


Inżynieria wiedzy i systemy ekspertowe

Logika rozmyta jako forma reprezentacji wiedzy
w systemach sterowania lub podejmowania decyzji.

©BM

Autor: Krzysztof Jarek	Temat projektu: Ocena jakości czerwonego wina na podstawie parametrów fizyko-chemicznych	Data oddania projektu: 5.12.21 r.
Identyfikator przedmiotu: iw-is-io-2021/2022		Prowadzący:  Dr hab. inż. Barbara Mrzygłód KISiM, WIMiIP, AGH mrzyglod@agh.edu.pl

1 WPROWADZENIE DO PROJEKTU

1.1. Ogólny opis wybranego problemu.

Zastanawiać mogłoby czy jest możliwe przewidywanie ocen jakości czerwonego wina na podstawie ich charakterystyki fizyko-chemicznej. W niniejszym projekcie została podjęta próba stworzenia takiego rozwiązania programistycznego, które mogłoby zarządzać temu: uprzedniemu znajdowaniu potencjalnej walidacji ocen ekspertów, konsumentów. W oparciu o tak zbudowane narzędzie mogłoby to być bardzo interesującym zwłaszcza dla producentów wina, rolników, inwestorów w celu pozyskiwania potrzebnej predykcji jakości produkowanych trunków.

1.2. Możliwość praktycznego zastosowania opracowanego systemu.

Tego typu rozwiązanie w sam raz nadawałoby się do wykorzystania w ramach aplikacji mobilnej, czy też desktopowej, w jakiej użytkownicy mogliby podawać parametry wina, a program zwracałby im prawdopodobną przyszłą ocenę. Taka aplikacja mogłaby powszechnie zainteresować winiarzy, większych producentów wina. Podobne rozwiązania, istniejące na rynku, to szczególnie aplikacja mobilna *Vivino: Buy the Right Wine*, ale w niej walidacja opiera się o oceny konsumentów, dodawane na bieżąco.

1.3. Charakterystyki opisowe wybranych do projektu zmiennych.

Zmienne wejściowe:

- volatile acidity - kwasowość lotna,
- citric acid - kwas cytrynowy,
- residual sugar - cukier resztkowy,
- free sulfur dioxide - wolny dwutlenek siarki,
- total sulfur dioxide - całkowity dwutlenek siarki,
- pH,
- sulphates - siarczany,
- alcohol - alkohol.

Zmienna wyjściowa:

- quality - ocena jakości.

Dane wejściowe mogą być pozyskiwane jedynie na drodze pomiaru za pomocą w miarę prostej aparatury bio-chemicznej, jaką musi posiadać użytkownik aplikacji (odpowiednie mierniki kwasowości, miernik pH, alkoholomierz itd.). Nie mniej to na użytkowniku spoczywać będzie samo wprowadzenie wartości dla uzyskania obliczonego rozwiązania zgodnie z modelem.

Na wyjściu otrzymywana jest wartość reprezentującą potencjalną ekspercką ocenę jakości wina.

1.4. Wybrany rozmyty model wnioskowania.

Sugeno wraz z ANFIS. Powodem wyboru tego modelu była duża złożoność i mnogość danych do uczenia z jaką trudno byłoby sobie poradzić z wykorzystaniem nie-wspomaganego podejścia. Oprócz tego potrzebny był mechanizm automatycznej generacji zbiorów rozmytych- z racji na wspomnianą złożoność, trudną do ogarnięcia konstruktorowi systemu.

1.5. Proces pozyskiwania wiedzy.

Dane do stworzenia modelu zostały pozyskane z portalu *kaggle.com*, a następnie przetworzone przez algorytm, tak by można było go zbudować.

2 OPRACOWANIE SILNIKA WNIOSKOWANIA ROZMYTEGO

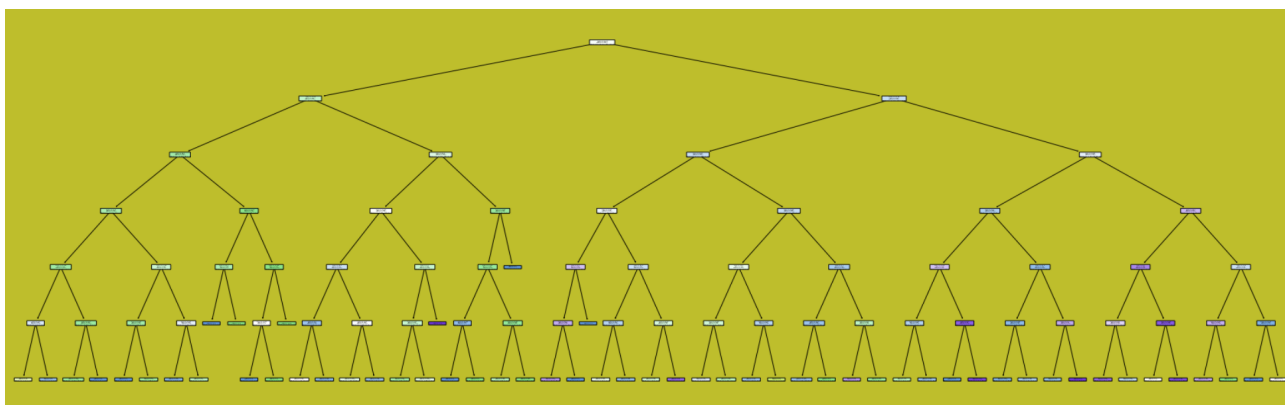
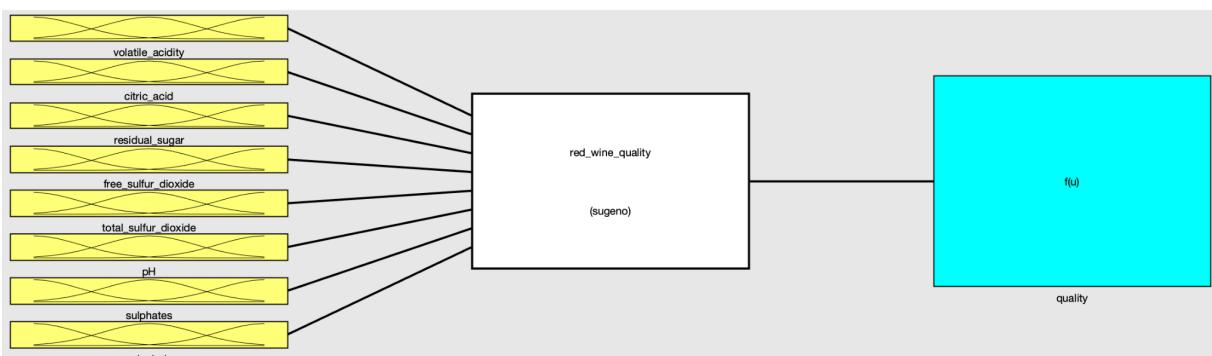
2.1. Reprezentacja wybranych zmiennych lingwistycznych w postaci zbiorów rozmytych.

