

Data oddania: \_\_\_\_\_

Ocena: \_\_\_\_\_

Julia Szymańska      224441  
Przemysław Zdrzałik    224466

## Projekt 2. Podsumowania lingwistyczne relacyjnych baz danych

### 1. Cel

Celem projektu jest stworzenie aplikacji pozwalającej na generowanie podsumowań lingwistycznych [1] w oparciu o kwantyfikatory rozmyte [2], co oznacza opisanie danych liczbowych ze zbioru danych [3] językiem quasi-naturalnym - pozornie naturalnym.

Przykładem podsumowania lingwistycznego [1] w formie

$$Q \ P \text{ jest } S \ [T] \quad (1)$$

jest: *Wiele wypadków jest przy ujemnej temperaturze  $[0, 76]$* , gdzie  $Q$  jest kwantyfikatorem lingwistycznym,  $P$  podmiotem podsumowań,  $S$  sumaryzatorem, a  $T[0, 1]$  stopniem prawdziwości.

Przykładem drugiego podsumowania lingwistycznego w formie:

$$Q \ P \text{ bedacych } W \text{ jest } S \ [T] \quad (2)$$

jest: *Wiele wypadków będących podczas deszczu, jest przy ujemnej temperaturze  $[0.68]$* , gdzie  $Q$  jest kwantyfikatorem lingwistycznym,  $P$  podmiotem podsumowań,  $S$  sumaryzatorem,  $W$  kwantyfikatorem reprezentującym dodatkowe własności obiektów, a  $T[0, 1]$  stopniem prawdziwości.

Analizowany zbiór danych zawiera liczbowe informacje o ponad 3 milionach wypadków samochodowych w 49 stanach Zjednoczonych Stanów Ameryki,

mających miejsce od lutego 2016 do grudnia 2020 [3]. Zbiór danych składa się z 47 kolumn. W tym celu wykonania podsumowania lingwistycznego zostaną wykorzystane metody logiki rozmytej [4]. Logika rozmyta pozwala na opisanie wartości zapisanych językiem naturalnym za pomocą zrozumiałych określeń jak: mało, dużo, około połowy. W projekcie zostaną wykorzystane kwantyfikatory lingwistyczne względne takie jak: niewiele, około połowy oraz kwantyfikatory lingwistyczne absolutne takie jak: około jednego, około stu.

## 2. Charakterystyka podsumowywanej bazy danych













W programie został użyty zbiór danych [3] znajdujący się w pliku CSV, który został przekształcony w bazę danych.

Zbiór danych zawiera informacje o ponad 3 milionach wypadków samochodowych w 49 stanach Zjednoczonych Stanów Ameryki, mających miejsce od lutego 2016 do grudnia 2020. Spośród 47 kolumn znajdujących się w zbiorze danych, wybraliśmy następujące 11 kolumn:

- Czas rozpoczęcia - Start\_Time - czas rozpoczęcia się wypadku w lokalnej strefie czasowej, przyjmuje wartości od 8 lutego 2016, do 31 grudnia 2020. Wartość kolumny zostanie zamieniona na wartość całkowitą oznaczającą liczbę sekund od początku 1970 roku.
- Czas zakończenia - End\_Time - czas zakończenia się wypadku w lokalnej strefie czasowej, przyjmuje wartości od 8 lutego 2016, do 1 stycznia 2021. Wartość kolumny zostanie zamieniona na wartość całkowitą oznaczającą liczbę sekund od początku 1970 roku.
- Odległość - Distance - długość odcinka ulicy wyrażony w milach, na którego miał wpływ wypadek. Przyjmuje wartości zmiennoprzecinkowe od 0 do 334, gdzie zdecydowana większość danych mieści się w przedziale od 0.00 do 4.00.
- Temperatura - Temperature - temperatura powietrza wyrażona w Fahrenheit'ach, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od -16.00 do 104.00. Temperature można opisać jako bardzo zimną, zimną, umiarkowaną, ciepłą, bardzo ciepłą. Oczywiście jest to opis subiektywny.
- Temperatura odczuwalna - Wind\_Chill - temperatura odczuwalna wyrażona w Fahrenheit'ach, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od -16.00 do 101.00. Temperaturę odczuwalną można opisać tak samo jak temperaturę.
- Wilgotność - Humidity - wilgotność powietrza wyrażona w procentach w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od 4.00 do 100.00.
- Ciśnienie - Pressure - ciśnienie powietrza wyrażone w inches, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od 27.00 do 32.00. Ciśnienie można opisać jako wysokie, umiarkowane lub niskie.
- Widoczność - Visibility - widoczność wyrażona w milach, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od 0.00 do 12.00. Widoczność można opisać jako dobrą, ograniczoną, słabą.

- Prędkość wiatru - Wind\_Speed - prędkość wiatru wyrażona w milach na godzinę, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od 0.00 do 40.00. Wiatr można opisać jako słaby, umiarkowany, silny.
- Ilość opadów - Precipitation - ilość opadów wyrażona w inches, w momencie, gdy zdarzył się wypadek. Jeśli opady nie występowały to kolumna przyjmuje wartość nan. Przyjmuje wartości zmiennoprzecinkowe od 0.00 do 0.50.

Atrybutom nadawane są opisane zwyczajowe wartości lingwistyczne ze względu na zwiększenie przystępności i ułatwienie szybkiego zrozumienia atrybutu przez człowieka, kiedy ten atrybut nie musi być dokładnie opisany. Przykładowo temperatura, mimo że zrozumiała dla człowieka w postaci liczbowej, jest łatwiejsza do szybszego zrozumienia w postaci tekstowej, a dla ludzi nie ma dużego znaczenia czy temperatura różni się o 1 czy 2 stopnie, wystarczy opisać ją słownie tak jak wcześniej podaliśmy jako bardzo zimną, zimną, umiarkowaną, ciepłą, bardzo ciepłą.

accidents	
 <b>id</b>	bigint
 <b>severity</b>	smallint
 <b>start_time</b>	timestamp with time zone
 <b>end_time</b>	timestamp with time zone
 <b>distance</b>	double precision
 <b>temperature</b>	double precision
 <b>wind_chill</b>	double precision
 <b>humidity</b>	double precision
 <b>pressure</b>	double precision
 <b>visibility</b>	double precision
 <b>wind_speed</b>	double precision
 <b>precipitation</b>	double precision

Powered by yFiles

Rysunek 1. Tabela reprezentująca omawiane dane wykonana w DBMS PostgreSQL

### 3. Atrybuty i liczności obiektów wyrażone zmiennymi lingwistycznymi

Poniżej zostaną przedstawione zmienne lingwistyczne [2] dla jedenastu atrybutów z bazy danych wraz z przypisanymi etykietami w formie funkcji przynależności oraz wzorów analitycznych.

#### 3.1. Czas utrudnień w ruchu drogowym spowodowanych przez wypadek

Na podstawie znajdujących się w bazie danych pól Czas rozpoczęcia (Start\_Time) oraz Czas zakończenia (End\_Time) zostanie obliczony Czas utrudnień w ruchu drogowym (Duration) spowodowanych przez wypadek według wzoru:

$$Duration = End\_Time - Start\_Time \quad (3)$$

Przedstawienie Czasu utrudnień w ruchu drogowym (Duration) spowodowanych przez wypadek jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: krótki, średni, długi.

$$\mu_{czasTrwaniaPonizejGodziny}(x) = \begin{cases} \frac{1-x}{1} & \text{dla } 0 < x \leq 1 \end{cases} \quad (4)$$

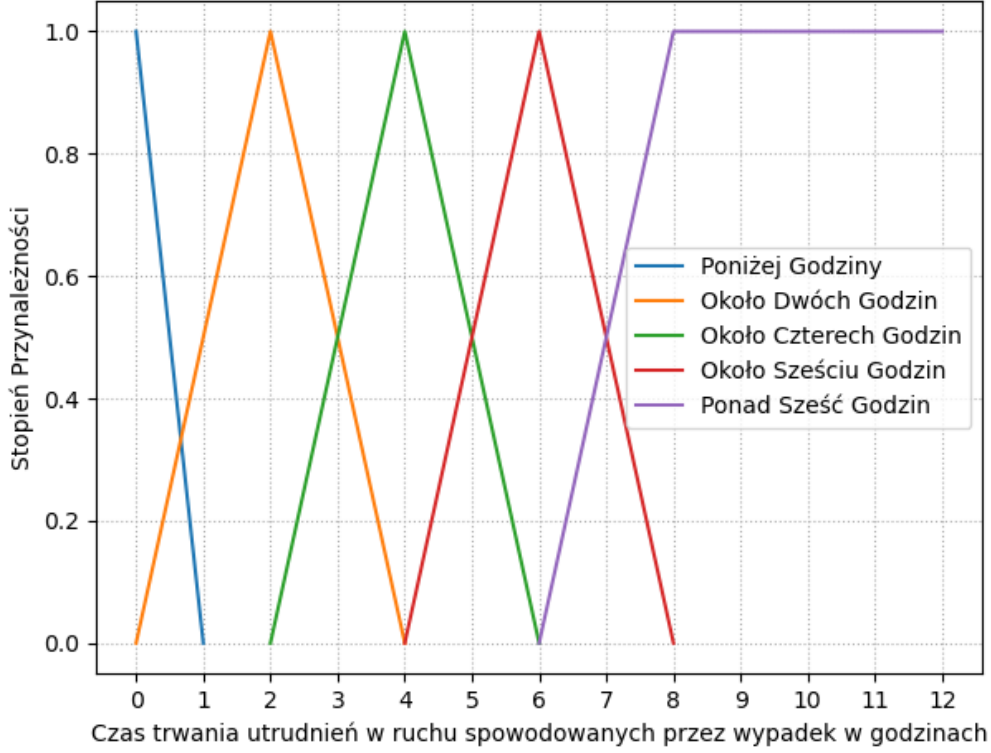
$$\mu_{czasTrwaniaOkoloDwochGodzin}(x) = \begin{cases} \frac{x}{2} & \text{dla } 0 < x \leq 2 \\ \frac{4-x}{2} & \text{dla } 2 < x \leq 4 \end{cases} \quad (5)$$

