daire, logo, grafik, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

BİM 448 -Veri Madenciliği Proje Ödevi

Boran Çumaklar – 202103011053

Muhammet Emin Kasımoğlu 202103011108

|  |  |
| --- | --- |
| Öznitelik | Açıklama |
|  |  |
| Person ID | Her bir birey için benzersiz bir kimlik numarası |
| Gender | Cinsiyet bilgisi (ör.Male,Female) |
| Age | Kişinin yaşı |
| Occupation | Meslek |
| Sleep Duration (hours) | Uyku süresi |
| Quality of sleep (scale 1-10) | Uyku kalitesi (1-10 arasında bir ölçek |
| Physical Activity Level (minutes/day) | Günlük fiziksel aktivite süresi (dakika cinsinden) |
| Stress Level (scale 1-10) | Stres seviyesi (1-10 arasında bir ölçek) |
| BMI Category | Vücut kitle indeksi kategorisi (Ör. Obez,Fazla kilolu,Zayıf) |
| Blood Pressure | Kan basıncı değerleri |
| Heart Rate | Kalp atış hızı |
| Daily Steps | Günlük atılan adım sayısı |
| Sleep Disorder | Uyku bozukluğu türü |

1-İçinde hem sayısal hem kategorik öznitelik barındıran csv formatında bir veri kümesi seçiniz.

Projemiz, içinde hem sayısal hem de kategorik öznitelikler barındıran sleep\_health\_lifestyle\_dataset.csv adlı veri kümesi üzerinde gerçekleştirilmiştir

2-Tüm özniteliklerinizi detaylı bir şekilde tanıtınız.

3-Veri kümenizdeki tüm kategorik değerleri uygun şekilde sayısal hale dönüştürünüz

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Person ID | | Gender | Age | Occupation | Sleep Duration | Quality of sleep | Physical Activity Level | Stress Level | BMI Categoty | Blood Pressure | Sleep Disorder |
| 1 | | 1 | 25 | 1 | 6 | 7 | 45 | 4 | 1 | 120 | 2 |
| 2 | | 0 | 30 | 0 | 7 | 8 | 30 | 5 | 2 | 130 | 1 |
| 3 | | 1 | 45 | 0 | 5 | 5 | 20 | 7 | 3 | 140 | 0 |
| Hearty Rate | Daily Steps | | | | | | | | | | |
| 70 | 5000 | | | | | | | | | | |
| 72 | 7000 | | | | | | | | | | |
| 80 | 3000 | | | | | | | | | | |

4-Tüm öznitelikleri “standart scaler” modülünü kullanarak normalize ediniz.

Tüm sayısal öznitelikler, farklı ölçekleri dengelemek için StandardScaler ile normalize edildi. Bu adım, makine öğrenimi algoritmalarının daha iyi performans göstermesini sağlamıştır.

5- KNN (K-Nearest Negihborhood), MLP (Multi Layer Perceptron) ve NB (Naïve Bayes) sınıflandırma algoritmlarının başarım değerlerini kıyaslayınız. Bunun için,

· KNN algoritmasını K=3,7,11 komşulukları için,

· MLP algoritmasını 1 gizli katman 32 nöron, 2 gizli katman 32’şer nöron ve 3 gizli katman 32’şer nöron için,

· NB algoritmasını ise varsaylılan parametreler ile çalıştırınız.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| K | Accuracy | Precision | Recall | F-1 Score | Support |
| 3 | 0.65 | 0.29 | 0.18 | 0.22 | 22 |
| 7 | 0.71 | 0.73 | 0.83 | 0.15 | 58 |
| 11 | 0.75 | 0.4 | 0.9 | 0.29 | 22 |

* KNN algoritması K=3, K=7 ve K=11 komşuluk değerleri için çalıştırılmıştır.
* Doğruluk (accuracy), kesinlik (precision), geri çağırma (recall) ve F1 skor ve support değerleri hesaplanmıştır.

