

SPRAWOZDANIE – LABORATORIUM 6

Karolina Kotłowska, 15 kwietnia 2023

6.1

Relation: autorzy

| No. | 1: author | 2: work | 3: content | 4: content_stemmed |
|-----|-----------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Reym... | Ziemia obiecana | I Łódź się budziła. Pierwszy wrzaskiwy św... | I Łódź Łódź się budzi pierwszy wrzaskiwy świst fabryczny rozedrzeć cisza ciszy wczesny poranek a za nim on we w wszystkie ws... |
| 2 | Reym... | Ziemia obiecana | - Zaraz będę budził, jeśli pan dyrektor każe... | - Zaraz zaraz być budzi jeśli pan dyrektor każe bo pan mówił wieczór wieczór że chce dzisiaj spać dużo dużo iść obudzić klu... |
| 3 | Reym... | Ziemia obiecana | - Ale spał się też fabryka Goldberga, na... | Al Ala Al Alo spać się też tenże fabryka Goldberg na cegielniny naszać nasz straż jeździć ale wszystko dobrze pójść ostatec ostatec... |
| 4 | Reym... | Ziemia obiecana | - Morycl - zawołal do drugiego pokoju - Ni... | zawołać do drugie drugi pokój nie on spać nie on spać cały noc dlaczego myśleć o ocean ojciec tym ten to nasz nasze nasi intere... |
| 5 | Reym... | Ziemia obiecana | Dziwilem się nawet, że tak długo zwleka, p... | dziwić się nawet że tak tak długo zwlekać przecieć procent mu on nie on dać od oda assekuracja Miałoby dużo towar Miałoby dużo zaas... |
| 6 | Reym... | Ziemia obiecana | Juz po piątej. Odpowiedź zagłuszyły świst... | juz po piąta piątą odpowiedź zagłuszyły świstawka który się rozległy rozleg jakby luz nad dom i ryczeć przez kilkanaście sekunda z Ł... |
| 7 | Reym... | Ziemia obiecana | A reszta wczoraj się z potęć Maks ty śle skrócić przez to ten zarode zart zniec zniec się diały i z wszystkie wszystkie mniemać niech... | A reszta wczoraj się z potęć Maks ty śle skrócić przez to ten zarode zart zniec zniec się diały i z wszystkie wszystkie mniemać niech... |
| 8 | Reym... | Ziemia obiecana | Coñal się do swojego pokoju i po chwili... | coñać się do swoje swój pokój i po chwila wynieść cały garderoba rzucić on na dywan i z wolny się ubierać ty my psuć psować int... |
| 9 | Reym... | Ziemia obiecana | - Dlaczego? - zapytał cicho i oparł się o st... | - Dlaczego? - zapytał cicho i oparł się o ocean ojciec stół ja ci ten ty powiedzieć jeśli ten to nie on rozumieć przed wszystko wszystkie... |
| 10 | Reym... | Ziemia obiecana | Morycl na odpowiedź usłyszaną odwrócił si... | na odpowiedzieć usłyszeć odwrócić się gwałtownie aż binokle binokl mu on spadły spaść i uderzyć w wiek poręcz poręcz krze... |
| 11 | Reym... | Ziemia obiecana | Będziemy mówić o nich wtedy, jak będzie... | być mówić o ocean ojciec on wtedy jak jaka być mieć mieć pieniądze a teraz trzeba siedzieć cicho on być my potrzebny no niech K... |
| 12 | Reym... | Ziemia obiecana | - Nasza wczorajsza rozmowa na czym sta... | - Nasza nasz wczorajszy rozmowa na czym co stanąć zapytać spokojnie juz zakładać fabryka tak taka ja nie on mama mamieć mieć... |
| 13 | Reym... | Ziemia obiecana | - powtórzył obaj - Co to, Goldberg się spa... | - powtórzył oba co co co to ten Goldberg się spać zapytać tak taka zrobić się bilans Mądry Mądry chłop zrobić milion albo alba s... |
| 14 | Reym... | Ziemia obiecana | Maks się nie odezwał. Świsławki znowu z... | Maks się nie on odezwać znowu zacząć podnosić swoje swój przenikliwy denenujący denenować glos i śpiewać coraz toponie... |
| 15 | Reym... | Ziemia obiecana | - Ty, Borowiecki, jesteś szlachoc, masz n... | ty Borowiecki, jesteś szlachoc mieć na bilet wójtowy herb Madrac kłaść nazywać swoje swój von a być duży wielki z my wsz... |
| 16 | Reym... | Ziemia obiecana | - rzuć! Ktoś stojący stać biec dalej dalej szepać i pójść wolnie wolnie wolno grać go on wąpiłwość hiszyc hiszyc myśleć myśle... | rzuć! Ktoś stojący stać biec dalej dalej szepać i pójść wolnie wolnie wolno grać go on wąpiłwość hiszyc hiszyc myśleć myśle... |
| 17 | Reym... | Ziemia obiecana | Białe dymy zaczęły biec z kominów i rozwie... | Biała Biała Biała Biała dym dym zacząć biec biec z komin i rozwieć się pomiędzy tym ten to potężny kamieniny las co co co... |
| 18 | Reym... | Ziemia obiecana | Pogładził nerwowo broda mokry od deszcz... | pogładzić nerwowo broda mokry od oda deszcz deszcz deszcz i oprzytomniać oprzytomnić co co co za głupstwo szepać niec... |
| 19 | Reym... | Ziemia obiecana | Murray, okręcony w długi, niebieski fartuch... | okręcić w wiek długi długi niebieski fartuch wysunąć się spoza rząd ruchomy koły kocół kocół w wiek który się gotować i robić far... |
| 20 | Reym... | Ziemia obiecana | - Pierwsze metry nieco lakowała. Przysłali... | pierwsze pierwsza pierwszy metr nieco lakować przysłać z centrala zamówienie zamówić na pięćset sztuk sztuka ten pański lama... |
| 21 | Reym... | Ziemia obiecana | - Chciałem panu coś powiedzieć... - Słuch... | chcieć pan panu coś powiedzieć słuchać słuchać szepać grzecznie ale z pewny pewien niechęć ująć go on pod ręką i odprowadzi... |
| 22 | Reym... | Ziemia obiecana | - Kupilem wczoraj meble - szepał cicho d... | kupić wczoraj mebel szepać cicho do ucha ucho Borowiecki uważać pan do salonik kupić żółty jedwabny w wiek styl stylo empire... |
| 23 | Reym... | Ziemia obiecana | - Tak jakby już. W niedziele właśnie powie... | tak taka jakby już wat niedziela właśnie powiedzieć mi ja jak jaka by chcieć mieć mieć mieć urządzać mieszkanie mieszkać wyp... |
| 24 | Reym... | Ziemia obiecana | Chciałem koniecznie usłyszeć pańskie zd... | chcieć koniecznie usłyszeć pańskie pański zdanie zdać o ocean ojciec ten mebel przysłać Al Ala Al Alo kiedy po obiad wrócić do far... |
| 25 | Reym... | Ziemia obiecana | Nie wiedział, kto wola. - Panie Borowiecki... | nie on wiedzieć kto wola wola wola Pan Pan Borowiecki digneć bo go on ująć pod ramię ramia a pan prezes szepać poznać wla... |
| 26 | Reym... | Ziemia obiecana | - Rób się - rzucił krótko i szedł naprzód F... | robić się rzucić krótko i iść naprzód fabrykant ucześć u on ramię więc się ogizko podparcie się gruba grubo laska laska i zparbony... |
| 27 | Reym... | Ziemia obiecana | - Panie Borowiecki - zawołał fabrykant do j... | Pan Pan Borowiecki zawołać fabrykant do jakiś ciek co co co co się wychylić z mgła ale to ten nie on być Borowiecki podmnieć się... |
| 28 | Reym... | Ziemia obiecana | Oderi i do wielkiej sali zaczął zaglądać, br... | deri dzień dnia dzień dniem i do wielki sala sala sala zacząć zaglądać brudny światło woskać się przez małe mała mała ma... |
| 29 | Reym... | Ziemia obiecana | Nie mógł żyć z dala od tego potężnego kró... | nie on móc żyć z dala od oda ten to potężny królestwo który stworzyć praca własny cały życie żyć i moc swoje swój geniusz pr... |
| 30 | Reym... | Ziemia obiecana | - Pan Welt dopiero z pół godziny jak od na... | Pan Welt dopiero z pół pół pół godzina jak jaka od oda my wyjść wyniść wyniść Tutaj być brać pięćdziesiąt sztuk sztuka biały bi... |
| 31 | Reym... | Ziemia obiecana | - Cóż słychać, panie Szwarc? - zapytał -... | cóż słychać panie pan pan szwarc zapytać się zalać zupełnie nie on wiadomo jeszcze ale ja myśleć myśleć że się być układać... |
| 32 | Reym... | Ziemia obiecana | - Jego szlag może trafić - szepnął któryś z... | on szlag szlaga może może trafić szepać któryś z pochylić przy robot robot być szkoda bardzo wielki niech bóg bronić bronić żyć sto... |
| 33 | Reym... | Ziemia obiecana | Na prost okien stał wielki, czteropiętrowy b... | na prost okno stać wielki czteropiętrowy budynek z nagi czerwona czerwony cegła i rzucić przynębiający przynębiać refleks do k... |
| 34 | Reym... | Ziemia obiecana | - Co panu jest? - zapytał go Borowiecki, k... | co co co co co pan pan zapytać go on Borowiecki który z nim on znać się blisko nic nica odeprzeć krótko i powlec nienawistny sp... |
| 35 | Reym... | Ziemia obiecana | Hem odeszła. Usiadł przy biurku i błądził s... | Hem odeszła usiąść przy biurko i błądzić spojrzenie spojrzeć po mur czerwony czerwona czerwone po okno za który biec się sto... |
| 36 | Reym... | Ziemia obiecana | - szepnęła cicho, opamięta wzrokiem te w... | szepnąć cicho ogarnąć wzrok te ten ty wszystkie głowa co co co co się podniosły podnieść znad pulpit schylić się pokornie do nog... |
| 37 | Reym... | Ziemia obiecana | - Pan się pyta? Ten cham umyślnie podło... | Pan się pyta pytać ten cham umyślnie podłożyć leć pod kolo kolo koło on się nie on chcieć pracować a on się chcieć okraść fabry... |
| 38 | Reym... | Ziemia obiecana | - wykrzykiwała przypadając mu do nog, cał... | wykrzykiwać przypadając mu on do noga całować go on po ręką wydrzeć się jej on i wyjść wyniść wyniść ale przystanąć w wiek wiel... |
| 39 | Reym... | Ziemia obiecana | - Biduję, wielmożny panie, biduję. Mie... | panie pan pan mieszkać na Bałut z jeden weber weber i cały tyz rubel płaca płacić za pomieszkać miesiecy półki mój nieboszc... |
| 40 | Reym... | Ziemia obiecana | Harował nieborak od świtu do nocy i jeś... | harować nieborak od oda świt do nocy i jeść nie on być co co co co żyć ki te ten ty dziady dziad ostatnie ostatni a nie on ki ki pies a... |

