**Projekt – Metody walidacji modeli statystycznych**

**Cel badania:**

Celem badania jest utworzenie optymalnego pod względem miar statystycznych modelu do klasyfikacji pacjentów przy wykorzystaniu wybranych algorytmów uczenia maszynowego. Model służyć będzie do określenia na podstawie zebranych predyktorów, **czy u danej osoby występuje zespół lęku napadowego (ang. panic disorder).**

**Opis zbioru badawczego:**

Zbiór danych ,,**Panic Disorder Detection Dataset**’’ pochodzi ze strony **Kaggle** (link: [Panic Disorder Detection Dataset](https://www.kaggle.com/datasets/muhammadshahidazeem/panic-disorder-detection-dataset/data)). Zawiera on 100 000 rekordów pacjentów w zbiorze zatytułowanym jako treningowy oraz 20 000 rekordów w zbiorze zatytułowanym jako testowy. W zbiorze znajduje się 17 zmiennych, z czego ostatnia dotyczy wystąpienia u pacjenta zespołu lęku napadowego. Dane zostały zebrane przez Muhammada Shahida Azeema, wykładowcę informatyki (ang. Computer Science) z Departamentu Szkolnictwa Wyższego Pendżab w Pakistanie (ang. Higher Education Department Punjab Pakistan), na podstawie analizy dokumentów z kliniki Ayeshy (ang. Ayesha Clinic).

**Informacje dotyczące zmiennych w wykorzystywanym zbiorze danych:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lp. | Nazwa zmiennej | Opis zmiennej | Typ zmiennej | Szczegóły |
| 1 | Participant.ID | identyfikator pacjenta | zmienna jakościowa, nominalna | Zmienna zawierająca unikatowy identyfikator każdego z pacjentów - od 1 do 100 000 w zbiorze treningowym oraz od 1 do 20 000 w zbiorze testowym. |
| 2 | Age | wiek pacjenta | zmienna jakościowa, przedziałowa | Przyjmuje 48 poziomów:  pełne lata od 18 lat do 65 lat |
| 3 | Gender | płeć pacjenta | zmienna jakościowa, nominalna | Przyjmuje 2 poziomy:   1. Female (kobieta) 2. Male (mężczyzna) |
| 4 | Family.History | informacja nt. tego, czy w przeszłości w rodzinie pacjenta obserwowane były zdarzenia lub czynniki wpływające na możliwość pojawienia się zespołu lęku napadowego (np. występowanie zespołu lęku napadowego u członka rodziny) | zmienna jakościowa, binarna | Przyjmuje 2 poziomy:   1. True (czynniki rodzinne występują) 2. False (czynniki rodzinne nie występują) |
| 5 | Personal.History | informacja nt. tego, czy w przeszłości u pacjenta obserwowane były zdarzenia lub czynniki wpływające na możliwość pojawienia się zespołu lęku napadowego | zmienna jakościowa, binarna | Przyjmuje 2 poziomy:   1. True (czynniki osobiste występują) 2. False (czynniki osobiste nie występują) |
| 6 | Current.Stressors | natężenie czynników stresogennych obecnie występujących u pacjenta | zmienna jakościowa, nominalna | Przyjmuje 3 poziomy:   1. High (wysokie) 2. Moderate (średnie) 3. Low (niskie) |
| 7 | Symptoms | symptomy występujące u pacjenta | zmienna jakościowa, nominalna | Przyjmuje 5 poziomów:   1. Chest pain (ból klatki piersiowej) 2. Dizziness (zawroty głowy) 3. Fear of losing control (strach przed utratą kontroli) 4. Panic attacks (ataki paniki) 5. Shortness of breath (duszności) |
| 8 | Severity | natężenie symptomów występujących u pacjenta | zmienna jakościowa, nominalna | Przyjmuje 3 poziomy:   1. Mild (łagodne) 2. Moderate (umiarkowane) 3. Severe (poważne) |
| 9 | Impact.on.Life | wpływ obserwowanych symptomów na codzienne życie pacjenta | zmienna jakościowa, nominalna | Przyjmuje 3 poziomy:   1. Mild (łagodny) 2. Moderate (umiarkowany) 3. Significant (znaczący) |
| 10 | Demographics | informacja demograficzna na temat miejsca zamieszkania pacjenta | zmienna jakościowa, nominalna | Przyjmuje 2 poziomy:   1. Rural (wiejske) 2. Urban (miejskie) |
| 11 | Medical.History | historia medyczna pacjenta, tj. inne choroby fizyczne występujące u pacjenta | zmienna jakościowa, nominalna | Przyjmuje 4 poziomy:   1. Asthma (astma) 2. Diabetes (cukrzyca) 3. Heart disease (choroba serca) 4. None (brak) |
| 12 | Psychiatric.History | historia psychiatryczna pacjenta, tj. inne choroby psychiczne występujące u pacjenta | zmienna jakościowa, nominalna | Przyjmuje 4 poziomy:   1. Anxiety disorder (zaburzenie lękowe) 2. Bipolar disorder (choroba afektywna dwubiegunowa) 3. Depressive disorder (zaburzenie depresyjne) 4. None (brak) |
| 13 | Substance.Use | rodzaj zażywanych przez pacjenta substancji psychoaktywnych | zmienna jakościowa, nominalna | Przyjmuje 3 poziomy:   1. Alcohol (alkohol) 2. Drugs (narkotyki) 3. None (brak) |
| 14 | Coping.Mechanisms | mechanizmy i strategie wykorzystywane przez pacjenta do radzenia sobie ze stresem lub objawami zespołu lęku napadowego | zmienna jakościowa, nominalna | Przyjmuje 4 poziomy:   1. Exercise (ćwiczenie fizyczne) 2. Meditation (medytacja) 3. Seeking therapy (poszukiwanie terapii, uczęszczanie na terapię) 4. Socializing (socjalizacja, kontakty towarzyskie) |
| 15 | Social.Support | natężenie i jakość wsparcia społecznego otrzymywanego przez pacjenta (np. od przyjaciół, rodziny, grup wsparcia) | zmienna jakościowa, nominalna | Przyjmuje 3 poziomy:   1. High (wysokie) 2. Moderate (umiarkowane) 3. Low (niskie) |
| 16 | Lifestyle.Factors | czynniki związane ze stylem życia pacjenta, które mogą wpływać na jego zdrowie psychiczne | zmienna jakościowa, nominalna | Przyjmuje 3 poziomy:   1. Diet (dieta) 2. Exercise (ćwiczenie fizyczne) 3. Sleep quality (jakość snu) |
| 17 | Panic.Disorder.Diagnosis | diagnoza pacjenta, tj. informacja o tym, czy u pacjenta występuje zespół lęku napadowego  **Zmienna zależna przeprowadzanego badania.** | zmienna jakościowa, binarna | Przyjmuje 2 wartości:   1. ,,0’’ – diagnoza negatywna (u pacjenta nie występuje zespół lęku napadowego) 2. ,,1’’ – diagnoza pozytywna (u pacjenta występuje zespół lęku napadowego) |