

Data oddania: _____

Ocena: _____

Mateusz Walczak 216911

Konrad Kajszczak 216790

Zadanie 2: Lingwistyczne podsumowania baz danych*

1. Cel

Praca w toku

2. Wprowadzenie

Praca w toku

3. Opis implementacji

Praca w toku

4. Materiały i metody

Wybrana przez nas baza danych zawiera historyczne pomiary pogodowe z Holandii [1]. Dane zostały zgromadzone przez KNMI (*Dutch weather institute* - Holenderski instytut pogody) na przestrzeni lat 1901-2018 i pochodziły z 50 różnych stacji pogowych znajdujących się na terenie całego kraju.

* SVN: <https://github.com/Walducha1908/KSR2>

Ze względu na fakt, iż oryginalna baza danych składa się z 804099 krotek, postanowiliśmy wybrać tylko niewielką część z dostępnych danych. Zdecydowaliśmy się na najnowsze dane pomiarowe - z lat 2016-2018. W ten sposób ograniczyliśmy liczbę wykorzystywanych krotek do 17000.

4.1. Wybór kolumn

W celu analizy bazy danych i tworzenia jej lingwistycznych podsumowań wybraliśmy następujące 10 kolumn z danymi liczbowymi:

- FG - średnia prędkość wiatru przez cały dzień [$0.1 \frac{m}{s}$].
- FHX - najwyższa średnia prędkość wiatru w ciągu jednej godziny [$0.1 \frac{m}{s}$].
- FHN - najniższa średnia prędkość wiatru w ciągu jednej godziny [$0.1 \frac{m}{s}$].
- FXX - najszybszy podmuch wiatru w ciągu całego dnia [$0.1 \frac{m}{s}$].
- TG - średnia dzienna temperatura [$0.1^{\circ}C$].
- TN - minimalna dzienna temperatura [$0.1^{\circ}C$].
- TX - maksymalna dzienna temperatura [$0.1^{\circ}C$].
- T10N - minimalna dzienna temperatura na wysokości 10 cm od poziomu gruntu [$0.1^{\circ}C$].
- Q - nasłonecznienie, energia słoneczna przypadająca na powierzchnię [$\frac{J}{cm^2}$].
- RH - suma opadów atmosferycznych w ciągu całego dnia [$0.1mm$].

Oprócz wyżej opisanych danych liczbowych, w naszej bazie znajdują się także dwie dodatkowe kolumny, służące do identyfikacji pomiaru:

- STN - numer stacji badawczej wykonującej pomiar.
- YYYYMMDD - data pomiaru w formacie opisanym przez nazwę kolumny.

	A	B	C	D	E	F	G	H	I	J	K	L
1	STN	YYYYMMDD	FG	FHX	FHN	FXX	TG	TN	TX	T10N	Q	RH
2	380	20181231	23	30	10	60	83	74	91	70	96	19
3	370	20181231	27	40	20	70	89	73	99	72	115	5
4	350	20181231	28	40	20	80	89	75	98	72	132	5
5	375	20181231	28	50	20	90	90	73	98	71	105	1
6	290	20181231	36	60	20	90	88	74	99	73	126	2
7	275	20181231	30	50	20	90	86	72	98	70	132	1
8	279	20181231	40	60	20	100	86	78	97	77	150	-1
9	260	20181231	27	40	20	100	89	77	100	76	137	1
10	269	20181231	36	50	20	80	87	78	100	75	155	-1
11	280	20181231	40	60	20	110	88	80	99	77	174	5
12	240	20181231	45	70	30	110	90	83	109	76	205	-1
13	344	20181231	35	50	20	90	91	79	104	77	207	-1
14	215	20181231	42	60	20	90	89	82	97	80	212	0
15	235	20181231	51	80	40	110	90	83	98	77	229	-1
16	270	20181231	50	80	30	120	87	82	97	78	236	-1
17	310	20181231	48	60	30	90	88	73	102	71	335	3
18	375	20181230	32	40	20	90	79	53	92	42	137	31
19	350	20181230	29	50	20	80	82	56	93	49	139	14
20	260	20181230	26	50	20	100	85	62	95	52	112	10
21	370	20181230	33	50	20	90	80	51	91	45	156	11
22	269	20181230	33	60	20	110	82	63	92	54	119	8
23	344	20181230	36	60	20	120	86	50	97	36	141	4
24	215	20181230	43	60	20	100	87	57	99	48	127	7
25	275	20181230	35	70	20	120	77	55	90	48	138	24
26	279	20181230	39	80	20	160	76	67	84	62	148	10