$$\mu_{czasTrwaniaOkoloCzterechGodzin}(x) = \begin{cases} \frac{x-2}{2} & \text{dla } 2 < x \leq 4 \\ \frac{6-x}{2} & \text{dla } 4 < x \leq 6 \end{cases} \quad (6)$$

$$\mu_{czasTrwaniaOkoloSzesciuGodzin}(x) = \begin{cases} \frac{x-4}{2} & \text{dla } 4 < x \leq 6 \\ \frac{8-x}{2} & \text{dla } 6 < x \leq 8 \end{cases} \quad (7)$$

$$\mu_{czasTrwaniaPonadSzescGodzin}(x) = \begin{cases} \frac{x-6}{2} & \text{dla } 6 < x \leq 8 \\ 1 & \text{dla } 8 \leq x \end{cases} \quad (8)$$

gdzie:  $\mu_{czasTrwaniaPonizejGodziny}$ ,  $\mu_{czasTrwaniaOkoloDwochGodzin}$ ,  $\mu_{czasTrwaniaOkoloCzterechGodzin}$ ,  $\mu_{czasTrwaniaOkoloSzesciuGodzin}$ ,  $\mu_{czasTrwaniaPonadSzescGodzin}$  - funkcje przynależności,  $x$  - czas trwania wypadku.



Rysunek 2. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej czas utrudnień w ruchu drogowym (Duration) spowodowanych przez wypadek.

### 3.2. Odległość

Przedstawienie odległości jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: krótki, długi.

$$\mu_{\text{OdlegloscPonizejPolMili}}(x) = \begin{cases} 1 & \text{dla } x \leq 0.5 \end{cases} \quad (9)$$

$$\mu_{\text{OdlegloscOkoloJednejMili}}(x) = \begin{cases} x & \text{dla } 0 < x \leq 1 \\ 2 - x & \text{dla } 1 < x \leq 2 \end{cases} \quad (10)$$

$$\mu_{\text{OdlegloscOkoloTrzechMili}}(x) = \begin{cases} \frac{x-1}{2} & \text{dla } 1 < x \leq 3 \\ \frac{5-x}{2} & \text{dla } 3 < x \leq 5 \end{cases} \quad (11)$$

$$\mu_{\text{OdlegloscPonadTrzyMile}}(x) = \begin{cases} \frac{x-3}{2} & \text{dla } 3 < x \leq 5 \\ 1 & \text{dla } 5 \leq x \end{cases} \quad (12)$$

gdzie:  $\mu_{\text{OdlegloscPonizejPolMili}}$ ,  $\mu_{\text{OdlegloscOkoloJednejMili}}$ ,  $\mu_{\text{OdlegloscOkoloTrzechMili}}$ ,  $\mu_{\text{OdlegloscPonadTrzyMile}}$  - funkcje przynależności,  $x$  - odległość.



Rysunek 3. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej odległości.

### 3.3. Tempertura i temperatura odczuwalna

Przedstawienie temperatury oraz temperatury odczuwalnej jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: bardzo zimno, zimno, umiarkowanie, ciepło, bardzo ciepło.

$$\mu_{temperaturaBardzoZimno}(x) = \begin{cases} 1 & \text{dla } x \leq 14 \\ \frac{23-x}{9} & \text{dla } 14 < x \leq 23 \end{cases} \quad (13)$$

$$\mu_{temperaturaZimno}(x) = \begin{cases} \frac{x-14}{9} & \text{dla } 14 < x \leq 23 \\ 1 & \text{dla } 23 < x < 44 \\ \frac{54-x}{10} & \text{dla } 44 < x \leq 54 \end{cases} \quad (14)$$

$$\mu_{temperaturaUmiarkowanie}(x) = \begin{cases} \frac{x-44}{10} & \text{dla } 44 < x \leq 54 \\ 1 & \text{dla } 54 < x < 63 \\ \frac{71-x}{8} & \text{dla } 63 < x \leq 71 \end{cases} \quad (15)$$

$$\mu_{temperaturaCieplo}(x) = \begin{cases} \frac{x-63}{8} & \text{dla } 63 < x \leq 71 \\ 1 & \text{dla } 71 < x < 80 \\ \frac{90-x}{10} & \text{dla } 80 < x \leq 90 \end{cases} \quad (16)$$

$$\mu_{temperaturaBardzoCieplo}(x) = \begin{cases} \frac{x-80}{10} & \text{dla } 80 < x \leq 90 \\ 1 & \text{dla } 90 \leq x \end{cases} \quad (17)$$

gdzie:  $\mu_{temperaturaBardzoZimna}$ ,  $\mu_{temperaturaZimna}$ ,  $\mu_{temperaturaUmiarkowana}$ ,  $\mu_{temperaturaCiepła}$ ,  $\mu_{temperaturaBardzoCiepła}$  - funkcje przynależności,  $x$  - temperatura, temperatura odczuwalna.



Rysunek 4. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej temperatury.



### 3.4. Wilgotność

Przedstawienie wilgotności jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: suche, umiarkowane, wilgotne.

$$\mu_{wilgotnoscBardzoSuche}(x) = \exp\left(\frac{-(x-4)^2}{128}\right) \quad (18)$$

$$\mu_{wilgotnoscSuche}(x) = \exp\left(\frac{-(x-25)^2}{128}\right) \quad (19)$$

$$\mu_{wilgotnoscUmiarkowane}(x) = \exp\left(\frac{-(x-50)^2}{128}\right) \quad (20)$$

$$\mu_{wilgotnoscWilgotne}(x) = \exp\left(\frac{-(x-75)^2}{128}\right) \quad (21)$$

$$\mu_{wilgotnoscBardzoWilgotne}(x) = \begin{cases} \frac{x-60}{10} & \text{dla } 60 < x \leq 70 \\ 1 & \text{dla } 70 < x < 80 \\ \frac{90-x}{10} & \text{dla } 80 < x \leq 90 \end{cases} \quad (22)$$

gdzie:  $\mu_{wilgotnoscBardzoSuche}$ ,  $\mu_{wilgotnoscSuche}$ ,  $\mu_{wilgotnoscUmiarkowane}$ ,  $\mu_{wilgotnoscWilgotne}$ ,  $\mu_{wilgotnoscBardzoWilgotne}$  - funkcje przynależności,  $x$  - wilgotność.



Rysunek 5. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej wilgotności.

### 3.5. Ciśnienie

Przedstawienie ciśnienia jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: niskie, umiarkowane, wysokie.

$$\mu_{cisnienieBardzoNiskie}(x) = \exp\left(\frac{-(x - 27)^2}{0.18}\right) \quad (23)$$

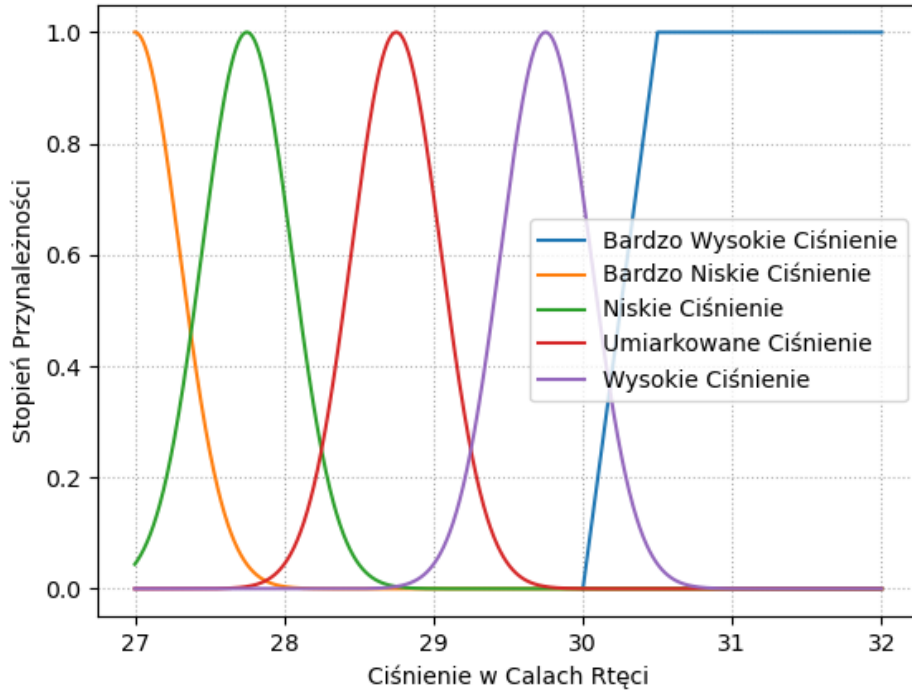
$$\mu_{cisnienieNiskie}(x) = \exp\left(\frac{-(x - 27.75)^2}{0.18}\right) \quad (24)$$

$$\mu_{cisnienieUmiarkowane}(x) = \exp\left(\frac{-(x - 28.75)^2}{0.18}\right) \quad (25)$$

$$\mu_{cisnienieWysokie}(x) = \exp\left(\frac{-(x - 29.75)^2}{0.18}\right) \quad (26)$$

$$\mu_{cisnienieBardzoWysokie}(x) = \begin{cases} \frac{x-30}{0.5} & \text{dla } 30 < x \leq 30.5 \\ 1 & \text{dla } 30.5 \leq x \end{cases} \quad (27)$$

gdzie:  $\mu_{cisnienieBardzoNiskie}$ ,  $\mu_{cisnienieNiskie}$ ,  $\mu_{cisnienieUmiarkowane}$ ,  $\mu_{cisnienieWysokie}$ ,  $\mu_{cisnienieBardzoWysokie}$  - funkcje przynależności,  $x$  - ciśnienie.



