

Porównanie różnych metod wyznaczania zawartości OWO na podstawie krzywych wynikowych geofizyki wiertniczej – metody standardowe i algorytmy uczenia maszynowego

Kinga Bobek¹

Comparison of different methods for determining TOC content based on well logs – standard methods and machine learning algorithms. Prz. Geol., 74:606–608; doi: 10.7306/2026.49

Redaktor prowadząca: Olga Rosowiecka

A b s t r a c t. The study compares different methods of estimating Total Organic Carbon (TOC) using well-log data and laboratory measurements. The analysis was performed on six boreholes in the Mid-Polish Trough, for data covered Jurassic and Cretaceous formations. The analysis used a traditional Passey's method and two machine learning methods: SVR and kNN. The calculations performed indicate that the accuracy of the investigated methods varies depending on the borehole. The kNN method produced the best results, exhibiting relatively high prediction accuracy for four out of six boreholes. The general accuracy of TOC estimation depends on the number and general reliability of the input data, the availability of particular well log curves, and wellbore stability.

Keywords: Total Organic Carbon, machine learning, well logs

Zawartość ogólnego węgla organicznego (OWO) jest kluczowym parametrem służącym do oceny potencjału naftowego basenów sedimentacyjnych (Bessereau i in., 1991; Myers, Jenkyns, 1992). Od wartości tego parametru zależy jakość skał macierzystych decydująca o opłacalności eksploatacji węglowodorów. Najbardziej wiarygodną metodą szacowania zawartości OWO jest analiza geochemiczna próbek rdzeniowych. Metoda ta jest jednak kosztowna i czasochłonna, co skutkuje uzyskaniem ograniczonych i nieciągłych danych, które w niewielkim stopniu odzwierciedlają badane formacje (Bessereau i in., 1991). Dlatego też kluczowe znaczenie dla poszukiwania i wydobywania węglowodorów miało opracowanie kilku metod szacowania zawartości OWO na podstawie ciągłych danych z profilowań geofizyki otworowej. W zależności od dostępnych profilowań geofizycznych są np. stosowane: metoda Passey'a (Passey i in., 1990) obliczania OWO na podstawie profilowań oporności (RT) i krzywych akustycznych (DT) oraz metody Schmokera (Schmoker, 1979; Schmoker, Hester, 1983) polegające na wykorzystywaniu do obliczeń pomierzonych wartości gęstości objętościowej skał (RHOB). Częstym problemem utrudniającym wykorzystanie tych tradycyjnych metod obliczeniowych jest brak danych z konkretnych profilowań, zwłaszcza jeśli są analizowane stare odwierty. Dlatego też w celu określenia zawartości OWO w skałach coraz częściej wykorzystuje się metody uczenia maszynowego – *Machine Learning* (np. Alizadeh i in., 2012; Lee i in., 2019; Amosu i in., 2021).

Celem niniejszego artykułu jest zaprezentowanie wyników prób wyznaczenia zawartości OWO w skałach pochodzących z odwiertów wykonanych przed rokiem 1990, których nie można było zbadać za pomocą większości tradycyjnych metod obliczeniowych ze względu na ograniczony zakres dostępnych krzywych geofizycznych. Analizy ukie-

nowego w skałach z profili odwiertów udokumentowanych archiwalnymi krzywymi geofizycznymi o stosunkowo słabej jakości.