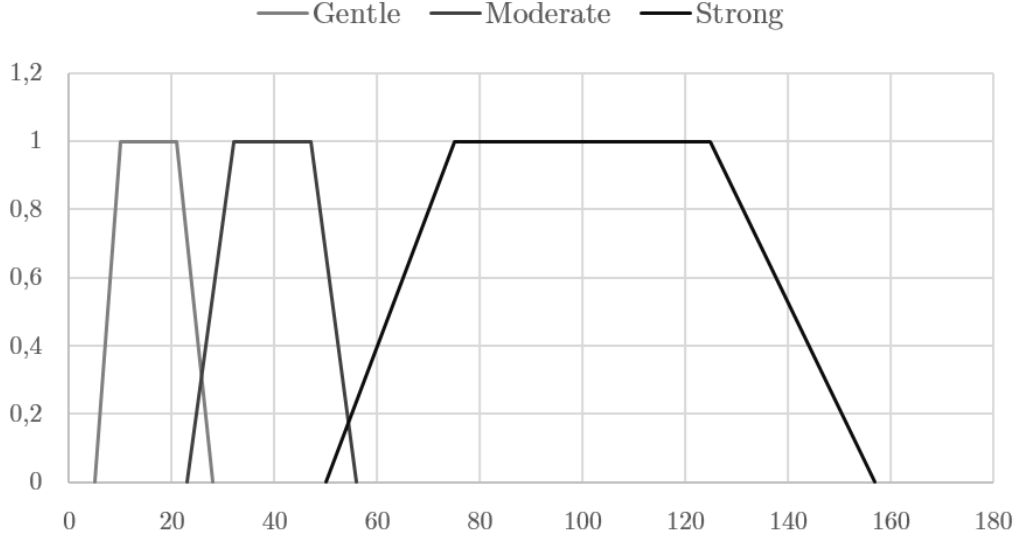
Rysunek 1. Fragment widoku bazy w formacie *xlsx*

4.2. Przykładowe zmienne lingwistyczne

W tym rozdziale przedstawimy wzory i wykresy opisujące zaproponowane przez nas zmienne lingwistyczne. We wszystkich przypadkach, wykorzystywanymi przez nas funkcjami przynależności są funkcje trapezoidalne.¹

4.2.1. Kolumna FG

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej wartości średniej prędkości wiatru przez cały dzień (FG), zamieszczono poniżej.



Rysunek 2. Wykres opisujący zmienną lingwistyczną dla kolumny FG.

Wzory opisujące przynależność do poszczególnych etykiet prezentują się następująco.

Dla etykiety *Gentle*:

$$FG_{GENTLE}(x) = \begin{cases} \frac{x-5}{5} & \text{jeśli } 5 \leq x < 10 \\ 1 & \text{jeśli } 10 \leq x \leq 21 \\ \frac{28-x}{7} & \text{jeśli } 21 < x \leq 28 \end{cases} \quad (1)$$

Dla etykiety *Moderate*:

$$FG_{MODERATE}(x) = \begin{cases} \frac{x-23}{9} & \text{jeśli } 23 \leq x < 32 \\ 1 & \text{jeśli } 32 \leq x \leq 47 \\ \frac{56-x}{9} & \text{jeśli } 47 < x \leq 56 \end{cases} \quad (2)$$

Dla etykiety *Strong*:

$$FG_{STRONG}(x) = \begin{cases} \frac{x-50}{25} & \text{jeśli } 50 \leq x < 75 \\ 1 & \text{jeśli } 75 \leq x \leq 125 \\ \frac{157-x}{34} & \text{jeśli } 125 < x \leq 157 \end{cases} \quad (3)$$