Zmienne wejściowe (obok jednostka i zakres):

- volatile acidity, g/100mL [0.12, 1.58]
- citric acid, g/L [0 1]
- residual sugar, g/L [0.9, 15.5]
- free sulfur dioxide, mg/L [1, 72]
- total sulfur dioxide, mg/L [6, 289]
- pH, - [2.74, 4.01]
- sulphates, dag/L [0.33, 2]
- alcohol. % [8.4, 14.9]

Zmienna wyjściowa:

- quality (zakres: [3, 8]).



Drzewo z ustawioną maksymalną głębokością korzeni na poziomie wynoszącym 6.

Pierwotna liczba kategorii opisujących wino była większa, w ramach zbioru danych, i wynosiła 11. Nie mniej została dokonana analiza, z wykorzystaniem drzewa decyzyjnego (korzystając z technologii *python3* i *sklearn*), z ustawioną maksymalną głębokością korzeni na poziomie wynoszącym 6.

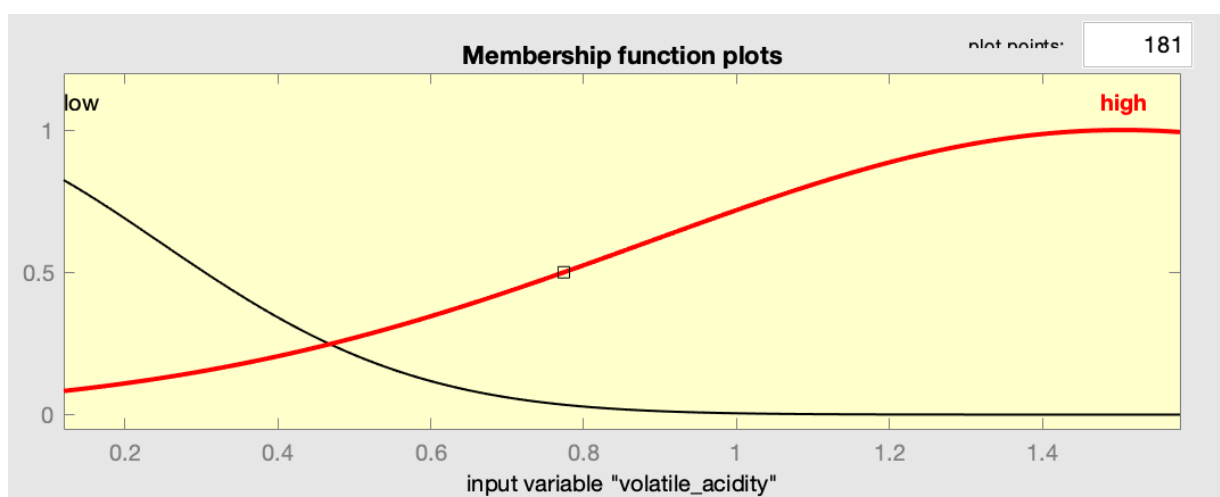
Poniżej zaprezentowano zestawienie znaczenia poszczególnych parametrów, pozyskane z pomocą pakietu. Spośród nich zostały usunięte z ważnością (ang. importance) mniejszą niż 0,02:

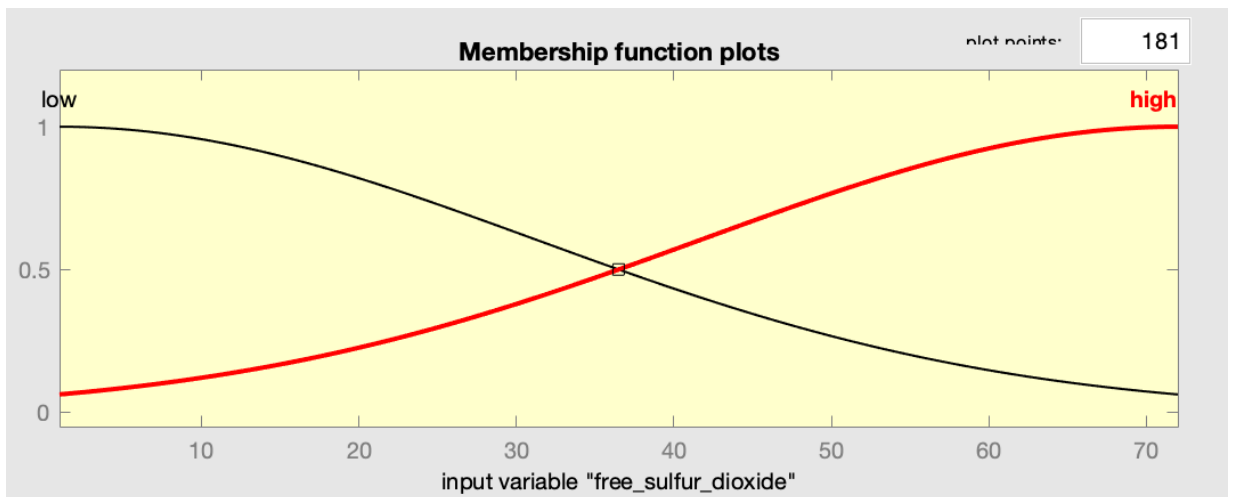
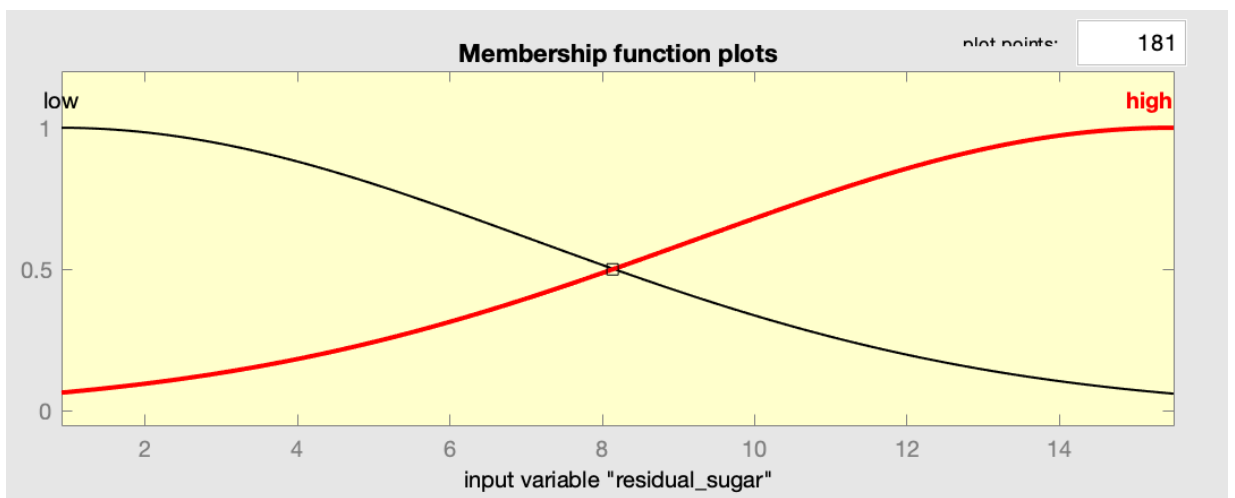
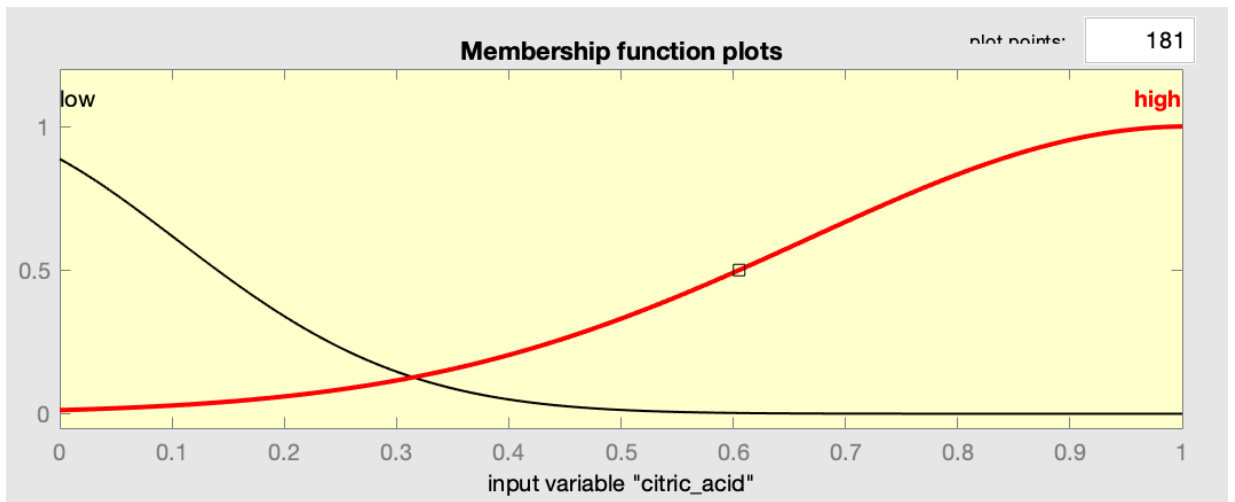
0. fixed acidity	0.016798655527215715
1. volatile acidity	0.08284727193663846
2. citric acid	0.045012301348718496
3. residual sugar	0.07812807553752027
4. chlorides	0.01941323585928673
5. free sulfur dioxide	0.06667040654784631
6. total sulfur dioxide	0.12476021466226984
7. density	0.019366817804518405
8. pH	0.07614962047335788
9. sulphates	0.13896052078344648
10. alcohol	0.3318928795191814