Rysunek 6. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej ciśnienia.

### 3.6. Widoczność

Przedstawienie widoczności jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: brak, słaba, ograniczona, dobra.

$$\mu_{widocznoscBrak}(x) = 1 \quad \text{dla } x = 0 \quad (28)$$

$$\mu_{widocznoscSlaba}(x) = \begin{cases} 1 & \text{dla } x \leq 0.1 \\ \frac{0.1-x}{0.2} & \text{dla } 0.1 < x \leq 0.3 \end{cases} \quad (29)$$

$$\mu_{widocznoscOgraniczona}(x) = \begin{cases} \frac{x-0.1}{0.2} & \text{dla } 0.1 < x \leq 0.3 \\ 1 & \text{dla } 0.3 < x < 0.7 \\ \frac{1-x}{0.3} & \text{dla } 0.7 < x \leq 1 \end{cases} \quad (30)$$

$$\mu_{widocznoscDobra}(x) = \begin{cases} \frac{x-0.7}{0.3} & \text{dla } 0.7 < x \leq 1 \\ 1 & \text{dla } 1 < x < 2 \\ \frac{3-x}{1} & \text{dla } 2 < x \leq 3 \end{cases} \quad (31)$$

$$\mu_{widocznoscPelna}(x) = \begin{cases} \frac{x-3}{1} & \text{dla } 2 < x \leq 3 \\ 1 & \text{dla } 3 \leq x \end{cases} \quad (32)$$

gdzie:  $\mu_{widocznoscBrak}$ ,  $\mu_{widocznoscSlaba}$ ,  $\mu_{widocznoscOgraniczona}$ ,  $\mu_{widocznoscDobra}$ ,  $\mu_{widocznoscPelna}$  - funkcje przynależności,  $x$  - widoczność.



Rysunek 7. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej widoczności.

### 3.7. Prędkość wiatru

Przedstawienie predkości wiatru jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: brak, słaby, umiarkowany, silny, wicher, huragan.

$$\mu_{wiatrBrak}(x) = 1 \quad dla \quad x = 0 \quad (33)$$

$$\mu_{wiatrSlaby}(x) = \begin{cases} 1 & dla \quad x \leq 3 \\ \frac{3-x}{0.5} & dla \quad 3 < x \leq 3.5 \end{cases} \quad (34)$$

$$\mu_{wiatrUmiarkowany}(x) = \begin{cases} \frac{x-3}{0.5} & dla \quad 3 < x \leq 3.5 \\ 1 & dla \quad 3.5 < x < 8 \\ \frac{9-x}{1} & dla \quad 8 < x \leq 9 \end{cases} \quad (35)$$

$$\mu_{wiatrSilny}(x) = \begin{cases} \frac{x-8}{1} & dla \quad 8 < x \leq 9 \\ 1 & dla \quad 9 < x < 17 \\ \frac{20-x}{3} & dla \quad 17 < x \leq 20 \end{cases} \quad (36)$$

$$\mu_{wiatrWicher}(x) = \begin{cases} \frac{x-17}{3} & dla \quad 17 < x \leq 20 \\ 1 & dla \quad 20 < x < 27 \\ \frac{30-x}{3} & dla \quad 27 < x \leq 30 \end{cases} \quad (37)$$

$$\mu_{wiatrHuragan}(x) = \begin{cases} \frac{x-40}{10} & dla \quad 30 < x \leq 40 \\ 1 & dla \quad 40 \leq x \end{cases} \quad (38)$$

gdzie:  $\mu_{wiatrBrak}$ ,  $\mu_{wiatrSlaby}$ ,  $\mu_{wiatrUmiarkowany}$ ,  $\mu_{wiatrSilny}$ ,  $\mu_{wiatrWicher}$ ,  $\mu_{wiatrHuragan}$  - funkcje przynależności,  $x$  - prędkość wiatru.



Rysunek 8. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej prędkości wiatru.

### 3.8. Opady

Przedstawienie opadów jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: brak, niewielkie, umiarkowane, duże.

$$\mu_{opadyBrak}(x) = 1 \quad \text{dla } x = 0 \quad (39)$$

$$\mu_{opadyNiewielkie}(x) = \begin{cases} 1 & \text{dla } x \leq 0.1 \\ \frac{0.1-x}{0.1} & \text{dla } 0.1 < x \leq 0.2 \end{cases} \quad (40)$$

$$\mu_{opadyUmiarkowane}(x) = \begin{cases} \frac{x-0.1}{0.1} & \text{dla } 0.1 < x \leq 0.2 \\ 1 & \text{dla } 0.2 < x < 0.3 \\ \frac{0.35-x}{0.05} & \text{dla } 0.3 < x \leq 0.35 \end{cases} \quad (41)$$

$$\mu_{opadyDuze}(x) = \begin{cases} \frac{x-0.35}{0.05} & \text{dla } 0.3 < x \leq 0.35 \\ 1 & \text{dla } 0.35 < x < 0.4 \\ \frac{0.45-x}{0.05} & \text{dla } 0.4 < x \leq 0.45 \end{cases} \quad (42)$$

$$\mu_{opadyBardzoDuze}(x) = \begin{cases} \frac{x-0.4}{0.05} & \text{dla } 0.4 < x \leq 0.45 \\ 1 & \text{dla } 0.45 \leq x \end{cases} \quad (43)$$

gdzie:  $\mu_{opadyBrak}$ ,  $\mu_{opadyNiewielkie}$ ,  $\mu_{opadyUmiarkowane}$ ,  $\mu_{opadyDuze}$ ,  $\mu_{opadyBardzoDuze}$  - funkcje przynależności,  $x$  - opady.



Rysunek 9. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej opadów.

### 3.9. Kwantyfikator lingwistyczny względny

Do kwantyfikatora lingwistycznego względnego [2] zostały dopasowane etykiety: niewiele, około 1/4, około połowy, większość, prawie wszystkie.

$$\mu_{niewiele}(x) = \exp\left(\frac{-x^2}{0.02}\right) \quad (44)$$

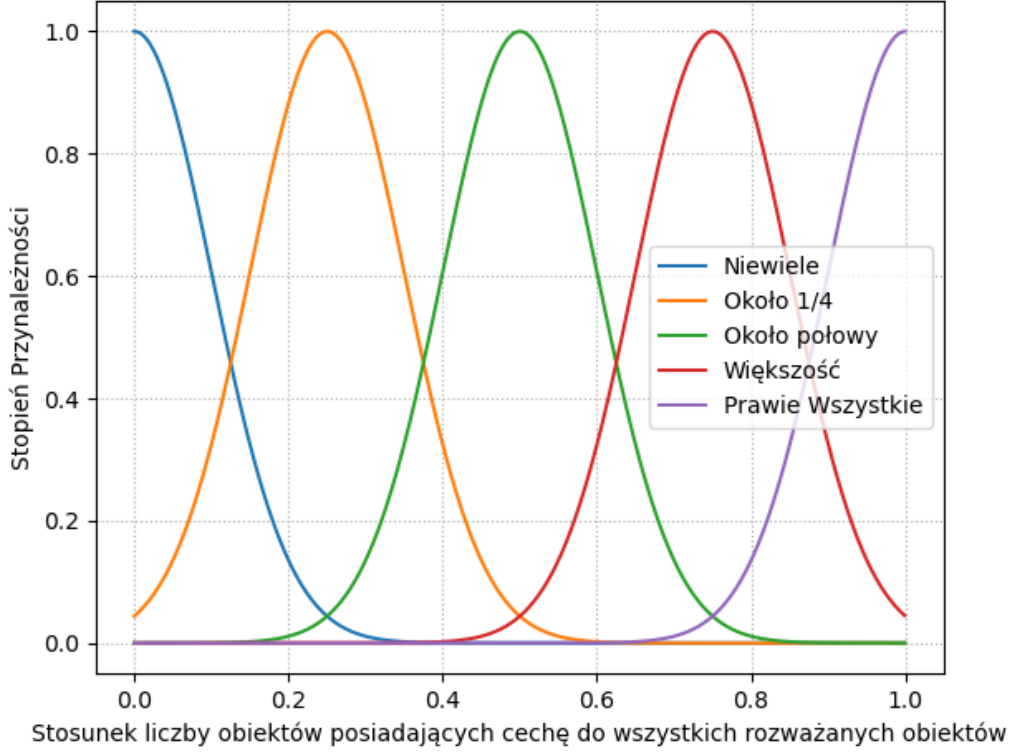
$$\mu_{okoloJednejCzwartej}(x) = \exp\left(\frac{-(x - 0.25)^2}{0.02}\right) \quad (45)$$

$$\mu_{okoloPolowy}(x) = \exp\left(\frac{-(x - 0.5)^2}{0.02}\right) \quad (46)$$

$$\mu_{wiekszosc}(x) = \exp\left(\frac{-(x - 0.75)^2}{0.02}\right) \quad (47)$$

$$\mu_{prawieWszystkie}(x) = \exp\left(\frac{-(x - 1)^2}{0.02}\right) \quad (48)$$

gdzie:  $\mu_{niewiele}$ ,  $\mu_{okoloJednejCzwartej}$ ,  $\mu_{okoloPolowy}$ ,  $\mu_{wiekszosc}$ ,  $\mu_{prawieWszystkie}$  - kwantyfikatory,  $x$  - stosunek liczby obiektów posiadających cechę do wszystkich rozważanych obiektów.



