



# A Framework for Predicting the Porosity and Permeability of Petro Physical Rock Types using Random Forest Classifier

P. S. Ezekiel<sup>1</sup>, O. E. Taylor<sup>2</sup> & M. O. Musa<sup>3</sup>

Department of Computer Science, Rivers State University, Port Harcourt, Nigeria.<sup>1,2</sup>

Department of Computer Science, University of Port Harcourt, Port Harcourt, Nigeria.<sup>3</sup>

**Abstract:** Porosity and permeability is plays an important effect in the production of oil and gas in a reservoir. The accumulation of oil and gas plays a very important role in the formation of gas reservoir. In other to have a better flowrate in a reservoir, this paper presents a smart system that will be used in estimating the permeability and porosity in carbonate rocks. The methodology or technique used in this work is Random Forest Classifier. Random Forest Classifier was used in training a machine model for classifying the porosity and permeability in carbonate rocks. The model was trained using a reservoir data consist of a total of 14 columns which was reduced to just three columns. The reduction was done by selecting just the most important feature by means of feature extraction. The model was trained using just the selected features with an estimator function of 100, the model for evaluate based on accuracy and precision. The result generated shows that the model achieved an accuracy of about 100%. The model was saved and deployed to production using python flask. The result of the deployed model shows a correct classification of the porosity and permeability in carbonate rocks.

**Keywords:** Porosity, Permeability, Carbonate Rocks, Oil Reservoir, Random Forest Classifier.

## I. INTRODUCTION

Routinely, oil and gas have been recuperated from sandstone or carbonate supplies where hydrocarbons are caught in well associated system of pores and cracks. Ongoing progressions in petrol advances, like parallel drilling and hydraulic cracking, oil and gas can likewise be recuperated whimsically from less-created mudstone (shale) repositories. Profound and tight stone formation of heterogeneous lithology and mineralogy with ineffectively associated nanometer-/micrometer-size pore systems. Among the critical factors in evaluating the oil/gas efficiency capability of shale supplies (likewise alluded as to shales) are the porosity and permeability of these oil-and additionally gas-bearing stone developments. Porosity of a rock is the volume of void space, which can be loaded up with various repository liquids (e.g., oil, gas, water) at different immersions, though porousness of a rock is the capacity of these liquids to stream inside and between the pore space. Shales are portrayed by exceptionally low porosity (commonly under 5%) and extremely low porousness (regularly under 1,000 nD), which make them trying in recuperating monetarily feasible hydrocarbons. Deciding the volume of oil and additionally gas present in a repository (oil-and additionally gas set up), and it's capability to course through supply pore/break system into the wellbore, helps oil industry to comprehend and advance the creation of a supply [1].

Foreseeing flow through porous medium is vital for a wide scope of scientific and engineering ventures including power cells, oil and gas recuperation, geologic carbon storage geothermal energy, material composites, and atomic garbage removal [2]. The inward construction of a porous medium characterizes its capacity to send liquid it's porousness. These porous shapes are regularly heterogeneous and range a few significant degrees in scale, especially in carbonate rocks, which are a plentiful topographical material for both oil and gas repositories and carbon stockpiling locales, making forecast of porousness in these cases particularly troublesome [3].

Generally, porousness has been estimated in center food tests utilizing Darcy's law, which depends on mass porosity estimations and the pressing factor drop across the center during stream. This technique can be costly and tedious. Besides, it gives little understanding into the nearby multi-scale underlying effects on a porousness, which is a very much perceived and huge test for upscaling carbonates since mass estimations don't give sufficient data to precisely extrapolate stream properties to different examples dependent on primary data. Additionally, stream estimations at the center scale are regularly not agent of stream at the repository scale because of enormous scope heterogeneities present in the supply that are not caught in a little cm-scale center example [4].

This paper aim in building a machine learning model, that will be used in estimating the porosity and permeability in Carbonate Rocks.



## II. RELATED WORKS

Routinely, oil and gas have been recuperated from sandstone or carbonate supplies where hydrocarbons are caught in well associated system of pores and cracks. Ongoing progressions in petrol advances, like parallel drilling and hydraulic cracking, oil and gas can likewise be recuperated whimsically from less-created mudstone (shale) repositories. Profound and tight stone formation of heterogeneous lithology and mineralogy with ineffectively associated nanometer-/micrometer-size pore systems. Among the critical factors in evaluating the oil/gas efficiency capability of shale supplies (likewise alluded as to shales) are the porosity and permeability of these oil-and additionally gas-bearing stone developments. Porosity of a rock is the volume of void space, which can be loaded up with various repository liquids (e.g., oil, gas, water) at different immersions, though porousness of a rock is the capacity of these liquids to stream inside and between the pore space. Shales are portrayed by exceptionally low porosity (commonly under 5%) and extremely low porousness (regularly under 1,000 nD), which make them trying in recuperating monetarily feasible hydrocarbons. Deciding the volume of oil and additionally gas present in a repository (oil-and additionally gas set up), and it's capability to course through supply pore/break system into the wellbore, helps oil industry to comprehend and advance the creation of a supply [1].