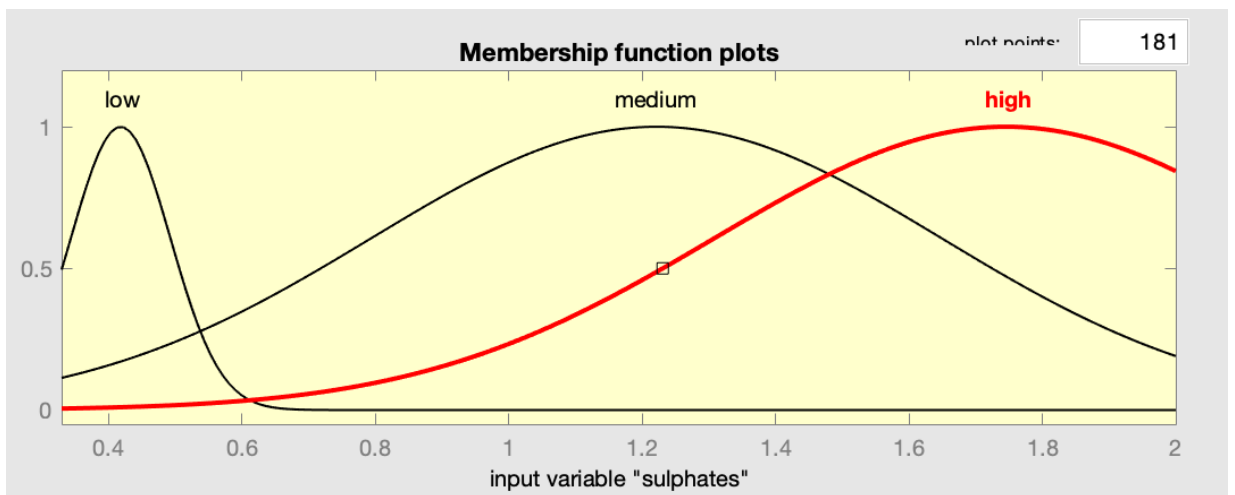
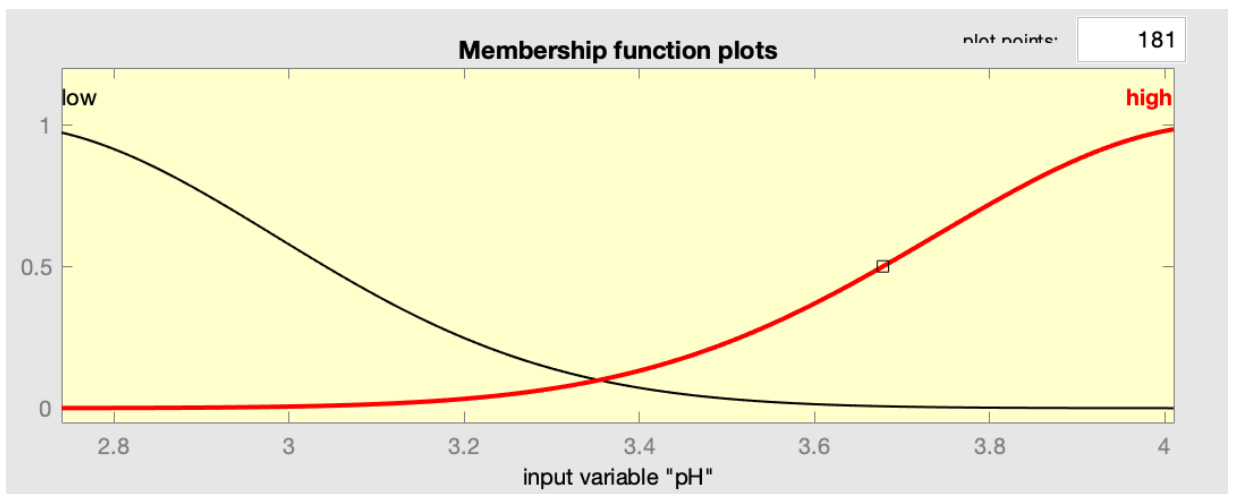
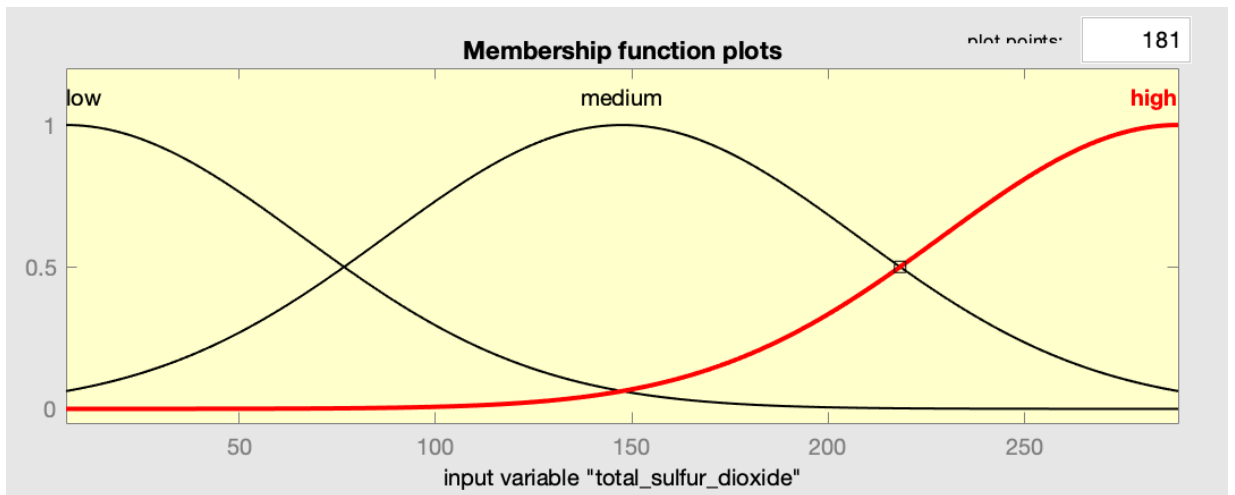
	volatile acidity	citric acid	residual sugar	free sulfur dioxide	total sulfur dioxide	pH	sulphates	alcohol	quality
volatile acidity	1.00	-0.55	0.00	-0.01	0.08	0.23	-0.26	-0.20	-0.39
citric acid	-0.55	1.00	0.14	-0.06	0.04	-0.54	0.31	0.11	0.23
residual sugar	0.00	0.14	1.00	0.19	0.20	-0.09	0.01	0.04	0.01
free sulfur dioxide	-0.01	-0.06	0.19	1.00	0.67	0.07	0.05	-0.07	-0.05
total sulfur dioxide	0.08	0.04	0.20	0.67	1.00	-0.07	0.04	-0.21	-0.19
pH	0.23	-0.54	-0.09	0.07	-0.07	1.00	-0.20	0.21	-0.06
sulphates	-0.26	0.31	0.01	0.05	0.04	-0.20	1.00	0.09	0.25
alcohol	-0.20	0.11	0.04	-0.07	-0.21	0.21	0.09	1.00	0.48
quality	-0.39	0.23	0.01	-0.05	-0.19	-0.06	0.25	0.48	1.00

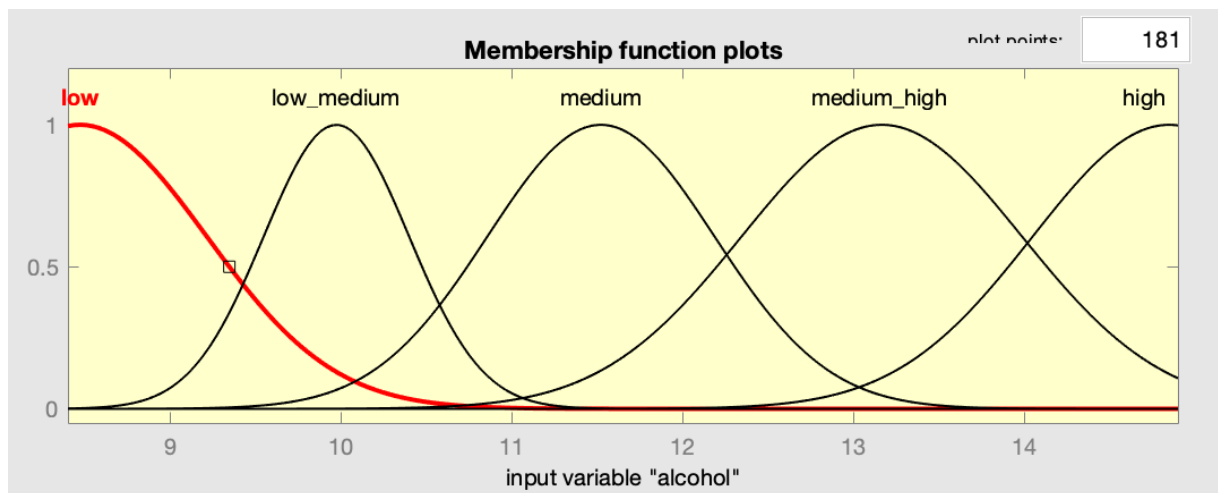
Macierz korelacji.