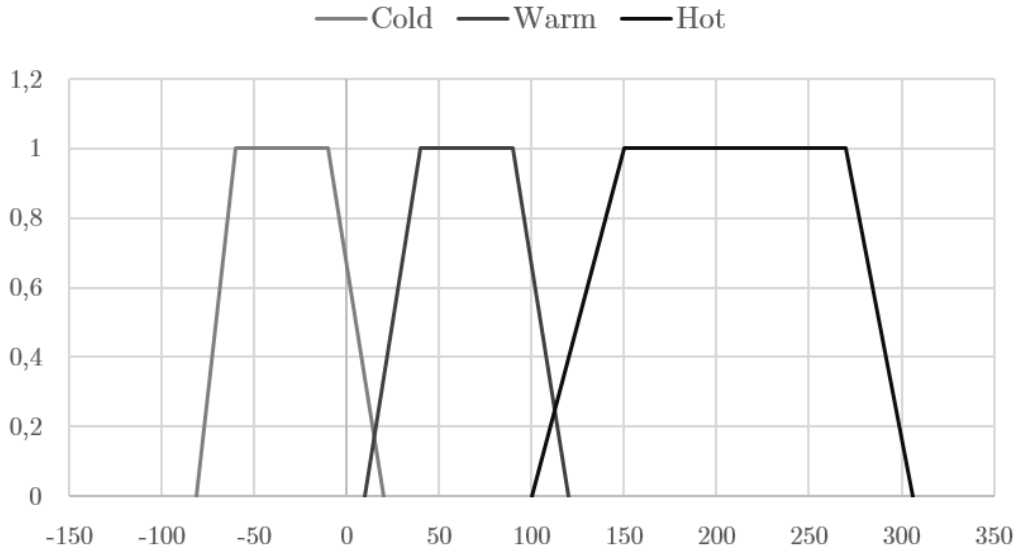
¹ Wartości prezentowane w tabelach są tylko propozycjami. Autorzy sprawozdania zastrzegają sobie możliwość do ich późniejszej modyfikacji

4.2.2. Kolumna TG

W przypadku średniej dziennej temperatury (TG), zdecydowaliśmy się podzielić nasze rozważania ze względu na pory roku. Dlatego też przyjęliśmy trzy różne warianty zmiennej lingwistycznej dla kolumny TG:

- TGW - dla pomiarów uzyskanych podczas astronomicznej zimy (litera *W* od *Winter*),
- TGSA - dla pomiarów uzyskanych podczas astronomicznej wiosny lub jesieni (*S* od *Spring*, *A* od *Autumn*),
- TGS - dla pomiarów uzyskanych podczas astronomicznego lata (litera *S* od *Summer*).

Rozpocznijmy od zmiennej lingwistycznej TGW.



Rysunek 3. Wykres opisujący zmienną lingwistyczną dla kolumny TG dla pomiarów wykonanych astronomiczną zimą.

Wzory opisujące przynależność do poszczególnych etykiet zmiennej lingwistycznej TGW prezentują się następująco.

Dla etykiety *Cold*:

$$TGW_{COLD}(x) = \begin{cases} \frac{x+81}{21} & \text{jeśli } -81 \leq x < -60 \\ 1 & \text{jeśli } -60 \leq x \leq -10 \\ \frac{20-x}{30} & \text{jeśli } -10 < x \leq 20 \end{cases} \quad (4)$$

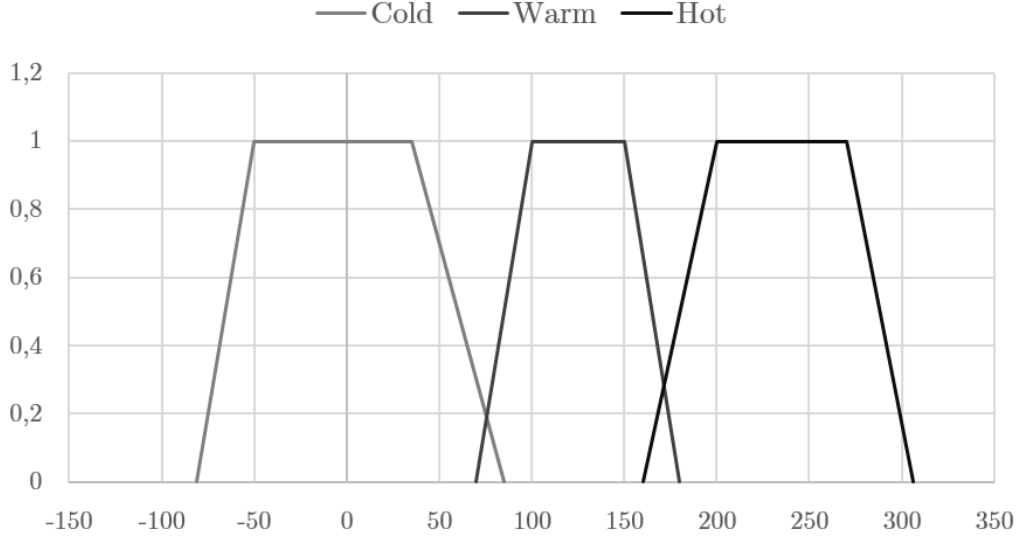
Dla etykiety *Warm*:

$$TGW_{WARM}(x) = \begin{cases} \frac{x-10}{30} & \text{jeśli } 10 \leq x < 40 \\ 1 & \text{jeśli } 40 \leq x \leq 90 \\ \frac{120-x}{30} & \text{jeśli } 90 < x \leq 120 \end{cases} \quad (5)$$