Foreseeing flow through porous medium is vital for a wide scope of scientific and engineering ventures including power cells, oil and gas recuperation, geologic carbon storage geothermal energy, material composites, and atomic garbage removal [2]. The inward construction of a porous medium characterizes its capacity to send liquid it's porousness. These porous shapes are regularly heterogeneous and range a few significant degrees in scale, especially in carbonate rocks, which are a plentiful topographical material for both oil and gas repositories and carbon stockpiling locales, making forecast of porousness in these cases particularly troublesome [3].

Generally, porousness has been estimated in center food tests utilizing Darcy's law, which depends on mass porosity estimations and the pressing factor drop across the center during stream. This technique can be costly and tedious. Besides, it gives little understanding into the nearby multi-scale underlying effects on a porousness, which is a very much perceived and huge test for upscaling carbonates since mass estimations don't give sufficient data to precisely extrapolate stream properties to different examples dependent on primary data. Additionally, stream estimations at the center scale are regularly not agent of stream at the repository scale because of enormous scope heterogeneities present in the supply that are not caught in a little cm-scale center example [4].

This paper aim in building a machine learning model, that will be used in estimating the porosity and permeability in Carbonate Rocks.

### 2. Literature Review

[5] built a prescient model, which depended on the results of routine principal survey data and information on central depth, top and base profundities of useful skylines. In other to have the best model, they compared different machine learning algorithms. The best performance of the model in terms of prescient capacity and generalizability for each of the three stone attributes together was that of a two hidden layers Neural network. Different Algorithms, like Support Vector Machine and Linear Regression, still performed well on the dataset. In general, the applied methodology permits anticipating the modification of porosity and porousness during desalination in permeable rocks and furthermore, assessing salt focus without direct estimations in a lab.

[6] Utilized probabilistic neural networks (PNNs) to display lithofacies successions as an element of well logging information to foresee discrete lithofacies conveyance at missing spans. The well log understandings considered for lithofacies characterization and penetrability displaying were neutron porosity, shale volume, and water immersion as an element of profundity. In any case, the deliberate discrete lithofacies types are sand, sandstone, and porosity. Precise lithofacies grouping was accomplished by the PNN as the complete percent right of the anticipated discrete lithofacies was 95.81%.

[1] carried out a survey on the effect of repository imprisonment (weight) on porosity and permeability of shales utilizing two advanced rock 3D models, which addressed nano porous natural/mineral microstructure of the Marcellus Shale. Five pressure situations were recreated for various profundities (2,000–6,000 feet) inside the creation time frame average oil/gas supply inside the Marcellus Shale play. Porosity and permeability of the pre-and post-pressure advanced stone 3D models were determined and thought about. A negligible impact of weight on porosity and penetrability was seen in both 3D models. The trial results have direct ramifications in deciding the oil-/gas set up and evaluating the creation capability of a shale supply under different pressure conditions

[7] present an original technique for upscaling multimodal porosity–penetrability connections utilizing machine learning based multivariate underlying relapse. Primary ascribes (porosity, stage availability, volume division, and so on) of each sub-volume were separated utilizing picture investigation apparatuses and afterward relapsed against the tackled Darcy–Brinkman–Stokes (DBS) penetrability utilizing an Extra-Trees relapse model to infer an upscaled porosity–porousness relationship. Ten experiments of 3603 voxels were then displayed utilizing Darcy-scale stream with this AI anticipated up scaled porosity–penetrability relationship and benchmarked against full Darcy–Brinkman–STOKES (DBS) reproductions. All mathematical reenactments were performed utilizing GeoChemFoam. They discovered great



arrangement between the full Darcy–Brinkman–Stokes (DBS) recreations and both the mathematical and AI up scaled models, with the AI model being multiple times less computationally costly.

[8] gathered 53 center examples with logging information from the Lower Cambrian dolomite supply in the Tarim Basinto set up solid strategies for penetrability and porosity forecast. Five normal porousness porosity relationships and six machine learning strategies were applied to the center information and logging information to assess the relevance or forecast execution of various techniques. Their proposed coordinated methodology incredibly further developed the porousness forecast precision with the most noteworthy R2 of 0.869, showing that the mix of the customary penetrability model and Machine algorithms could decrease the impact of heterogeneity on penetrability expectation.