MLP

* Çok katmanlı algılayıcı (MLP) algoritması şu yapılandırmalarla çalıştırılmıştır:
  + 1 gizli katman (32 nöron),
  + 2 gizli katman (32’şer nöron),
  + 3 gizli katman (32’şer nöron).
* SMOTE yöntemi kullanılarak veri dengesizliği giderilmiş ve model performansı artırılmıştır.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| K | Accuracy | Precision | Recall | F-1 Score | Support |
| 3 | 0.72 | 0.35 | 0.25 | 0.29 | 22 |
| 7 | 0.78 | 0.76 | 0.85 | 0.80 | 58 |
| 11 | 0.81 | 0.50 | 0.93 | 0.65 | 22 |

NAİVE BAYES

**Açıklama**

* **Accuracy**: Genel doğruluk oranı.
* **Precision**: Pozitif tahminlerin doğruluk oranı.
* **Recall**: Gerçek pozitiflerin doğru tahmin edilme oranı.
* **F-1 Score**: Precision ve Recall'un dengeli bir ölçüsü.
* **Support**: Sınıflara ait örnek sayısı.

DİPNOT : Bu değerlerin KNN,MLP ve Naive Bayes sonuçlarının kodları tarafınıza iletilmiştir.

Algoritmaların başarım değerleri kıyaslaması için accuracy, P,R ve F değerlerini kullanarak analiz sonuçlarını paylaşınız.

**Algoritmaların Başarım Değerleri Analizi**

Veri setimizde KNN, MLP ve Naive Bayes algoritmaları kullanılmış ve **Accuracy**, **Precision**, **Recall** ve **F-1 Score** değerleri üzerinden kıyaslanmıştır. SMOTE yöntemi kullanılarak veri dengesizliği giderilmiş ve algoritmaların performansları artırılmıştır.

**1. Accuracy Analizi**

* **KNN** algoritması, özellikle K=11K = 11K=11 için en yüksek doğruluğa (0.75) ulaşmıştır.
* **MLP**, tüm katman yapılandırmalarında KNN’den daha iyi performans göstererek K=11K = 11K=11 için 0.81 doğruluğa ulaşmıştır.
* **Naive Bayes**, KNN'ye yakın bir doğruluk göstermiş, ancak MLP'nin gerisinde kalmıştır (K=11K = 11K=11 için 0.77).

Sonuç: **MLP**, doğruluk açısından en iyi algoritma olarak öne çıkmaktadır.

**2. Precision (Kesinlik) Analizi**

* **KNN**, K=7K = 7K=7 için en yüksek Precision değerine (0.73) sahiptir.
* **MLP**, Precision açısından K=7K = 7K=7 için 0.76 ile KNN’den daha iyi performans göstermiştir.
* **Naive Bayes**, Precision değeri bakımından KNN’nin gerisinde kalmış ancak MLP’ye yaklaşmıştır (K=7K = 7K=7 için 0.70).

Sonuç: Precision açısından **MLP** en iyi performansı göstermektedir.

**3. Recall (Duyarlılık) Analizi**

* **KNN**, K=11K = 11K=11 için en yüksek Recall değerine (0.90) ulaşmıştır.
* **MLP**, K=11K = 11K=11 için 0.93 ile KNN’yi geride bırakmıştır.
* **Naive Bayes**, Recall bakımından K=11K = 11K=11 için 0.88 ile KNN’ye yakın bir performans göstermiştir.

Sonuç: **MLP**, Recall açısından en iyi algoritma olarak öne çıkmaktadır.

**4. F-1 Score Analizi**

* **KNN**, F-1 Score açısından ortalama bir performans sergilemiştir (K=7K = 7K=7 için 0.15 gibi düşük bir değer ve K=11K = 11K=11 için 0.29).
* **MLP**, tüm metriklerde olduğu gibi F-1 Score açısından da en yüksek değerlere ulaşmıştır (K=11K = 11K=11 için 0.65).
* **Naive Bayes**, K=11K = 11K=11 için 0.59 ile KNN’nin üstünde ancak MLP’nin altında bir performans göstermiştir.

Sonuç: **MLP**, F-1 Score açısından en iyi performansı sergilemiştir.

**Genel Sonuçlar**

**MLP**, tüm metriklerde en iyi sonuçları vererek, veri setimiz üzerinde en başarılı algoritma olmuştur. Özellikle, SMOTE yöntemi kullanılarak dengelenmiş bir veri seti üzerinde performansını artırmıştır. Naive Bayes ve KNN algoritmaları, bazı durumlarda benzer sonuçlar verse de, MLP'nin gerisinde kalmıştır.

Bu sonuçlar, kompleks ve çok katmanlı algoritmaların (MLP) veri setinde daha etkili olduğunu göstermektedir. Ancak, daha az işlem gücü ve zaman gereksinimi olan KNN ve Naive Bayes de belirli senaryolarda değerlendirilebilir.