DANE I METODY

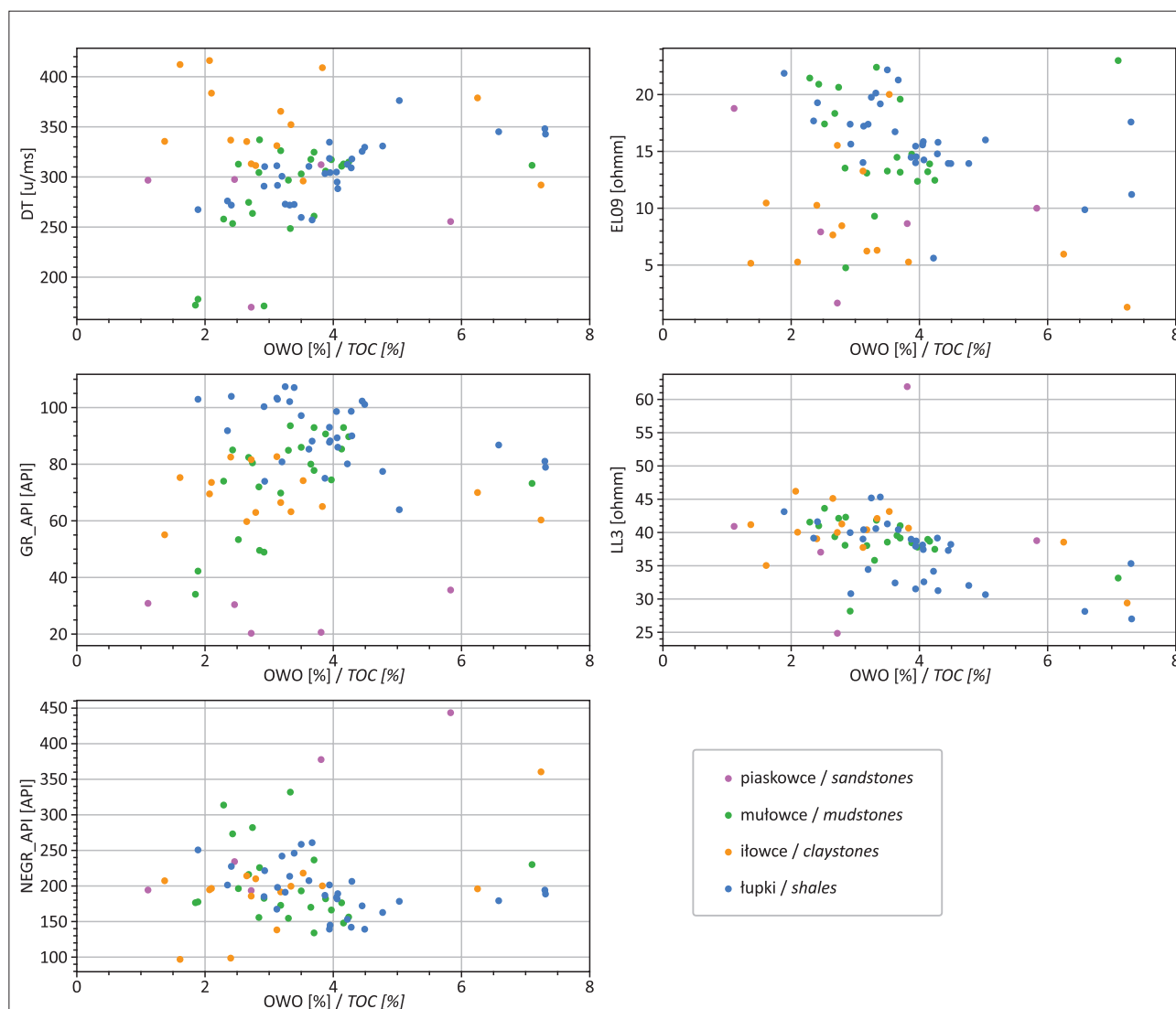
Zawartość OWO oznaczono w próbkach skał z 6 rdzeni wiertniczych z odwiertów wykonanych na Niżu Polskim w latach 1966–1989 (Gostynin IG 1/A, Brześć Kujawski IG 2, Szwejki IG 3, Ciechocinek IG 1, Chociwel 2, Wojszyce IG 3). Profilowania geofizyczne tych otworów prowadzono w przedziale głębokości obejmującym formacje jurajskie i kredowe. Liczba próbek skał pobranych z każdego z tych rdzeni, w których zawartość OWO wyznaczono laboratoryjną metodą *Rock-Eval*, wynosi od 24 do 125. Do estymacji wartości OWO wybrano standardową metodę Passey'a. Obliczenia wykonano z wykorzystaniem algorytmu wbudowanego w oprogramowanie *TechLog*, w którym do wyznaczenia OWO konieczna jest znajomość wartości oporności (krzywa LL3), czasu przebiegu fali P (krzywa DT) oraz bazowych wartości tych parametrów w poszczególnych interwałach głębokościowych. Dane umożliwiające wykorzystanie tej metody uzyskano z 4 otworów wiertniczych spośród 6 analizowanych.

W ramach testów metod maszynowych wykorzystano jeden z modeli liniowych, jakim jest maszyna wektorów nośnych (SVM – *Support Vector Machine*), oraz model nieliniowy – k najbliższych sąsiadów (kNN – *k Nearest Neighbors*). Maszyna wektorów nośnych jest algorytmem samouczącym, który można wykorzystać do rozwiązywania problemów dotyczących klasyfikacji i regresji (Cristianini i in., 2000). Optymalne parametry zmiennych funkcji danej maszyny są wyznaczane poprzez znalezienie hiperpłaszczyzny rozdzielającej punkty należące do poszczególnych klas w N-wymiarowej przestrzeni z jak największym marginesem. Kluczowym elementem maszyny wektorów jest tzw.

