

## Projekt: Podsumowania lingwistyczne opisujące modele predykcyjne i dane akustyczne zebrane ze smartphone'ów osób cierpiących na chorobę afektywną dwubiegunową

Dla danych data\_MoodMon\_project.csv należy:

- (1) Przygotować funkcję „*generate\_descriptive\_linguistic\_summaries*”, która jako parametry wejściowe przyjmować będzie m.in. (a) ramkę z przygotowanymi danymi; (b) podmiot podsumowania; (c) kwalifikator; (d) **lista** nazw parametru dla którego generowane będą podsumowania; (e) kwantyfikator lingwistyczny; (f) t- norma  
Zdefiniować niezbędne zmienne lingwistyczne i zapisać wykresy do plików (png lub pdf).  
Dane wyjściowe: treść podsumowań, stopień prawdy

Przykład:

For **almost none/ few/ around half/ most/ almost every** patients with **low/medium/high** localjitter, energy\_mean is **low/medium/high**.

For **almost none/ few/ around half/ most/ almost every** patients in depression with **low/medium/high** localjitter, energy\_mean is **low/medium/high**.

- (2) Podzielić dane na zbiór treningowy oraz zbiór testowy (np. 0.8 zbiór treningowy, 0.2 zbiór testowy), przy niestandardowym wyborze zbiorów uargumentować wybór. Zbudować model predykcyjny klasyfikujący pacjentów w depresji i bez (podział na klasy 0 – pacjenci w stanie eutymii oraz klasa 1 – pacjencie w stanie depresji) bazując na zmiennych:

**tel\_usage\_times** – liczba użyć telefonu w trakcie danego dnia,  
**sleep\_length** - długość snu dla danego dnia,  
**brief\_pauses** – liczba krótkich pauz w rozmowach telefonicznych,  
**long\_pauses** – liczba długich pauz w rozmowach telefonicznych,  
**energy\_mean, energy\_std, f0\_slope\_mean, f0\_slope\_std, localjitter, localshimmer, pitch\_mean, pitch\_std** - wartości opisujące głos zebrane podczas rozmów telefonicznych,  
**state** – zmienna odpowiedzi opisująca stan pacjenta (0 eutymia, 1 depresja).

Na podstawie finalnego modelu wyliczyć wartości Shapleya dla zbioru testowego.

Utworzyć zestawienie z metryką F1 dla finalnego modelu, tabelę kontyngencji oraz beeswarm plot korzystając z wyliczonych wcześniej wartości Shapleya dla zbioru testowego.

- (3) Przygotować funkcję „*generate\_predictive\_linguistic\_summaries*”, która jako parametry wejściowe przyjmować będzie m.in. (a) ramkę z przygotowanymi danymi akustycznymi; (b) **model predykcyjny lub ramkę z wartościami Shapley’a**; (c) kwalifikator; (d) nazwę parametru dla którego generowane będą podsumowania; (e) kwantyfikator lingwistyczny; (f) t- norma  
Zdefiniować niezbędne zmienne lingwistyczne i zapisać wykresy do plików (png lub pdf). **W analizie rozważać predykcje dla tylko jednego stanu, np. choroby (depresja).**  
Dane wyjściowe: treść podsumowań, stopień prawdy

Przykład:

For **almost none/ few/ around half/ most/ almost every** records in test set **with low/medium/high** localjitter, contribution of energy\_mean is **strongly negative/around zero/ strongly positive**.

For **almost none/ few/ around half/ most/ almost every** records in test set in depression **with low/medium/high** localjitter, contribution of energy\_mean is **strongly negative/around zero/strongly positive**.

For **almost none/ few/ around half/ most/ almost every** records in test set in euthymia **with low/medium/high** localjitter, contribution of energy\_mean is **strongly negative/around zero/ strongly positive**.

- (4) Utworzyć wszystkie możliwe podsumowania lingwistyczne z zadań (1) i (3) i zapisać w pliku „all\_summaries.csv”.
- (5) Przygotować prezentację z najważniejszymi wynikami. Przedstawić w niej także rozwiązywany problem oraz przyjęte założenia.

### Zasady zaliczenia

Rezultaty z projektu	Punkty
Kod realizujący pkt.(1)	2 pkt
Kod realizujący pkt.(2) i pliki wynikowe	2 pkt
Kod realizujący pkt.(3)	2 pkt
Pliki wynikowe z podsumowaniami z pkt. (4) oraz pliki graficzne	2 pkt
Prezentacja z najważniejszymi wnioskami z przeprowadzonych analiz i obliczeń (10-minutowa)	2 pkt
<b>Razem</b>	<b>10 pkt</b>

Termin nadsyłania projektów: **09.06.2025 r. godz. 10:00.**

Projekt należy wysłać na adres: [k.kaczmarek@ibspan.waw.pl](mailto:k.kaczmarek@ibspan.waw.pl).

Termin prezentacji projektów (forma zdalna): 09.06.2025 r. o godzinie 10:00 – 11:30.

Link do spotkania:

[https://teams.microsoft.com/l/meetup-join/19%3ameeting\\_ZDNhOGIwOGUtNWM2Yi00MTc0LWE3YjctYTNkYWE2NTQ4OWE3%40thread.v2/0?context=%7b%22Tid%22%3a%22788ff3aa-65d0-4fbd-ad6e-2f93881731e4%22%2c%22Oid%22%3a%2285b46941-fe2c-470c-a6f3-2f0fdccdae0a%22%7d](https://teams.microsoft.com/l/meetup-join/19%3ameeting_ZDNhOGIwOGUtNWM2Yi00MTc0LWE3YjctYTNkYWE2NTQ4OWE3%40thread.v2/0?context=%7b%22Tid%22%3a%22788ff3aa-65d0-4fbd-ad6e-2f93881731e4%22%2c%22Oid%22%3a%2285b46941-fe2c-470c-a6f3-2f0fdccdae0a%22%7d)

