

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

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

Daria Rogowska 229989

Mateusz Srebnik 230004

## Projekt 1. Klasyfikacja dokumentów tekstowych

### 1. Cel projektu

Celem projektu jest implementacja algorytmu  $k$  najbliższych sąsiadów w celu klasyfikacji zbioru ponad 9 tysięcy dokumentów tekstowych [3] względem kraju, którego dany tekst dotyczy. Aplikacja podzielona jest na moduł ekstrakcji cech tekstu oraz klasyfikatora. Ideą części badawczej projektu jest zbadanie wpływu wielkości parametru  $k$  na dokładność dopasowania etykiety, również porównanie metryk i miary podobieństwa tekstów.

### 2. Klasyfikacja nadzorowana metodą $k$ -NN

Algorytm  $k$  najbliższych sąsiadów (ang.  $k$  Nearest Neighbours) wyznacza  $k$  sąsiadów do których badany element ma najbliżej dla wybranej metryki. Parametrami wejściowymi algorytmu jest wartość stałej  $k$ , zbiór uczący elementów spośród, którego wyznaczani będą sąsiedzi oraz zbiór badany, dla którego elementów chcemy przypisać określone klasy. Wynikiem działania algorytmu jest przyporządkowanie elementu (dokumentu tekstowego) do jednej z możliwych klas (państwa, o którym mowa jest w danym dokumencie) na podstawie najbardziej powszechnej klasy wśród wyznaczonego zbioru  $k$  sąsiadów [1].

## 2.1. Ekstrakcja cech, wektory cech

Każdy dokument został poddany procesowi ekstrakcji cech, potrzebnych do procesu klasyfikacji. Poniżej znajduje się lista określonych cech liczbowych i tekstowych:

1. cecha tekstowa K1 - Pierwsza występująca nazwa geograficzna w danym tekście (nazwa geograficzna  $k_1$  taka, że  $k_1 \in DK1$ )
2. cecha tekstowa F2 - Pierwszy występująca nazwa/symbol waluty w danym tekście (nazwa/symbol waluty  $f_2$  taki, że  $f_2 \in DF2$ )
3. cecha tekstowa P1 - Najczęściej występująca nazwa partii politycznej w danym tekście (nazwa partii politycznej  $p_1$  taka, że  $p_1 \in DP1$ )
4. cecha liczbową N1 - Liczba nazw znanych zachodnioniemieckich budowli w danym tekście (liczba takich nazw znanych budowli  $n_1$ , że  $n_1 \in DN1$ )
5. cecha liczbową N2 - Liczba nazw znanych amerykańskich budowli w danym tekście (liczba takich nazw znanych budowli  $n_2$ , że  $n_2 \in DN2$ )
6. cecha liczbową N3 - Liczba nazw znanych japońskich budowli w danym tekście (liczba takich nazw znanych budowli  $n_3$ , że  $n_3 \in DN3$ )
7. cecha liczbową N4 - Liczba nazw znanych francuskich budowli w danym tekście (liczba takich nazw znanych budowli  $n_4$ , że  $n_4 \in DN4$ )
8. cecha liczbową N5 - Liczba nazw znanych angielskich budowli w danym tekście (liczba takich nazw znanych budowli  $n_5$ , że  $n_5 \in DN5$ )
9. cecha liczbową N6 - Liczba nazw znanych kanadyjskich budowli w danym tekście (liczba takich nazw znanych budowli  $n_6$ , że  $n_6 \in DN6$ )
10. cecha tekstowa R1 - Najczęściej występująca nazwa kraju w danym tekście (nazwa kraju  $r_1$  taka, że  $r_1 \in DR1$ )
11. cecha tekstowa M1 - Godzina opublikowania dokumentu w formacie tekstowym (wartość cechy dla danego dokumentu jest zawarta w jego sekcji "DATE")

gdzie:

—  $DK1 = \{\text{Alabama, Alaska, American Samoa, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Guam, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Minor Outlying Islands, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Northern Mariana Islands, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, U.S. Virgin Islands, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming, Cologne, Trier, Mainz, Marburg, Aachen, Benkastel-Kues, Düsseldorf, Tiergarten, Kreuzberg, Charlottenburg, Wilmersdorf, Reinickendorf, Spandau Steglitz-Zehlendorf, Neukölln, Schöneberg, Tempelhof, Wedding, Hokkaido, Tohoku, Kanto, Chubu, Kinki, Chugoku, Shikoku, Kyushu-Okinawa, Tokyo, Yokohama, Osaka, Nagoya, Sapporo, Kobe, Kyoto, Fukuoka, New York, Los Angeles, Chicago, Houston, Phoenix, Philadelphia, San Antonio, San Diego, Dallas, San Jose, Detroit, Jacksonville, Indianapolis, San Francisco, Columbus, Austin, Memphis, Fort Worth, Baltimore, Charlotte, El Paso, Boston, Seattle, Washington, Milwaukee, Denver, Louisville, Las Vegas, Nashville-Davidson, Oklahoma}$

- City, Portland, Auvergne, Rhône-Alpes, Bretagne, Brittany, Bourgogne, Franche-Comté, Corse, Corsica, Val de Loire, Grand Est, Hauts de France, Ile de France, Nouvelle Aquitaine, Normandie, Occitanie, Pays de la Loire, Provence, Cote d’Azur, Paris, Marseille, Lyon, Toulouse, Nice, Strasbourg, Nantes, Bordeaux, North East, North West, Yorkshire and The Humber, East Midlands, West Midlands, East of England, London, South East, South West, London , Manchester, Birmingham-Wolverhampton, Leeds-Bradford, Glasgow, Southampton-Portsmouth, Liverpool, Newcastle, Nottingham, Sheffield, Bristol, Belfast, Leicester, Edinburgh, Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland, Labrador, Nova Scotia, Ontario, Prince Edward Island, Quebec, Saskatchewan, Toronto, Montreal, Calgary, Ottawa, Edmonton, Winnipeg, Mississauga, Vancouver, Brampton}
- $DF2 = \{\text{yen, ¥, JPY, dollar, \$, USD, Deutsche Mark, Deutsche Mark, DM, D-Mark, DEM, FRF, franc, Fr, FF, NF, GBP, £, pound, sterling, British pound, CAD, ¢, C$, dollar canadien}\}$
  - $DP1 = \{\text{Conservative and Unionist Party, Labour Party, Democratic socialism, Scottish National Party, Liberal Democrats, Democratic Unionist Party, Sinn Féin, Plaid Cymru, Social Democratic and Labour Party, Ulster Unionist Party, Scottish Greens, Green Party Northern Ireland, Liberal Party, Fellowship Party, Vanguard Unionist Progressive Party, Democratic Labour, Social Democratic Alliance, United Ulster Unionist Party, Ecology Party, Fancy Dress Party, Ulster Popular Unionist Party, Social Democratic Party, Alliance of Germans, German Party, German Conservative Party - German Right Party, German Reich Party, All-German Bloc/League of Expellees and Deprived of Rights, All-German People’s Party, Communist Party of Germany, German Peace Union, Liberal Party of Canada, New Democratic Party, Green Party of Canada, Communist Party of Canada, Marxist–Leninist Party of Canada, Libertarian Party of Canada, Christian Heritage Party of Canada, Conservative Party of Canada, Anti-Confederation Party, Nationalist, Nationalist Conservative, Equal Rights, Patrons of Industry, Protestant Protective Association, McCarthyite, Socialist Party of Canada, Non-Partisan League, Progressive Party of Canada, Liberal-Progressive, Liberal Protectionist, Progressive-Conservative, Labour Party of Canada, Socialist Party of Canada, Co-operative Commonwealth Federation, Reconstruction Party of Canada, Social Credit Party of Canada, United Reform, Bloc populaire, Labor-Progressive Party, Democratic, Socialist Labour Party, Ralliement créditiste, Parti de la Démocratisation Économique, Republican Party, Union Populaire, Parti Nationaliste du Quebec, Confederation of Regions Party of Canada, Party for the Commonwealth of Canada, Reform Party of Canada, Liberal Democratic Party, Constitutional Democratic Party, Komeito, Japan Innovation Party, Democratic Party for the People ,Japanese Communist Party, Reiwa Shinsengumi, Social Democratic Party, Party of Hope, The Party to Protect the People from NHK, La République, The Republicans, Socialist Party, Territories of Progress, Democratic Movement MoDem, En Commun, La France Insoumise, Radical Party, French Communist Party, Agir,$