[9] introduced a case ten investigation testing the viability of artificial neural networks and genetic algorithms for predicting eleven of penetrability in close carbonate rocks. The most minimal RMSE from the ordinary porousness twenty-one conditions was from the RGPZ condition, which anticipated the test dataset with a RMSE score of 220.458%, while the most noteworthy RMSE came from the Berg condition, with a RMSE of 2.368. By doing a correlation of twenty-three, the RMSE for the genetic algorithm and artificial neural network algorithms were 240.433% and 0.38%, separately. They credited the better execution of 25 machine learning methods over traditional ways to deal with their upgraded capacity to display the availability of pore microstructures brought about by mutually dependent and contending diagenetic measures.

[10] propose a model for foreseeing porosity and penetrability of oil supplies by utilizing a machine learning idea and petrophysical logs. A thorough machine learning procedure called Least Squares Support Vector Machine (LSSVM) was utilized. The Least Square Support Vector approach was developed and tested by means of sampled data recorded from northern Persian Gulf oil repositories. The result acquired from the Least Square Support Vector model were contrasted with the pertinent genuine petro actual information and the yields accomplished from different techniques utilized in their experiment. The normal relative total deviation between the methodology assessments and the significant genuine information was observed to be under 1% for the least squares SVM approach.

[11] propose two new relationships, with and without gathering analysis, for penetrability forecast. The outcomes are promising, as they arrived at the midpoint of 16 score 17 got with 50 repetition is bigger than 0.96. The authors gave a significant instrument to penetrability forecast dependent on mathematical techniques that recognize the pore structure by considering 19 fundamental patterns in the estimations.

[12] used 400 completely saline solution-soaked 3D miniature CT pictures of Bentheimer and Clashach miniature center attachments. They applied different three-measurement picture investigation methods to evaluate the rock properties (for example porosity, total penetrability) and to separate pore structure data, for example, pore throat conveyance, pore availability and pore harshness from these pictures. The entire picture dataset is isolated into two separate subsets, 80% for training purpose and 20% for testing purpose. The two subsets are taken care of to a computerized reasoning-based model to check and approve the outcomes. To work on the exactness of the model, k-overlay approval strategy is carried out. The precision of the created model can be improved by fusing extra data, for instance rock mineralogy. In any case, the created model is restricted just previously mentioned rock types can be effectively stretched out to other stone sorts gave sufficient miniature CT pictures are accessible.

### III. METHODOLOGY

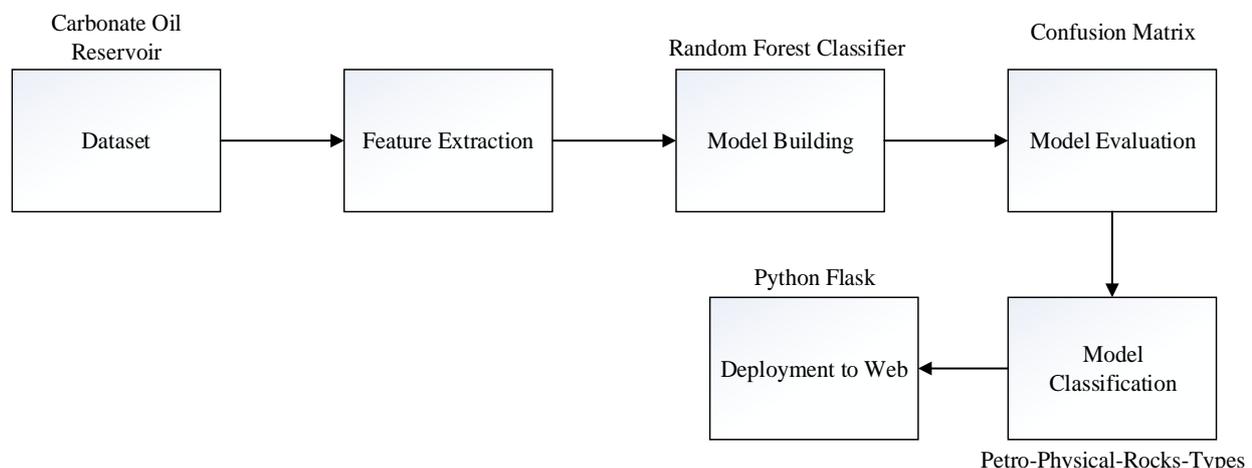


Fig. 1 Architecture of the Proposed System

**Dataset:** The dataset utilized in this work is the final sampled data collected by clerke last from Arab-D carbonate oil supply. Clerke gained almost 450 High Pressure Mercury Injection Capillary Pressure (HPMI) estimations in the Arab D



repository; be that as it may, Clerke's last examples were haphazardly chosen from 1,000's of pre-qualified center examples guaranteeing an expansive conveyance and portrayal of all Petrophysical properties. Clerke fit a Thomeer hyperbolas to each pore framework in each example to produce the distributed Thomeer Capillary Pressure boundaries including Bulk Volume Occupied (BVi), shape of Capillary Pressure bend identified with the inconstancy of pore throats (Gi) and the Initial Displacement Pressures (Pdi) for each pore framework I. From these information Clerke set up the PRTs dependent on the Initial Displacement Pressures for each pore framework and the quantity of pore frameworks present in each sampled data. [13]. Figure 2 shows the first 10 rows of the dataset

