

Data oddania: _____

Ocena: _____

Krzysztof Dwużnik 229870

Wojciech Majchrzak 229947

Projekt 2. Podsumowania lingwistyczne relacyjnych baz danych

1. Cel

Celem projektu jest zaprojektowanie i stworzenie aplikacji, której celem jest generowanie podsumowań lingwistycznych [5] postaci:

Q które są $W_1, W_2 \dots W_n$ jest $S_1, S_2 \dots S_n$.

gdzie: Q - kwantyfikator ilościowy względny lub absolutny, $W_1, W_2 \dots W_n$ - zbiór n kwalifikatorów lingwistycznych, $S_1, S_2 \dots S_N$ - zbiór n sumaryzatorów lingwistycznych.

Podsumowania będą generowane dla wybranej bazy opisującej domy w Australii (patrz rozdział 2.1) [1][2]. W zadaniu będziemy korzystać z 13 kolumn, z czego 11 będziemy rozmywać. Zadaniem aplikacji będzie, na podstawie wyborów użytkownika, przedstawienie wyników w postaci opisów, czyli inaczej podsumowań lingwistycznych. Tak wygenerowane podsumowania mogą się przydać przy poszukiwaniach domu w Australii.

Przykładowo jako wynik możemy otrzymać:

Prawie żaden z domów jest tani, przestrzenny, i nowy.

2. Baza danych, zmienne lingwistyczne, kwantyfikatory lingwistyczne

2.1. Charakterystyka podsumowywanej bazy danych

Do analizy danych przez naszą aplikację wybraliśmy bazę perth-house-prices [2] ze strony Kaggle [1]. Baza opiera się na danych mieszkań znajdujących się w Australii. Baza zawiera 19208 unikalnych rekordów oraz 14 kolumn z czego 12 to kolumny liczbowe, a 2 to kolumny tekstowe. Atrybuty które będziemy rozmywać to:

1. PRICE (Cena) - wartości z przedziału [52000-2440000]
2. LAND AREA (Powierzchnia działki) - wartości z przedziału [61-999999]
3. FLOOR AREA (Powierzchnia mieszkania) - wartości z przedziału [1-849]
4. BUILD YEAR (Rok wybudowania) - wartości z przedziału [1870-2017]
5. CBD DIST (Odległość do centrum miasta) - wartości z przedziału [1300-56900]
6. NEAREST STN DIST (Odległość do najbliższej stacji) - wartości z przedziału [46-34300]
7. LAST SOLD TIME (Czas od ostatniej sprzedaży [dni]) - wartości z przedziału [528-12369]
8. LATITUDE (Wysokość geograficzna) - wartości z przedziału [od -32.455550 do -31.595910]
9. LONGITUDE (Szerokość geograficzna) - wartości z przedziału [od 115.675834 do 116.341780]
10. NEAREST SCH DIST (Odległość do najbliższej szkoły) wartości z przedziału [0.070912 do 20.720911]
11. NEAREST SCH RANK (Ranga najbliższej szkoły) - wartości z przedziału [1-139]

Pozostałe wartości to wartości opisowe jednoznacznie identyfikujące dany dom:

1. ADDRESS (adres domu)
2. SUBURB (nazwa dzielnicy w jakiej dany dom się znajduje)

Dodatkowo znajduje się też kolumna BEDROOMS określająca ilość sypialni w domu. Posłuży ona nam do podziału na podmioty.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ADDRESS	SUBURB	PRICE	LAND_AREA	FLOOR_AREA	BUILD_YEAR	CBD_DIST	NEAREST_STN_DIST	LATITUDE	LONGITUDE	NEAREST_SCH_DIST	NEAREST_SCH_RANK	LAST_SOLD_TIME
2	1 Addis Way	Wandi	365000	351	139	2013	26900	4900	-32,19347028	115,8595535	5,524324354	129	1197
3	1 Ainsley Court	Camillo	287000	719	86	1979	22600	1900	-32,12057813	115,993579	1,649178183	113	2538
4	1 Arundel Street	Bayswater	685000	552	126	1999	5900	508	-31,91788	115,90705	0,9362432649	29	955
5	1 Ashcott Gate	Butler	367500	398	158	2003	36300	2100	-31,65428	115,7022	0,6808433268	39	1289
6	1 Ashendon Boulevard	Hammond Park	535000	704	247	2002	23100	3900	-32,15959	115,84948	2,220642666	129	1047
7	1 August Court	Bull Creek	685000	706	196	1985	12500	961	-32,06458	115,86046	1,183362603	89	1654
8	1 Avonlea Place	Bullsbrook	350000	663	154	1994	36700	26000	-31,66068818	116,0407159	14,01354839	98	1654
9	1 Babbler Court	Maida Vale	570000	568	225	2003	15300	7000	-31,95242	116,0207	2,74677214	137	589
10	1 Banken Court	Forrestdale	387500	782	111	1981	23300	6600	-32,14885	115,94308	4,204627996	62	1868
11	1 Barcroft Court	Attwell	485000	450	173	2000	20700	1600	-32,13768	115,86697	0,7471080892	129	1746
12	1 Batoni Way	Ashby	515000	388	160	2007	25000	3300	-31,7328966	115,7996371	3,141444909	131	2811
13	1 Bayley Street	Woodbridge	610000	550	121	1920	14600	275	-31,88948	115,99427	0,5195732638	130	3329
14	1 Beach Street	Bitcon	1950000	855	305	1981	10800	2900	-32,02366	115,78189	1,428825658	10	1320
15	1 Beacon Way	Singleton	420000	600	295	2006	54500	10400	-32,434584	115,758448	2,13203052	35	2689
16	1 Benowa Drive	Glen Forrest	600000	2028	222	1988	24000	10400	-31,91039	116,10841	1,634924182	78	2477
17	1 Bianco Place	Secret Harbour	400000	672	186	2002	50400	7900	-32,39686	115,75637	2,197314944	35	985
18	1 Bignell Place	Herne Hill	350000	797	177	1974	20900	7600	-31,82521	116,02121	3,788223563	93	3207
19	1 Bindoon Close	Gwelup	835000	480	170	1998	10800	2800	-31,86994	115,79768	1,509007988	135	2172
20	1 Binnia Mews	Harrisdale	482000	469	152	2008	19000	5700	-32,11598945	115,9173273	1,784415086	62	558
21	1 Binns Court	North Fremantle	1260000	354	227	2000	12900	387	-32,02689	115,75382	2,427821859	17	2019
22	1 Bishops Close	Quinns Rocks	405000	781	156	1989	34800	3800	-31,66895	115,70455	0,7749138173	55	1169
23	1 Bittern Close	Maida Vale	465000	620	143	2002	14800	7300	-31,95549	116,01601	2,295611114	137	1047
24	1 Bohemia Way	Parkwood	380000	979	122	1978	11300	4400	-32,0459	115,90519	1,12289063	96	2203
25	1 Boncath Road	Butler	360000	397	130	2006	38000	372	-31,638689	115,700206	1,652201534	39	528
26	1 Bonita Way	Tapping	540000	641	221	2005	26900	4800	-31,71607737	115,7997715	1,959016413	92	1534
27	1 Boronia Court	Thornlie	420000	680	163	1991	16200	3400	-32,07705	115,94632	1,735659976	125	711
28	1 Borough Road	Beldingville	330000	495	171	2011	45500	5600	-32,36170962	115,8136691	2,668401882	62	528
29	1 Bourke Street	Kensington	705000	774	132	1952	4500	3200	-31,98792	115,88137	1,093286046	93	1685
30	1 Bradbury Road	Hamilton Hill	565000	775	151	1960	15500	4800	-32,07693	115,78611	0,8307663583	91	742

Rysunek 1. Fragment wykorzystanej bazy danych

2.2. Zmienne lingwistyczne (atrybuty/własności obiektów)

Weźmy pod uwagę wszystkie 11 atrybutów liczbowych i przekształćmy je na zmienne lingwistyczne. Poniżej przedstawiamy przestrzenie rozważań dla każdej zmiennej, wszystkie wzory analityczne oraz wykresy funkcje przynależności dla każdej zmiennej lingwistycznej.

1. Cena

Przestrzeń rozważań zmiennej lingwistycznej: 50000 - 2500000 [AUD]

a) Bardzo tani

$$\mu_{bTani}(x) = \begin{cases} 1 & \text{gdy } x \in [50000, 244200] \\ (437400 - x)/193200 & \text{gdy } x \in (244200, 437400) \end{cases} \quad (1)$$

b) Tani

$$\mu_{tani}(x) = \begin{cases} (x - 244200)/193200 & \text{gdy } x \in (244200, 437400), \\ 1 & \text{gdy } x \in [437400, 534000] \\ (727200 - x)/193200 & \text{gdy } x \in (534000, 727200) \end{cases} \quad (2)$$

c) W rozsądnej cenie

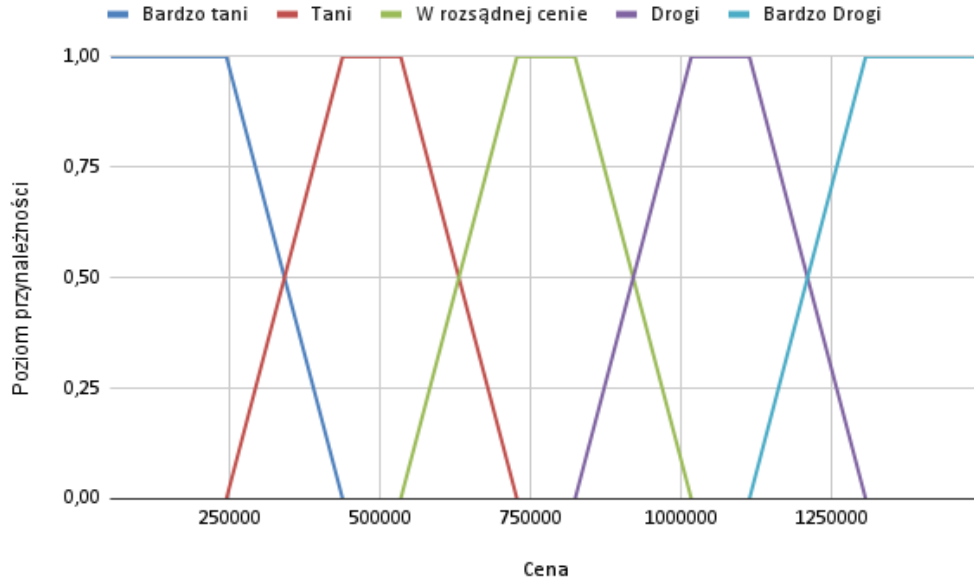
$$\mu_{rozCena}(x) = \begin{cases} (x - 534000)/193200 & \text{gdy } x \in (534000, 727200), \\ 1 & \text{gdy } x \in [727200, 823800] \\ (1017000 - x)/193200 & \text{gdy } x \in (823800, 1017000) \end{cases} \quad (3)$$

d) Drogi

$$\mu_{drogi}(x) = \begin{cases} (x - 823800)/193200 & \text{gdy } x \in (823800, 1017000), \\ 1 & \text{gdy } x \in [1017000, 1113600] \\ (1306800 - x)/193200 & \text{gdy } x \in (1113600, 1306800) \end{cases} \quad (4)$$

e) Bardzo drogi

$$\mu_{bDrogi}(x) = \begin{cases} (x - 1113600)/193200 & \text{gdy } x \in (1113600, 1306800), \\ 1 & \text{gdy } x \in [1306800, 2500000] \end{cases} \quad (5)$$