Rysunek 10. Wykres funkcji przynależności kwantyfikatorów lingwistycznych względnych.

### 3.10. Kwantyfikator lingwistyczny absolutny

Do kwantyfikatora lingwistycznego absolutnego [2] zostały dopasowane etykiety: poniżej 10, około 50, około 100, między 100 a 200, około 200, ponad 200.

$$\mu_{ponizej10}(x) = \begin{cases} \frac{10-x}{10} & \text{dla } 0 < x \leq 10 \end{cases} \quad (49)$$

$$\mu_{okolo20}(x) = \begin{cases} \frac{x}{20} & \text{dla } 0 < x \leq 20 \\ \frac{40-x}{20} & \text{dla } 20 < x \leq 40 \end{cases} \quad (50)$$

$$\mu_{okolo50}(x) = \begin{cases} \frac{x-10}{30} & \text{dla } 10 < x \leq 40 \\ 1 & \text{dla } 40 < x < 60 \\ \frac{90-x}{30} & \text{dla } 60 < x \leq 90 \end{cases} \quad (51)$$

$$\mu_{okolo100}(x) = \begin{cases} \frac{x-50}{40} & \text{dla } 50 < x \leq 90 \\ 1 & \text{dla } 90 < x < 110 \\ \frac{150-x}{50} & \text{dla } 110 < x \leq 150 \end{cases} \quad (52)$$

$$\mu_{miedzy100A200}(x) = \begin{cases} \frac{x}{50} & \text{dla } 100 < x \leq 150 \\ \frac{200-x}{50} & \text{dla } 150 < x \leq 200 \end{cases} \quad (53)$$

$$\mu_{okolo200}(x) = \begin{cases} \frac{x-150}{40} & \text{dla } 150 < x \leq 190 \\ 1 & \text{dla } 190 < x < 210 \\ \frac{250-x}{40} & \text{dla } 210 < x \leq 250 \end{cases} \quad (54)$$

$$\mu_{Ponad200}(x) = \begin{cases} \frac{x-200}{50} & \text{dla } 200 < x \leq 250 \\ 1 & \text{dla } 250 < x < 300 \\ \frac{300-x}{50} & \text{dla } 300 < x \leq 350 \end{cases} \quad (55)$$

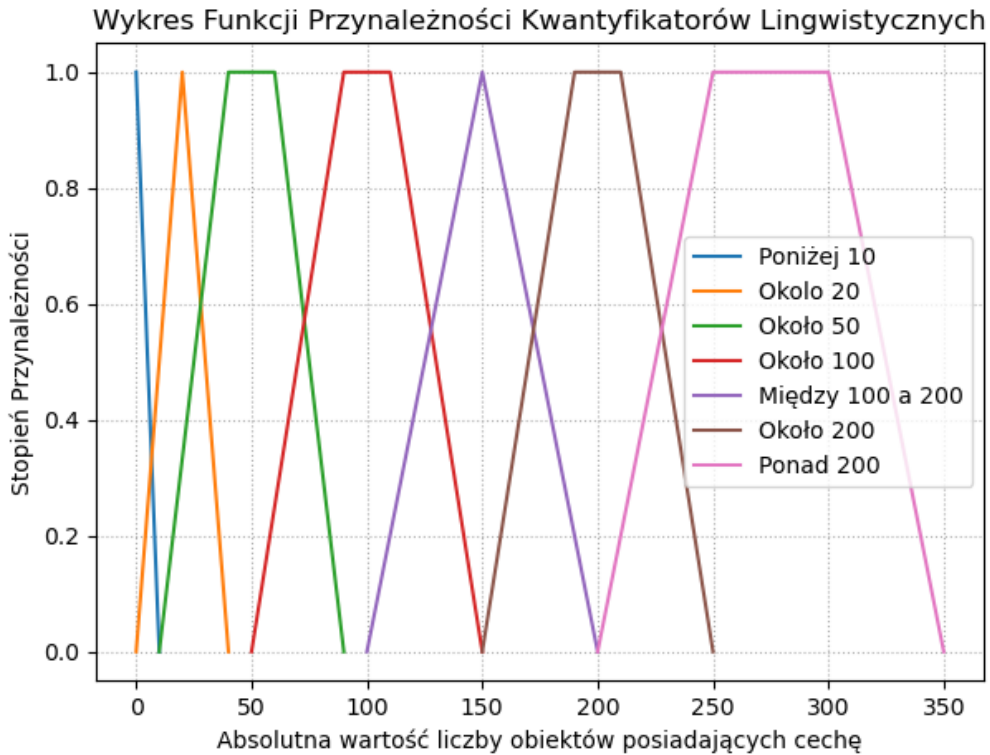
$$\mu_{Okolo500}(x) = \begin{cases} \frac{x-300}{200} & \text{dla } 300 < x \leq 400 \\ 1 & \text{dla } 400 < x < 600 \\ \frac{600-x}{100} & \text{dla } 600 < x \leq 700 \end{cases} \quad (56)$$

$$\mu_{Okolo1000}(x) = \begin{cases} \frac{x-600}{200} & \text{dla } 600 < x \leq 800 \\ 1 & \text{dla } 800 < x < 1200 \\ \frac{1200-x}{200} & \text{dla } 1200 < x \leq 1400 \end{cases} \quad (57)$$

$$\mu_{Okolo2000}(x) = \begin{cases} \frac{x-1200}{400} & \text{dla } 1200 < x \leq 1600 \\ 1 & \text{dla } 1600 < x < 2400 \\ \frac{2400-x}{400} & \text{dla } 2400 < x \leq 2800 \end{cases} \quad (58)$$

$$\mu_{Okolo3500}(x) = \begin{cases} \frac{x-2400}{400} & \text{dla } 2400 < x \leq 2800 \\ 1 & \text{dla } 2800 < x < 4200 \\ \frac{4200-x}{400} & \text{dla } 4200 < x \leq 4600 \end{cases} \quad (59)$$

gdzie:  $\mu_{ponizej10}$ ,  $\mu_{okolo20}$ ,  $\mu_{okolo50}$ ,  $\mu_{okolo100}$ ,  $\mu_{miedzy100A200}$ ,  $\mu_{okolo200}$ ,  $\mu_{Ponad200}$ ,  $\mu_{Okolo500}$ ,  $\mu_{Okolo1000}$ ,  $\mu_{Okolo2000}$ ,  $\mu_{Okolo3500}$ , - kwantyfikatory absolutne,  $x$  - absolutna wartość liczby obiektów posiadających cechę.



Rysunek 11. Wykres funkcji przynależności kwantyfikatorów lingwistycznych absolutnych od wartości 0 do 350.

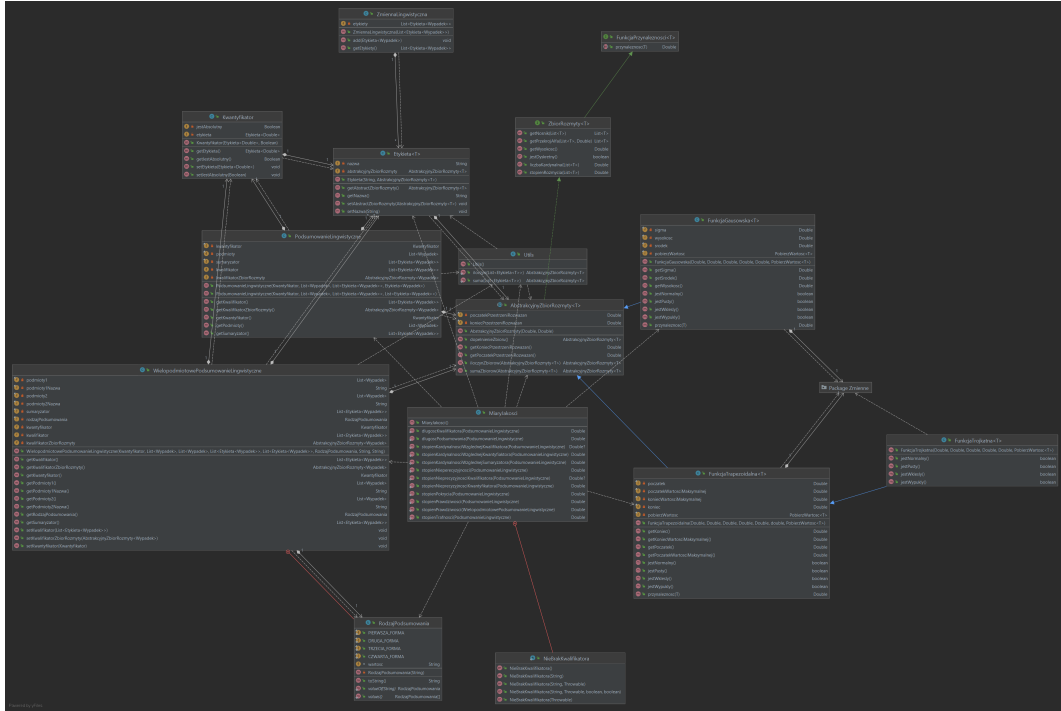




Rysunek 12. Wykres funkcji przynależności kwantyfikatorów lingwistycznych absolutnych.