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose: NaiveBayes

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds: 10

☐ Percentage split % 66

More options...

(Nom) author

Start Stop

Result list (right-click for options)

11:28:56 - bayes NaiveBayes

Classifier output

Summary

| Metric | Value | Percentage |
|----------------------------------|-----------|------------|
| Correctly Classified Instances | 4288 | 96.4679 % |
| Incorrectly Classified Instances | 157 | 3.5321 % |
| Kappa statistic | 0.9441 | |
| Mean absolute error | 0.0183 | |
| Root mean squared error | 0.1222 | |
| Relative absolute error | 5.7834 % | |
| Root relative squared error | 30.7609 % | |
| Total Number of Instances | 4445 | |

Detailed Accuracy By Class

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| | 0,977 | 0,009 | 0,980 | 0,977 | 0,979 | 0,969 | 0,999 | 0,998 | Reymont |
| | 0,926 | 0,007 | 0,929 | 0,926 | 0,928 | 0,921 | 0,998 | 0,981 | Żuławski |
| | 0,977 | 0,023 | 0,977 | 0,977 | 0,977 | 0,954 | 0,998 | 0,998 | Sienkiewicz |
| | 0,896 | 0,013 | 0,884 | 0,896 | 0,890 | 0,878 | 0,994 | 0,960 | Żeromski |
| Weighted Avg. | 0,965 | 0,016 | 0,965 | 0,965 | 0,965 | 0,948 | 0,998 | 0,993 | |

Confusion Matrix

| a | b | c | d | Classified as |
|------|-----|------|-----|-----------------|
| 1334 | 15 | 7 | 9 | a = Reymont |
| 22 | 365 | 4 | 3 | b = Żuławski |
| 1 | 12 | 2201 | 39 | c = Sienkiewicz |
| 4 | 1 | 40 | 388 | d = Żeromski |