	Depth	PERMEABILITY	POROSITY	G1	PD1	G2	PD2	BV1	BV2	Mode	PRT	ROCK_INDEX	DATA_SOURCE
0	1	4800.0000	0.25810	0.49	1.29	0.15	1000.0	25.81	0.0	94.16	M	1	Rosetta
1	2	1550.0000	0.30050	0.69	1.55	0.15	1000.0	30.05	0.0	62.44	M	1	Rosetta
2	3	520.9935	0.30712	0.74	3.37	0.15	1000.0	30.71	0.0	27.12	M	1	Rosetta
3	4	234.0000	0.22080	0.64	4.39	0.15	1000.0	22.08	0.0	23.36	M	1	Rosetta
4	5	432.2024	0.24874	0.53	4.46	0.15	1000.0	24.87	0.0	26.20	M	1	Rosetta
5	6	63.0000	0.28290	0.83	7.49	0.15	1000.0	28.29	0.0	10.94	M	1	Rosetta
6	7	28.7200	0.24580	0.73	11.21	0.15	1000.0	24.58	0.0	8.20	M	1	Rosetta
7	8	23.6613	0.26601	0.37	22.70	0.15	1000.0	26.60	0.0	6.16	M	1	Rosetta
8	9	25.6819	0.17585	0.55	23.00	0.15	1000.0	17.59	0.0	4.94	M	1	Rosetta
9	10	6.1260	0.17163	0.33	32.06	0.15	1000.0	17.16	0.0	4.56	M	1	Rosetta

Fig. 2 Training Data

**Feature Extraction:** The dataset consists of 14 columns ranging from depth, permeability, porosity to rock index. Here we applied feature extraction in reducing the dataset columns from 14 columns to 3 columns, in other to have the best training samples of the proposed model. The reduced columns used in training the proposed model, are permeability, porosity and rock index column., while other columns/features were used for analysis.

**Model Building:** The model was trained using Random Forest Classifier. 80% of the Reservoir data to the random forest classifier and 20% of the dataset are used for testing. In other to get a better training accuracy, we changed the number of estimators until we find a better training result.

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2 \dots (1) \text{ Mean Squared Error}$$

Where

N represents the number of points in the data,

$f_i$  is the value returned by the model and

$y_i$  is the original value for data point  $i$ .

$$Gini = 1 - \sum_{i=1}^C (p_i)^2 \dots (2) \text{ Classification Problem}$$

Where  $p_i$  represents the relative frequency of the classes reservoir data and  $c$  represents the numbers of classes

**Model Evaluation:** Cross-Validation is an extremely integral asset. It helps in giving a superior information, and it give substantially more data about the exhibition of the Random Forest model. In complex machine learning models, it's occasionally simple not give sufficient consideration and utilize similar information in various strides of the pipeline. Accordingly, In other to get a more clear image of the Petro-Physical-Rock-Types different people in their local locations and workplaces, we investigate the warm solace information via completing some statistical examination like plotting of diagrams, histograms, order report and disarray measurements.

**Model Classification:** This has to do with the classification of the Petro-Physical-Rock-Types (Rock index) using the just the permeability and porosity data gotten from the carbonate reservoir database. The classification includes:

1. Macro with Meso Grains: Clean carbonate grain stone with colour cyan
2. Macro with Micro Grains: Clean clay size carbonate stone with colour dodge-blue
3. Macro with Meso and Micro Grains: clean carbonate grain stone with clay size carbonate with Blue colour.
4. Meso Porous Rock: Highest ratio of porous rock with colour yellow
5. Meso and Micro Porous Rock: Second highest ratio of a porous rock with colour orange
6. Micro Porous Rock: Smallest ratio of a porous rock with colour brown.