#### 4. Narzędzia obliczeniowe: projekt (wybór, implementacja) i diagram UML pakietu obliczeń rozmytych. Diagram UML generatora podsumowań

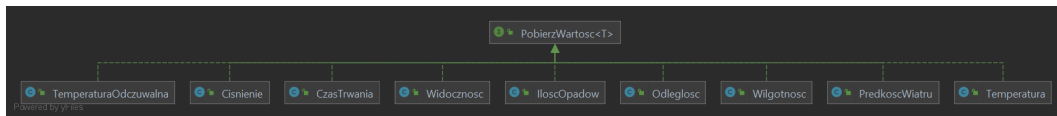
##### 4.1. Diagram pakietu obliczeń rozmytych



Rysunek 13. Diagram uml pakietu obliczeń rozmytych.

Program został wykonany w języku Java 11 LTS z użyciem narzędzia Maven. W celu wykonania projektu stworzyliśmy własny pakiet obliczeń rozmytych - obliczeniaRozmyte. Pakiet ten jest przystosowany do pracy przy użyciu naszego dyskretnego zbioru danych o wypadkach samochodowych w Stanach Zjednoczonych Ameryki. Pakiet posiada klasy oraz interfejsy niezbędne do wykonywania obliczeń rozmytych i generowania podsumowań lingwistycznych. Interfejs FunkcjaPrzynależności posiada jedną metodę pozwalającą na obliczenie przynależności obiektu do zbioru [9]. Interfejs ten rozszerzany jest przez interfejs ZbiórRozmyty który pozwala na obliczenie różnych własności zbioru rozmytego: wysokość [4], liczbę kardynalna [4], stopień rozmycia [5] oraz pozwala na pobranie zbiorów nierozmytych będących odpowiednio nośnikami [9] i przekrojem alfa zbioru rozmytego [9]. Interfejs ten jest implementowany przez abstrakcyjną klasę AbstrakcyjnyZbiórRozmyty posiadającą informacje o początku i końcu przestrzeni rozważań [4], i pozwalającą na wyznaczenie dopełnienia zbioru oraz iloczynu i sumy zbioru z innym zbiorem rozmytym [4]. Klasa ta implementowana jest przez klasy odpowiadające odpowiednio funkcji gaussowskiej [9] i trapezoidalnej [9], a po funkcji trapezoidalnej następnie dziedziczy funkcja trójkątna [9]. W celu pobrania odpowiedniego pola z obiektu w celu obliczenia przynależności klasy implementujące AbstrakcyjnyZbiórRozmyty posiadają jako swoje pole interfejs PobierzWartość, który wybiera z klasy odpowiednie pole podlegające podsumowaniu. Odpowiednie implementacje klasy AbstrakcyjnyZbiórRozmyty posiadają też informacje o tym czy zbiór przez nie reprezentowany jest normalny [9], pusty [9], wklęsły [9] lub wypukły [9]. Zbiory rozmyte są natępnie używane przy tworzeniu etykiety [5], gdzie każda posiada swoją nazwę i odpowiedni zbiór

rozmyty. Używając Etykiet reprezentujemy Kwantyfikatory [5], kwalifikatory [5] oraz sumaryzatory [5]. Do obliczania iloczynu lub sumy wielu zbiorów rozmytych używamy klasy Utils posiadającą odpowiednie statyczne metody. Kwantyfikator posiada etykietę oraz informacje czy jest on absolutny [5]. Klasa zmienna lingwistyczna reprezentuje zmienną lingwistyczną poprzez listę Etykiet. Klasa podsumowanie lingwistyczne reprezentuje podsumowanie lingwistyczne poprzez kwantyfikator, klasyfikator, sumaryzator oraz podmioty. Dla każdego podsumowania możemy policzyć miary T1 - T11 korzystając z klasy MiaryJakości.



Rysunek 14. Diagram uml pakietu Zmienne.

W programie został zaimplementowany pakiet Zmienne posiadający interfejs PobierzWartosc, k który jest implementowany przez 9 klas reprezentujących analizowane pola z obiektu Wypadek.

#### 4.2.1. Diagram UML



W celu stworzenia GUI zostały stworzone dwa pakiety gui oraz predefiniowany. Jedną z dwóch klas znajdujących się w pakiecie predefiniowane jest klasa PredefiniowaneKwantyfikatory. Jest to klasa reprezentująca singleton, zawiera listę predefiniowanych kwantyfikatorów oraz metodę zwracającą wspomnianą listę. Posiada również metodę pozwalającą na dodanie kwantyfikatora do listy, która jest wykorzystywana, gdy użytkownik zaawansowany dodaje własny kwantyfikator. Drugą klasą w wymienionym pakiecie jest klasa PredefiniowaneKwalifikatorySumaryzatory, która również jest klasą reprezentującą singleton. Zawiera pola będące zmiennymi lingwistycznymi dla każdego z atrybutów wypadku, a także metody zwracające każdą ze zmiennych, bądź wszystkie zmienne w liście. Klasa zawiera metody pozwalające na dodanie kwalifikatorów lub sumaryzatorów do listy, która jest wykorzystywana, gdy użytkownik zaawansowany dodaje własny kwalifikator lub sumaryzator. W pakiecie gui znajduje się klasa PodstawowyUzytkownik, która jest kontrolerem łączącym się z oknem wyświetlanym użytkownikowi. Znajdują się w niej pola reprezentujące elementy wyświetlane w interfejsie użytkownika oraz niezbędne listy zmiennych lingwistycznych bądź Stringów, zawierające elementy do wyświetlenia w interfejsie użytkownika. Klasa posiada również metody służące wybraniu wskazanych kwantyfikatorów, kwalifikatorów oraz sumaryzatorów w celu wykonania podsumowania lingwistycznego. Klasa zawiera metodę doZaawansowanego, która otwiera okno dla zaawansowanego użytkownika. Klasa, do generowania podsumowania lingwistycznego wykorzystuje klasę MiaryJakosciWagi oraz PodsumowanieLingwistyczneIMiary. Klasa MiaryJakosciWagi zawiera pola zawierające wartości wag dla poszczególnych miar jakości. Klasa PodsumowanieLingwistyczneIMiary zawiera metodę getGlownaMiaraJakosci zwracającą w postaci Stringa wartość głównej miary jakości na podstawie wag miar jakości oraz wartości miar jakości. W celu obliczenia głównej miary jakości konstruktor klasy wykorzystuje metodę calculateGlownaMiaraJakosci, która oblicza wspomnianą miarę. W pakiecie gui znajduje się również klasa ZaawansowanyUzytkownik, będąca kontrolerem, w której znajdują się pola dla każdego z elementów wyświetlanych w interfejsie użytkownika. Klasa pozwala na definiowanie własnych kwantyfikatorów, kwalifikatorów i sumaryzatorów. Metody setEtykietaKWantyfikikator oraz setEtykietaKwalifikatorSumaryzator pozwalają na stworzenie obiektu Etykieta odpowiednio dla kwantyfikatora, kwalifikatora lub sumaryzatora. Metody zapiszKwantyfikikator oraz zapiszKwalifikatorSumaryzator pozwalają na zapis odpowiednio kwantyfikatora, kwalifikatora lub sumaryzatora do listy dostępnych dla użytkownika kwantyfikatorów, kwalifikatorów oraz sumaryzatorów. Klasa zawiera metodę doGlownego, która otwiera okno dla podstawowego użytkownika.

## 4.2.2. Instrukcja obsługi

The screenshot shows a software interface with the following components:

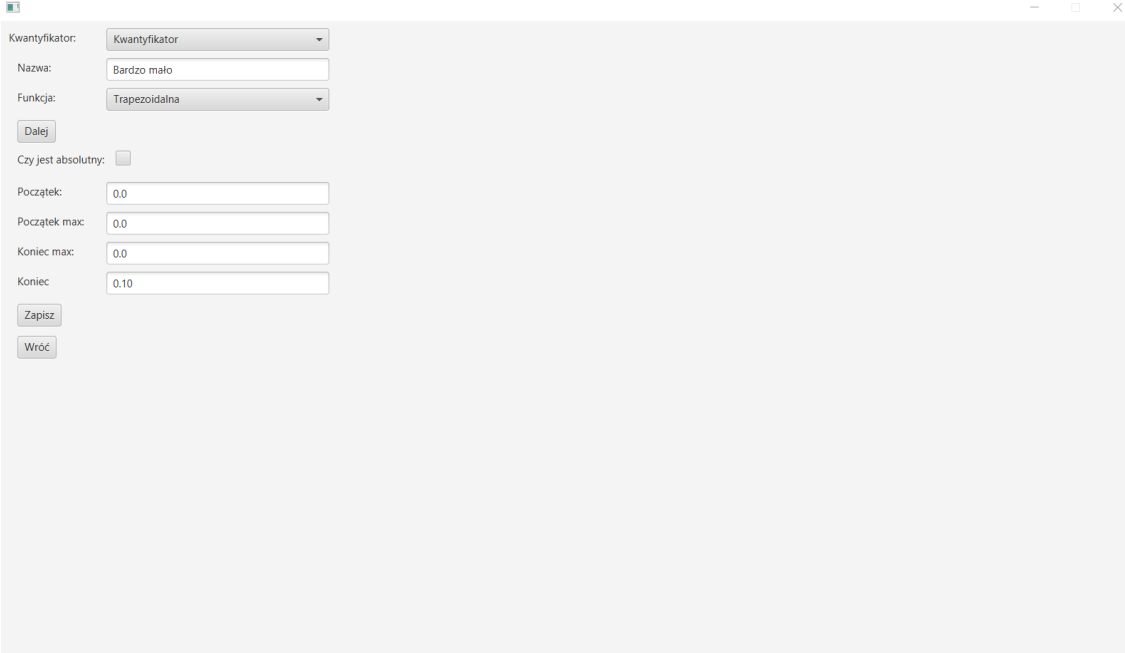
- Kwantyfikatory:** A dropdown menu set to "Okolo 200 (Absolutny)".
- Kwalifikatory:** A dropdown menu set to "przy silnym wietrze".
- Dostępne sumaryzatory:** A dropdown menu set to "przy umiarkowanie wilgotnym powietrzu".
- Wybrane sumaryzatory:** A dropdown menu set to "wywołuje utrudnienia w ruchu trwające około czterech godzin".
- Buttons:** "Dodaj", "Usuń", "Wybierz", and "Do zaawansowanego".
- Weights (Wagi):** A grid of input fields for T1 through T11, with values like 0.5, 0.25, 0.04, 0.03, 0.00.
- Table:** A table with columns: Tekst, T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, Suma, and Zapis. It contains five rows of data with calculated sums and a "Zapisz" button for each row.