Poniżej zbiory rozmyte zmiennych lingwistycznych:









Jako funkcje przynależności dla wszystkich zbiorów rozmytych zostały wybrane funkcje gaussowskie, z racji na płynność zmian wartości jakie zwracają- ciągłość wykresu.

Nazwa zbioru	Typ funkcji	Liczba zbiorów
volatile acidity	gaussmf	2
citric acid	gaussmf	2
residual sugar	gaussmf	2
free sulfur dioxide	gaussmf	2
total sulfur dioxide	gaussmf	3
pH	gaussmf	2
sulphates	gaussmf	3
alcohol	gaussmf	5

2.2. Rozmyta baza wiedzy

W silniku wnioskowania rozmytego pojawiły się następujące relacje:



Zostały one automatycznie wygenerowane przez algorytm ANFIS. Łączna liczba ich wyniosła 1440 (nie dało się więcej, co poprawiłoby wyniki, gdyż liczba próbek wynosiła 1599, i należało przestać poniżej tego).

2.3. Uzasadnienie przyjętych operatorów logicznych

Jako operatory dla wnioskowania zostały przyjęte operatory typu *product* ($prod: \mu(A) \cdot \mu(B)$), z racji na swoje uniwersalne zastosowanie, dobre dla podejmowanego problemu.

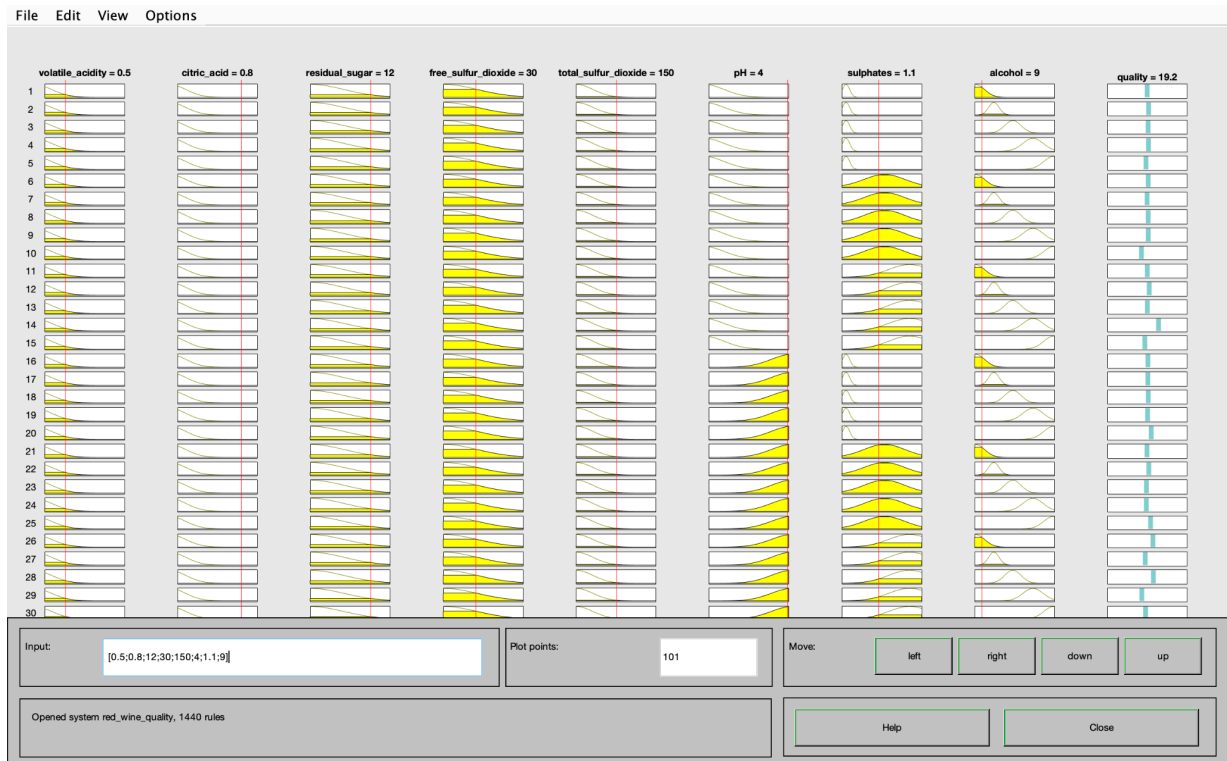
2.4. Wnioskowanie rozmyte – analiza

W celu dokonania analizy pracy modelu przeprowadzono próbne scenariusze wnioskowania dla wybranych wartości zmiennych wejściowych. Ogólnie, to początkowe wartości jakie można było znaleźć w środowisku były dostarczonymi wartościami uśrednionymi: [0.85;0.5;8.2;36.5;147.5;3.375;1.165;11.65].