Status

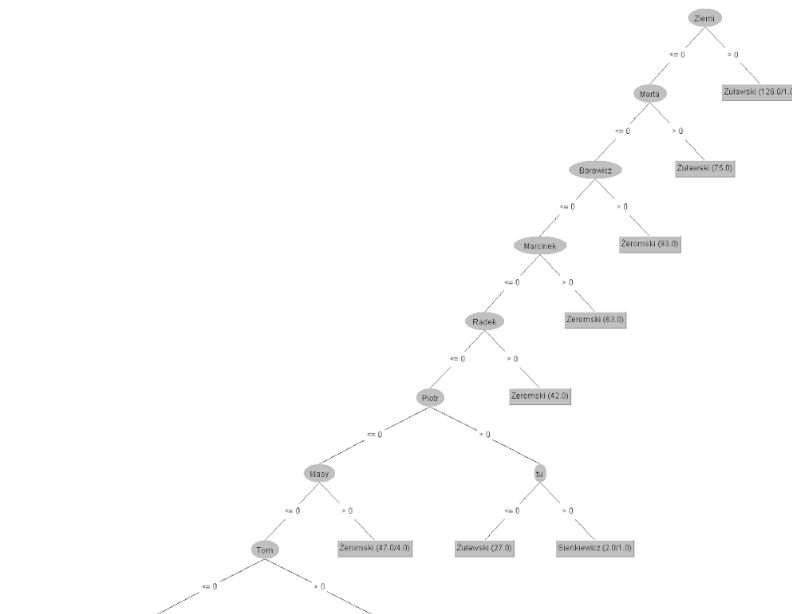
OK

Log x0

• Informatyka I (EAIIB)

• Studio projektowe 1

6.2



Najczęściej występującymi wartościami było słowo takie jak: 'ziemia' oraz imiona bohaterów danych książek.

Classifier
Choose J48 - C 0.25 - M 2

Test options
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds 10
☐ Percentage split % 66
 More options...

Classifier output
 Time taken to build model: 91.47 seconds
 --- Evaluation on test split ---
 Time taken to test model on test split: 0.11 seconds
 --- Summary ---
 Correctly Classified Instances 1328 87.888 %
 Incorrectly Classified Instances 183 12.112 %
 Kappa statistic 0.8981
 Mean absolute error 0.0647
 Root mean squared error 0.2303
 Relative absolute error 20.5064 %
 Root relative squared error 58.0956 %
 Total Number of Instances 1511
 --- Detailed Accuracy By Class ---

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| | 0.897 | 0.069 | 0.863 | 0.897 | 0.878 | 0.819 | 0.959 | 0.869 | Reymont |
| | 0.754 | 0.021 | 0.777 | 0.754 | 0.765 | 0.743 | 0.902 | 0.692 | Zulawski |
| | 0.917 | 0.051 | 0.948 | 0.917 | 0.933 | 0.867 | 0.962 | 0.956 | Sienkiewicz |
| | 0.715 | 0.033 | 0.674 | 0.715 | 0.694 | 0.665 | 0.915 | 0.719 | Zeromski |
| Weighted Avg. | 0.879 | 0.052 | 0.881 | 0.879 | 0.880 | 0.823 | 0.945 | 0.884 | |

 --- Confusion Matrix ---
 a b c d <-- classified as
 436 16 17 17 | a = Reymont
 21 101 8 4 | b = Zulawski
 20 11 698 24 | c = Sienkiewicz
 22 2 13 93 | d = Zeromski

Result list (right click for options)
 112856 - bayes.NaiveBayes
 113136 - trees.J48
 114025 - trees.J48

Status
 OK Log x0

J48 wykonywał się około 90 sekund, natomiast Naive Bayes wykonywał się 5 razy krócej.

6.3

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds 5

☐ Percentage split % 66

More options

(Nom) author

Start Stop

Result list (right-click for options)

11:28:56 - bayes NaiveBayes

11:31:36 - trees J48

11:40:25 - trees J48

11:48:29 - bayes NaiveBayes

11:49:19 - bayes NaiveBayes

Classifier output

Time taken to build model: 3.42 seconds

=== Stratified cross-validation ===

=== Summary ===

| Metric | Value | Percentage |
|----------------------------------|-----------|------------|
| Correctly Classified Instances | 4172 | 93.8583 % |
| Incorrectly Classified Instances | 273 | 6.1417 % |
| Kappa statistic | 0.9038 | |
| Mean absolute error | 0.0308 | |
| Root mean squared error | 0.1651 | |
| Relative absolute error | 9.7416 % | |
| Root relative squared error | 41.5457 % | |
| Total Number of Instances | 4445 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|-------------|-------|
| 0,955 | 0,035 | 0,923 | 0,955 | 0,939 | 0,911 | 0,994 | 0,987 | Reymont | |
| 0,924 | 0,014 | 0,863 | 0,924 | 0,892 | 0,882 | 0,996 | 0,965 | Zulaewski | |
| 0,943 | 0,028 | 0,972 | 0,943 | 0,957 | 0,915 | 0,993 | 0,994 | Sienkiewicz | |
| 0,875 | 0,011 | 0,896 | 0,875 | 0,886 | 0,873 | 0,991 | 0,951 | Żeromski | |
| Weighted Avg. | 0,939 | 0,028 | 0,940 | 0,939 | 0,939 | 0,907 | 0,994 | 0,985 | |

=== Confusion Matrix ===

| | a | b | c | d | <-- classified as |
|------|-----|------|-----|---|-------------------|
| 1304 | 11 | 34 | 16 | | a = Reymont |
| 23 | 364 | 7 | 0 | | b = Zulaewski |
| 56 | 44 | 2129 | 28 | | c = Sienkiewicz |
| 30 | 3 | 21 | 379 | | d = Żeromski |

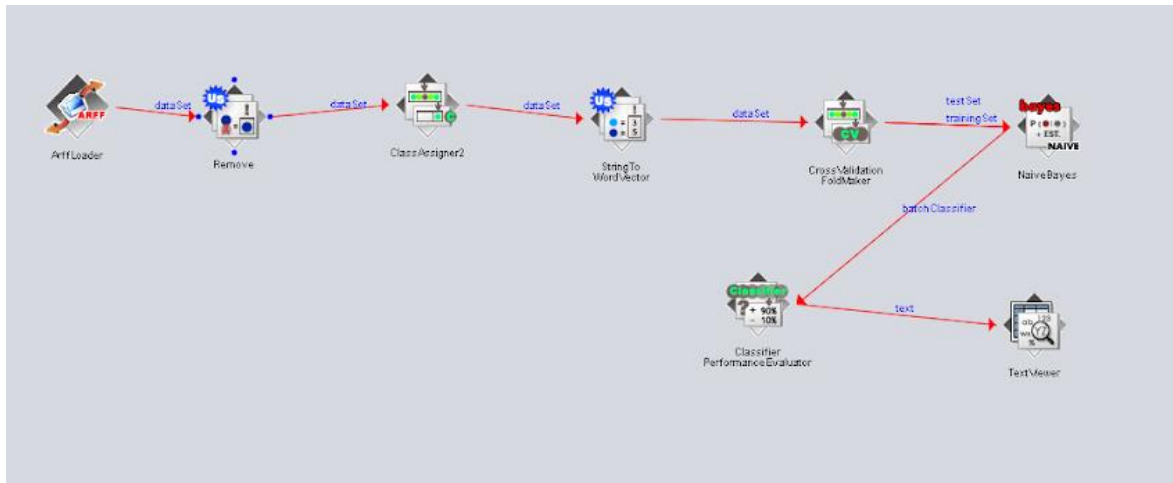
Status

OK Log x 0

Porównanie wyników:

| Precision | Recall | F-Measure | Precision | Recall | F-Measure |
|-----------|--------|-----------|-----------|--------|-----------|
| 0,860 | 0,897 | 0,878 | 0,923 | 0,955 | 0,939 |
| 0,777 | 0,754 | 0,765 | 0,863 | 0,924 | 0,892 |
| 0,948 | 0,917 | 0,933 | 0,972 | 0,943 | 0,957 |
| 0,674 | 0,715 | 0,694 | 0,896 | 0,875 | 0,886 |
| 0,881 | 0,879 | 0,880 | 0,940 | 0,939 | 0,939 |