Tekst	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	Suma	Zapis
Ponizej 10 wypadków będących przy wietrze, jest przy dużych opadach	0,00	1,00	0,00	0,00	0,00	0,00	0,00	1,00	0,96	0,97	1,00	0,34	Zapisz
Okolo 1/4 wypadków będących przy dużych opadach, jest przy zimnej temperaturze odczuwalnej	0,28	0,67	0,20	0,20	0,00	0,00	0,92	0,75	1,00	1,00	1,00	0,43	Zapisz
Większość wypadków będących przy dużych opadach, jest przy zimnej temperaturze odczuwalnej	0,00	0,67	0,20	0,20	0,00	0,00	0,96	0,75	1,00	1,00	1,00	0,29	Zapisz
Niewiele wypadków będących przy dużych opadach, jest przy zimnej temperaturze odczuwalnej	0,66	0,67	0,20	0,20	0,00	0,00	0,31	0,75	1,00	1,00	1,00	0,60	Zapisz
Okolo 200 wypadków będących przy silnym wietrze, jest przy wilgotnym powietrzu, jest wywołuje utrudnienia w ruchu trwające około cztere...	0,00	1,00	0,00	0,00	0,00	0,00	1,00	1,00	0,68	0,69	1,00	0,35	Zapisz

Rysunek 16. Przykład wyglądu interfejsu podstawowego użytkownika

Na rysunku 16. został przedstawiony podstawowy wygląd interfejsu użytkownika. Zaczynając od lewego górnego rogu najpierw użytkownik z rozwijanej listy wybiera kwantyfikator, następnie z rozwijanej listy wybierany jest kwalifikator. Kolejnym krokiem jest wybranie jednego bądź wielu sumaryzatorów, w tym celu użytkownik wybiera pierwszy wybrany sumaryzator i naciska przycisk Dodaj znajdujący się po prawej stronie od rozwijanej listy sumaryzatorów, za pomocą którego wybrany sumaryzator jest dodawany do listy wybranych przez użytkownika sumaryzatorów. Aby wybrać więcej niż jeden sumaryzator użytkownik musi powtórzyć wcześniejsze czynności zaczynając od wybrania z rozwijanej listy pożądanego sumaryzatora. Istnieje możliwość usunięcia sumaryzatora z listy wybranych sumaryzatorów, w tym celu użytkownik wybiera z rozwijanej listy, znajdującej się pod dostępnymi sumaryzatorami, sumaryzator do usunięcia. Następnie przyciska przycisk Usuń. W celu usunięcia więcej niż jednego sumaryzatora, użytkownik musi powtórzyć opisane czynności zaczynając od wyboru sumaryzatora z listy wybranych sumaryzatorów. W celu wykonania podsumowania lignwistycznego należy przycisnąć przycisk Wybierz znajdujący się pod rozwijaną listą z wybranymi sumaryzatorami, następnie w tabeli na dole ekranu pojawi się wygenerowany tekst podsumowania oraz wartości miar jakości. W ostatniej kolumnie tabeli znajduje się przycisk Zapisz, dla każdego wygenerowanego podsumowania. Po przyciśnięciu opisanego przycisku dane podsumowanie zostanie dopisane do pliku txt. W programie istnieje możliwość obliczenia głównej miary podsumowania, w tym celu należy wpisać odpowiednie wagi, dla każdej z miar w polach znajdujących się na środku ekranu. Pod polami znajduje się przycisk Wybierz, którego kliknięcie spowoduje policzenie

głównej miary podsumowań we wszystkich podsumowaniach znajdujących się w tabeli. Kliknięcie przycisku Do zaawansowanego spowoduje przejście do ekranu użytkownika zaawansowanego.



The screenshot shows a web application window titled "Kwantyfikator" (Quantifier). It contains the following elements:

- A dropdown menu labeled "Kwantyfikator:" with "Kwantyfikator" selected.
- A text input field labeled "Nazwa:" containing the text "Bardzo mało".
- A dropdown menu labeled "Funkcja:" with "Trapezoidalna" selected.
- A "Dalej" (Next) button.
- A checkbox labeled "Czy jest absolutny:" which is currently unchecked.
- Four text input fields for numerical values:
  - "Początek:" with value "0.0"
  - "Początek max:" with value "0.0"
  - "Koniec max:" with value "0.0"
  - "Koniec:" with value "0.10"
- "Zapisz" (Save) and "Wróć" (Back) buttons at the bottom.

Rysunek 17. Przykład wyglądu interfejsu zaawansowanego użytkownika

Na rysunku 17. został przedstawiony wygląd interfejsu zaawansowanego użytkownika, który służy do tworzenia własnych kwantyfikatorów, klasyfikatorów oraz sumaryzatorów. Zaczynając od lewego górnego rogu na użytkownik wybiera, który z elementów chce utworzyć: kwantyfikator, kwalifikator, sumaryzator. Następnie poniżej należy wpisać nazwę elementu oraz wybrać z rozwijanej listy funkcję przynależności. Kliknięcie przycisku Dalej powoduje pojawienie się pól do wprowadzenia kolejnych danych w zależności od wcześniej wybranych wartości. Dla wartości wybranych na rysunku 17 po kliknięciu przycisku Dalej pojawiają się następujące pola: checkbox do zaznaczenia czy kwantyfikator jest absolutny, początek funkcji przynależności, początek i koniec maksymalnej wartości, dla których funkcja przyjmuje wartość maksymalną oraz wartość końcową. Kliknięcie przycisku Zapisz spowoduje zapisanie elementu do listy dostępnych elementów. Przeciśnięcie przycisku Wróć wróci do ekranu użytkownika podstawowego.

Wymagania, które muszą zostać spełnione by uruchomić program na własnym komputerze to: wersja Javy 11.

## 5. Jednopo­dmiotowe podsumowania lingwistyczne. Miary jakości, podsumowanie optymalne

### 5.1. Eksperyment 1

Kwantyfikatory:

Poniżej 10 (Absolutny)

Kwalifikatory:

Brak

Dostępne sumaryzatory:

wywołuje utrudnienia w ruchu o długości ponad trzech mili

Dodaj

Wybrane sumaryzatory:

przy słabym wietrze

Usuń

Wybierz

Do zaawansowanego

Wagi:

T1: 0.65

T7: 0.03

T2: 0.10

T8: 0.03

T3: 0.04

T9: 0.03

T4: 0.03

T10: 0.03

T5: 0.03

T11: 0.00

T6: 0.03

Wybierz

Tekst	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	Suma*	Zapisz
Okolo 3500 wypadków, jest przy słabym wietrze	0.94	0.71	0.29	0.29	1.00	0.85	1.00	0.71	null	null	0	0.81	Zapisz
Niewiele wypadków, jest przy słabym wietrze, jest przy umiarkowanych opadach	1.00	0.94	0.00	0.00	0.50	0.00	0.00	0.95	null	null	0	0.79	Zapisz
Okolo 1/4 wypadków, jest przy słabym wietrze	0.92	0.71	0.29	0.29	1.00	0.00	0.97	0.71	null	null	0	0.77	Zapisz
Okolo 20 wypadków, jest przy słabym wietrze, jest przy umiarkowanych opadach	0.74	0.94	0.00	0.00	0.50	1.00	1.00	0.95	null	null	0	0.68	Zapisz
Okolo 50 wypadków, jest przy słabym wietrze, jest przy umiarkowanych opadach	0.16	0.94	0.00	0.00	0.50	0.99	1.00	0.95	null	null	0	0.30	Zapisz
Okolo 2000 wypadków, jest przy słabym wietrze	0.00	0.71	0.29	0.29	1.00	0.89	1.00	0.71	null	null	0	0.20	Zapisz
Poniżej 10 wypadków, jest przy słabym wietrze, jest przy umiarkowanych opadach	0.00	0.94	0.00	0.00	0.50	1.00	1.00	0.95	null	null	0	0.20	Zapisz
Niewiele wypadków, jest przy słabym wietrze	0.01	0.71	0.29	0.29	1.00	0.00	0.29	0.71	null	null	0	0.16	Zapisz

Rysunek 18. Tabela przedstawia wyniki podsumowań lingwistycznych dla eksperymentu 1. Kolumna Tekst zawiera tekst wygenerowanego podsumowania lingwistycznego. Kolumny T1-T11 są miarami jakości podsumowania lingwistycznego [5]. Kolumna Suma zawiera wartości jakości optymalnego podsumowania lingwistycznego, które są średnią ważoną miar jakości od T1 do T11. Ostatnia kolumna, Zapisz, zawiera przycisk zapisujący dane podsumowanie lingwistyczne do pliku. Każdy rząd zawiera jedno wygenerowane podsumowanie lingwistyczne. Rzędy zostały posortowane po kolumnie Suma malejąco.