Rysunek 2. Zmienna lingwistyczna cena

2. Powierzchnia działki

Przestrzeń rozważań zmiennej lingwistycznej: 50 - 1100000 [m²]

a) Bardzo mała

$$\mu_{bMala}(x) = \begin{cases} 1 & \text{gdy } x \in [50, 5050] \\ (10050 - x)/5000 & \text{gdy } x \in (5050, 10050) \end{cases} \quad (6)$$

b) Mała

$$\mu_{mala}(x) = \begin{cases} (x - 5050)/5000 & \text{gdy } x \in (5050, 10050), \\ 1 & \text{gdy } x \in [10050, 12550] \\ (17550 - x)/5000 & \text{gdy } x \in (12550, 17550) \end{cases} \quad (7)$$

c) Średnia

$$\mu_{srednia}(x) = \begin{cases} (x - 244200)/5000 & \text{gdy } x \in (12550, 17550), \\ 1 & \text{gdy } x \in [17550, 20050] \\ (25050 - x)/5000 & \text{gdy } x \in (20050, 25050) \end{cases} \quad (8)$$

d) Duża

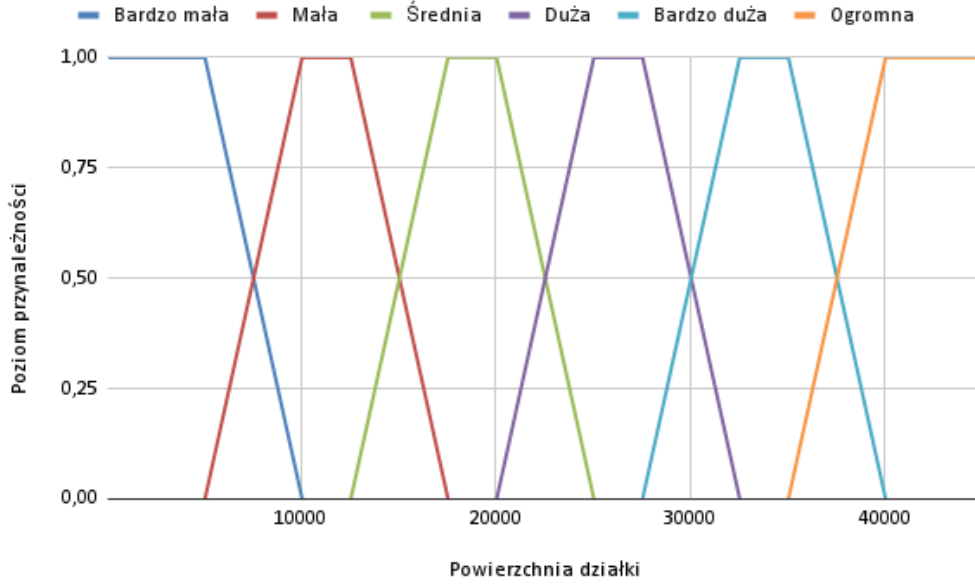
$$\mu_{duza}(x) = \begin{cases} (x - 20050)/5000 & \text{gdy } x \in (20050, 25050), \\ 1 & \text{gdy } x \in [25050, 27550] \\ (32550 - x)/5000 & \text{gdy } x \in (27550, 32550) \end{cases} \quad (9)$$

e) Bardzo duża

$$\mu_{bDuza}(x) = \begin{cases} (x - 27550)/5000 & \text{gdy } x \in (27550, 32550), \\ 1 & \text{gdy } x \in [32550, 35050] \\ (40050 - x)/5000 & \text{gdy } x \in (35050, 40050) \end{cases} \quad (10)$$

f) Ogromna

$$\mu_{ogromna}(x) = \begin{cases} (x - 35050)/5000 & \text{gdy } x \in (35050, 40050), \\ 1 & \text{gdy } x \in [40050, 1100000] \end{cases} \quad (11)$$



Rysunek 3. Zmienna lingwistyczna powierzchnia działki

3. Powierzchnia mieszkania

Przestrzeń rozważań zmiennej lingwistycznej: 1 - 1000 [m²]

a) Klaustrofobiczne

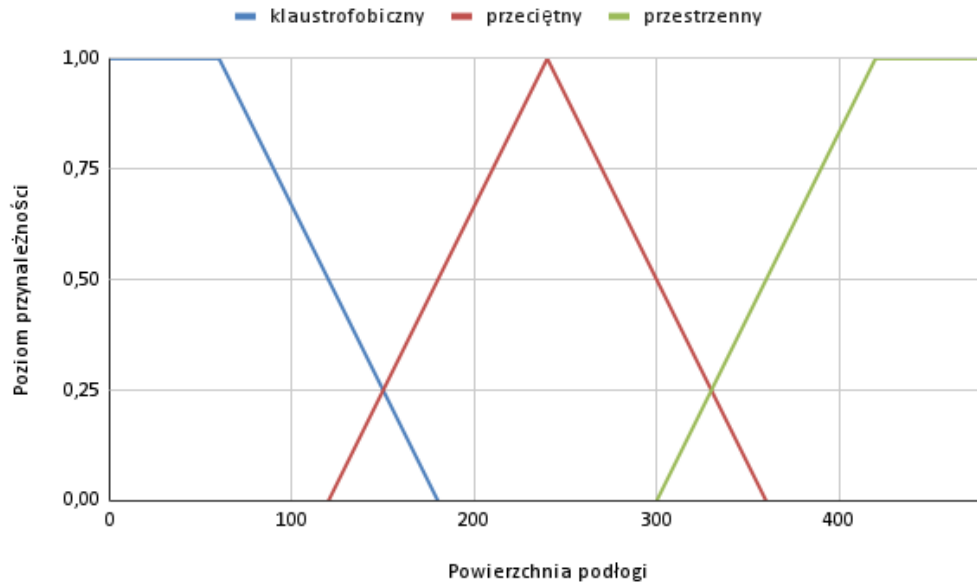
$$\mu_{klaustrofobiczne}(x) = \begin{cases} 1 & \text{gdy } x \in [0, 60] \\ (180 - x)/120 & \text{gdy } x \in (60, 180) \end{cases} \quad (12)$$

b) Przeciętne

$$\mu_{przecietne}(x) = \begin{cases} (x - 60)/120 & \text{gdy } x \in (120, 240), \\ 1 & \text{gdy } x = 240 \\ (360 - x)/120 & \text{gdy } x \in (240, 360) \end{cases} \quad (13)$$

c) Przestrzenne

$$\mu_{przestrzenne}(x) = \begin{cases} (x - 27550)/120 & \text{gdy } x \in (300, 420), \\ 1 & \text{gdy } x \in [420, 1000] \end{cases} \quad (14)$$



Rysunek 4. Zmienna lingwistyczna powierzchnia mieszkania

4. Rok wybudowania

Przestrzeń rozważań zmiennej lingwistycznej: 1850 - 2022

a) Bardzo stary

$$\mu_{bStary}(x) = \begin{cases} 1 & \text{gdy } x \in [1850, 1860] \\ (1900 - x)/40 & \text{gdy } x \in (1860, 1900) \end{cases} \quad (15)$$

b) Stary

$$\mu_{stary}(x) = \begin{cases} (x - 1860)/40 & \text{gdy } x \in (1860, 1900), \\ 1 & \text{gdy } x = 1900 \\ (1940 - x)/40 & \text{gdy } x \in (1900, 1940) \end{cases} \quad (16)$$

c) W średnim wieku

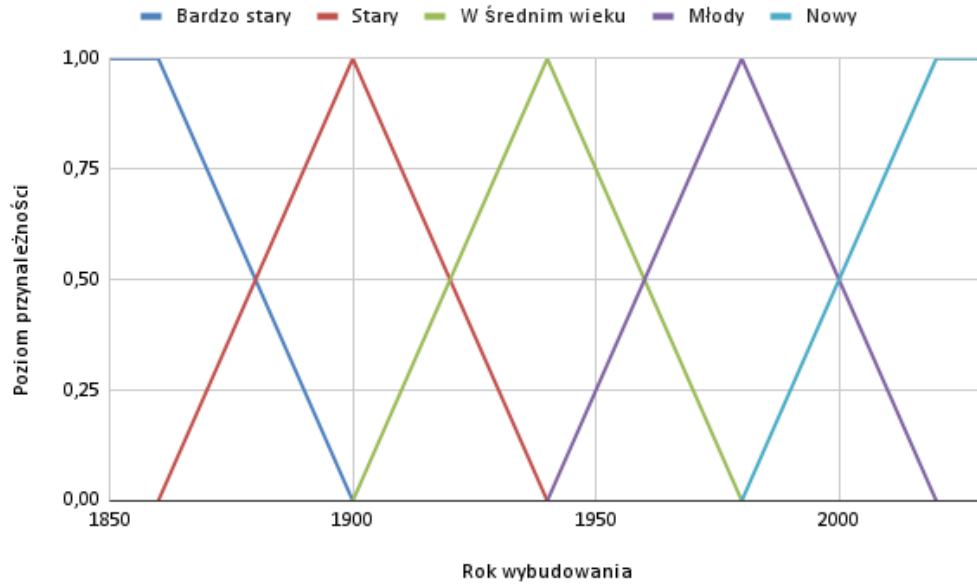
$$\mu_{sredniWiek}(x) = \begin{cases} (x - 1900)/40 & \text{gdy } x \in (1900, 1940), \\ 1 & \text{gdy } x = 1940 \\ (1980 - x)/40 & \text{gdy } x \in (1940, 1980) \end{cases} \quad (17)$$

d) Młody

$$\mu_{mlody}(x) = \begin{cases} (x - 1940)/40 & \text{gdy } x \in (1940, 1980), \\ 1 & \text{gdy } x = 1980 \\ (2000 - x)/40 & \text{gdy } x \in (1980, 2020) \end{cases} \quad (18)$$

e) Nowy

$$\mu_{nowy}(x) = \begin{cases} (x - 1980)/40 & \text{gdy } x \in (1980, 2020), \\ 1 & \text{gdy } x \in [2020, 2022] \end{cases} \quad (19)$$



Rysunek 5. Zmienna lingwistyczna rok wybudowania

5. Odległość do centrum

Przestrzeń rozważań zmiennej lingwistycznej: 500 - 70000 [m]

a) Bardzo blisko centrum

$$\mu_{bbCentrum}(x) = \begin{cases} 1 & \text{gdy } x \in [500, 5500] \\ (15500 - x)/10000 & \text{gdy } x \in (5500, 15500) \end{cases} \quad (20)$$

b) Nieopodal centrum

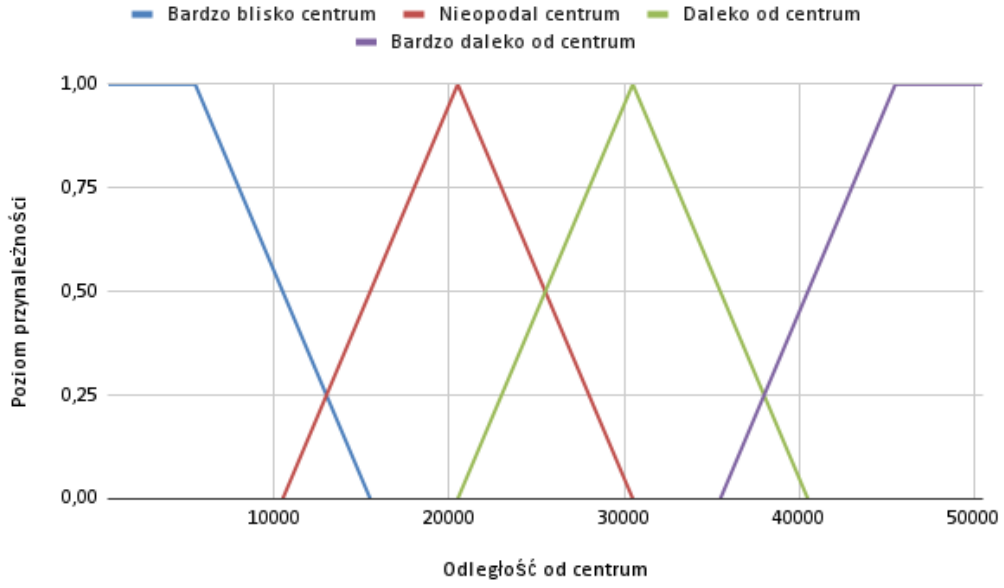
$$\mu_{nCentrum}(x) = \begin{cases} (x - 10500)/10000 & \text{gdy } x \in (10500, 20500), \\ 1 & \text{gdy } x = 20500 \\ (30500 - x)/10000 & \text{gdy } x \in (20500, 30500) \end{cases} \quad (21)$$

c) Daleko od centrum

$$\mu_{dCentrum}(x) = \begin{cases} (x - 20500)/10000 & \text{gdy } x \in (20500, 30500), \\ 1 & \text{gdy } x = 30500 \\ (40500 - x)/10000 & \text{gdy } x \in (30500, 40500) \end{cases} \quad (22)$$

d) Bardzo daleko od centrum

$$\mu_{bdCentrum}(x) = \begin{cases} (x - 35500)/10000 & \text{gdy } x \in (35500, 45500), \\ 1 & \text{gdy } x \in [45500, 70000] \end{cases} \quad (23)$$