Algorithm of the proposed model

Step 1: Load carbonate Reservoir data

Step 2: Apply feature extraction to select few features



```

Step 3: //Split the data into training (80%) and test (30%)set
Step4: X_train, X_test, y_train, y_test = train_test_split
Step5: (x_process, y_process, test size = 0.20)
Step6: //Import Random Forest Classifier
Step7: Rfc = RandomForestClassifier()
Step8: Rfc.fit (X_train, y train)
Step9: predicT = Random.predict(X test)
Step10: //Evaluate model accuracy
Step11: Evaluate(Random)
Step12: Return Random

```

#### IV. RESULTS AND DISCUSSION

This system present a smart system for predicting the porosity and permeability of Petro-Physical-Rock-Types from a carbonate reservoir using Random Forest Classifier. The system starts by using an Arab D carbonate reservoir dataset, which was fine tuned by [13]. The dataset consist of fourteen columns which comprises of Depth, permeability, porosity, pore throats, displacement pressure, bulk volume, mode, and rock index which can be seen in figure 2 . We applied feature extraction by selecting just three features from the dataset. The selected features are permeability, porosity and rock index columns. This columns were selected so as to train a model that predicted the porosity and permeability of Petro-Physical-Rock-types using permeability and porosity data. We prepared the dataset by assigning the permeability and porosity column to x variable and the rock index column as y variable. The training data was then used in building a model using Random Forest Classifier which we imported from sklearn.ensemble. We trained our proposed petro-physical-rock-types model using  $n\_estimators=100$ . By  $n\_estimators$ , we mean the number of nodes present in the Random Forest Classifier. We set the number of nodes in the Random Forest Classifier to be 100. We trained our model by passing the training data to the Random Forest Algorithm. The trained model showed a promising accuracy of about 100% approximately. Figure 5 and 6 shows the evaluation performance of the model in terms of classification report and confusion matrix. Figure 3 shows a bar chart that represent the different classes of Petro-Physical-Rock-Types and the various pore size they belong to and the specific colour they posses. Figure 4 shows a scatter plot of permeability against porosity, which shows that the increase in permeability determine the increase in porosity. This indicates that the higher the greater the permeability and porosity, the easier it becomes to extract oil or natural gas from the rock. Figure 7 shows the deployment of the trained model to web using python flask in building a smart system that predicted the porosity and permeability of Petro-Physical-Rock-types by collecting input data from the user. Figure 8 shows the predicted result shows a meso porous rock.

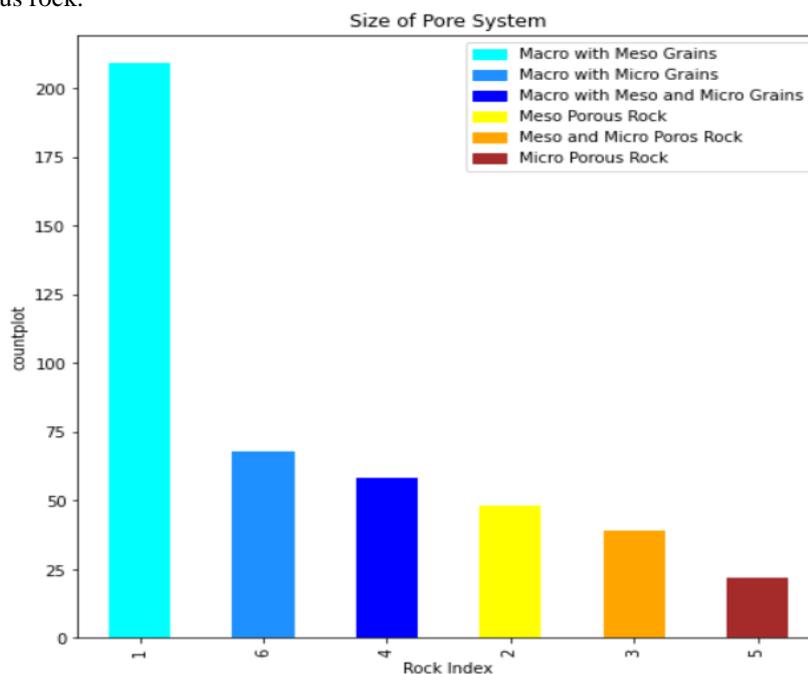


Fig. 3: Petro-Physical-Rock-Types



This shows the pore size of the different classes of Petro-Physical-Rock-Types starting from macro with meso grains, macro with micro grains, macro with meso and micro grains, meso porous rock, meso and micro porous rock to micro porous rock. The classified classes are six.