W pierwszym eksperymencie wygenerowaliśmy jednopo­dmiotowe podsumowania lingwistyczne w pierwszej formie. Do obliczenia miary jakości podsumowania optymalnego zostały użyte następujące wagi odpowiednio dla miar od T1 do T11: 0.65, 0.10, 0.04, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.00. W kolumnach T9 i T10 wszystkie wartości przyjmują wartość null, ponieważ są to miary jakości

Na podstawie tabeli 18 można zauważyć, że zdanie 'Niewiele wypadków, jest przy słabym wietrze, jest przy umiarkowanych opadach' jest zdaniem najbardziej zgodnym z prawdą dla kwantfikatorów względnych, ponieważ nie ma wielu wypadków, które były przy słabym wietrze i umiarkowanych opadach. Natomiast patrząc na zdanie, które jest najmniej zgodne z prawdą, 'Niewiele wypadków, jest przy słabym wietrze', wnioskujemy, że jest wiele wypadków, które nastąpiły przy słabym wietrze. Zdanie 'Okolo 3500 wypadków, jest przy słabym wietrze' jest zdaniem najbardziej zgodnym z prawdą.



## 5.2. Eksperyment 2

Kwantyfikatory:

Większość (Względny)

Kwalifikatory:

przy dobrej widoczności

Dostępne sumaryzatory:

przy bardzo zimnej temperaturze

Wybrane sumaryzatory:

przy niewielkich opadach

Wybierz

Do zaawansowanego

Wagi:

T1: 0.65

T2: 0.10

T3: 0.04

T4: 0.03

T5: 0.03

T6: 0.03

T7: 0.03

T8: 0.03

T9: 0.03

T10: 0.03

T11: 0.00

Tekst	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	Suma*	Zapis
Około połowy wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy zimnej temperaturze	0.99	0.41	0.61	0.61	0.50	0.00	0.96	0.49	0.95	0.96	1.00	0.84	Zapisz
Niewiele wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy suchym powietrzu	1.00	0.00	0.94	0.94	0.50	0.00	0.16	0.69	0.95	0.96	1.00	0.81	Zapisz
Prawie wszystkie wypadków będących przy dobrej widoczności, jest przy niewielkich opadach	0.69	0.01	0.94	0.94	1.00	0.00	0.01	0.01	0.95	0.96	1.00	0.61	Zapisz
Większość wypadków będących przy dobrej widoczności, jest przy niewielkich opadach	0.26	0.01	0.94	0.94	1.00	0.00	0.95	0.01	0.95	0.96	1.00	0.35	Zapisz
Około 1/4 wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy zimnej temperaturze	0.06	0.41	0.61	0.61	0.50	0.00	0.93	0.49	0.95	0.96	1.00	0.24	Zapisz
Około 1/4 wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy suchym powietrzu	0.05	0.00	0.94	0.94	0.50	0.00	0.91	0.69	0.95	0.96	1.00	0.22	Zapisz
Większość wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy zimnej temperaturze	0.03	0.41	0.61	0.61	0.50	0.00	0.95	0.49	0.95	0.96	1.00	0.22	Zapisz
Większość wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy suchym powietrzu	0.00	0.00	0.94	0.94	0.50	0.00	0.96	0.69	0.95	0.96	1.00	0.19	Zapisz
Około połowy wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy suchym powietrzu	0.00	0.00	0.94	0.94	0.50	0.00	0.97	0.69	0.95	0.96	1.00	0.19	Zapisz
Niewiele wypadków będących przy dobrej widoczności, jest przy niewielkich opadach	0.00	0.01	0.94	0.94	1.00	0.00	0.99	0.01	0.95	0.96	1.00	0.18	Zapisz
Niewiele wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy zimnej temperaturze	0.00	0.41	0.61	0.61	0.50	0.00	0.33	0.49	0.95	0.96	1.00	0.18	Zapisz

Rysunek 19. Tabela przedstawia wyniki podsumowań lingwistycznych dla eksperymentu 2. Kolumna Tekst zawiera tekst wygenerowanego podsumowania lingwistycznego. Kolumny T1-T11 są miarami jakości podsumowania lingwistycznego [5]. Kolumna Suma zawiera wartości jakości optymalnego podsumowania lingwistycznego, które są średnią ważoną miar jakości od T1 do T11. Ostatnia kolumna, Zapisz, zawiera przycisk zapisujący dane podsumowanie lingwistyczne do pliku. Każdy rząd zawiera jedno wygenerowane podsumowanie lingwistyczne. Rzędy zostały posortowane po kolumnie Suma malejąco.

W drugim eksperymencie wygenerowaliśmy jednopodmiotowe podsumowania lingwistyczne w drugiej formie. Do obliczenia miary jakości podsumowania optymalnego zostały użyte następujące wagi odpowiednio dla miar od T1 do T11: 0.65, 0.10, 0.04, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.00.

Na podstawie tabeli 19 można zauważyć, że zdanie 'Około połowy wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy zimnej temperaturze' jest zdaniem najbardziej zgodnym z prawdą, a jednocześnie zdanie 'Niewiele wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy zimnej temperaturze' jest zdaniem najmniej zgodnym z prawdą. Można zauważyć, że te zdania zaprzeczają sobie nawzajem, więc wyniki miar jakości podsumowania optymalnego są uzasadnione. Zgodnie z naszymi oczekiwaniami 'Prawie wszystkie wypadki będące przy dobrej widoczności, jest przy niewielkich opadach', jest trzecim zdaniem najbardziej zgodnym z prawdą, ponieważ opady powodują złą widoczność, więc jeśli widoczność była dobra oznacza to, że opadów nie było, bądź były niewielkie. Natomiast zdanie 'Niewiele wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy suchym powietrzu' nie było zgodne z naszymi oczekiwaniami, ponieważ spodziewaliśmy się, że przy większości wypadków jeśli jest dobra widoczność, to nie ma opadów, a co za tym idzie jest suche powietrze, co okazało się być zdaniem nieprawdziwym.

### 5.3. Eksperyment 3

Kwantyfikatory:

Okolo 1/4 (Względny)

Kwalifikatory:

wywołuje utrudnienia w ruchu o długości poniżej pół mili

Dostępne sumaryzatory:

przy bardzo ciepłej temperaturze odczuwalnej

Wybrane sumaryzatory:

wywołuje utrudnienia w ruchu trwające około sześciu godzin

Wybierz

Do zaawansowanego

Wagi:

T1: 0.65

T2: 0.10

T3: 0.04

T4: 0.03

T5: 0.03

T6: 0.03

T7: 0.03

T8: 0.03

T9: 0.03

T10: 0.03

T11: 0.00

Tekst	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	Suma*	Zapis
Okolo połowy wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około dwóch godzin	1,00	0,11	0,91	0,91	1,00	0,00	0,73	0,53	0,19	0,28	1,00	0,81	Zapisz
Niewiele wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około sześciu godzin	0,99	0,95	0,04	0,04	1,00	0,00	0,04	0,97	0,19	0,28	1,00	0,81	Zapisz
Okolo 1/4 wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające poniżej godziny	0,66	0,61	0,42	0,42	1,00	0,00	0,87	0,86	0,19	0,28	1,00	0,62	Zapisz
Niewiele wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około czterech godzin	0,70	0,79	0,19	0,19	1,00	0,00	0,17	0,91	0,19	0,28	1,00	0,62	Zapisz
Okolo 1/4 wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około czterech godzin	0,26	0,79	0,19	0,19	1,00	0,00	0,91	0,91	0,19	0,28	1,00	0,36	Zapisz
Niewiele wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające poniżej godziny	0,28	0,61	0,42	0,42	1,00	0,00	0,32	0,86	0,19	0,28	1,00	0,35	Zapisz
Okolo 1/4 wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około sześciu godzin	0,06	0,95	0,04	0,04	1,00	0,00	0,95	0,97	0,19	0,28	1,00	0,24	Zapisz
Większość wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około dwóch godzin	0,05	0,11	0,91	0,91	1,00	0,00	0,78	0,53	0,19	0,28	1,00	0,19	Zapisz
Okolo połowy wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające poniżej godziny	0,00	0,61	0,42	0,42	1,00	0,00	0,80	0,86	0,19	0,28	1,00	0,19	Zapisz
Okolo 1/4 wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około dwóch godzin	0,04	0,11	0,91	0,91	1,00	0,00	0,72	0,53	0,19	0,28	1,00	0,18	Zapisz
Niewiele wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około dwóch godzin	0,00	0,11	0,91	0,91	1,00	0,00	0,86	0,53	0,19	0,28	1,00	0,16	Zapisz

Rysunek 20. Tabela przedstawia wyniki podsumowań lingwistycznych dla eksperymentu 3. Kolumna Tekst zawiera tekst wygenerowanego podsumowania lingwistycznego. Kolumny T1-T11 są miarami jakości podsumowania lingwistycznego [5]. Kolumna Suma zawiera wartości jakości optymalnego podsumowania lingwistycznego, które są średnią ważoną miar jakości od T1 do T11. Ostatnia kolumna, Zapisz, zawiera przycisk zapisujący dane podsumowanie lingwistyczne do pliku. Każdy rząd zawiera jedno wygenerowane podsumowanie lingwistyczne. Rzędy zostały posortowane po kolumnie Suma malejąco.