6.4



five-books-all-1000-1-stem.arff

Attribute summary | Scatter plot matrix | SQL Viewer | Simple CLI

Untitled1 x lab6-kk x

Text Viewer

Result list

12:07:19.193 - NaiveBay

Text

Scheme: NaiveBayes
Relation: autorzy-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.attribute.StringToWordVector-R2-W1000-prune-r

Correctly Classified Instances 29926 67.3615 %
Incorrectly Classified Instances 14500 32.6385 %
Kappa statistic 0.443
Mean absolute error 0.2013
Root mean squared error 0.3366
Relative absolute error 63.7698 %
Root relative squared error 84.7183 %
Total Number of Instances 44426

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| | 0.613 | 0.082 | 0.765 | 0.613 | 0.682 | 0.570 | 0.369 | 0.786 | Reymont |
| | 0.354 | 0.063 | 0.352 | 0.354 | 0.353 | 0.290 | 0.798 | 0.336 | Zulawski |
| | 0.883 | 0.409 | 0.689 | 0.883 | 0.774 | 0.496 | 0.849 | 0.838 | Stenkiewicz |
| | 0.065 | 0.011 | 0.383 | 0.065 | 0.111 | 0.125 | 0.757 | 0.245 | Zeromski |
| Weighted Avg. | 0.674 | 0.239 | 0.654 | 0.674 | 0.644 | 0.464 | 0.841 | 0.720 | |

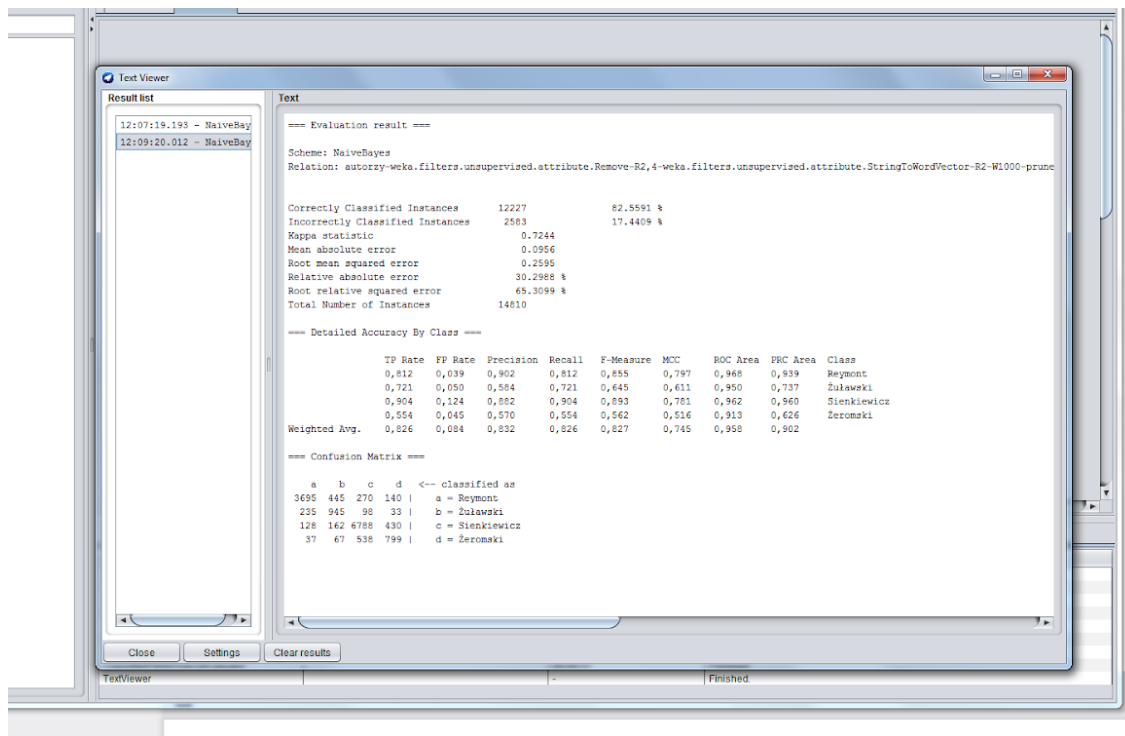
=== Confusion Matrix ===

| | a | b | c | d | <-- classified as |
|------|------|-------|-----|---|-------------------|
| 8372 | 1202 | 3899 | 176 | | a = Reymont |
| 891 | 1391 | 1596 | 54 | | b = Zulawski |
| 1299 | 1118 | 19883 | 222 | | c = Stenkiewicz |
| 337 | 237 | 3469 | 280 | | d = Zeromski |