Rysunek 6. Zmienna lingwistyczna odległość od centrum

6. Odległość do najbliższej stacji

Przestrzeń rozważań zmiennej lingwistycznej: 30 - 35000 [m]

a) Bardzo blisko

$$\mu_{bbStacji}(x) = \begin{cases} 1 & \text{gdy } x \in [30, 4030] \\ (8030 - x)/4000 & \text{gdy } x \in (4030, 8030) \end{cases} \quad (24)$$

b) W pobliżu

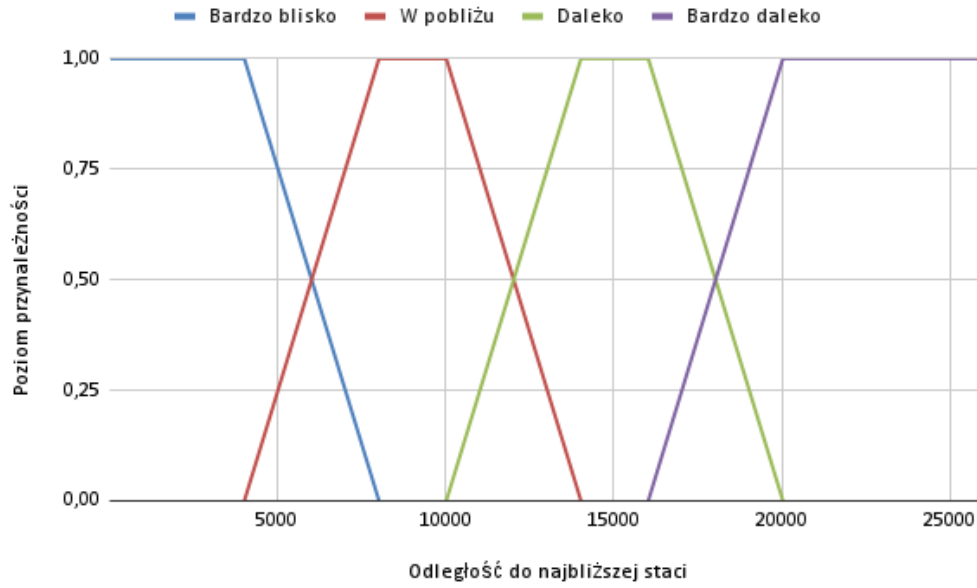
$$\mu_{wPobliżu}(x) = \begin{cases} (x - 4030)/4000 & \text{gdy } x \in (4030, 8030), \\ 1 & \text{gdy } x \in [8030, 10030] \\ (14030 - x)/4000 & \text{gdy } x \in (10030, 14030) \end{cases} \quad (25)$$

c) Daleko

$$\mu_{dStacji}(x) = \begin{cases} (x - 10030)/4000 & \text{gdy } x \in (10030, 14030), \\ 1 & \text{gdy } x \in [14030, 16030] \\ (20030 - x)/4000 & \text{gdy } x \in (16030, 20030) \end{cases} \quad (26)$$

d) Bardzo daleko

$$\mu_{bdStacji}(x) = \begin{cases} (x - 16030)/4000 & \text{gdy } x \in (16030, 20030), \\ 1 & \text{gdy } x \in [20030, 35000] \end{cases} \quad (27)$$



Rysunek 7. Zmienna lingwistyczna odległość do najbliższej stacji

7. Czas od ostatniej sprzedaży

Przestrzeń rozważań zmiennej lingwistycznej: 30 - 15000 [dni]

a) Ledwie co

$$\mu_{ledCo}(x) = \begin{cases} 1 & \text{gdy } x \in [30, 780], \\ \frac{1}{4800\sqrt{2\pi}} e^{-\frac{1}{2} \frac{x-780}{4800^2}} & \text{gdy } x \in (780, 6480) \end{cases} \quad (28)$$

b) Ostatnio

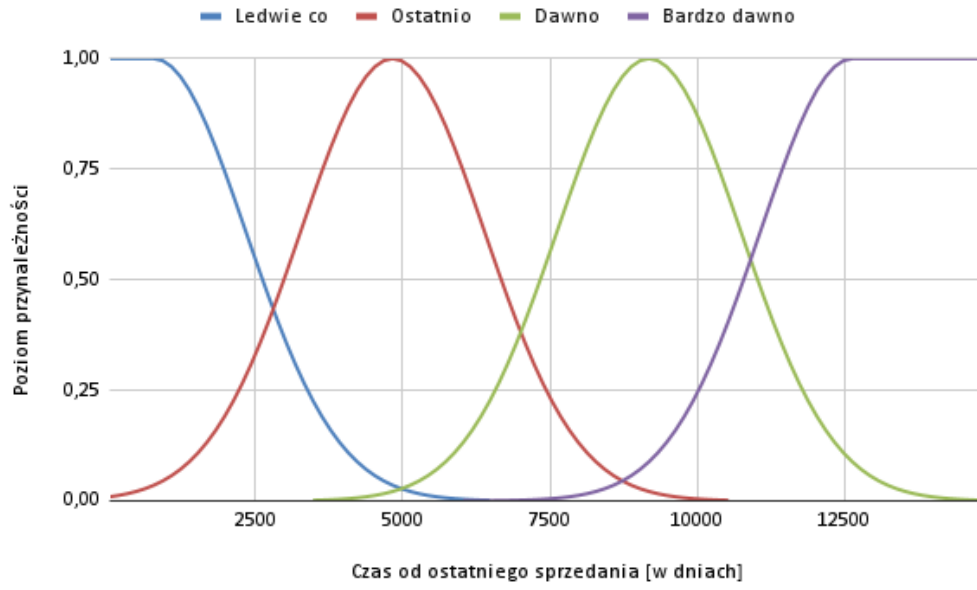
$$\mu_{ostatnio}(x) = \frac{1}{4800\sqrt{2\pi}} e^{-\frac{1}{2} \frac{x-4830}{4800^2}} \quad \text{gdy } x \in (30, 10530) \quad (29)$$

c) Dawno

$$\mu_{dawno}(x) = \frac{1}{4800\sqrt{2\pi}} e^{-\frac{1}{2} \frac{x-9180}{4800^2}} \quad \text{gdy } x \in (3480, 14880) \quad (30)$$

d) Bardzo dawno

$$\mu_{bardzoDawno}(x) = \begin{cases} \frac{1}{4800\sqrt{2\pi}} e^{-\frac{1}{2} \frac{x-12630}{4800^2}} & \text{gdy } x \in (6030, 12630) \\ 1 & \text{gdy } x \in [12630, 15000] \end{cases} \quad (31)$$



Rysunek 8. Zmienna lingwistyczna czas od ostatniej sprzedaży

8. Wysokość geograficzna

Przestrzeń rozważań zmiennej lingwistycznej: $-32.5 - -31.5$ [$^{\circ}$]

a) Na południu

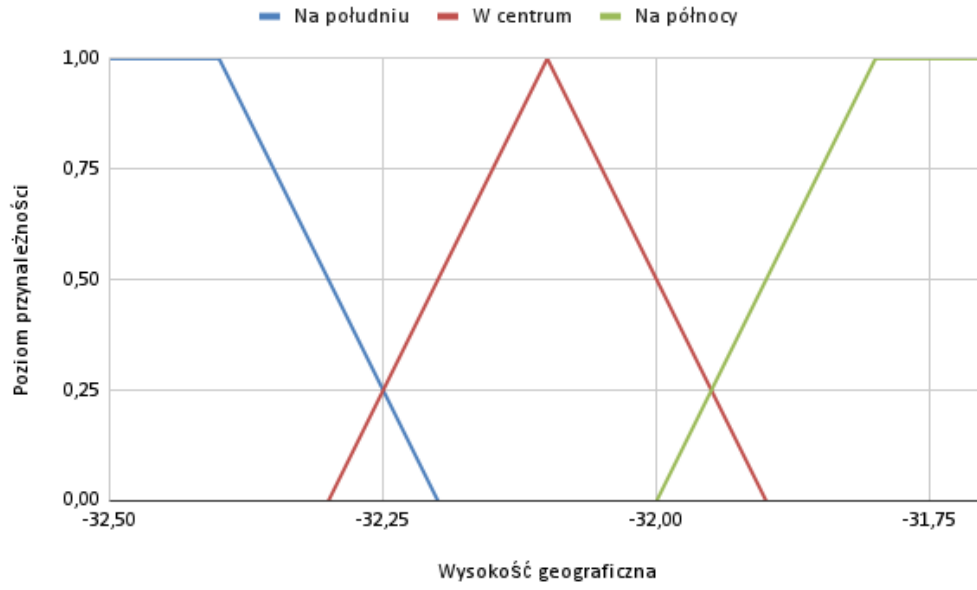
$$\mu_{naPoludniu}(x) = \begin{cases} 1 & \text{gdy } x \in [-32.5, -32.4], \\ (-32.2 - x)/0.2 & \text{gdy } x \in (-32.4, -32.2) \end{cases} \quad (32)$$

b) W centrum

$$\mu_{wCentrumWys}(x) = \begin{cases} (x - -32.3)/0.2 & \text{gdy } x \in (-32.3, -32.1), \\ 1 & \text{gdy } x = -32.1 \\ (-31.9 - x)/0.2 & \text{gdy } x \in (-32.1, -31.9) \end{cases} \quad (33)$$

c) Na północy

$$\mu_{naPolnocy}(x) = \begin{cases} (x - -32)/0.2 & \text{gdy } x \in (-32, -31.8), \\ 1 & \text{gdy } x \in [-31.8, -31.5] \end{cases} \quad (34)$$