Left Party, National Rally, The Centrists, Centrist Alliance, The New Democrats, Soyons libres, Radical Party of the Left, Ecologist Party, Centrist Alliance, Democratic European Force, Mouvement des citoyens, Reconquête R!, Horizons, Ensemble!, Movement of Progressives, Public Place, Debout la France, Freedom Ecology Fraternity, The Patriots, Gauche républicaine et socialiste, Citizen and Republican Movement, Résistons!, Initiative démocratique de gauche, Europe Ecology – The Greens, La Force du 13, Modern Left, Génération citoyens, Démocratie et République, Mouvement libéral et modéré, New Deal, Cap, Génération.s, Federalist Party, Anti-Administration party, Democratic-Republican Party, Toleration Party, National Republican Party, Anti-Masonic Party, Nullifier Party, Working Men’s Party, Whig Party, Liberty Party, Law and Order Party of Rhode Island, American Republican Party, Democratic-Republican Party, Free Soil Party, Southern Rights Party, Unionist Party, American Party, Opposition Party, Opposition Party, Constitutional Union Party, Unconditional Union Party, National Union Party, Radical Democracy Party, Readjuster Party, People’s Party, Liberal Party, Liberal Republican Party, Greenback Party, Socialist Labor Party of America, Anti-Monopoly Party, People’s Party, Silver Party, National Democratic Party, Silver Republican Party, Social Democracy of America, United Christian Party, Social Democratic Party, Home Rule Party of Hawaii, Socialist Party of America, Independence Party, Single Tax Party, Progressive Party, National Woman’s Party, Nonpartisan League, National Party, Minnesota Farmer–Labor Party, Labor Party of the United States, Farmer–Labor Party, Proletarian Party of America, Workers Party of America, Puerto Rican Nationalist Party, American Party, Progressive Party, Communist League of America, American Labor Party, American Workers Party, Workers Party of the United States, Union Party, American Labor Party, America First Party, Progressive Democratic Party, American Vegetarian Party, States’ Rights Democratic Party, Progressive Party, National Renaissance Party, Constitution Party, National States’ Rights Party, Puerto Rican Socialist Party, Mississippi Freedom Democratic Party, Black Panther Party, Patriot Party, Youth International Party, Marxist–Leninist Party, Red Guard Party, Communist Workers Party, American Party, National Socialist Party of America, Raza Unida Party, People’s Party, New Union Party, U.S. Labor Party, International Socialist Organization, Citizens Party, New Alliance Party, Labor–Farm Party of Wisconsin, Populist Party, Grassroots Party, Illinois Solidarity Party, Republican Moderate Party of Alaska}

- $DN1 = \{\text{Brandenburg Gate, Berlin Wall, Reichstag, Cologne Cathedral, Dachau Concentration Camp, Lubeck Town Hall, Holstentor, Würzburg Residence, Munich Frauenkirche, Neuschwanstein, Heidelberg castle, Elbphilharmonie, Frankfurter Römer, Hohenzollern Castle, Ulmer Münster, Sanssouci, Charlottenburg palace}\}$
- $DN2 = \{\text{Empire State, Mount Rushmore, Brooklyn Bridge, The White House, Washington National Cathedral, Jefferson Memorial, Golden Gate Bridge, Golden Gate Bridge, Lincoln Memorial, Biltmore Estate, Chrysler Building, Vietnam Veterans Memorial, St. Patrick’s Cathedral, Washing-}$