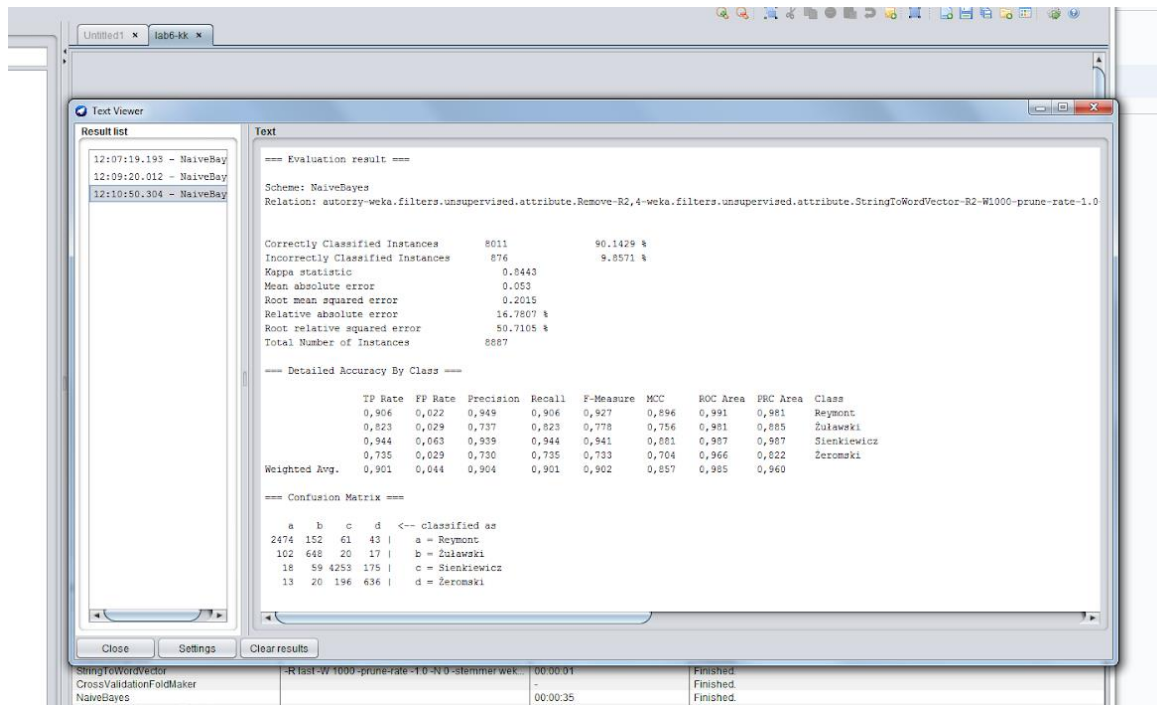
Close Settings Clear results

| Step | Operation | Time | Status |
|------|--------------------------|----------|----------|
| 1 | ArffLoader | 00:00:01 | Finished |
| 2 | Remove | - | Finished |
| 3 | ClassAssigner2 | - | Finished |
| 4 | StringToWordVector | 00:00:02 | Finished |
| 5 | CrossValidationFoldMaker | - | Finished |
| 6 | NaiveBayes | 00:01:59 | Finished |

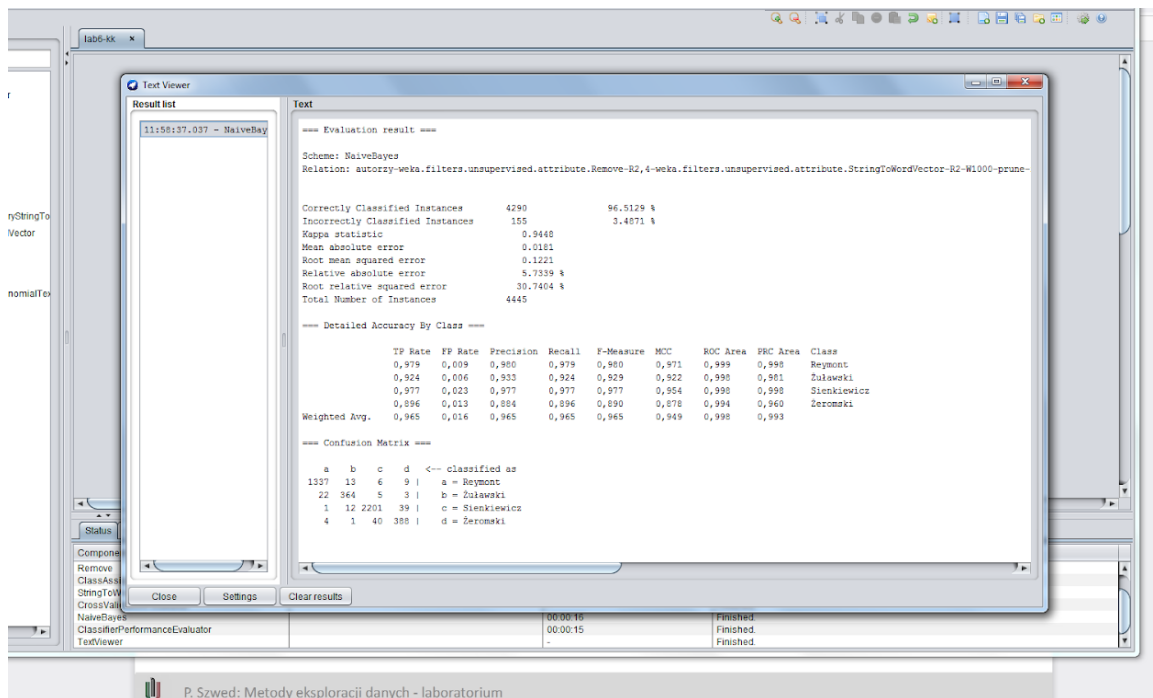
five-books-all-1000-3-stem.arff



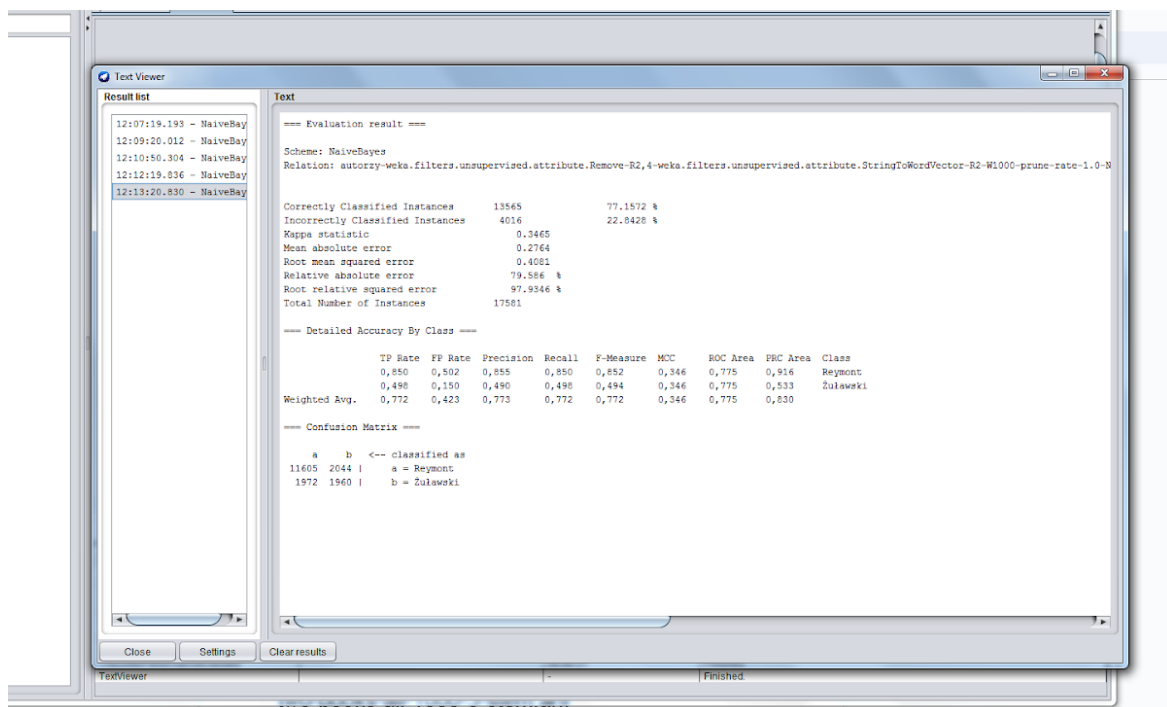
five-books-all-1000-5-stem.arff



five-books-all-1000-10-stem.arff



two-books-all-1000-1-stem.arff



two-books-all-1000-3-stem.arff

The screenshot shows the Weka GUI with the 'Text Viewer' window open. The 'Result list' on the left shows several NaiveBayes models. The 'Text' window on the right displays the evaluation results for the selected model.