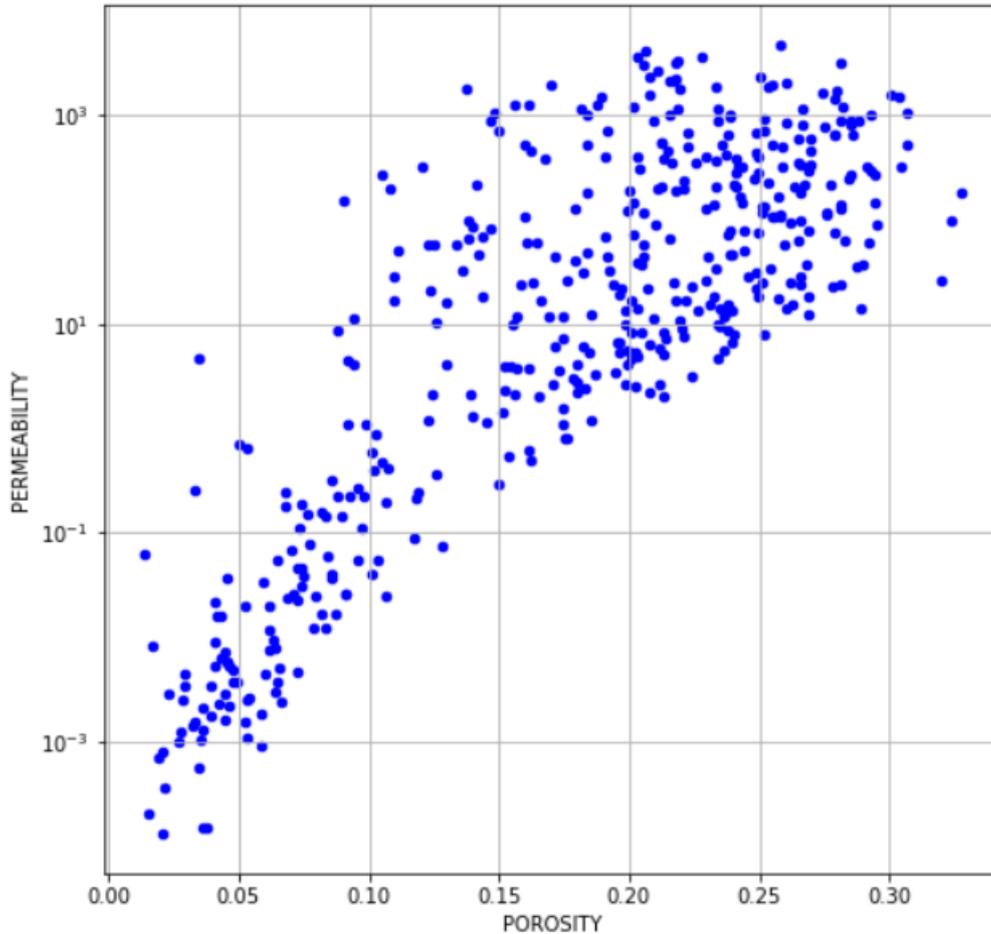


Fig. 4: Scatter Plot showing the increase in permeability and porosity of a Petro-Physical-Rock-Type.

	precision	recall	f1-score	support
1	1.00	1.00	1.00	192
2	1.00	1.00	1.00	42
3	1.00	1.00	1.00	35
4	1.00	1.00	1.00	52
5	1.00	1.00	1.00	20
6	1.00	1.00	1.00	58
accuracy			1.00	399
macro avg	1.00	1.00	1.00	399
weighted avg	1.00	1.00	1.00	399

Fig. 5: Classification Report of Our Proposed Model.

These shows a promise performance of our model with a training accuracy of about 100%, precision 100%, the support shows the total number of prediction carried out by our model