Dla etykiety *Hot*:

$$TGW_{HOT}(x) = \begin{cases} \frac{x-100}{50} & \text{jeśli } 100 \leq x < 150 \\ 1 & \text{jeśli } 150 \leq x \leq 270 \\ \frac{306-x}{36} & \text{jeśli } 270 < x \leq 306 \end{cases} \quad (6)$$

Następną prezentowaną zmienną, będzie zmienna lingwistyczna TGSA.



Rysunek 4. Wykres opisujący zmienną lingwistyczną dla kolumny TG dla pomiarów wykonanych astronomiczną wiosną i jesienią.

Wzory opisujące przynależność do poszczególnych etykiet zmiennej lingwistycznej TGSA prezentują się następująco.

Dla etykiety *Cold*:

$$TGSA_{COLD}(x) = \begin{cases} \frac{x+81}{31} & \text{jeśli } -81 \leq x < -50 \\ 1 & \text{jeśli } -50 \leq x \leq 35 \\ \frac{85-x}{50} & \text{jeśli } 35 < x \leq 85 \end{cases} \quad (7)$$

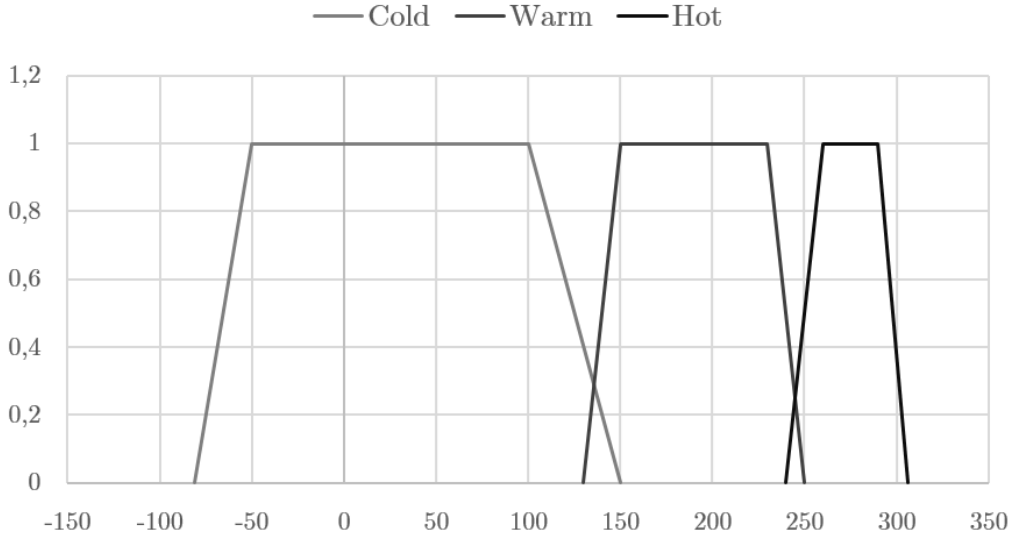
Dla etykiety *Warm*:

$$TGSA_{WARM}(x) = \begin{cases} \frac{x-70}{30} & \text{jeśli } 70 \leq x < 100 \\ 1 & \text{jeśli } 100 \leq x \leq 150 \\ \frac{180-x}{30} & \text{jeśli } 150 < x \leq 180 \end{cases} \quad (8)$$

Dla etykiety *Hot*:

$$TGSA_{HOT}(x) = \begin{cases} \frac{x-160}{40} & \text{jeśli } 160 \leq x < 200 \\ 1 & \text{jeśli } 200 \leq x \leq 270 \\ \frac{306-x}{36} & \text{jeśli } 270 < x \leq 306 \end{cases} \quad (9)$$

Ostatnią zmienną dla kolumny TG będzie zmienna dotycząca pomiarów letnich - TGS.



Rysunek 5. Wykres opisujący zmienną lingwistyczną dla kolumny TG dla pomiarów wykonanych astronomicznym latem.

Wzory opisujące przynależność do poszczególnych etykiet zmiennej lingwistycznej TGS prezentują się następująco.