Result list:

- 12:07:19.193 - NaiveBayes
- 12:09:20.012 - NaiveBayes
- 12:10:50.304 - NaiveBayes
- 12:12:19.836 - NaiveBayes
- 12:13:20.830 - NaiveBayes
- 12:14:56.142 - NaiveBayes

Text:

=== Evaluation result ===

Scheme: NaiveBayes
Relation: autorzy-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.attribute.StringToWordVector-R2-W1000-prune-r

| Metric | Value | Percentage |
|----------------------------------|-----------|------------|
| Correctly Classified Instances | 5049 | 86.1457 % |
| Incorrectly Classified Instances | 812 | 13.8543 % |
| Kappa statistic | 0.6311 | |
| Mean absolute error | 0.148 | |
| Root mean squared error | 0.3333 | |
| Relative absolute error | 42.5969 % | |
| Root relative squared error | 79.9766 % | |
| Total Number of Instances | 5861 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PBC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|----------|
| | 0,878 | 0,195 | 0,940 | 0,878 | 0,908 | 0,637 | 0,926 | 0,975 | Reymont |
| | 0,805 | 0,122 | 0,655 | 0,805 | 0,722 | 0,637 | 0,926 | 0,830 | Zulawski |
| Weighted Avg. | 0,861 | 0,179 | 0,876 | 0,861 | 0,866 | 0,637 | 0,926 | 0,943 | |

=== Confusion Matrix ===

| a | b | -- classified as |
|------|------|------------------|
| 3994 | 556 | a = Reymont |
| 256 | 1055 | b = Zulawski |

Close Settings Clear results

NaiveBayes 00:00:14 Finished
ClassifierPerformanceEvaluator 00:00:11 Finished
TextViewer - Finished

two-books-all-1000-5-stem.arff

The screenshot shows the Weka GUI with the 'Text Viewer' window open. The 'Result list' on the left shows several NaiveBayes models. The 'Text' window on the right displays the evaluation results for the selected model.

Result list:

- 12:07:19.193 - NaiveBayes
- 12:09:20.012 - NaiveBayes
- 12:10:50.304 - NaiveBayes
- 12:12:19.836 - NaiveBayes
- 12:13:20.830 - NaiveBayes
- 12:14:56.142 - NaiveBayes
- 12:15:54.836 - NaiveBayes

Text:

=== Evaluation result ===

Scheme: NaiveBayes
Relation: autorzy-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.attribute.StringToWordVector-R2-W1000-prune-r

| Metric | Value | Percentage |
|----------------------------------|-----------|------------|
| Correctly Classified Instances | 3272 | 93.0338 % |
| Incorrectly Classified Instances | 245 | 6.9662 % |
| Kappa statistic | 0.805 | |
| Mean absolute error | 0.0752 | |
| Root mean squared error | 0.2426 | |
| Relative absolute error | 21.6316 % | |
| Root relative squared error | 58.215 % | |
| Total Number of Instances | 3517 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PBC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|----------|
| | 0,944 | 0,116 | 0,966 | 0,944 | 0,955 | 0,806 | 0,975 | 0,992 | Reymont |
| | 0,884 | 0,056 | 0,819 | 0,884 | 0,850 | 0,806 | 0,975 | 0,940 | Zulawski |
| Weighted Avg. | 0,930 | 0,102 | 0,933 | 0,930 | 0,931 | 0,806 | 0,975 | 0,980 | |

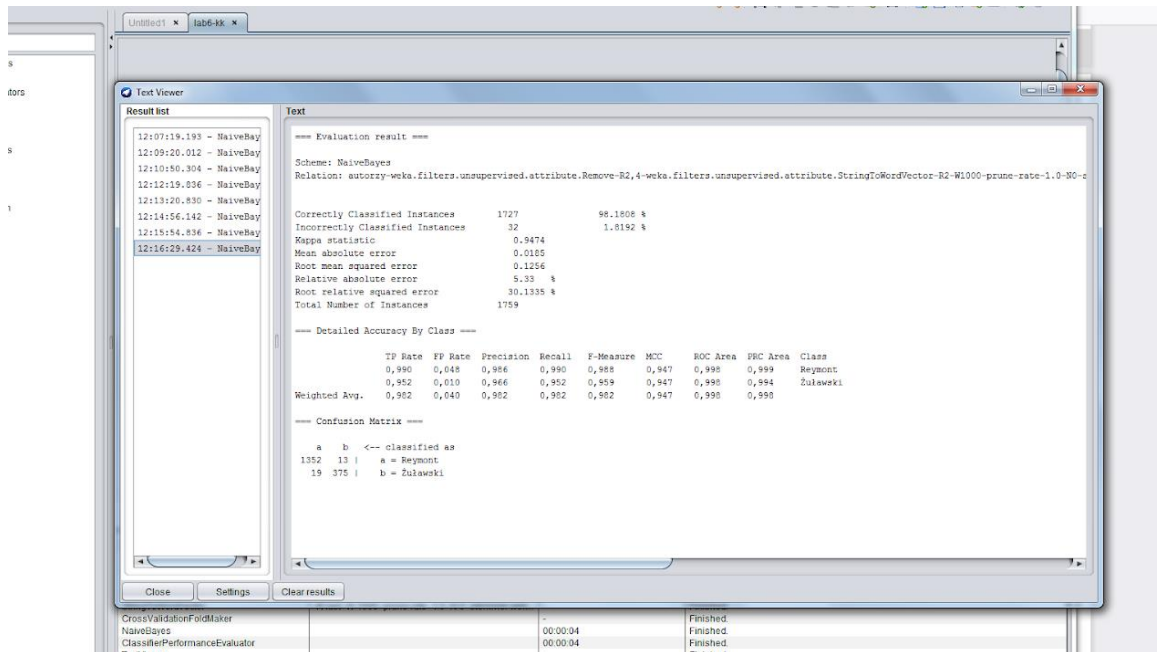
=== Confusion Matrix ===

| a | b | -- classified as |
|------|-----|------------------|
| 2576 | 154 | a = Reymont |
| 91 | 696 | b = Zulawski |