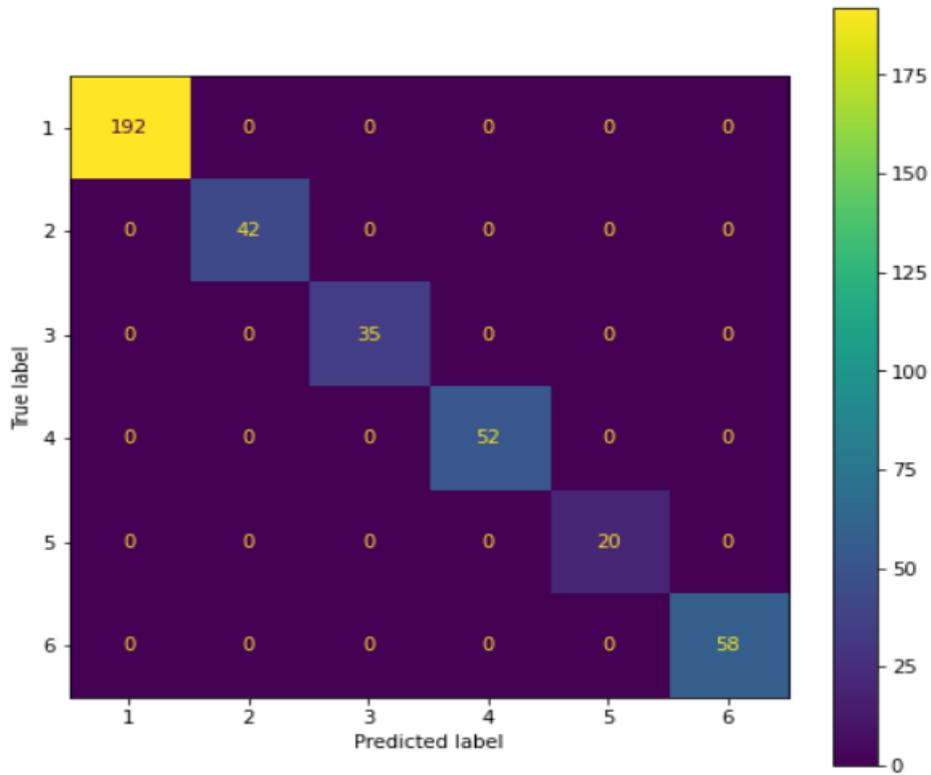


Fig. 6: Confusion matrix

This shows the correct prediction of our trained model

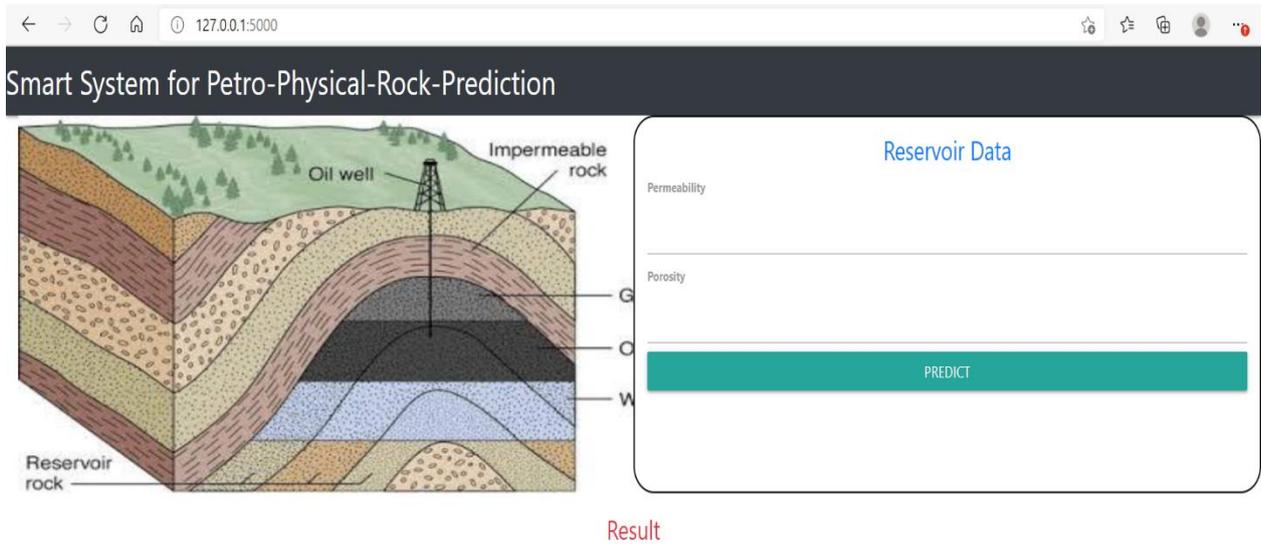
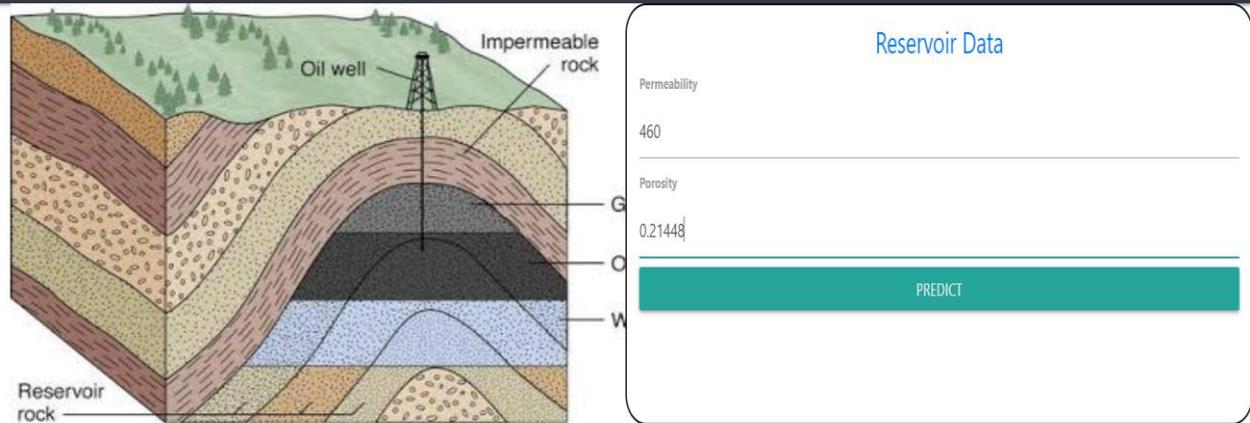