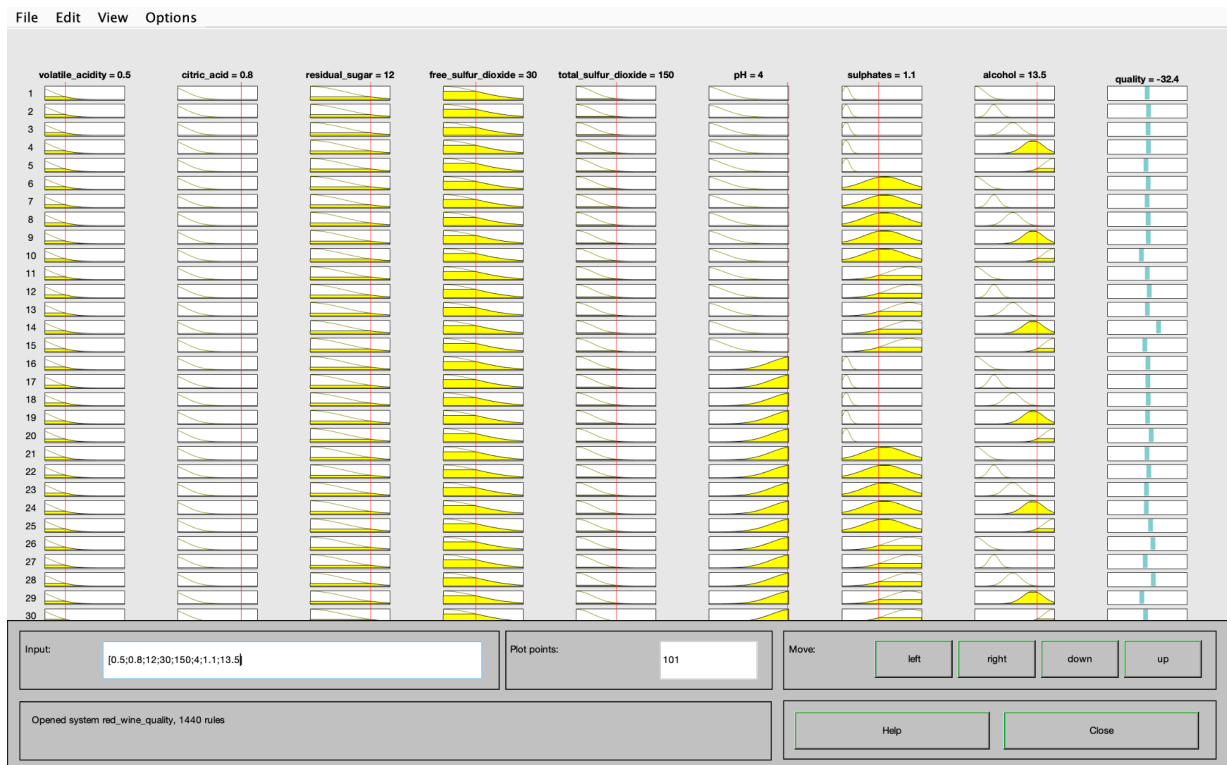
Dla sprawdzenia poszczególnych wartości zmiennych dokonano pewnej manipulacji na wejściu- początkowo sprawdzając znaczenie zmiennej

lingwistycznej “alcohol”- wpływ jej zmienności dla wyniku- i zmiennej “total sulfur dioxide”:

- [0.5;0.8;8.2;36.5;150;3.375;1.165;9] - wartość wykroczyła poza zakres wielkości przypisywanych “quality” ([3, 8]):



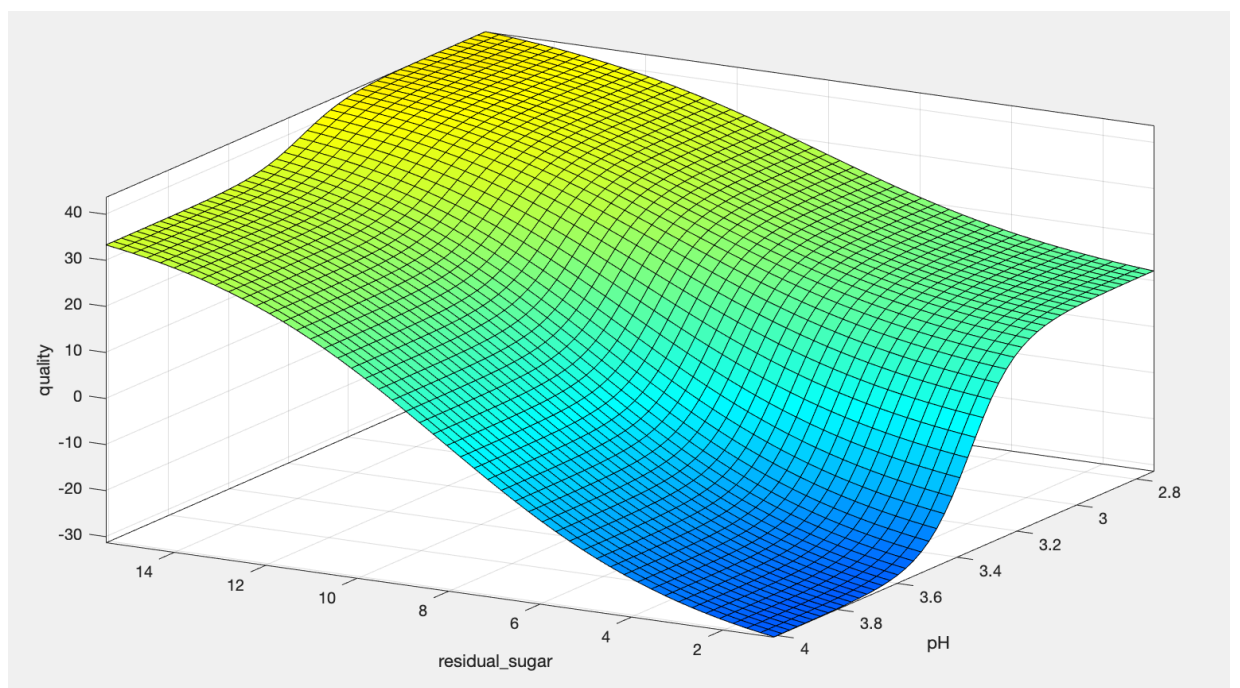
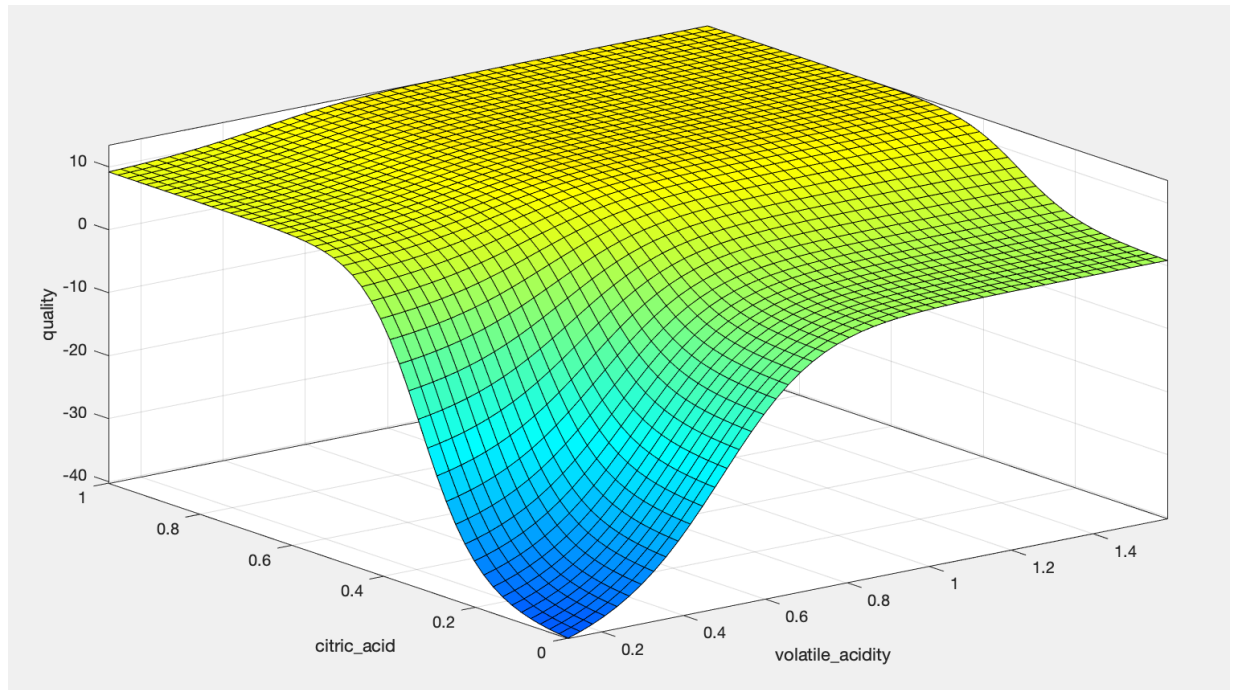
- [0.5;0.8;8.2;36.5;150;3.375;1.165;13.5] - wartość wykroczyła poza zakres wielkości przypisywanych “quality”, ku zaskoczeniu przyjmując wartość ujemną:

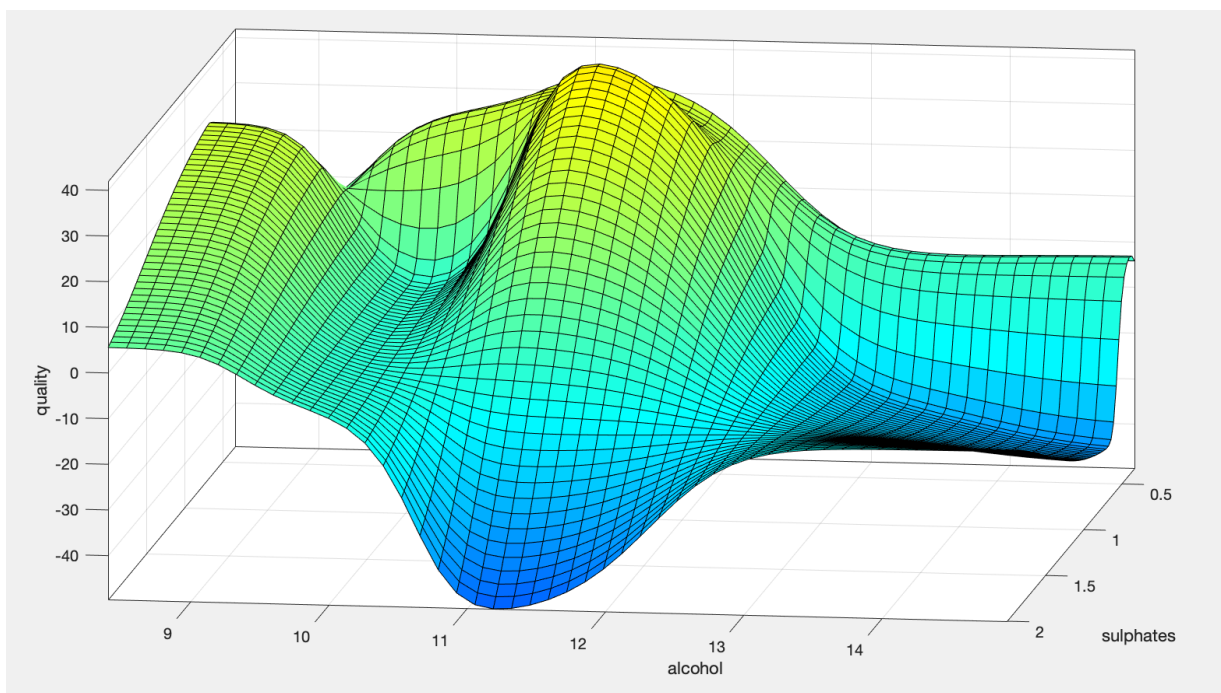
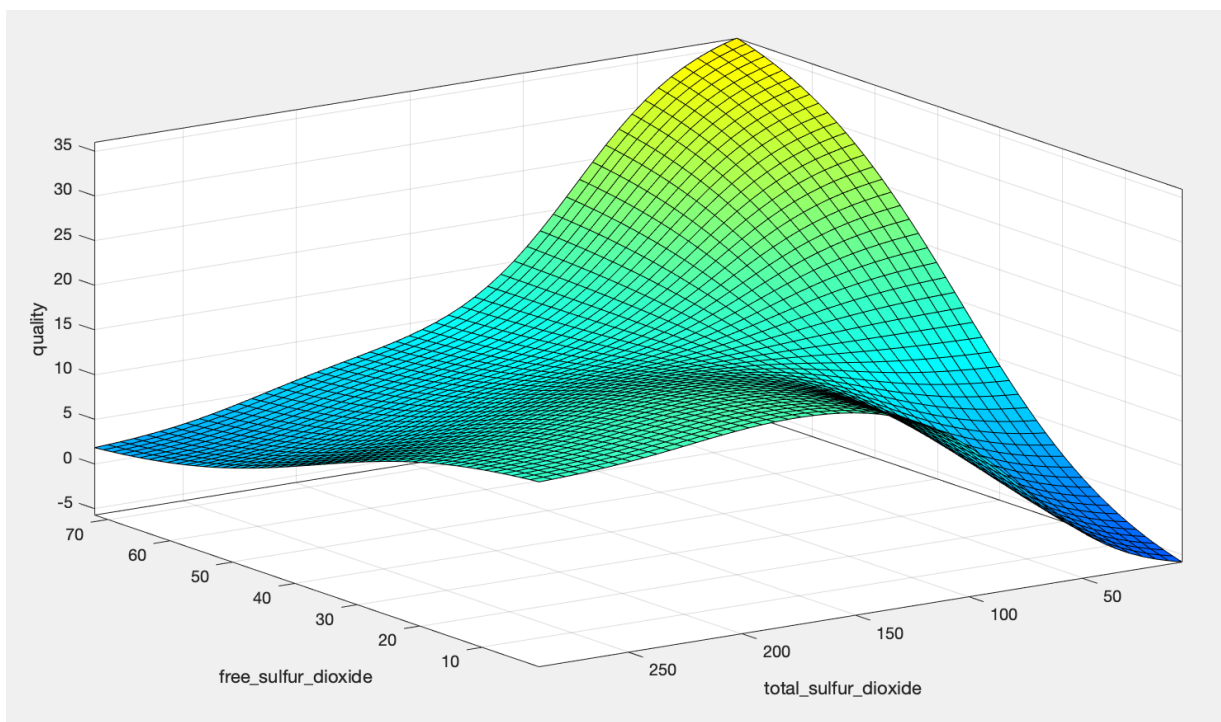