Rysunek 9. Zmienna lingwistyczna wysokość geograficzna

9. Szerokość geograficzna

Przestrzeń rozważań zmiennej lingwistycznej: 115.6 - 116.4 [°]

a) Na wschodzie

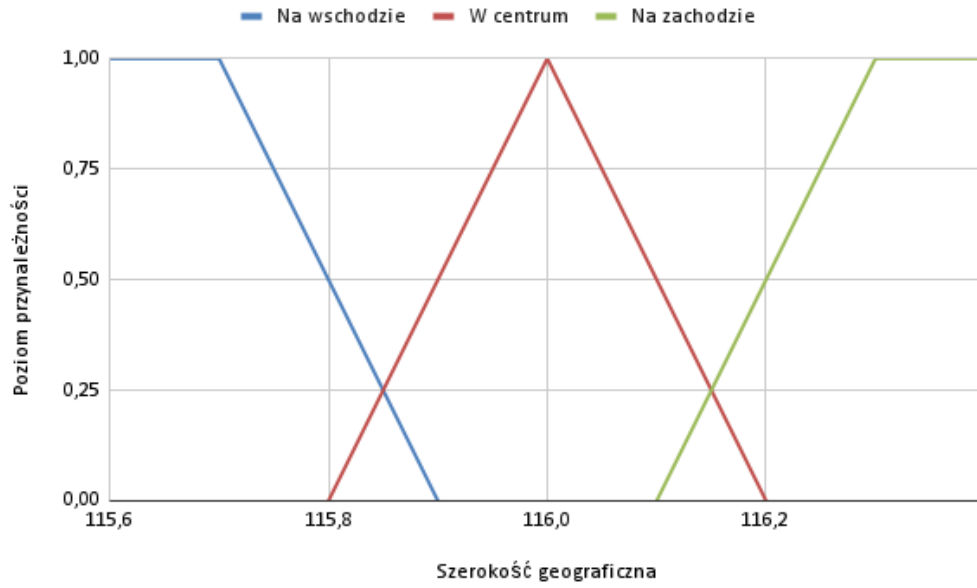
$$\mu_{naWschodzie}(x) = \begin{cases} 1 & \text{gdy } x \in [115.6, 115.7], \\ (115.9 - x)/0.2 & \text{gdy } x \in (115.7, 115.9) \end{cases} \quad (35)$$

b) W centrum

$$\mu_{wCentrumSzer}(x) = \begin{cases} (x - 115.8)/0.2 & \text{gdy } x \in (115.8, 116), \\ 1 & \text{gdy } x = 116 \\ (116.2 - x)/0.2 & \text{gdy } x \in (116, 116.2) \end{cases} \quad (36)$$

c) Na zachodzie

$$\mu_{naZachodzie}(x) = \begin{cases} (x - 116.1)/0.2 & \text{gdy } x \in (116.1, 116.3), \\ 1 & \text{gdy } x \in [116.3, 116.4] \end{cases} \quad (37)$$



Rysunek 10. Zmienna lingwistyczna szerokość geograficzna

10. Odległość do szkoły

Przestrzeń rozważań zmiennej lingwistycznej: 0.05 - 25 [km]

a) Bardzo blisko

$$\mu_{bbSzkoly}(x) = \begin{cases} 1 & \text{gdy } x \in [0.05, 2.05] \\ (6.05 - x)/4 & \text{gdy } x \in (2.05, 6.05) \end{cases} \quad (38)$$

b) W pobliżu

$$\mu_{wPobliżuSzkoly}(x) = \begin{cases} (x - 2.05)/4 & \text{gdy } x \in (2.05, 6.05), \\ 1 & \text{gdy } x \in [6.05, 8.05] \\ (12.05 - x)/4 & \text{gdy } x \in (8.05, 12.05) \end{cases} \quad (39)$$

c) Daleko

$$\mu_{dalekoSzkoly}(x) = \begin{cases} (x - 8.05)/4 & \text{gdy } x \in (8.05, 12.05), \\ 1 & \text{gdy } x \in [12.05, 14.05] \\ (18.05 - x)/4 & \text{gdy } x \in (14.05, 18.05) \end{cases} \quad (40)$$

d) Bardzo daleko

$$\mu_{bdSzkoly}(x) = \begin{cases} (x - 14.05)/4 & \text{gdy } x \in (14.05, 18.05), \\ 1 & \text{gdy } x \in [18.05, 25] \end{cases} \quad (41)$$



Rysunek 11. Zmienna lingwistyczna odległość do szkoły

11. Ranga szkoły

Przestrzeń rozważań zmiennej lingwistycznej: 1-150

a) Wybitna szkoła

$$\mu_{wybSzkola}(x) = \begin{cases} 1 & \text{gdy } x \in [1, 11] \\ (31 - x)/20 & \text{gdy } x \in (11, 31) \end{cases} \quad (42)$$

b) Dobra szkoła

$$\mu_{dSzkola}(x) = \begin{cases} (x - 11)/20 & \text{gdy } x \in (11, 31), \\ 1 & \text{gdy } x \in [31, 41] \\ (61 - x)/20 & \text{gdy } x \in (41, 61) \end{cases} \quad (43)$$

c) Średnia szkoła

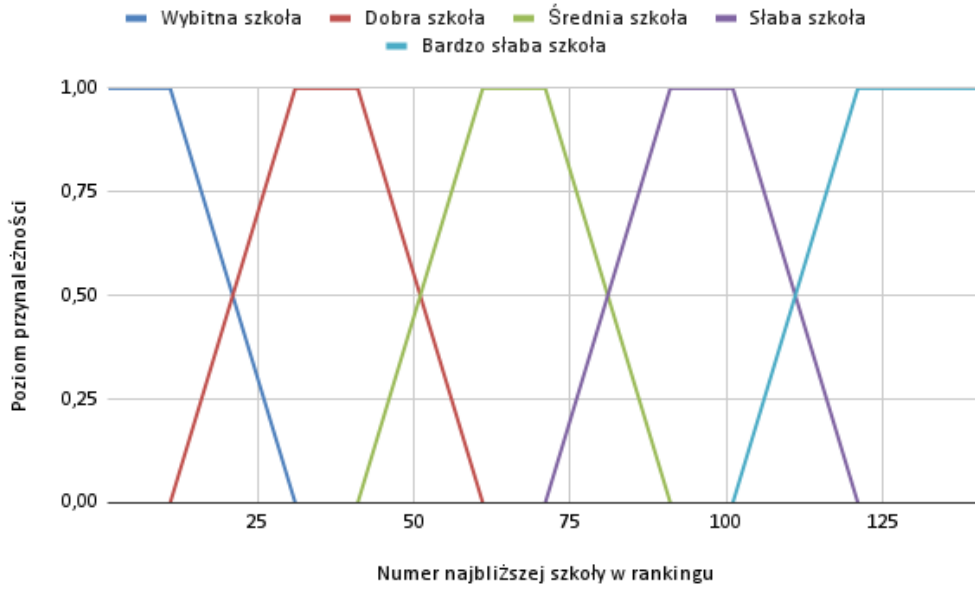
$$\mu_{srSzkola}(x) = \begin{cases} (x - 41)/20 & \text{gdy } x \in (41, 61), \\ 1 & \text{gdy } x \in [61, 71] \\ (91 - x)/20 & \text{gdy } x \in (71, 91) \end{cases} \quad (44)$$

d) Słaba szkoła

$$\mu_{slabaSzkola}(x) = \begin{cases} (x - 71)/20 & \text{gdy } x \in (71, 91), \\ 1 & \text{gdy } x \in [91, 101] \\ (121 - x)/20 & \text{gdy } x \in (101, 121) \end{cases} \quad (45)$$

e) Bardzo słaba szkoła

$$\mu_{bslabaSzkola}(x) = \begin{cases} (x - 101)/20 & \text{gdy } x \in (101, 121), \\ 1 & \text{gdy } x \in [121, 150] \end{cases} \quad (46)$$



Rysunek 12. Zmienna lingwistyczna ranga szkoły

2.3. Kwantyfikatory lingwistyczne (liczności obiektów)

W projekcie będziemy korzystać z 5 kwantyfikatorów względnych:

1. Prawie żaden

$$\mu_{k1}(x) = \begin{cases} 1 & \text{gdy } x \in [0, 768] \\ (3073 - x)/2305 & \text{gdy } x \in (768, 3073) \end{cases} \quad (47)$$

2. Trochę

$$\mu_{k2}(x) = \begin{cases} (x - 2305)/768 & \text{gdy } x \in (2305, 3073), \\ 1 & \text{gdy } x \in [3073, 6147] \\ (7683 - x)/1536 & \text{gdy } x \in (6147, 7683) \end{cases} \quad (48)$$

3. Około połowy

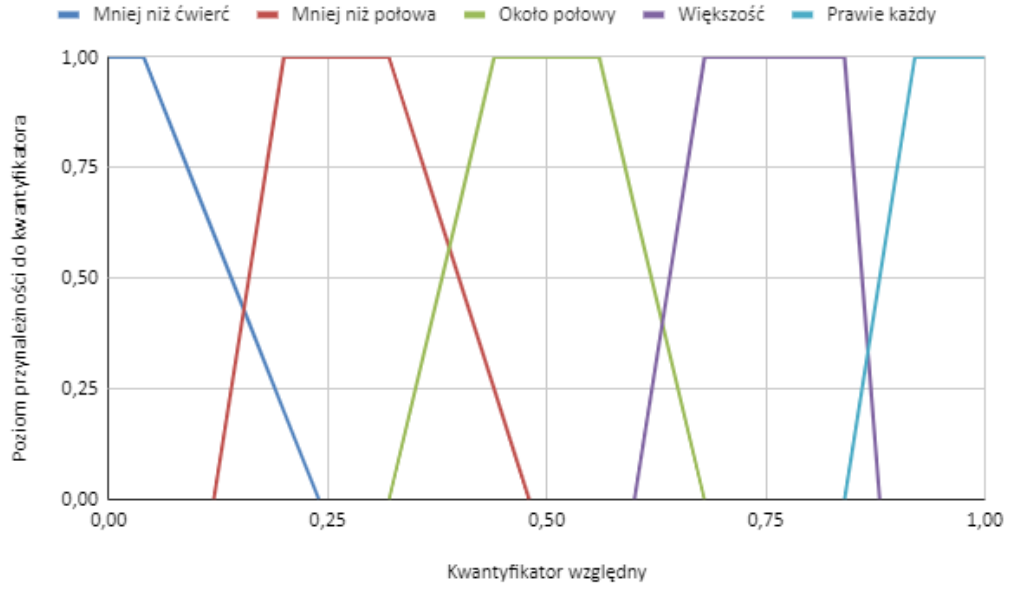
$$\mu_{k3}(x) = \begin{cases} (x - 6147)/2305 & \text{gdy } x \in (6147, 8452), \\ 1 & \text{gdy } x \in [8452, 10756] \\ (13061 - x)/2305 & \text{gdy } x \in (10756, 13061) \end{cases} \quad (49)$$

4. Sporo

$$\mu_{k4}(x) = \begin{cases} (x - 11525)/1536 & \text{gdy } x \in (11525, 13061), \\ 1 & \text{gdy } x \in [13061, 16135] \\ (16903 - x)/768 & \text{gdy } x \in (16135, 16903) \end{cases} \quad (50)$$

5. Prawie wszystkie

$$\mu_{k5}(x) = \begin{cases} (x - 16135)/2305 & \text{gdy } x \in (16135, 18440), \\ 1 & \text{gdy } x \in [18440, 19208] \end{cases} \quad (51)$$



Rysunek 13. Wykresy funkcji przynależności dla kwalifikatorów względnych

Będziemy również korzystać z następujących kwantyfikatorów absolutnych:

1. Mniej niż 3000

$$\mu_{k1}(x) = \begin{cases} 1 & \text{gdy } x \in [0, 3600] \\ (6200 - x)/2600 & \text{gdy } x \in (3600, 6200) \end{cases} \quad (52)$$