Dla etykiety *Cold*:

$$TGS_{COLD}(x) = \begin{cases} \frac{x+81}{31} & \text{jeśli } -81 \leq x < -50 \\ 1 & \text{jeśli } -50 \leq x \leq 100 \\ \frac{150-x}{50} & \text{jeśli } 100 < x \leq 150 \end{cases} \quad (10)$$

Dla etykiety *Warm*:

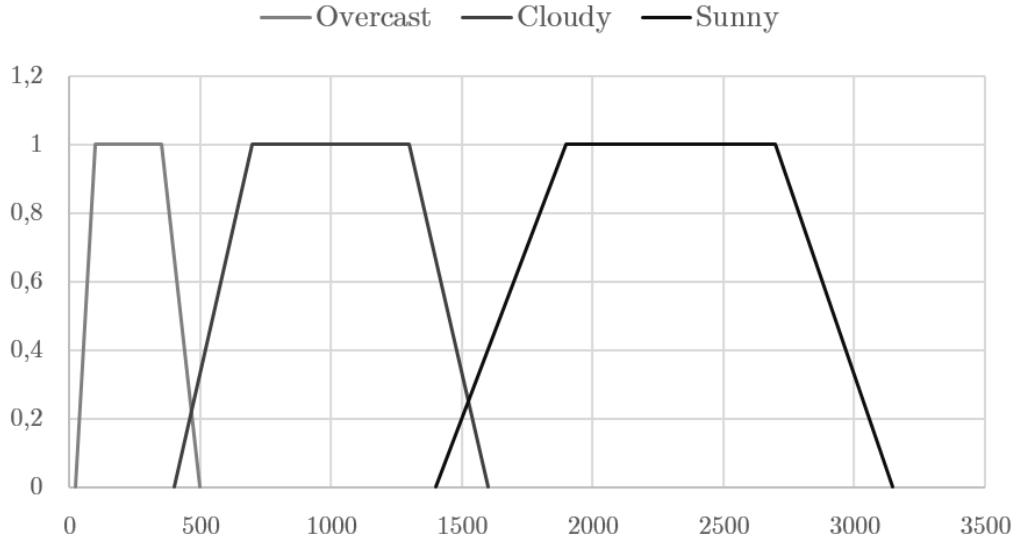
$$TGS_{WARM}(x) = \begin{cases} \frac{x-130}{20} & \text{jeśli } 130 \leq x < 150 \\ 1 & \text{jeśli } 150 \leq x \leq 230 \\ \frac{250-x}{20} & \text{jeśli } 230 < x \leq 250 \end{cases} \quad (11)$$

Dla etykiety *Hot*:

$$TGS_{HOT}(x) = \begin{cases} \frac{x-240}{20} & \text{jeśli } 240 \leq x < 260 \\ 1 & \text{jeśli } 260 \leq x \leq 290 \\ \frac{306-x}{16} & \text{jeśli } 290 < x \leq 306 \end{cases} \quad (12)$$

4.2.3. Kolumna Q

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej wartości nasłonecznienia (Q), zamieszczono poniżej.



Rysunek 6. Wykres opisujący zmienną lingwistyczną dla kolumny Q

Wzory opisujące przynależność do poszczególnych etykiet zmiennej lingwistycznej Q prezentują się następująco.

Dla etykiety *Overcast*:

$$Q_{OVERCAST}(x) = \begin{cases} \frac{x-24}{76} & \text{jeśli } 24 \leq x < 100 \\ 1 & \text{jeśli } 100 \leq x \leq 350 \\ \frac{500-x}{150} & \text{jeśli } 350 < x \leq 500 \end{cases} \quad (13)$$

Dla etykiety *Cloudy*:

$$Q_{CLOUDY}(x) = \begin{cases} \frac{x-400}{300} & \text{jeśli } 400 \leq x < 700 \\ 1 & \text{jeśli } 700 \leq x \leq 1300 \\ \frac{1600-x}{300} & \text{jeśli } 1300 < x \leq 1600 \end{cases} \quad (14)$$

Dla etykiety *Sunny*:

$$Q_{SUNNY}(x) = \begin{cases} \frac{x-1400}{500} & \text{jeśli } 1400 \leq x < 1900 \\ 1 & \text{jeśli } 1900 \leq x \leq 2700 \\ \frac{3145-x}{445} & \text{jeśli } 2700 < x \leq 3145 \end{cases} \quad (15)$$

5. Wyniki

Praca w toku

6. Dyskusja

Praca w toku

7. Wnioski

Praca w toku

Literatura

- [1] Baza danych - "*Historical weather in the Netherlands 1901-2018*"