- [0.5;0.8;8.2;36.5;210;3.375;1.165;11.65] - operowanie na zmiennej “total sulfur dioxide” pozwoliło uzyskać wartość wyjściową, która mieściła się w zadanym przedziale:



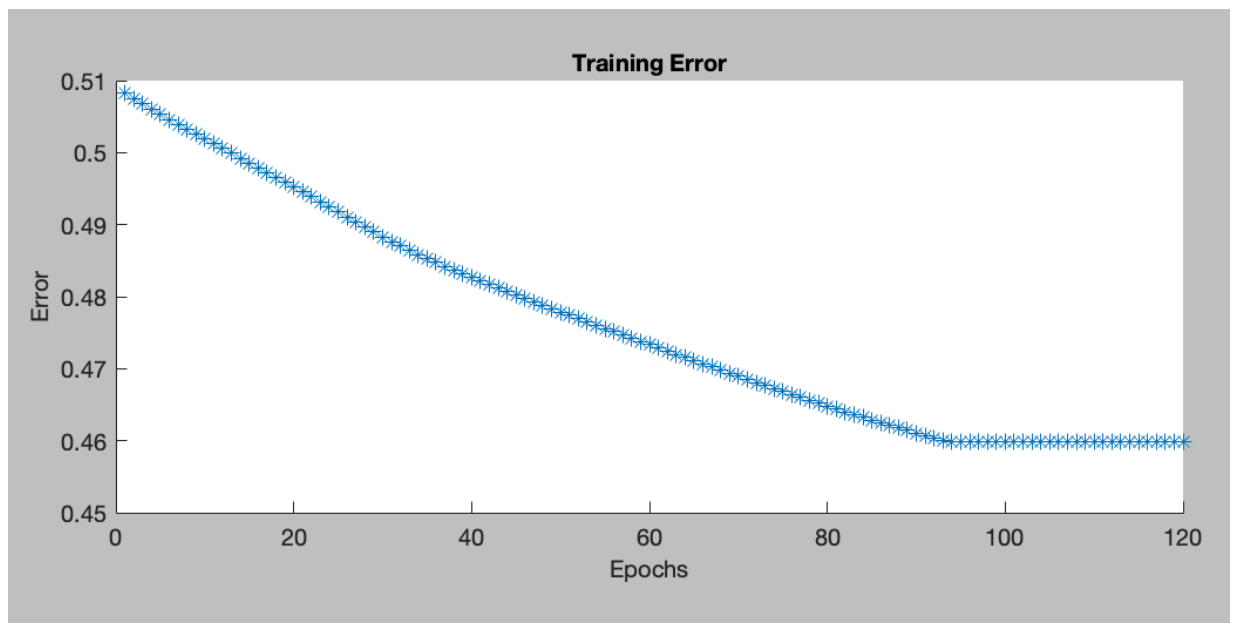
W celu dalszej wizualizacji problemu umieszczono pod spodem wybrane ilustracje działania systemu wygenerowane przez środowisko MATLAB:





Natomiast proces uczenia, dostrajania, wyglądał następująco, co prezentując wykres poniżej. Był on przeprowadzany dla 120 epok, gdzie ostatecznie osiągnął błąd RMSE wynoszący: 0,4598.

Epoch 120:error= 0.4598



Problem jaki tu się pojawił polegał na tym, że w pewnym momencie uczenia dochodziło do pojawienia zatrzymania procesu poprawy wyników. Zaproponowany jest jednym z możliwie lepszych, który też nie przekroczył liczbą reguł liczby próbek.

3 PODSUMOWANIE I WNIOSKI

Niestety, patrząc na błąd uczenia, ciężko chcieć postulować wdrożenie stworzonego modelu dla wykorzystania w predykcji jakości trunku. Teoretycznie taka aplikacja mogłaby zainteresować producentów, czy też uprawiających wino. Dalej też, przyszłością jej mogło by być rozszerzenie jej o analizę kolejnych napojów alkoholowych.

Możliwe, że poprawa mogłaby nastąpić wraz z zastosowaniem metody *oversampling'u*- wówczas można by wygenerować więcej reguł- ale autor już tego się nie podjął.

4 IMPLEMENTACJA

W kwestii implementacji autor poprzestał na pracy w środowisku MATLAB, nie tworząc implementacji w innych językach korzystających z stworzonego modelu. Dołączone: plik *.fis* wraz ze sprawozdaniem.