2. Około 7500

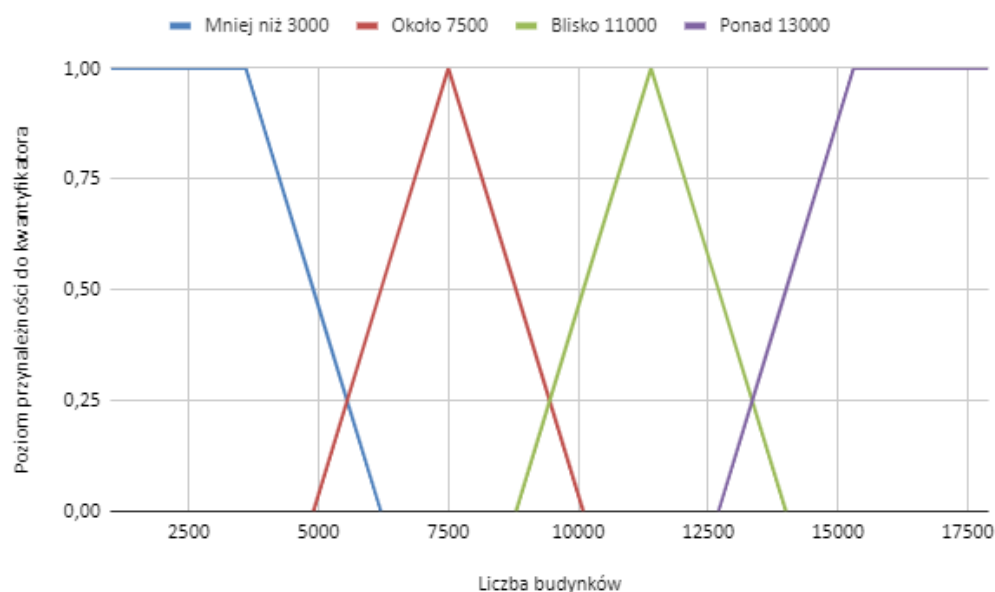
$$\mu_{k2}(x) = \begin{cases} (x - 4900)/2600 & \text{gdy } x \in (4900, 7500) \\ 1 & \text{gdy } x = 7500 \\ (10100 - x)/2600 & \text{gdy } x \in (7500, 10100) \end{cases} \quad (53)$$

3. Blisko 11000

$$\mu_{k3}(x) = \begin{cases} (x - 8800)/2600 & \text{gdy } x \in (8800, 11400) \\ 1 & \text{gdy } x = 11400 \\ (14000 - x)/2600 & \text{gdy } x \in (11400, 14000) \end{cases} \quad (54)$$

4. Ponad 13000

$$\mu_{k5}(x) = \begin{cases} (x - 12700)/2600 & \text{gdy } x \in (12700, 15300), \\ 1 & \text{gdy } x \in [15300, 19208] \end{cases} \quad (55)$$



Rysunek 14. Wykresy funkcji przynależności dla kwalifikatorów absolutnych

3. Narzędzia obliczeniowe: wybór/implementacja. Diagram UML pakietu obliczeń rozmytych i generatora podsumowań. Instrukcja użytkownika

Kwantyfikatory

- ☐ Względne
 - ☐ Mniej niż ćwierć domów
 - ☐ Mniej niż połowa domów
 - ☐ Około połowy domów
 - ☐ Większość domów
 - ☐ Prawie każdy z domów
- ☐ Bezwzględne
 - ☐ Mniej niż 3000 domów
 - ☐ Około 7500 domów
 - ☐ Blisko 11000 z domów
 - ☐ Ponad 13000 domów

Kwalifikatory

- ☐ Cena
 - ☐ bardzo tani
 - ☐ tani
 - ☐ w rozsądnej cenie
 - ☐ drogi
 - ☐ bardzo drogi
- ☐ Powierzchnia działki
 - ☐ bardzo mała
 - ☐ mała
 - ☐ średnia
 - ☐ duża
 - ☐ bardzo duża
 - ☐ ogromna
- ☐ Powierzchnia mieszkania

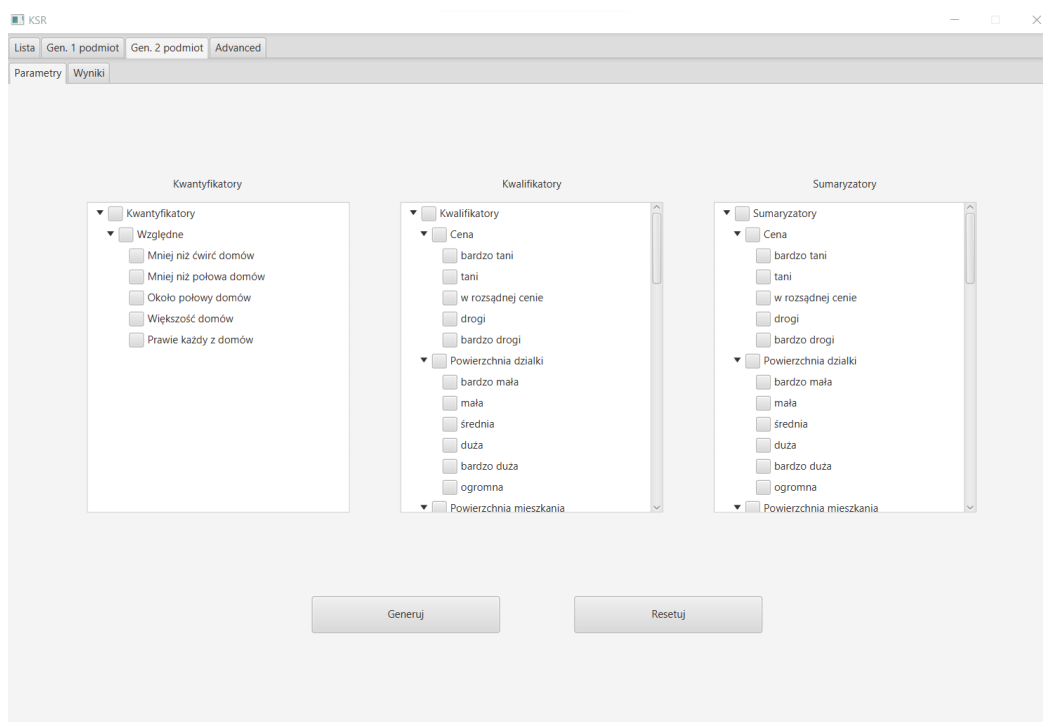
Sumaryzatory

- ☐ Cena
 - ☐ bardzo tani
 - ☐ tani
 - ☐ w rozsądnej cenie
 - ☐ drogi
 - ☐ bardzo drogi
- ☐ Powierzchnia działki
 - ☐ bardzo mała
 - ☐ mała
 - ☐ średnia
 - ☐ duża
 - ☐ bardzo duża
 - ☐ ogromna
- ☐ Powierzchnia mieszkania

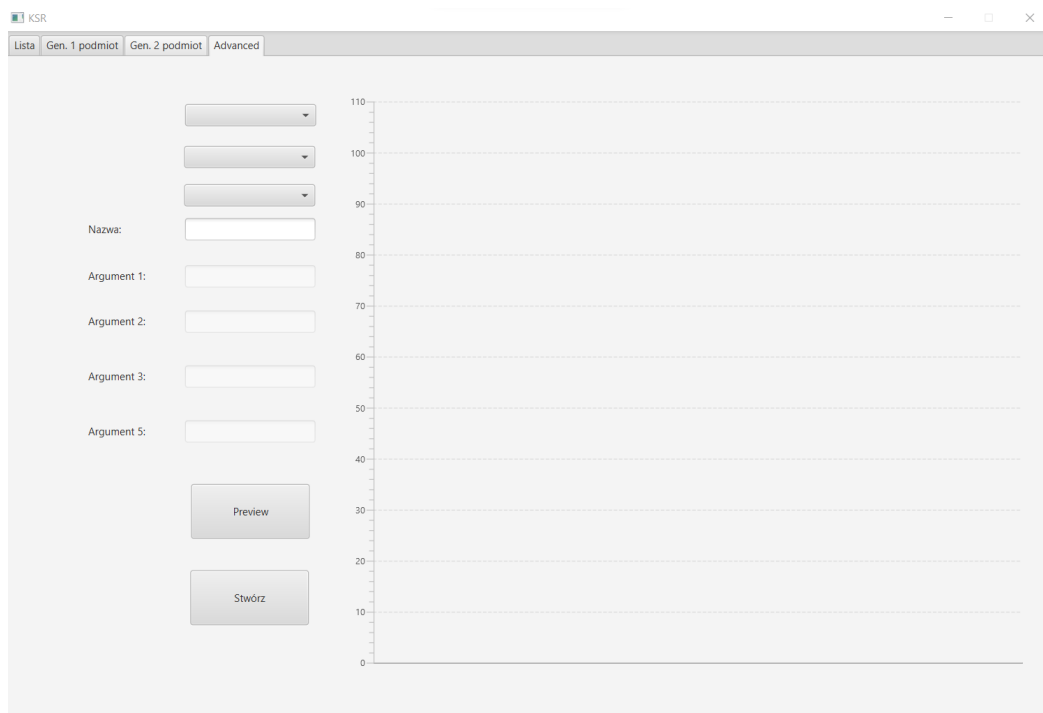
Wagi miar

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.5	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

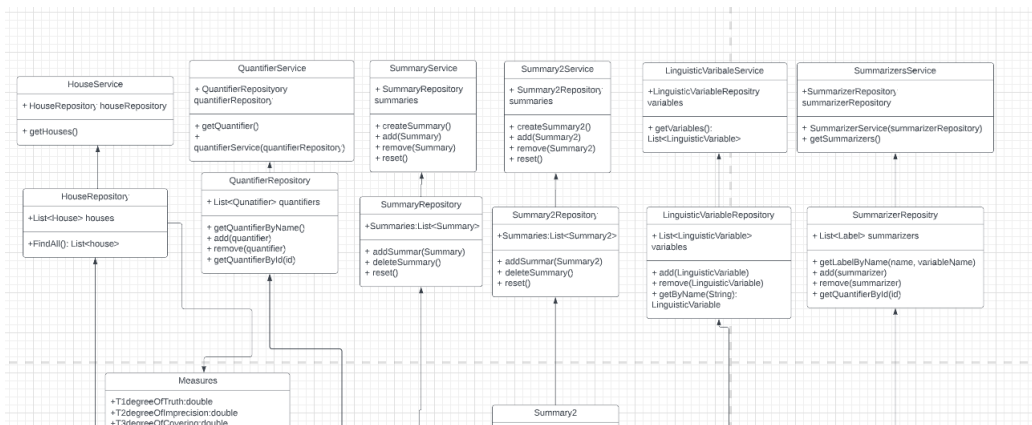
Rysunek 15. Interfejs użytkownika do generowania podsumowań jednopodmiotowych.



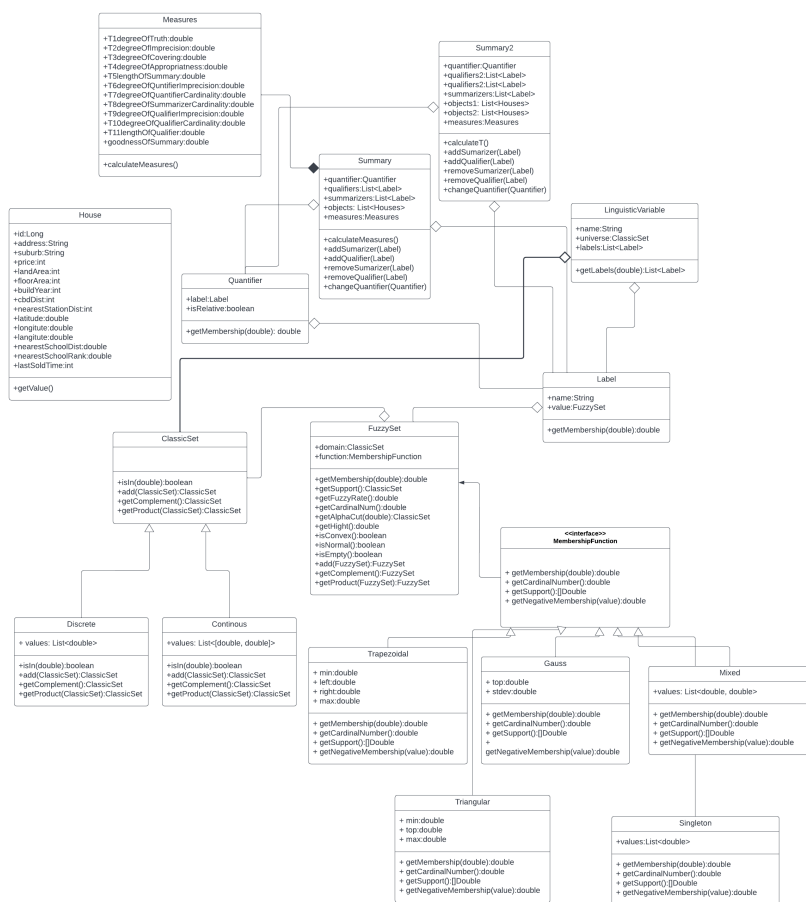
Rysunek 16. Interfejs użytkownika do generowania podsumowań wielopodmiotowych.