- ton Monument, Grand Central Terminal, Gateway Arch, Supreme Court of the United States, St. Regis, Metropolitan Museum of Art, Hotel Del Coronado, World Trade Center, Brooklyn Bridge, Philadelphia City Hall, Bellagio Hotel and Casino, Cathedral of St. John the Divine, Philadelphia Museum of Art}
- $DN3 = \{\text{Kinkaku-ji, Fushimi Inari-Taisha, Tokyo Tower, Tokyo Skytree, Ōsaka castle, Sensō-ji, Tokyo Metropolitan Government Building, Art Triangle Roppongi, Kōtoku-in, Meiji-jingū, Meiji Shrine, Kiyomizu-dera, Shōfukuji Temple, Sennari, Jusandi, Kobe Port Tower, Kabuki-za Theatre, Meiji-Mura, Prada Tokyo Aoyama, Tokyo Metropolitan Art Museum, Asahi Beer Hall}\}$
  - $DN4 = \{\text{Luxembourg Palace, Pompidou Centre, Les Invalides, Paris Opera house, Panthéon, Sacré-Cœur Basilica, Notre-Dame Cathedral, Palais du Louvre, Arc de Triomphe, Eiffel Tower, Eiffel, Palace of Versailles, Millau Viaduct, Abbey of Fontenay, Pont du Gard, Musée d'Orsay}\}$
  - $DN5 = \{\text{Buckingham Palace, Westminster Abbey, Big Ben, St Paul's Cathedral, York Minster, St George's Hall Liverpool, Cathedral Church of the Blessed Virgin Mary of Lincoln, Windsor castle, Windsor, Rievaulx Abbey, All Saints Church, Hadrian's Wall, Albert Dock, Coventry Catedry, Blackpool Tower, Warwick castle, British Museum, Chatsworth House, Blenheim Palace, The Shard, Royal Pavilion, Spinnaker Tower, National Gallery, Bankside Power Station, Bath Abbey, Royal Albert Hall, Highclere Castle}\}$
  - $DN6 = \{\text{CN Tower, Habitat 67, Canada Place, Science World, Gooderham Building, Fairmont Banff Springs, Château Frontenac, Art Gallery of Ontario, Parliament of Canda}\}$
  - $DR1 = \{\text{Germany, USA, America, United States, Japan, France, England, Great Britain, United Kingdom, Canada}\}$

Po procesie ekstrakcji cech, każdemu tekstowi zostaje przypisany wektor cech postaci:  $\vec{W} = (K1, F2, P1, N1, N2, N3, N4, N5, N6, R1, M1)$

## 2.2. Miary jakości klasyfikacji

W celu oceny jakości klasyfikacji dokumentów tekstowych [3] przez algorytm, przyjęto konwencje nazewnictw  $A_{bc}$ , potrzebną do wyznaczenia miar jakości klasyfikacji, gdzie:

- $A$  - wartość oznaczająca czy poprawnie zaklasyfikowano dokument (możliwe wartości =  $\{F, P\}$ )
- $b$  - wartość oznaczająca kraj, o którym mowa w dokumencie (możliwe wartości =  $\{\text{France, Canda, UK, USA, West-Germany, Japan}\}$ )
- $c$  - wartość oznaczająca kraj, do którego został zaklasyfikowany dokument (możliwe wartości =  $\{\text{France, Canda, UK, USA, West-Germany, Japan}\}$ )
- $A = F \iff b \neq c$
- $A = P \iff b = c$

Przykładowo:

- Wartość  $F_{UK\ Canada}$  oznacza liczbę dokumentów traktujących o Wielkiej Brytanii, które niepoprawnie zostały zaklasyfikowane jako dokumenty traktujące o Kanadzie
- Zapis  $P_{UK\ UK} = 12$  oznacza 12 dokumentów traktujących o Wielkiej Brytanii zostało zaklasyfikowane poprawnie

Pewne kombinacje wartości  $A$ ,  $b$  oraz  $c$  są logicznie niepoprawne i nie będą wykorzystywane w sprawozdaniu, przykładowo:

- Zapis  $F_{UK\ UK}$  jest logicznie sprzeczny ( $A = F \iff b \neq c$ )

Ocena klasyfikacji przez wykorzystany algorytm została wykonana przy pomocy następujących miar [4] (wzory miar wykorzystują powyższą konwencję nazewnictwa wartości):

1. miara dokładności ACC (ang. accuracy) określana dla całości wyników klasyfikacji (iloraz liczby dokumentów zaklasyfikowanych poprawnie oraz liczby wszystkich dokumentów podlegających klasyfikacji). Wskaźnik ten pozwala ocenić jakość dokonanej klasyfikacji tekstu. Przyjmuje on wartości z przedziału  $[0,1]$ , przy czym wartość 1 oznacza, że wszystkie badane elementy zostały zaklasyfikowane poprawnie. Przykładowo miara  $ACC = 0.85$ , oznacza, że 85% wszystkich badanych elementów zostało poprawnie zaklasyfikowanych.

$$ACC = \sum_x P_{x\ x} / (\sum_x P_{x\ x} + \sum_{xy} F_{x\ y}), \text{ gdzie } x \neq y$$