Close Settings Clear results

StringToWordVector 00:00:01 Finished
CrossValidationFoldMaker - Finished
NaiveBayes 00:00:06 Finished

two-books-all-1000-10-stem.arff

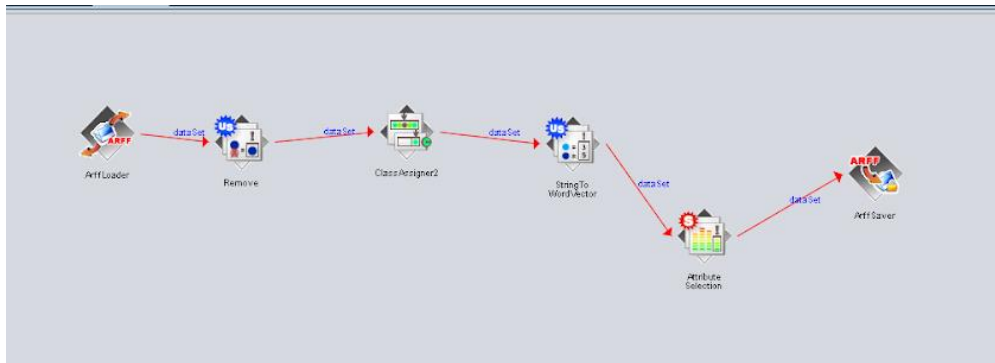


| Zbiór danych | Precision | Recall | F-measure |
|----------------------------------|-----------|--------|-----------|
| five-books-all-1000-1-stem.arff | 0,654 | 0,674 | 0,464 |
| five-books-all-1000-3-stem.arff | 0,832 | 0,826 | 0,827 |
| five-books-all-1000-5-stem.arff | 0,904 | 0,901 | 0,902 |
| five-books-all-1000-10-stem.arff | 0,965 | 0,965 | 0,965 |
| two-books-all-1000-1-stem.arff | 0,773 | 0,772 | 0,772 |
| two-books-all-1000-3-stem.arff | 0,876 | 0,861 | 0,866 |
| two-books-all-1000-5-stem.arff | 0,933 | 0,930 | 0,931 |
| two-books-all-1000-10-stem.arff | 0,982 | 0,982 | 0,982 |

Wnioski: Większa liczba danych w zbiorze danych, zwiększa prawdopodobieństwa uzyskania lepszego i dokładniejszego wyniku klasyfikacji. Wraz ze wzrostem liczby klas, wyniki klasyfikacji mogą okazywać się coraz mniej dokładne. Zastosowanie narzędzi klasyfikacji tekstów można spotkać takie jak: analiza sentymentu (służąca do określania tego, czy dany tekst jest pozytywny, negatywny czy neutralny), filtracja spamu (wykorzystywane w poczcie elektronicznej czy na portalach internetowych), personalizacja reklam (dzięki analizie tekstów można określić zainteresowania użytkownika i dopasować do niego odpowiednie reklamy).

6.5

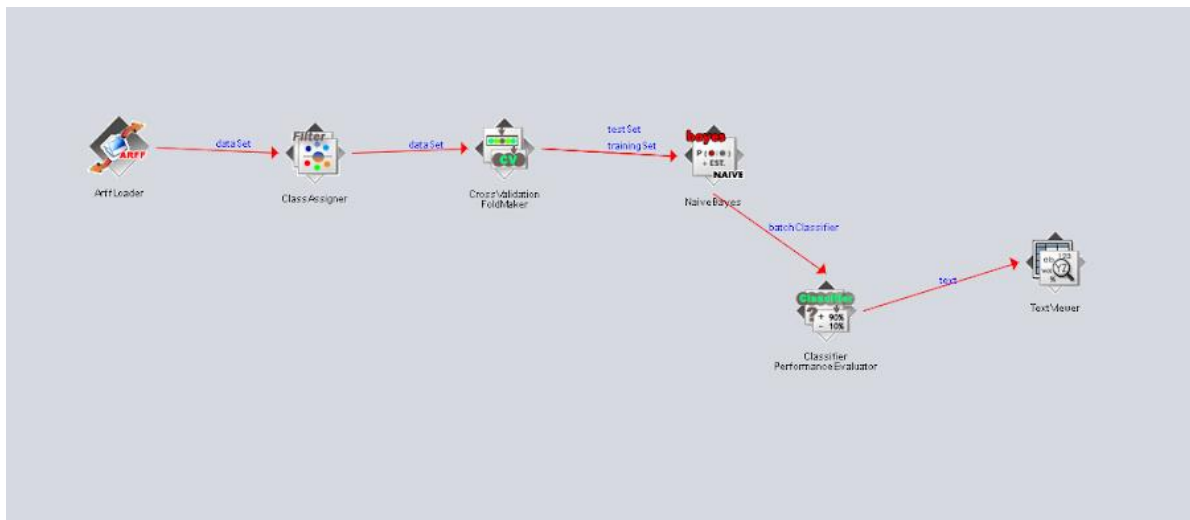
five-books-all-1000-1-stem.arff



| All | | None | Invert | Pattern |
|-----|---------------------------------------|--------|--------|---------|
| No. | Name | | | |
| 1 | <input checked="" type="checkbox"/> * | | | |
| 2 | <input type="checkbox"/> - | | | |
| 3 | <input type="checkbox"/> Ada | | | |
| 4 | <input type="checkbox"/> Anka | | | |
| 5 | <input type="checkbox"/> Borowiecki | | | |
| 6 | <input type="checkbox"/> Człowieku | | | |
| 7 | <input type="checkbox"/> Horn | | | |
| 8 | <input type="checkbox"/> Jązio | | | |
| 9 | <input type="checkbox"/> Karol | | | |
| 10 | <input type="checkbox"/> Kocham | | | |
| 11 | <input type="checkbox"/> Książę | | | |
| 12 | <input type="checkbox"/> Marta | | | |
| 13 | <input type="checkbox"/> Marty | | | |
| 14 | <input type="checkbox"/> Mela | | | |
| 15 | <input type="checkbox"/> Moryc | | | |
| 16 | <input type="checkbox"/> Nina | | | |
| 17 | <input type="checkbox"/> Pan | | | |
| 18 | <input type="checkbox"/> Piotr | | | |
| 19 | <input type="checkbox"/> Prawda | | | |
| 20 | <input type="checkbox"/> Tom | | | |
| 21 | <input type="checkbox"/> Toma | | | |
| 22 | <input type="checkbox"/> Ziemi | | | |
| 23 | <input type="checkbox"/> Ziemia | | | |
| 24 | <input type="checkbox"/> Ziemia™ | | | |
| 25 | <input type="checkbox"/> choŃ | | | |
| 26 | <input type="checkbox"/> fabryka | | | |
| 27 | <input type="checkbox"/> fabryki | | | |
| 28 | <input type="checkbox"/> godzin | | | |
| 29 | <input type="checkbox"/> juŃ | | | |
| 30 | <input type="checkbox"/> ku | | | |
| 31 | <input type="checkbox"/> mimo | | | |
| 32 | <input type="checkbox"/> mnie | | | |
| 33 | <input type="checkbox"/> mu | | | |
| 34 | <input type="checkbox"/> nam | | | |
| 35 | <input type="checkbox"/> nami | | | |
| 36 | <input type="checkbox"/> nas | | | |
| 37 | <input type="checkbox"/> oczami | | | |
| 38 | <input type="checkbox"/> pan | | | |
| 39 | <input type="checkbox"/> pana | | | |
| 40 | <input type="checkbox"/> pani | | | |
| 41 | <input type="checkbox"/> procent | | | |
| 42 | <input type="checkbox"/> wŃz | | | |
| 43 | <input type="checkbox"/> zaraz | | | |
| | | Remove | | |