Rysunek 17. Interfejs użytkownika do dodawania nowych kwantyfikatorów i sumaryzatorów.



Rysunek 18. Diagram UML części obliczeniowej aplikacji do tworzenia podsumowań lingwistycznych.



Rysunek 19. Diagram UML części logicznej modelu aplikacji do tworzenia podsumowań lingwistycznych

Opis diagramu UML

- Jedną z głównych klas warstwy logicznej jest klasa FuzzySet definiująca obiekt zbioru rozmytego. Zawiera w sobie obiekt zbioru klasycznego ClassicSet oraz obiekt funkcji przynależności.

- Funkcja przynależności zdefiniowana jest jako interfejs wykorzystujący różne funkcje:
 - Gaussa
 - Trapezowa
 - Trójkątna
 - Singleton
 - Mieszana
- Klasa ClassicSet wykorzystuje dwie klasy Discrete oraz Crisp do definiowania kolejno przestrzeni rozważań: dyskretnej oraz gęstej.
- Klasa LinguisticVariable definiuje zmienną lingwistyczną np. "cena". Zawiera ona listę etykiet definiowanych przez klasę Label. Etykietami danej zmiennej lingwistycznej np. "cena" są: "bardzo tani", "tani", "w rozsądnej cenie", "drogi", "bardzo drogi".
- Klasa Quantifier definiuje kwantyfikatory lingwistyczne. Klasa przechowuje w sobie również obiekt klasy Label. Definiuje on jedno z pięciu słów kluczowych sugerujących liczbę obiektów: "prawie żaden", "trochę", "około połowy", "sporo", "prawie wszystkie".
- Klasa Summary to klasa definiująca podsumowania lingwistyczne jedno-podmiotowe zawierająca kwantifikator, listę kwalifikatorów, sumaryzatorów, listę obiektów, oraz miary jakości podsumowań lingwistycznych. Zawiera ona metody modyfikacji parametrów podsumowań takie jak: addQualifier() czy removeSummarizer()
- Klasa Summary2 to klasa definiująca podsumowania lingwistyczne dwu-podmiotowe zawierająca kwantifikator, 2 listy kwalifikatorów, listę sumaryzatorów, 2 listy obiektów, oraz miarę jakości podsumowania. Zawiera ona metody modyfikacji parametrów podsumowań takie jak: addQualifier() czy removeSummarizer()
- Wszystkie podsumowania i operacje z nimi związane przechowywane są w obiekcie klasy SummaryRepository, czyli repozytorium na podsumowania.
- Klasa Measures definiuje metody obliczania miar jakości podsumowań lingwistycznych $T_1 - T_{11}$ oraz miarę ogólnej dobroci podsumowania goodnessOfSummary T .
- Klasy repozytorium takie jak: SummarizerRepository, LinguisticVariableRepository, SummaryRepository, QuantifierRepository oraz HouseRepository odpowiedzialne są za przechowywanie konkretnych wartości takich jak: kwantyfikatory, podsumowania czy zmienne lingwistyczne
- Klasy serwis: SummarizersService, LinguisticVariableService, SummaryService, SummaryService2, QuantifierService oraz HouseService odpowiedzialne są za operacje takie jak: dodawanie, usuwanie, czy pobieranie obiektów trzymanych w repozytorium

4. Jednopo­dmiotowe podsumowania lingwistyczne.

Miary jakości, podsumowanie optymalne

Przykład podsumowania lingwistycznego jednopo­dmiotowego, na podstawie którego opisane są poniższe miary jakości podsumowań lingwistycznych. Jako przykładowe parametry wzięliśmy:

- Kwantyfikator:
 - Mniej niż ćwierć domów
- Kwalifikator:
 - Klaustrofobiczne (Powierzchnia Mieszkania)
- Sumaryzator:
 - stary (Rok wybudowania)
- Mniej niż ćwierć domów które są/mają klaustrofobiczne Powierzchnia mieszkania jest/ma stary rok wybudowania

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.83	1.0	0.53	0.05	0.01	1.0	0.75	0.86	0.77	0.82	0.88	1.0

Do oceny jakości podsumowań jedno-podmiotowych korzystaliśmy z poszczególnych miar [3]:

1. *T1* Degree of truth (Stopień prawdziwości) - określa jak bardzo prawdziwe jest dane podsumowanie lingwistyczne. Na podstawie przykładu powyżej widzimy że podsumowanie jest prawdziwe w stopniu 1. Wysoka wartość oznacza wysoki stopień prawdziwości naszego podsumowania.
2. *T2* Degree of imprecision (Stopień (nie)precyzyjności) - określa jak bardzo precyzyjne (nieprecyzyjne) jest dane podsumowanie na podstawie sumaryzatorów. Na podstawie przykładu powyżej oznacza to, że podsumowanie na podstawie sumaryzatora "stary" jest precyzyjne w stopniu 0.53.
3. *T3* Degree of covering (Stopień pokrycia) - określa jak dużo domów (obiektów z bazy danych) pasującymi do kwalifikatora, pasuje do całości podsumowania. Na podstawie przykładu powyżej można stwierdzić że mało obiektów/domów z bazy pokrywa się z kwalifikatorem określającym powierzchnie mieszkania "klaustrofobiczny".
4. *T4* Degree of appropriateness (Stopień adekwatności) - miara zależna od miary *T3*, określa ogólną przydatność danego podsumowania. Na podstawie przykładu powyżej że wynik przydatności podsumowania wynosi 0.01, wynika to z niskiego pokrycia z miarą *T3* oraz ich powiązaniem. Oznacza to niski stopień przydatności naszego przykładowego podsumowania.
5. *T5* Length of summary (długość podsumowania) - określa długość podsumowania na podstawie ilości sumaryzatorów. Na podstawie przykładu powyżej widzimy że wynik wynosi 1.0 ze względu na niską długość podsumowania (wybór jednego sumaryzatora do podsumowania). Wysoka wartość miary oznacza małą długość naszego przykładowego podsumowania.
6. *T6* Degree of quantifier imprecision (Stopień (nie)precyzyjności kwantyfikatora) - Miara analogiczna do miary *T2*, określa (nie)precyzyjność danego podsumowania na podstawie kwantyfikatora. Na podstawie przykładu

powyżej widzimy że stopień (nie)precyzyjności dla kwantyfikatora "mniej niż ćwierć domów" wynosi 0.75. Oznacza to, że nasze podsumowanie ze względu na wybrany kwantyfikator jest precyzyjne w stopniu 0.75.

7. *T7 Degree of quantifier cardinality* (Stopień kardynalności kwantyfikatora) - opiera się na liczbie kardynalnej kwantyfikatora. Reprezentuje jakość kwantyfikatora w podsumowaniu. Na podstawie przykładu powyżej widzimy że jakość wybranego kwantyfikatora w przykładowym podsumowaniu określona jest wartością 0.86. Ze względu na dość wysokie wyniki miary *T7* w naszym przykładowym podsumowaniu oznacza to, że wybrany kwantyfikator jest dobrej jakości.
8. *T8 Degree of summarizer cardinality* (Stopień kardynalności sumaryzatora) - Określa precyzyjność podsumowania na podstawie kardynalności sumaryzatorów. Im większa liczba kardynalna sumaryzatora tym gorsza jakość podsumowania. Na podstawie przykładu powyżej widzimy że jakość podsumowania określona na podstawie wybranego sumaryzatora "stary" wynosi 0.77. Oznacza to w naszym przykładowym podsumowaniu wysoką jakość wybranego sumaryzatora ze względu na wysoką wartość miary *T8*.
9. *T9 Degree of qualifier imprecision* (Stopień (nie)precyzyjności kwalifikatora) - Miara analogiczna do miary *T2*, określa (nie)precyzyjność danego podsumowania na podstawie kwalifikatorów. Na podstawie przykładu powyżej widzimy że stopień (nie)precyzyjności podsumowania na podstawie kwalifikatora "klastrofobiczny" określona jest wartością 0.82. Oznacza to wysoką precyzję wybranego kwalifikatora w naszym przykładowym podsumowaniu.
10. *T10 Degree of qualifier cardinality* (Stopień kardynalności kwalifikatora) - analogicznie do miary *T8*, określa precyzyjność podsumowania na podstawie kardynalności kwalifikatorów. Im większa liczba kardynalna kwalifikatora tym gorsza jakość podsumowania. Na podstawie przykładu powyżej widzimy że stopień kardynalności kwalifikatora wynosi 0.88. Oznacza to, że w przykładowym podsumowaniu można wywnioskować o dobrej jakości podsumowania pod kątem wybranego kwalifikatora.
11. *T11 Length of a qualifier* (Długość kwalifikatora) - analogicznie do miary *T5* określa długość podsumowania na podstawie ilości kwalifikatorów. Na podstawie przykładu powyżej widzimy że długość kwalifikatora wynosi 1.0. Oznacza to niską długość przykładowego podsumowania, ze względu na wybór tylko jednego kwalifikatora.
12. *T Ogólna dobroć podsumowania* - jest to średnia ważona z wszystkich wymienionych powyżej miar.

Jako najlepsze podsumowania uważamy te które mają największą wartość miary *T* czyli ogólnej dobroci podsumowania.

Jeżeli chodzi o reprezentacje wyników podsumowań lingwistycznych jednopodmiotowych to zastosowaliśmy podsumowania z kwalifikatorem oraz bez. Wyniki podsumowań generowane są dla różnych kwantyfikatorów względnych. Jeżeli chodzi o kwantyfikatory absolutne to w podsumowaniu brane pod uwagę są jedynie sumaryzatory.

Dane wejściowe dla przykładowego wyniku poniżej

- Kwantyfikator:
 - Mniej niż połowa
- Kwalifikator:
 - tani (Cena)
- Sumaryzator:
 - średnia (Powierzchnia działki)

Jako wynik z powyższymi danymi wygenerowały nam się dwa podsumowania. Jedno z sumaryzatorem oraz uwzględnieniem kwalifikatora:

- Mniej niż ćwierć domów które są/mają **średnią** powierzchnie działki jest/ma **tani** cena

Drugim wynikiem było podsumowanie jednopodmiotowe uwzględniające jedynie sumaryzator

- Mniej niż ćwierć domów jest/ma **tani** cena

Poniżej kilka przykładowych podsumowań przy stałym kwalifikatorze - "tani" (Cena) oraz sumaryzatorze - "średnia" (Powierzchnia działki) dla różnych kwantyfikatorów:

1. S_1 - "**Mniej niż ćwierć domów** jest/ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.73	1.0	0.99	0.03	0.0	1.0	0.75	0.86	0.99	0.0	0.0	0.0

2. S_2 - "**Mniej niż ćwierć domów** które są/mają tani cena jest/ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.87	1.0	0.99	0.02	0.01	1.0	0.75	0.86	0.99	0.8	0.88	1.0

3. S_3 - "**Mniej niż połowa domów** jest/ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.22	0.0	0.99	0.03	0.0	1.0	0.62	0.75	0.99	0.0	0.0	0.0

4. S_4 - "**Mniej niż połowa domów** które są/mają niska cena jest/ma bardzo średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.35	0.0	0.99	0.02	0.01	1.0	0.62	0.75	0.99	0.8	0.88	1.0

5. S_5 - "**Około połowy domów** jest/ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.22	0.0	0.99	0.03	0.0	1.0	0.64	0.76	0.99	0.0	0.0	0.0