2. miara precyzji PPV (ang. precision) określana dla poszczególnych klas krajów, do których możemy zaklasyfikować dokumenty (iloraz liczby dokumentów zaklasyfikowanych poprawnie do danego kraju "x" i sumy liczby dokumentów zaklasyfikowanych poprawnie do kraju "x" oraz liczby dokumentów niepoprawnie zaklasyfikowanych do kraju "x") Wskaźnik ten obrazuje ile wśród przykładów zaklasyfikowanych pozytywnie jest rzeczywiście prawidłowo dokonana. Precyzja przyjmuje maksymalną wartość równą 1, co oznacza, że wszystkie poprawnie zaklasyfikowane wektory rzeczywiście należą do danej klasy. Przykładowo miara  $PPV = 0.55$ , oznacza, że 55% wszystkich pozytywnie zaklasyfikowanych elementów rzeczywiście należy do przyporządkowanej klasy.

$$PPV(x) = P_{x\ x} / (P_{x\ x} + \sum_y F_{y\ x}), \text{ gdzie } x \neq y$$

3. miara czułości TPR (ang. recall) określana dla poszczególnych klas krajów, do których możemy zaklasyfikować dokumenty (stosunek liczby dokumentów zaklasyfikowanych poprawnie do danego kraju "x" do sumy liczby dokumentów zaklasyfikowanych poprawnie do kraju "x" oraz liczby dokumentów niepoprawnie zaklasyfikowanych do kraju innego niż "x") Miara ta pozwala ocenić jaki jest udział prawidłowo rozpoznanych przy-

padków wśród wszystkich przypadków pozytywnych. Czułość jest również parametrem, który powinien przyjmować wartości jak najbliższe 1. Przykładowo, kiedy miara TPR osiągnęła wartość 0.30, oznacza to, że 30% wszystkich elementów, które powinny do danej klasy należeć, rzeczywiście do niej należy.

$$TPR(x) = P_{xx} / (P_{xx} + \sum_y F_{xy}) , \text{ gdzie } x \neq y$$

4. miara precyzji  $PPV_c$ , określana dla całości wyników klasyfikacji, obliczana jest jako średnia ważona poszczególnych wartości miary  $PPV$  dla danej klasy i jej liczebności. Miara ta przyjmuje wartości z zakresu analogicznego jak miara  $PPV$ . W najlepszym przypadku przyjmuje wartość 1, jest to sytuacja kiedy dla wszystkich rozważanych klas poszczególne miary  $PPV$  osiągnęły wartość maksymalną.

$$PPV_c = (\sum_x PPV_x \cdot n_x) / \sum_x n_x,$$

gdzie  $n$  jest liczebnością danej klasy,  $n \in C$ .

5. miara czułości  $TPR_c$ , określana dla całości wyników klasyfikacji, obliczana jest jako średnia ważona poszczególnych wartości miary  $TPR$  dla danej klasy i jej liczebności. Miara ta przyjmuje wartości z zakresu analogicznego jak miara  $TPR$ . W najlepszym przypadku przyjmuje wartość 1, jest to sytuacja kiedy dla wszystkich rozważanych klas poszczególne miary  $TPR$  osiągnęły wartość maksymalną.

$$TPR_c = (\sum_x PPV_x \cdot n_x) / \sum_x n_x,$$

gdzie  $n$  jest liczebnością danej klasy,  $n \in C$ .

6. miara  $F1$  ( $F1 - score$ ), określana dla każdej z klas z osobna, obliczana jako średnia harmoniczna pomiędzy precyzją i czułością. Im otrzymana wartość jest bliższa jedynce, tym lepiej to świadczy o dokładności algorytmu klasyfikującego. W najlepszym przypadku przyjmuje ona wartość 1, zaś w najgorszym - 0. Miara ta pozwala na wyznaczenie dokładności algorytmu, bazując tylko wyznaczonej wcześniej jego dokładności i precyzji.

$$F1(x) = 2 \cdot \frac{PPV_x \cdot TPR_x}{PPV_x + TPR_x}, \text{ gdzie } x \text{ określa daną klasę.}$$

7. miara  $F1_c$ , określana dla całości wyników klasyfikacji, obliczana jako średnia ważona wartości miary  $F1$  dla każdej z klas. Miara ta przyjmuje wartości z zakresu analogicznego jak w przypadku miary  $F1$ .

$$F1_c = (\sum_x F1_x \cdot n_x) / \sum_x n_x, \text{ gdzie } n \text{ jest liczebnością danej klasy.}$$