¹ Państwowy Instytut Geologiczny – Państwowy Instytut Badawczy, ul. Rakowiecka 4, 00-975 Warszawa; kbob@pgi.gov.pl; ORCID ID: 0000-0002-6137-9945

Tab. 1. Wartości współczynnika determinacji (R^2) i błędu standardowego (SE) – parametrów użytych do oceny skuteczności predykcji modeli otworów wiertniczych**Table 1.** Parameters used to evaluate the prediction accuracy of particular models for individual wells: coefficient of determination (R^2) and standard error (SE)

Otwór wiertniczy <i>Borehole</i>	Metoda Passeya <i>Passey's Method</i>		SVM		kNN	
	R^2 [-]	SE [%]	R^2 [-]	SE [%]	R^2 [-]	SE [%]
Gostynin IG 1/A	–	–	0,193	1,37	0,212	0,894
Brześć Kujawski IG 2	0,107	1,942	0,565	1,884	0,816	1,047
Szwejk IG 3	0,796	0,935	0,016	2,381	0,717	1,285
Ciechocinek IG 1	0,005	2,203	0,003	1,847	0,688	1,341
Chociwel 2	–	–	0,446	0,809	0,177	1,191
Wojszyce IG 3	0,444	1,398	0,133	1,625	0,468	0,968

SVM – maszyna wektorów nośnych / *Support Vector Machine* – model liniowykNN – k najbliższych sąsiadów / *k Nearest Neighbours* – model nieliniowy**Ryc. 1.** Zestawienie wartości krzywych geofizycznych (DT – akustyka, EL₀₉ – gradientowe profilowanie oporności, GR – profilowanie gamma, LL₃ – profilowanie oporności i NEGR – profilowanie neutron-gamma) z zawartością OWO w skałach z profili 6 otworów wiertniczych na Niziu Polskim (Brześć Kujawski IG 2, Szwejk IG 3, Ciechocinek IG 1, Wojszyce IG 3) wyliczoną w toku analiz laboratoryjnych**Fig. 1.** Correlation of well-log values (DT – sonic, EL₀₉ – gradient resistivity logging, GR – gamma ray logging, LL₃ – resistivity logging (laterolog), and NEGR – neutron-gamma logging) with the TOC value in rocks from the profiles of six drillholes in the Polish Lowlands (Brześć Kujawski IG 2, Szwejk IG 3, Ciechocinek IG 1, Wojszyce IG 3) obtained from laboratory analyses

funkcja jądra (*kernel function*), która może być funkcją liniową lub radialną w zależności od problemu i potrzebnej przestrzeni do rozmieszczenia punktów. W procesie uczenia dodatkowym parametrem, jaki należy uwzględnić, jest parametr regulujący model, który zapobiega zbyt silnemu dopasowaniu do konkretnych wartości danych (*overfit*). Wartości wymienionych parametrów dobierano za pomocą przeszukiwania siatki wartości (*GridSearch*). W modelu typu kNN proces uczenia i predykcji odbywa się poprzez wyznaczenie dystansu pomiędzy wartościami atrybutów danego parametru dla k punktów zlokalizowanych najbliżej punktu, którego wartość próbujemy oszacować (Laaksonen, Oja, 1996).

Algorytmy wybranych metod uczenia maszynowego zaimplementowano do oprogramowania *TechLog*, wykorzystując możliwość tworzenia dodatkowych modułów w języku *Python* (*Utility* -> *Python 3 Editor*) oraz pakiet *Scikit-Learn* (Pedregosa i in., 2011). W celu zwiększenia dokładności predykcji utworzono osobne modele wszystkich wydzielonych kompleksów skał o zbliżonej litologii, a dane wejściowe znormalizowano za pomocą metody *min-max*. Jako zestaw treningowy wybrano 3 najlepiej opróbowane otwory, które miały pełen zestaw danych geofizycznych, po wcześniejszym skalibrowaniu próbek pobranych do analizy *Rock-Eval* z danymi geofizycznymi pod kątem dopasowania głębokości w miejscach, gdzie było to możliwe. Modele pozostałych otworów tworzone następnie na podstawie dostępnych krzywych geofizycznych, takich jak: DT (akustyka), GR (profilowanie gamma), NEGR (profilowanie neutron-gamma), EL09 (gradientowe profilowanie oporności) i LL3 (profilowanie oporności), w różnych konfiguracjach.

Do estymacji zawartości OWO wykorzystano wszystkie metody, jakie można było zastosować w danym otworze. Dokładność wyników porównano poprzez zestawienie wartości obliczonych na podstawie krzywych geofizycznych z wartościami uzyskanymi w toku analizy *Rock-Eval* odpowiadających punktom znajdującym się w profilach otworów wiertniczych na tej samej głębokości, wykorzystując takie parametry, jak współczynnik determinacji (R_2) i wartość błędu standardowego (SE).