6. S_6 - "**Około połowy domów** które są/mają tani cena jest/ma w średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.35	0.0	0.99	0.02	0.01	1.0	0.64	0.76	0.99	0.8	0.88	1.0

7. S_7 - "**Większość domów** jest/ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.23	0.0	0.99	0.3	0.0	1.0	0.72	0.78	0.99	0.0	0.0	0.0

8. S_8 - "**Większość domów** które/są mają tani cena jest/ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.36	0.0	0.99	0.02	0.01	1.0	0.72	0.78	0.99	0.8	0.88	1.0

9. S_9 - "**Prawie każdy z domów** jest/ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.24	0.0	0.99	0.03	0.0	1.0	0.84	0.88	0.99	0.0	0.0	0.0

10. S_{10} - "**Prawie każdy z domów** które są/mają tani cena jest/ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.37	0.0	0.99	0.02	0.01	1.0	0.84	0.88	0.99	0.8	0.88	1.0

11. S_{11} - "**Mniej niż 3000 domów** jest ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.72	1.0	0.99	0.03	0.0	1.0	0.68	0.74	0.99	0.0	0.0	0.0

12. S_{12} - "**Około 7500 domów** jest ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.23	0.0	0.99	0.03	0.0	1.0	0.73	0.86	0.99	0.0	0.0	0.0

13. S_{13} - "**Blisko 11000 domów** jest ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.23	0.0	0.99	0.03	0.0	1.0	0.73	0.86	0.99	0.0	0.0	0.0

14. S_{14} - "**Ponad 13000 domów** jest ma średnia powierzchnia działki"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.22	0.0	0.99	0.03	0.0	1.0	0.66	0.73	0.99	0.0	0.0	0.0

Poniżej przedstawiona jest lista przykładowych podsumowań lingwistycznych jednopodmiotowych dla stałego kwantyfikatora "Mniej niż połowa domów", sumaryzatora "młody" (Rok wybudowania) oraz zmiennym kwalifikatorze dla zmiennej lingwistycznej cena.

1. S_{15} (Bez kwalifikatora) - "Mniej niż połowa domów jest/ma młody rok wybudowania"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.23	0.0	0.53	0.95	0.0	1.0	0.62	0.75	0.77	0.0	0.0	0.0

2. S_{16} - "Mniej niż połowa domów które/są maja **bardzo tani** cena jest/ma młody rok wybudowania"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.37	0.0	0.53	0.99	0.04	1.0	0.62	0.75	0.77	0.84	0.88	1.0

3. S_{17} - "Mniej niż połowa domów które/są maja **tani** cena jest/ma młody rok wybudowania"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.37	0.0	0.53	0.98	0.03	1.0	0.62	0.75	0.77	0.8	0.88	1.0

4. S_{18} - "Mniej niż połowa domów które/są maja **w rozsądnej cenie** cena jest/ma młody rok wybudowania"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.37	0.0	0.53	0.95	0.01	1.0	0.62	0.75	0.77	0.8	0.88	1.0

5. S_{19} - "Mniej niż połowa domów które/są maja **drogi** cena jest/ma młody rok wybudowania"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.37	0.0	0.53	0.89	0.06	1.0	0.62	0.75	0.77	0.8	0.88	1.0

6. S_{20} - "Mniej niż połowa domów które/są maja **bardzo drogi** cena jest/ma młody rok wybudowania"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.41	0.16	0.53	0.87	0.08	1.0	0.62	0.75	0.77	0.43	0.47	1.0

Poniżej przedstawiona jest lista przykładowych podsumowań lingwistycznych jednopodmiotowych dla stałego kwantyfikatora "Mniej niż połowa domów", kwalifikatora "drogi" (Cena) oraz zmiennym sumaryzatorze dla zmiennej lingwistycznej odległość od szkoły.

1. S_{21} - "Mniej niż połowa domów jest/ma **bardzo blisko** odległość od szkoły"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.25	0.0	0.76	0.98	0.0	1.0	0.62	0.75	0.84	0.0	0.0	0.0

2. S_{22} - "Mniej niż połowa domów które/są mają drogi cena jest/ma **bardzo blisko** odległość od szkoły"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.38	0.0	0.76	0.97	0.01	1.0	0.62	0.75	0.84	0.8	0.88	1.0

3. S_{23} - "Mniej niż połowa domów jest/ma **w pobliżu** odległość od szkoły"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.2	0.0	0.6	0.22	0.0	1.0	0.62	0.75	0.76	0.0	0.0	0.0

4. S_{24} - "Mniej niż połowa domów które/są mają drogi cena jest/ma **w pobliżu** odległość od szkoły"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.33	0.0	0.6	0.19	0.03	1.0	0.62	0.75	0.76	0.8	0.88	1.0

5. S_{25} - "Mniej niż połowa domów jest/ma **daleko** odległość od szkoły"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.19	0.0	0.6	0.01	0.0	1.0	0.62	0.75	0.76	0.0	0.0	0.0

6. S_{26} - "Mniej niż połowa domów które/są mają drogi cena jest/ma **daleko** odległość od szkoły"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.32	0.0	0.6	0.01	0.0	1.0	0.62	0.75	0.76	0.8	0.88	1.0

7. S_{27} - "Mniej niż połowa domów jest/ma **bardzo daleko** odległość od szkoły"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.18	0.0	0.56	0.0	0.0	1.0	0.62	0.75	0.64	0.0	0.0	0.0

8. S_{28} - "Mniej niż połowa domów które/są mają drogi cena jest/ma **bardzo daleko** odległość od szkoły"

T	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
0.31	0.0	0.56	0.0	0.0	1.0	0.62	0.75	0.64	0.8	0.88	1.0

5. Wielopodmiotowe podsumowania lingwistyczne i ich miary jakości

Do oceny jakości podsumowań wielo-podmiotowych korzystaliśmy z miary T [3] dla podsumowań wielopodmiotowych, przyjmującej wartości z zakresu $[0\ 1]$. Miara ta określa stopień prawdziwości podsumowania (jak bardzo jest bliskie prawdzie). Podmioty wykorzystywane w naszych podsumowaniach to:

- Domy mające więcej niż 3 sypialnie
- Domy mające 3 lub mniej sypialnie

Nasza aplikacja implementuje wszystkie 4 formy podsumowań lingwistycznych wielopodmiotowych, których implementacje przeprowadziliśmy na podstawie wzorów zawartych w książce [3]. Poniżej przedstawione są podsumowania wykorzystujące różne kwantyfikatory względne oraz przykłady podsumowań dla każdej z form oraz wynik miary T dla każdego podsumowania.

Formy podsumowań wielopodmiotowych

- Pierwsza forma - nie zawiera kwalifikatora w podsumowaniu.
- Druga forma - podsumowanie uwzględnia kwalifikator który dotyczy drugiego podmiotu w zdaniu.
- Trzecia forma - podsumowanie uwzględnia kwalifikator, który dotyczy pierwszego podmiotu w zdaniu.
- Czwarta forma - forma uwzględniająca hierarchie ilościowa pomiędzy dwoma podmiotami na podstawie słowa kluczowego "Więcej".

Pierwsza forma

Przykładowe podsumowania wielopodmiotowe dla pierwszej formy bez kwalifikatora dla różnych kwantyfikatorów względnych.

1. S_{29} - "**Mniej niż ćwierć domów** mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma na południu wysokość geograficzna" $[0.0]$
2. S_{30} - "**Mniej niż ćwierć domów** mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma na południu wysokość geograficzna" $[0.0]$
3. S_{31} - "**Mniej niż połowa domów** mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma na południu wysokość geograficzna" $[0.44]$
4. S_{32} - "**Mniej niż połowa domów** mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma na południu wysokość geograficzna" $[0.0]$
5. S_{33} - "**Około połowy domów** mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma na południu wysokość geograficzna" $[0.85]$
6. S_{34} - "**Około połowy domów** mających więcej niż 3 sypialnie w po-

równaniu do tych mających 3 lub mniej sypialnie jest/ma na południu wysokość geograficzna” [0.85]

7. S_{35} - ”**Większość domów** mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma na południu wysokość geograficzna” [0.0]
8. S_{36} - **Większość domów** mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma na południu wysokość geograficzna” [0.0]

Przykładowe podsumowania wielopodmiotowe dla stałego kwantyfikatora ”Okolo połowy domów” przy zmiennym sumaryzatorze dla zmiennej lingwistycznej wysokość geograficzna.

1. S_{37} - ”Okolo połowy domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **na południu** wysokość geograficzna” [0.85]
2. S_{38} - ”Okolo połowy domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **na południu** wysokość geograficzna” [0.85]
3. S_{39} - ”Okolo połowy domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **w centrum** wysokość geograficzna” [1.0]
4. S_{40} - ”Okolo połowy domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **w centrum** wysokość geograficzna” [1.0]
5. S_{41} - ”Okolo połowy domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **na północy** wysokość geograficzna” [1.0]
6. S_{42} - ”Okolo połowy domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **na północy** wysokość geograficzna” [1.0]

Druga forma

Przykładowe podsumowania wielopodmiotowe dla drugiej formy przy stałym kwantyfikatorze ”Większość domów”, sumaryzatorze ”tani” dla zmiennej cena oraz zmiennym kwalifikatorze dla zmiennej lingwistycznej powierzchnia mieszkania.

1. S_{43} - ”Większość domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają **klaustrofobiczne** powierzchnia mieszkania jest/ma tani cena” [0.0]
2. S_{44} - ”Większość domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają **klaustrofobiczne** powierzchnia mieszkania jest/ma tani cena” [0.25]
3. S_{45} - ”Większość domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają **przeciętne** powierzchnia mieszkania jest/ma tani cena” [0.63]

4. S_{46} - "Większość domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają **przeciętne** powierzchnia mieszkania jest/ma tani cena" [1.0]
5. S_{47} - "Większość domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają **przestrzenne** powierzchnia mieszkania jest/ma tani cena" [0.0]
6. S_{48} - "Większość domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają **przestrzenne** powierzchnia mieszkania jest/ma tani cena" [0.0]

Przykładowe podsumowania wielopodmiotowe dla drugiej formy przy zmiennym kwantyfikatorze dla stałego kwalifikatora "drogi" dla zmiennej cena oraz sumaryzatora "bardzo mała" dla zmiennej powierzchnia działki.

1. S_{49} - "**Mniej niż ćwierć domów** mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [0.0]
2. S_{50} - "**Mniej niż ćwierć domów** mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [0.0]
3. S_{51} - "**Mniej niż połowa domów** mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [0.0]
4. S_{52} - "**Mniej niż połowa domów** mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [0.0]
5. S_{53} - "**Około połowy domów** mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [0.0]
6. S_{54} - "**Około połowy domów** mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [0.0]
7. S_{55} - "**Większość domów** mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [0.0]
8. S_{56} - "**Większość domów** mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [0.0]
9. S_{57} - "**Prawie każdy z domów** mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [0.86]
10. S_{58} - "**Prawie każdy z domów** mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają drogi cena jest/ma bardzo mała powierzchnia działki" [1.0]

Przykładowe podsumowania wielopodmiotowe dla drugiej formy przy stałym kwantyfikatorze "Około połowy domów", kwalifikatorze "drogi" zmiennej cena oraz zmiennym sumaryzatorze zmiennej powierzchnia mieszkania.