Fig. 7: Homepage of the Web App

## Smart System for Petro-Physical-Rock-Prediction



Result

Meso Porous Rock

Fig. 8: Predicted Result

This shows that the rock has a high porosity. Therefore, it is good to extract oil or gas

## V. CONCLUSION

This paper presents a smart system for predicting the porosity and permeability of petro-physical-rock-types. The system starts by training a model using just the permeability and porosity data gotten from a reservoir database. The reservoir data consist of a total of 14 columns which was reduced to just three columns. The reduction was done by selecting just the most important feature by means of feature extraction. The model was trained using just the selected features. The model was used using a Random Forest Classifier with number of estimators to be 100. The model achieved a better training performance by obtaining a training accuracy and precision of about 100% each. We evaluated the model performance by plotting a classification report and confusion matrix. The trained model was then saved and deployed to web using python flask for future prediction of porosity and permeability in carbonate rocks. The results generated by the trained model on web is very impressive. This was done by collecting just permeability and porosity data from users through an input form. This paper can be improved by predicting a continuous flow of porosity in a carbonate rock using a well log data.

## REFERENCES

- [1]. J. Goral, P. Panja, M. Deo, M. Andrew, S. Linden, J. Schwarz, A. Wiegmann, "Confinement Effect on Porosity and Permeability of Shales", *Scientific Reports* 10(1), 1-11, 2020.
- [2]. G. Zhang, B. Xie, Z. Bao, Z. Niu, K. Jiao, "Multi-phase simulation of proton exchange membrane fuel cell with 3D fine mesh flow field". *Int. J. Energy Res.* 42, 4697–4709 2018.
- [3]. A.S. Al-Menhali, H.P. Menke, M. Blunt, S.C. Krevor, "Pore scale observations of trapped CO<sub>2</sub> in mixed-wet carbonate rock: Applications to storage in oil fields". *Environ. Sci. Technol.* 50, 10282–10290 2016.
- [4]. S.E. Kaczmarek, S.M. Fullmer, F.J.A. Hasiuk, "A universal classification scheme for the microcrystals that host limestone micro porosity". *J. Sediment. Res.* 85, 1197-1212 2015.
- [5]. A. Erofeev, D. Orlov, A. Ryzhov, D. Koroteev, "Prediction of Porosity and Permeability Alteration Based on Machine Learning Algorithms", arXiv: 1902.06525v1 [cs.LG] 18 Feb 2019.
- [6]. W.J. Al-Mudhafar "Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms". *J Petrol Explor Prod Technol* 7, 1023–1033 (2017). <https://doi.org/10.1007/s13202-017-0360-0>.
- [7]. H. P. Menke, J. Maes, S. Geiger, "Upscaling the porosity–permeability relationship of a microporous carbonate for Darcy-scale flow with machine learning", *Scientific Reports* 11(1), 1-10, 2021.
- [8]. Z. Zhang, H. Zhang, J. Li, Z. Cai, "Permeability and porosity prediction using logging data in a heterogeneous dolomite reservoir: An integrated approach", *Journal of Natural Gas Science and Engineering*, 86, 1-16, 2021.



- [9]. H.A. Khalifah, p.w. Glover, P. Lorinczi, “Permeability Prediction and Diagenesis in Tight Carbonates Using Machine Learning Techniques”, *Marine and Petroleum Geology*, 112, 1-33, 2020.
- [10]. M.A. Ahmadi, Z. Chen, “Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs”. *Petroleum* 2018, doi:10.1016/j.petlm.2018.06.002.
- [11]. Tran H., A. Kasha, A. Sakhaee-Pour, & i. Hussein, I, “Predicting carbonate formation permeability using machine learnin”g. *Journal of Petroleum Science and Engineering*, 107581. doi:10.1016/j.petrol.2020.107581.
- [12]. A.R. Shaik, A.A. Al-Ratrou, A.M. AlSumaiti, A.K. Jilani, “Rock Classification Based on Micro-CT Images using Machine Learning Techniques”. *Abu Dhabi International Petroleum Exhibition & Conference*. doi:10.2118/197651-ms.
- [13]. Clerke, E. A., Mueller, H. W., Phillips, E. C., Eyvazzadeh, R. Y., Jones, D. H., Ramamoorthy, R., Srivastava, A., “Application of Thomeer Hyperbolas to decode the pore systems, facies and reservoir properties of the Upper Jurassic Arab D Limestone, Ghawar field, Saudi Arabia: A Rosetta Stone approach”, *GeoArabia*, Vol. 13, No. 4, p. 113-160, October, 2008.