WYNIKI

Wyniki predykcji zawartości OWO wyznaczonej różnymi metodami wskazują na znacząco zmienność skuteczności tych metod w odniesieniu do analizowanych 6 otworów wiertniczych (tab. 1). W standardowym modelu Passey'a skuteczność predykcji była generalnie niska w związku ze złym stanem otworów i słabą wiarygodnością danych akustycznych (krzywa DT). Metoda ta wykazała najwyższą skuteczność predykcji jedynie w otworze Szwejki IG-3. Model SVM również wykazał słabą skuteczność w analizowanych otworach, a wysoką skuteczność predykcji tego modelu odnotowano tylko w przypadku otworu Chociwel-2. Najskuteczniejszym modelem predykcyjnym okazał się model typu kNN, który wykazał najwyższą skuteczność predykcji w 4 z 6 analizowanych otworów.

WNIOSKI

Możliwą przyczyną wysokiej dokładności predykcji ogólnie mało złożonego modelu typu kNN jest najprawdopodobniej brak wyraźnych zależności liniowych pomiędzy wartościami DT, GR, NEGR, EL09 i LL3 a wyznaczoną laboratoryjnie zawartością OWO (ryc. 1). Brak oczekiwanych zależności prawdopodobnie jest spowodowany małą wiarygodnością wyników pomiarów, związaną głównie ze zwiększoną średnicą otworu oraz słabą korelacją głębokości danych z rdzenia i krzywych geofizycznych, które generują losowe i trudne do usunięcia błędy. Skuteczność zastosowanych modeli zależy od kilku czynników:

- 1) wiarygodności danych wykorzystanych do uczenia modelu;
- 2) dostępności krzywych, których zmienność odzwierciedla zmienność zawartości OWO;
- 3) stabilności otworu oraz
- 4) liczby dostępnych danych.

Serdecznie dziękuję anonimowemu Recenzentowi za uwagi, które znacząco przyczyniły się do poprawy jakości tekstu.

LITERATURA

- ALIZADEH B., NAJJARI S., KADKHODAIE-ILKHCHI A. 2012 – Artificial neural network modeling and cluster analysis for organic facies and burial history estimation using well log data: A case study of the South Pars Gas Field, Persian Gulf, Iran. *Computers & Geosciences*, 45: 261–269.
- AMOSU A., IMSALEM M., SUN Y. 2021 – Effective machine learning identification of TOC-rich zones in the Eagle Ford Shale. *Journal of Applied Geophysics*, 188: 104311.
- BESSEREAU G., CARPENTIER B., HUC A.Y. 1991 – Wireline Logging and Source Rocks – Estimation of Organic Carbon Content by the CARBOLOG Method. *The Log Analyst*, 32: 279–297.
- CRISTIANINI N., SHAW-TAYLOR J., SHAW-TAYLOR D. 2000 – An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, London.
- LAAKSONEN J., OJA E. 1996 – Classification with learning k-nearest neighbors. *IEEE*, 3: 1480–1483.
- LEE T.R., WOOD W.T., PHRAMPUS B.J. 2019 – A Machine Learning (kNN) Approach to Predicting Global Seafloor Total Organic Carbon. *Global Biogeochemical Cycles*, 33: 37–46.
- MYERS K.J., JENKINS K.F. 1992 – Determining total organic carbon contents from well logs: an intercomparison of GST data and a new density log method. Geological Society, London, Special Publications, 65: 369–376.
- PASSEY Q.R., CREANEY S., KULLA J.B., MORETTI F.J., STROUD J.D. 1990: A Practical Model for Organic Richness from Porosity and Resistivity Logs. *AAPG Bulletin*, 74: 1777–1794.
- PEDREGOSA F., VAROQUAUX G., GRAMFORT A., MICHEL V., THIRION B., GRISEL O., BLONDEL M., PRETTEHOFER P., WEISS R., DUBOURG V., VANDERPLAS J., PASSOS A., COURNAPEAU D., BRUCHER M. 2011 – Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830.
- SCHMOKER J.W. 1979 – Determination of organic content of Appalachian Devonian shales from formation-density logs. *American Association of Petroleum Geologists Bulletin*, 63: 1504–1509.
- SCHMOKER J.W., HESTER T.C. 1983 – Organic Carbon in Bakken Formation, United States Portion of Williston Basin. *AAPG Bulletin*, 67: 2165–2174.

Praca wpłynęła do redakcji 24.06.2026 r.

Akceptowano do druku 28.07.2026 r.