6.6

Naive Bayes



Text Viewer

Result list

12:35:50,569 - NaiveBay

Text

==== Evaluation result ====

Scheme: NaiveBayes

Relation: autorzy-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.attribute.StringToWordVector-R2-W1000-prune-rate-1.0-NC-st

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 29726 | 66.9113 % |
| Incorrectly Classified Instances | 14700 | 33.0887 % |
| Kappa statistic | 0.3889 | |
| Mean absolute error | 0.2226 | |
| Root mean squared error | 0.3364 | |
| Relative absolute error | 70.5174 % | |
| Root relative squared error | 84.6796 % | |
| Total Number of Instances | 44426 | |

==== Detailed Accuracy By Class ====

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PSC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| | 0.538 | 0.034 | 0.876 | 0.538 | 0.667 | 0.595 | 0.867 | 0.749 | Reymont |
| | 0.114 | 0.017 | 0.397 | 0.114 | 0.178 | 0.176 | 0.736 | 0.212 | Zulawski |
| | 0.974 | 0.592 | 0.628 | 0.974 | 0.764 | 0.465 | 0.033 | 0.019 | Sienkiewicz |
| | 0.000 | 0.000 | ? | 0.000 | ? | ? | 0.706 | 0.239 | Zeromski |
| Weighted Avg. | 0.649 | 0.312 | ? | 0.669 | ? | ? | 0.822 | 0.687 | |

==== Confusion Matrix ====

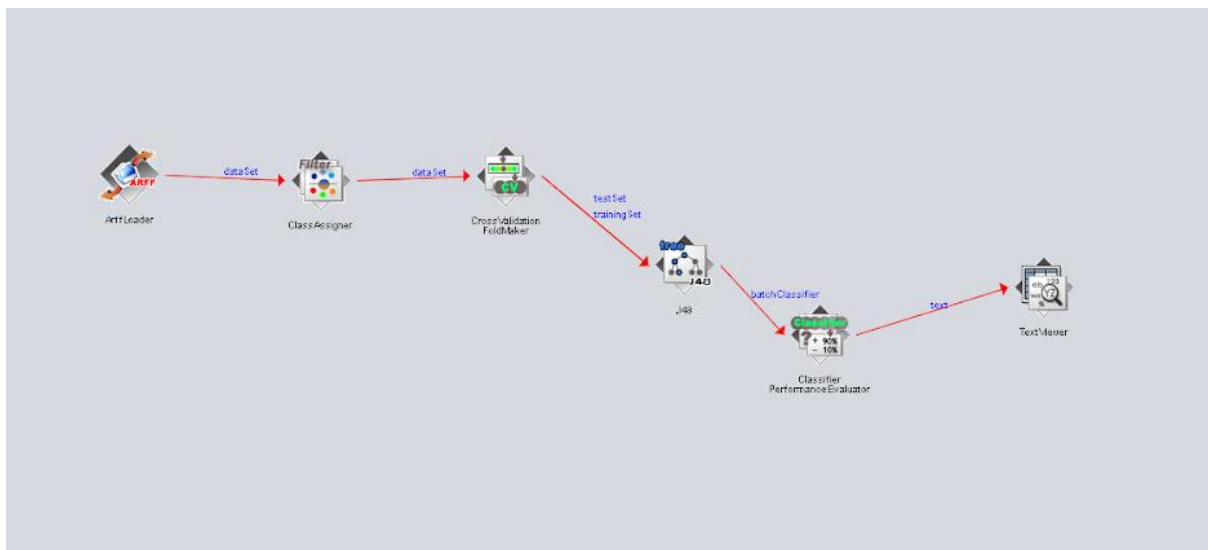
| | a | b | c | d | <-- classified as |
|------|-----|-------|---|---|-------------------|
| 7348 | 385 | 5916 | 0 | 0 | a = Reymont |
| 703 | 450 | 2779 | 0 | 0 | b = Zulawski |
| 312 | 282 | 21928 | 0 | 0 | c = Sienkiewicz |
| 29 | 16 | 4278 | 0 | 0 | d = Zeromski |

Close Settings Clear results

CrossValidationFoldMaker Finished

NaiveBayes Finished

00:00:03



Text Viewer

Result list

- 12:35:50.569 - NaiveBay
- 12:39:30.161 - J48

Text

```

==== Evaluation result ====

Scheme: J48
Options: -C 0.25 -M 2
Relation: autocorrelation-2,4-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.attribute.StringToWordVector-R2-W1000-prune-rate-1.0-NO-

Correctly Classified Instances      31114           70.0356 %
Incorrectly Classified Instances    13312           29.9644 %
Kappa statistic                    0.4475
Mean absolute error                 0.2078
Root mean squared error             0.3226
Relative absolute error             65.8204 %
Root relative squared error         81.1919 %
Total Number of Instances          44426

==== Detailed Accuracy By Class ====

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0.607    0.041    0.868    0.607    0.714    0.636    0.870    0.767    Reymont
0.091    0.001    0.937    0.091    0.166    0.279    0.718    0.238    Zulawski
0.966    0.548    0.649    0.966    0.703    0.520    0.835    0.823    Sienkiewicz
0.064    0.001    0.911    0.064    0.120    0.228    0.685    0.218    Zeromski
Weighted Avg.    0.700    0.290    0.767    0.700    0.643    0.506    0.821    0.695

==== Confusion Matrix ====

      a      b      c      d  <-- Classified as
8281   14   5337   17 |   a = Reymont
 820   359   2752    1 |   b = Zulawski
 308    9  22196    9 |   c = Sienkiewicz
 133    1   3911   278 |   d = Zeromski
  
```

Buttons: Close, Settings, Clear results

TaskView: Classifier Performance: autocorrelation-2,4-weka.filters.unsupervised.attribute.Remove-R2,4-weka.filters.unsupervised.attribute.StringToWordVector-R2-W1000-prune-rate-1.0-NO- Finished.