W trzecim eksperymencie wygenerowaliśmy jednopodmiotowe podsumowania lingwistyczne w drugiej formie. Do obliczenia miary jakości podsumowania optymalnego zostały użyte następujące wagi odpowiednio dla miar od T1 do T11: 0.65, 0.10, 0.04, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.00.

Wykonany eksperyment miał na celu sprawdzenie zależności długości odcinka drogi, na którym nastąpiły utrudnienia w ruchu spowodowane przez wypadek od czasu trwania utrudnień w ruchu drogowym przez ten wypadek. Uzyskane wyniki potwierdzają postawioną tezę. Na podstawie tabeli 20 można zauważyć, że zdanie 'Okolo połowy wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około dwóch godzin' jest najbardziej prawdziwym zdaniem. Jednocześnie możemy zauważyć, że to samo zdanie z kwantyfikatorem 'Niewiele' jest najmniej prawdziwym zdaniem. To samo zdanie z kwantyfikatorami 'Okolo 1/4' lub 'Większość' są zdecydowanie mniej prawdziwe niż pierwsze opisane zdanie. Kolejnym zdaniem potwierdzającym tezę jest: 'Niewiele wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające około sześciu godzin', ponieważ, odcinek poniżej pół mili jest krótkim odcinkiem, a czas trwania sześć godzin jest znacznie długim czasem. Ze zdania możemy odczytać, że nie wiele wypadków będących na krótkim odcinku drogi wywołuje tak długi czas trwania utrudnień w ruchu drogowym.

## 6. Wielopodmiotowe podsumowania lingwistyczne i ich miary jakości

Wyniki kolejnych eksperymentów wg punktów 2.-4. opisu projektu 2. Uzasadnienie i metoda podziału zbioru danych na rozłączne podmioty. Listy podsumowań wielopodmiotowych i tabele/rankingi podsumowań dla danych atrybutów obowiązkowe i dokładnie opisane w „captions” (tytułach), konieczny opis kolumn i wierszy tabel. Konieczne uwzględnienie wszystkich 4-ch form podsumowań wielopodmiotowych.

\*\* Możliwe sformułowanie zagadnienia wielopodmiotowego podsumowania optymalnego \*\*.

\*\*Ewentualne wyniki realizacji punktu „na ocenę 5.0” wg opisu Projektu 2. i ich porównanie do wyników z części obowiązkowej\*\*.

**Sekcja uzupełniona jako efekt zadania Tydzień 12 wg Harmonogramu Zajęć na WIKAMP KSR.**

## 7. Dyskusja, wnioski

### 7.1. Jednopodmiotowe podsumowania lingwistyczne

W sekcji 5 przeprowadziliśmy trzy eksperymenty. Sekcja 5.1 była zapoznaniem z przeprowadzaniem podsumowań lingwistycznych. Zostały wygenerowane podsumowania lingwistyczne w pierwszej formie, badające związek ilości wypadków przy słabym wietrze oraz przy słabym wietrze i umiarkowanych opadach. Przeprowadzone podsumowanie sprawdzające związek pomiędzy liczbą wypadków przy słabym wietrze, zgodnie z naszymi założeniami, około 1/3 wypadków miałyby miejsce przy słabym wietrze, co zostało potwierdzone przez podsumowanie 'Około 1/4 wypadków jest przy słabym wietrze', dla którego miara T1 osiągnęła wartość 0.92. Podczas przeprowadzania tego eksperymentu eksperymentowaliśmy z ustawieniem wag dla miar jakości podsumowań lingwistycznych, które ustawiliśmy na 0.65, 0.10, 0.04, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.00 odpowiednio dla miar od T1 do T11. Waga dla miary T11 została ustawiona na 0.00, ponieważ wszystkie porównywane podsumowania lingwistyczne posiadają tę samą wartość 1.0. Natomiast miara T1 posiada największą wagę, ponieważ według nas jest najważniejszą miarą oceny jakości podsumowania. 'Około 20 wypadków jest przy słabym wietrze, jest przy umiarkowanych opadach' podsumowanie to osiągnęło również wysoką wartość w prawdziwości zdania, oznacza to, że w naszym zbiorze danych występuje mało wypadków przy wcześniej wymienionych warunkach.

W sekcji 5.2 wygenerowaliśmy jednopodmiotowe podsumowania lingwistyczne w drugiej formie - z klasyfikatorem. Przyjęliśmy takie wagi jakie zostały wymienione w dyskusji poprzedniego eksperymentu. W tym ekspery-

mencie skupiliśmy się na powiązaniu liczby wypadków z widocznością i ilością opadów. Zweryfikowaliśmy tezę, że wypadki będące przy dobrej widoczności są przy niewielkich opadach. Tą tezę udowodniło podsumowanie lingwistyczne: 'Prawie wszystkie wypadków będących przy dobrej widoczności, jest przy niewielkich opadach', dla którego miara T1 osiągnęła wartość 0.69, a średnia ważona miar 0.61. Oznacza to, że tak jak założyliśmy, zdecydowana większość wypadków w naszym zbiorze danych, które były przy dobrej widoczności były przy dobrych opadach, czyli istnieje związek pomiędzy ilością opadów a widocznością. Następnie chcieliśmy zweryfikować tezę, że wcześniej opiane warunki można powiązać z temperaturą oraz wilgotnością powietrza. Zgodnie z wynikiem podsumowania 'Okolo połowy wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy zimnej temperaturze', które osiągnęło bardzo wysokie wyniki miar jakości, dla miary T1 wynosiło 0.99, a dla średniej ważonej miar jakości 0.84. Wnioskujemy, że dobra widoczność oraz niskie opady idą w parze z niską temperaturą. Wbrew naszym oczekiwaniom nasza teza dotycząca zależności z wilgotnością powietrza nie potwierdziła się. Zgodnie z wynikami miar jakości podsumowania prawdą jest, że : 'Niewiele wypadków będących przy dobrej widoczności, jest przy niewielkich opadach, jest przy suchym powietrzu'.

W sekcji 5.3 wygenerowaliśmy jednopodmiotowe podsumowania lingwistyczne w drugiej formie - z klasyfikatorem. Przyjęliśmy takie wagi jakie zostały wymienione w dyskusji poprzedniego eksperymentu. W tym eksperymencie skupiliśmy się na powiązaniu długości ulicy posiadającej utrudnienia w ruchu drogowym spowodowanym przez wypadek a czasem trwania utrudnień ruchu drogowego spowodowanymi przez wypadek. Przewidywaliśmy, że krótki odcinek drogi powoduje krótki czas utrudnień w ruchu drogowym. Teza została potwierdzona przez trzy podsumowania lingwistyczne: 'Okolo połowy wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające okolo dwóch godzin', 'Niewiele wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające okolo sześciu godzin', 'Okolo 1/4 wypadków będących wywołuje utrudnienia w ruchu o długości poniżej pół mili, jest wywołuje utrudnienia w ruchu trwające poniżej godziny' zgodnie z przewidywaniami podsumowania te uzyskały wysokie miary jakości podsumowania. Tak jak przewidywaliśmy w naszym zbiorze danych istnieje związek pomiędzy długością drogi dotkniętą utrudnieniami w ruchu drogowym a czasem trwania utrudnień w ruchu drogowym.

## 8. Braki w realizacji projektu 2.

Wymienić wg opisu Projektu 2. wszystkie niezrealizowane obowiązkowe elementy projektu, ewentualnie podać merytoryczne (ale nie czasowe) przyczyny tych braków.

## Literatura

- [1] I. Superson, A. Niewiadomski, POZYSKIWANIE WIEDZY Z RELACYJNYCH BAZ DANYCH: WIELOPODMIOTOWE PODSUMOWANIA LINGWISTYCZNE, Politechnika Łódzka
- [2] A. Niewiadomski, Rozmyte metody inteligentnej interpretacji danych, tom 10, 2006, 546-547
- [3] 2021 Kaggle Inc [internetowa społeczność związana z analizą danych], US Accidents (3 million records – updated) A Countrywide Traffic Accident Dataset (2016 - 2020) [przeglądany 24 kwietnia 2021], Dostępny w: <https://www.kaggle.com/sobhanmoosavi/us-accidents> Oficyna Wydawnicza EXIT, Warszawa, 2008.
- [4] Zadeh L.A., A computational approach to fuzzy quantifiers in natural languages. Computers and Maths with Applications, nr 9, 1983, 149-183
- [5] A. Niewiadomski, Zbiory rozmyte typu 2. Zastosowania w reprezentowaniu informacji. Seria „Problemy współczesnej informatyki” pod redakcją L. Rutkowskiego. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2019.
- [6] S. Zadrozny, Zapytania nieprecyzyjne i lingwistyczne podsumowania baz danych, EXIT, 2006, Warszawa
- [7] A. Niewiadomski, Methods for the Linguistic Summarization of Data: Applications of Fuzzy Sets and Their Extensions, Akademicka
- [8] Zadeh, L. A.: 1965, ‘Fuzzy sets’. Inf. and Control 8, 338–353.
- [9] dr inż. Anna Bryniarska, Podstawy teorii systemów rozmytych z zadaniami, Dostępny w: [https://a.bryniarska.po.opole.pl/images/nsi/skrypt\\_r2\\_28listopada.pdf](https://a.bryniarska.po.opole.pl/images/nsi/skrypt_r2_28listopada.pdf),
- [10] Niewiadomski, Adam. Zbiory rozmyte typu 2. Zastosowania w reprezentacji informacji. Red. . : Akademicka Oficyna Wydawnicza EXIT Andrzej Lang, 2019, 195 s. ISBN 978-83-7837-595-1

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