### 3. Klasyfikacja z użyciem metryk i miar podobieństwa tekstów

Wzory, znaczenia i opisy symboli zastosowanych metryk z przykładami. Wzory, opisy i znaczenia miar podobieństwa tekstów zastosowanych w obliczaniu metryk dla wektorów cech z przykładami dla każdej miary [2]. Oznaczenia jednolite w obrębie całego sprawozdania. Wstępne wyniki miary Accuracy dla próbnych klasyfikacji na ograniczonym zbiorze tekstów (podać parametry i kryteria wyboru wg punktów 3.-8. z opisu Projektu 1.). **Podaj metryki i miary podobieństwa nie z literatury (te wystarczy zacytować linkiem), ale konkretne ich postaci stosowane w zadaniu. Jakie zakresy wartości przyjmują te miary i metryki, co oznaczają ich wartości? Podaj przykładowe wartości dla przykładowych wektorów cech.**

**Sekcja uzupełniona jako efekt zadania Tydzień 04 wg Harmonogramu Zajęć na WIKAMP KSR.**

### 4. Budowa aplikacji

#### 4.1. Diagramy UML

Diagramy UML i zwięzłe opisy: idei aplikacji, modułu ekstrakcji i modułu klasyfikatora.

**Sekcja uzupełniona jako efekt zadania Tydzień 03 wg Harmonogramu Zajęć na WIKAMP KSR.**

#### 4.2. Prezentacja wyników, interfejs użytkownika

Krótki ilustrowany opis jak użytkownik może korzystać z aplikacji, w szczególności wprowadzać parametry klasyfikacji i odczytywać wyniki. Wersja JRE i inne wymogi niezbędne do uruchomienia aplikacji przez użytkownika na własnym komputerze.

**Sekcja uzupełniona jako efekt zadania Tydzień 04 wg Harmonogramu Zajęć na WIKAMP KSR.**

### 5. Wyniki klasyfikacji dla różnych parametrów wejściowych

Wyniki kolejnych eksperymentów wg punktów 2.-8. opisu projektu 1. Wykresy i tabele obowiązkowe, dokładnie opisane w „captions” (tytułach), konieczny opis osi i jednostek wykresów oraz kolumn i wierszy tabel.

**\*\*Ewentualne wyniki realizacji punktu 9. opisu Projektu 1., czyli „na ocenę 5.0” i ich porównanie do wyników z części obowiązkowej\*\*.**

**Sekcja uzupełniona jako efekt zadania Tydzień 05 wg Harmonogramu Zajęć na WIKAMP KSR.**



## 6. Dyskusja, wnioski

Dokładne interpretacje uzyskanych wyników w zależności od parametrów klasyfikacji opisanych w punktach 3.-8 opisu Projektu 1. Szczególnie istotne są wnioski o charakterze uniwersalnym, istotne dla podobnych zadań. Omówić i wyjaśnić napotkane problemy (jeśli były). Każdy wniosek/problem powinien mieć poparcie w przeprowadzonych eksperymentach (odwołania do konkretnych wyników: wykresów, tabel).

Dla końcowej oceny jest to najważniejsza sekcja sprawozdania, gdyż prezentuje poziom zrozumienia rozwiązywanego problemu.

\*\* Możliwości kontynuacji prac w obszarze systemów rozpoznawania, zwłaszcza w kontekście pracy inżynierskiej, magisterskiej, naukowej, itp. \*\*

**Sekcja uzupełniona jako efekt zadania Tydzień 06 wg Harmonogramu Zajęć na WIKAMP KSR.**

## 7. Braki w realizacji projektu 1.

Wymienić wg opisu Projektu 1. wszystkie niezrealizowane obowiązkowe elementy projektu, ewentualnie podać merytoryczne (ale nie czasowe) przyczyny tych braków.

## Literatura

- [1] R. Tadeusiewicz: Rozpoznawanie obrazów, PWN, Warszawa, 1991.
- [2] A. Niewiadomski, Methods for the Linguistic Summarization of Data: Applications of Fuzzy Sets and Their Extensions, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2008.
- [3] <http://archive.ics.uci.edu/ml/datasets/Reuters-21578+Text+Categorization+Collection>
- [4] [https://pl.wikipedia.org/wiki/Tablica\\_pomylek](https://pl.wikipedia.org/wiki/Tablica_pomylek)

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