1. S_{59} - "Okolo połowy domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają drogi cena jest/ma **klaustrofobiczne** powierzchnia mieszkania" [0.0]
2. S_{60} - "Okolo połowy domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają drogi cena jest/ma **klaustrofobiczne** powierzchnia mieszkania" [0.0]
3. S_{61} - "Okolo połowy domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają drogi cena jest/ma **przeciętne** powierzchnia mieszkania" [0.0]
4. S_{62} - "Okolo połowy domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają drogi cena jest/ma **przeciętne** powierzchnia mieszkania" [0.0]
5. S_{63} - "Okolo połowy domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie które są/mają drogi cena jest/ma **przestrzenne** powierzchnia mieszkania" [0.59]
6. S_{64} - "Okolo połowy domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie które są/mają drogi cena jest/ma **przestrzenne** powierzchnia mieszkania" [0.0]

Trzecia forma

Przykładowe podsumowania wielopodmiotowe dla trzeciej formy przy stałym kwantyfikatorze "Większość domów", sumaryzatorze "tani" dla zmiennej cena oraz zmiennym kwalifikatorze dla zmiennej lingwistycznej powierzchnia mieszkania. Analogiczne do zestawu przykładów dla drugiej formy lecz teraz związanych z pierwszym podmiotem.

1. S_{65} - "Większość domów mających 3 lub mniej sypialnie które są/mają **klaustrofobiczne** powierzchnia mieszkania w porównaniu do tych mających więcej niż 3 sypialnie jest/ma tani cena" [0.0]
2. S_{66} - "Większość domów mających więcej niż 3 sypialnie które są/mają **klaustrofobiczne** powierzchnia mieszkania w porównaniu do tych mających 3 lub mniej sypialnie jest/ma tani cena" [0.0]
3. S_{67} - "Większość domów mających 3 lub mniej sypialnie które są/mają **przeciętne** powierzchnia mieszkania w porównaniu do tych mających więcej niż 3 sypialnie jest/ma tani cena" [0.0]
4. S_{68} - "Większość domów mających więcej niż 3 sypialnie które są/mają **przeciętne** powierzchnia mieszkania w porównaniu do tych mających 3 lub mniej sypialnie jest/ma tani cena" [0.0]
5. S_{69} - "Większość domów mających 3 lub mniej sypialnie które są/mają **przestrzenne** powierzchnia mieszkania w porównaniu do tych mających więcej niż 3 sypialnie jest/ma tani cena" [0.0]
6. S_{70} - "Większość domów mających więcej niż 3 sypialnie które są/mają **przestrzenne** powierzchnia mieszkania w porównaniu do tych mających 3 lub mniej sypialnie jest/ma tani cena" [0.0]

Przykładowe podsumowania wielopodmiotowe dla trzeciej formy przy

zmiennym kwantyfikatorze dla stałego kwalifikatora "drogi" dla zmiennej cena oraz sumaryzatora "bardzo mała" dla zmiennej powierzchnia działki.

1. S_{71} - "**Mniej niż ćwierć domów** mających 3 lub mniej sypialnie które są/mają drogi cena w porównaniu do tych mających więcej niż 3 sypialnie jest/ma bardzo mała powierzchnia działki" [0.84]
2. S_{72} - "**Mniej niż ćwierć domów** mających więcej niż 3 sypialnie które są/mają drogi cena w porównaniu do tych mających 3 lub mniej sypialnie jest/ma bardzo mała powierzchnia działki" [0.76]
3. S_{73} - "**Mniej niż połowa domów** mających 3 lub mniej sypialnie które są/mają drogi cena w porównaniu do tych mających więcej niż 3 sypialnie jest/ma bardzo mała powierzchnia działki" [0.0]
4. S_{74} - "**Mniej niż połowa domów** mających więcej niż 3 sypialnie które są/mają drogi cena w porównaniu do tych mających 3 lub mniej sypialnie jest/ma bardzo mała powierzchnia działki" [0.0]
5. S_{75} - "**Około połowy domów** mających 3 lub mniej sypialnie które są/mają drogi cena w porównaniu do tych mających więcej niż 3 sypialnie jest/ma bardzo mała powierzchnia działki" [0.0]
6. S_{76} - "**Około połowy domów** mających więcej niż 3 sypialnie które są/mają drogi cena w porównaniu do tych mających 3 lub mniej sypialnie jest/ma bardzo mała powierzchnia działki" [0.0]
7. S_{77} - "**Większość domów** mających 3 lub mniej sypialnie które są/mają drogi cena w porównaniu do tych mających więcej niż 3 sypialnie jest/ma bardzo mała powierzchnia działki" [0.0]
8. S_{78} - "**Większość domów** mających więcej niż 3 sypialnie które są/mają drogi cena w porównaniu do tych mających 3 lub mniej sypialnie jest/ma bardzo mała powierzchnia działki" [0.0]
9. S_{79} - "**Prawie każdy z domów** mających 3 lub mniej sypialnie które są/mają drogi cena w porównaniu do tych mających więcej niż 3 sypialnie jest/ma bardzo mała powierzchnia działki" [0.0]
10. S_{80} - "**Prawie każdy z domów** mających więcej niż 3 sypialnie które są/mają drogi cena w porównaniu do tych mających 3 lub mniej sypialnie jest/ma bardzo mała powierzchnia działki" [0.0]

Przykładowe podsumowania wielopodmiotowe dla trzeciej formy przy stałym kwantyfikatorze "Około połowy domów", kwalifikatorze "drogi" zmiennej cena oraz zmiennym sumaryzatorze zmiennej powierzchnia mieszkania.

1. S_{81} - "Około połowy domów mających 3 lub mniej sypialnie które są/mają drogi cena w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **klaustrofobiczne** powierzchnia mieszkania" [0.0]
2. S_{82} - "Około połowy domów mających więcej niż 3 sypialnie które są/mają drogi cena w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **klaustrofobiczne** powierzchnia mieszkania" [0.0]
3. S_{83} - "Około połowy domów mających 3 lub mniej sypialnie które są/mają drogi cena w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **przeciętne** powierzchnia mieszkania" [0.0]

4. S_{84} - "Okolo połowy domów mających więcej niż 3 sypialnie które są/mają drogi cena w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **przeciętne** powierzchnia mieszkania" [0.0]
5. S_{85} - "Okolo połowy domów mających 3 lub mniej sypialnie które są/mają drogi cena w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **przestrzenne** powierzchnia mieszkania" [0.0]
6. S_{86} - "Okolo połowy domów mających więcej niż 3 sypialnie które są/mają drogi cena w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **przestrzenne** powierzchnia mieszkania" [0.59]

Czwarta forma

Przykładowe podsumowania wielopodmiotowe dla czwartej formy przy zmiennym sumaryzatorze dla zmiennej odległość od szkoły.

1. S_{87} - "Większość domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **bardzo blisko** odległość do szkoły" [0.56]
2. S_{88} - "Większość domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **bardzo blisko** odległość do szkoły" [0.37]
3. S_{89} - "Większość domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **w pobliżu** odległość do szkoły" [0.05]
4. S_{90} - "Większość domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **w pobliżu** odległość do szkoły" [0.02]
5. S_{91} - "Większość domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **daleko** odległość do szkoły" [0.0]
6. S_{92} - "Większość domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **daleko** odległość do szkoły" [0.0]
7. S_{93} - "Większość domów mających 3 lub mniej sypialnie w porównaniu do tych mających więcej niż 3 sypialnie jest/ma **bardzo daleko** odległość do szkoły" [0.0]
8. S_{94} - "Większość domów mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej sypialnie jest/ma **bardzo daleko** odległość do szkoły" [0.0]

6. Dyskusja, wnioski

- Wnioski podsumowań jednopodmiotowych (wyniki sekcja 4):
 - Podsumowania lingwistyczne z kwalifikatorem mają zazwyczaj lepszą ogólną miarę dobroci podsumowania przez uwzględnienie miar T_9 , T_{10} oraz T_{11} .

- Zmiany kwantyfikatora wskazują nam jedynie wyniki dla poszczególnych przedziałów co do ilości obiektów bazy. Na podstawie przykładów $S_1 - S_{14}$ widzimy że najbliższej prawdy jest to że mniej niż ćwierć domów w naszej bazie które są tanie mają średnią powierzchnię działki.
- Kwantyfikator absolutny "Mniej niż 3000 domów" w podsumowaniu S_{11} , w porównaniu do innych wyróżnia się wysokim stopniem prawdziwości T_1 co wskazuje na wysoką trafność naszego przykładu co do ilości obiektów naszej bazy.
- W przykładach $S_{21} - S_{28}$ widzimy pewną zależność między miarą T_3 a zmianami związanymi z modyfikacją sumaryzatora. Widzimy że wraz ze wzrostem odległości od szkoły spada miara pokrycia z kwalifikatorem. Oznacza to w naszym przypadku, że drogie domy charakteryzują się mniejszą odległością do szkoły.
- Wnioski podsumowań wielopodmiotowych (wyniki sekcja 5):
 - Zmiana kwantyfikatora względnego w pierwszej formie oraz miara wskazuje nam na stopień prawdziwości co do ilości obiektów w bazie. Na przykładach S_{33} oraz S_{34} widzimy że miary T przyjmują największą, jednakową wartość. Oznacza to że prawdziwe jest to, że około połowy domów przyjmujących zależności między podmiotami znajduje się na południu.
 - Na podstawie przykładów $S_{37} - S_{42}$ widzimy że kwantyfikator "Okolo połowy domów" okazał się bardzo trafny w określaniu wysokości geograficznej w każdym kierunku. Jednak większym stopniem prawdziwości charakteryzuje się centrum oraz północ.
 - Na podstawie podsumowań $S_{43} - S_{48}$ oraz modyfikacją kwalifikatora jesteśmy w stanie określić stopień prawdziwości dla danej ilości obiektów bazy. Widzimy, że najbliższe prawdy jest to, że większość domów z bazy mających 3 lub mniej/więcej niż 3 sypialnie w porównaniu do takich mających 3 lub mniej/więcej niż 3 sypialnie i mają przeciętną powierzchnię mieszkania jest tanich.
 - Na podstawie przykładów $S_{49} - S_{58}$ widzimy że wysoka cena nie idzie w parze z dużą powierzchnią działki.
 - Na podstawie podsumowań $S_{59} - S_{64}$ widzimy że najlepsze dopasowanie ma przykład S_{63} . Dysproporcja w przykładzie S_{64} dla tego samego sumaryzatora wynika odwrotnym, w tym przypadku gorszym pokryciem w kwalifikatorze. Z tych przykładów wynika, że bliższe prawdzie jest to że domy mające więcej niż 3 sypialnie zazwyczaj są droższe, a za tym idzie przestrzenna powierzchnia mieszkania.
 - W przykładach trzeciej formy $S_{71} - S_{80}$ widzimy odwrotną zależność co do prawdziwości wyników z przykładów drugiej formy $S_{49} - S_{58}$ dla analogicznych parametrów. Wynika to z różnicy, z którym podmiotem związany jest kwalifikator opisujący cenę domu.
 - Jeżeli chodzi o czwartą formę to możemy wywnioskować, że najwięcej domów z bazy jest tych które mają bardzo bliską odległość. Natomiast mimo wszystko wśród tych domów mających bliską odległość jest więcej domów mających 3 lub mniej sypialni w porównaniu do

tych mających więcej niż 3 od tych mających więcej niż 3 sypialnie w porównaniu do tych mających 3 lub mniej.

Literatura

- [1] Strona internetowa kaggle.com [online] [dostęp 13.05.2022] <https://www.kaggle.com>
- [2] Baza danych perth-house-prices. [online] [dostęp 13.05.2022] <https://www.kaggle.com/datasets/syuzai/perth-house-prices>
- [3] A. Niewiadomski, Zbiory rozmyte typu 2. Zastosowania w reprezentowaniu informacji. Seria „Problemy współczesnej informatyki” pod redakcją L. Rutkowskiego. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2019.
- [4] S. Zadrozny, Zapytania nieprecyzyjne i lingwistyczne podsumowania baz danych, EXIT, 2006, Warszawa
- [5] A. Niewiadomski, Methods for the Linguistic Summarization of Data: Applications of Fuzzy Sets and Their Extensions, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2008.

Literatura zawiera wyłącznie źródła recenzowane i/lub o potwierdzonej wiarygodności, możliwe do weryfikacji i cytowane w sprawozdaniu.