyoruz. Ardından, verileri uygun bir şekilde ölçeklendiriyoruz, böylece özelliklerin farklı ölçeklere sahip olmasından kaynaklanan yan etkileri ortadan kaldırıyoruz.

II. Çevresel Kütlenin Hesaplanması

A. Müonun Dört Momentumunun Hesaplanması

Bu adımda, veri setindeki muonların dört momentumunu hesaplıyoruz. Muonun PT, Eta ve Phi değerlerinden yola çıkarak, enerji bileşenlerini hesaplıyoruz ve DataFrame'e enerjiyi ekliyoruz.

B. Jetlerin Dört Momentumunun Hesaplanması

Veri setindeki jetlerin dört momentumunu hesaplıyoruz. Jetlerin PT, Eta ve Phi değerlerini kullanarak enerji bileşenlerini hesaplıyoruz ve toplam dört momentumu elde ediyoruz.

C. Çevresel Kütle Hesaplama Algoritması

Bu adımda, çevresel kütle hesaplamak için gerekli olan adımları gerçekleştiriyoruz. Muona yakın olan jetleri belirliyoruz, toplam jet dört momentumunu hesaplıyoruz ve eksik enerjiyi (Missing ET) hesaplıyoruz. Ardından, çevresel kütleyi hesaplamak için bu değerleri kullanıyoruz. Elde ettiğimiz çevresel kütleyi, muonun enerjisinden çıkararak nihai sonucu elde ediyoruz.

III. Makine Öğrenimi Modellerinin Uygulanması

A. Veri Setinin Bölünmesi: Eğitim ve Test Setleri

Veri setimizi eğitim ve test setlerine bölmek için `train_test_split` fonksiyonunu kullanıyoruz. Böylece, makine öğrenimi modellerimizi eğitirken verilerin doğruluklarını değerlendirebileceğiz.

B. Model Seçimi ve Tanımlama

Çeşitli makine öğrenimi modellerini tanımlıyoruz ve bu modeller arasından kullanmak istediğimiz modelleri seçiyoruz. Örneğin, `RandomForestRegressor`, `LinearRegression`, `SVR` vb.

C. Model Eğitimi ve Performans Değerlendirmesi

Seçtiğimiz modelleri eğitiyoruz ve performanslarını değerlendiriyoruz. Eğitim için eğitim setini kullanıyoruz ve performansı değerlendirmek için test setini kullanıyoruz. Değerlendirme için çeşitli metrikler kullanabiliriz, örneğin ortalama mutlak hata (`mean_absolute_error`), medyan mutlak hata (`median_absolute_error`), R-kare (`r2_score`) vb.

IV. Özellik Önem Analizi

A. Müonun Dört Momentumunun Hesaplanması

Modelin performansını etkileyen özelliklerin önemini değerlendirmek için permutasyon önem değerlendirmesi yöntemini kullanıyoruz. Bu yöntem, her bir özelliği sırasıyla karıştırarak modelin performansındaki değişimi ölçer.

B. Jetlerin Dört Momentumunun Hesaplanması

Permutasyon önem değerlendirmesi sonucunda elde ettiğimiz değerleri kullanarak özellik önem sıralamasını belirliyoruz. Özelliklerin önem sıralaması, modelin performansına olan katkılarına göre belirlenir.

I. Sonular ve Tartıřma

A. Makine ğrenimi Modellerinin Performansı

Her bir makine ğrenimi modelinin performansını deęerlendiriyoruz. Elde ettięimiz sonuları karřılařtırıyor ve hangi modelin daha iyi bir performans gsterdięini belirliyoruz.

B. zellik nem Analizinin Sonuları

zellik nem analizi sonularını deęerlendiriyoruz. Hangi zelliklerin modelin performansında daha nemli olduęunu belirliyoruz.

C. Bulguların Yorumlanması ve Tartıřılması

Elde ettięimiz sonuları yorumluyoruz ve bulgularımızı tartıřıyoruz. Hangi zelliklerin ve modellerin daha etkili olduęunu analiz ediyoruz.

SONUÇ VE ÖNERİLER

Bu projede, farklı makine öğrenimi modelleri kullanarak bir parçacık fiziği veri kümesi üzerinde tahminlerde bulunduk ve özellik önem analizini gerçekleştirdik. Elde ettiğimiz sonuçlar şu şekildedir:

1. Makine Öğrenimi Modellerinin Performansı:

- En düşük ortalama mutlak hata (mean_absolute_error) değeri CatBoostRegressor modelinde elde edildi, bu da modelin gerçek değerlerle tahminleri arasındaki ortalama mutlak farkın en düşük olduğunu gösterir.
- R-kare (r2_score) değerine göre en yüksek performansı XGBRegressor modeli gösterdi. Bu, modelin bağımsız değişkenlerin bağımlı değişken üzerindeki varyansının büyük bir kısmını açıkladığını gösterir.

1. Özellik Önem Analizi:

- Tüm modellerde Muon/Muon.PT özelliği diğer özelliklere göre daha yüksek bir öneme sahiptir. Bu, parçacığın momentumunun tahmininde bu özelliğin en belirleyici olduğunu gösterir.
- Muon/Muon.Eta özelliği Linear Regression ve Support Vector Regression modellerinde de belirli bir öneme sahiptir. Ancak Random Forest modelinde daha düşük bir önem değerine sahiptir.
- Muon/Muon.Phi özelliği tüm modellerde diğer özelliklere göre daha düşük bir öneme sahiptir. Bu, parçacığın açısal bilgisinin tahmininde diğer özelliklere göre daha az belirleyici olduğunu gösterir.

Sonuç olarak, CatBoostRegressor modeli en düşük hata değerlerini elde ederken, XGBRegressor modeli en yüksek varyansı açıklama yeteneğine sahiptir. Muon/Muon.PT özelliği tüm modellerde en önemli özellik olarak belirlenmiştir.

SONUÇ VE ÖNERİLER

Öneriler:

- Daha fazla veri toplanabilir veya mevcut veri seti genişletilebilir. Bu, modellerin daha güvenilir ve genelleştirilebilir sonuçlar üretmesine yardımcı olabilir.
- Farklı makine öğrenimi modelleri kullanılarak daha kapsamlı bir karşılaştırma yapılabilir. Örneğin, daha karmaşık derin öğrenme modelleri veya diğer ensemble yöntemleri denenebilir.
- Özellik mühendisliği teknikleri uygulanarak yeni özellikler oluşturulabilir ve modellerin performansı artırılabilir.
- Model eğitim sürecinde hiperparametre ayarlaması yapılabilir ve farklı parametre kombinasyonları denenerek daha iyi sonuçlar elde edilebilir.
- Diğer değerlendirme metrikleri kullanılabilir ve modellerin farklı açılardan performansı değerlendirilebilir.

Bu önerilerin dikkate alınması, parçacık fiziği veri kümesi üzerinde daha iyi tahminler elde etmek ve model performansını artırmak için faydalı olabilir.

KAYNAKLAR

Projenin Kodlarının Bulunduğu Link. URL: <https://github.com/masanbasa3k/ML-FOR-ENVIRONMENT-MASS-CALCULATIONS-OF-MUONS>

CERN. (2023). Ana Sayfa. CERN Resmi İnternet Sitesi. URL: <https://root.cern/>

OpenAI. (2023). ChatGPT. OpenAI. URL: <https://openai.com/>