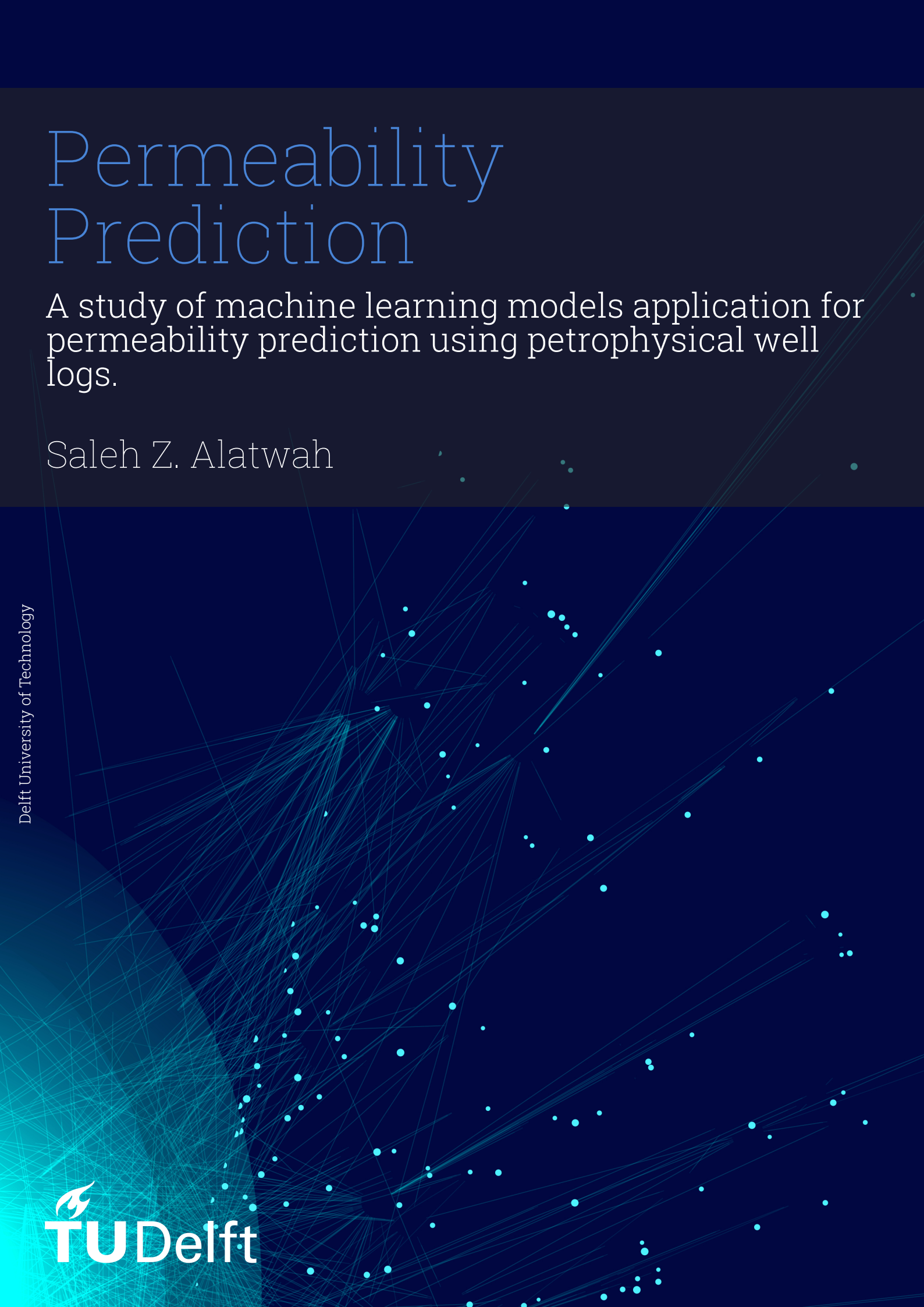
Permeability Prediction

A study of machine learning models application for permeability prediction using petrophysical well logs.

Saleh Z. Alatwah

Delft University of Technology



Permeability Prediction

A study of machine learning models application for permeability prediction using petrophysical well logs.

Thesis report

by

Saleh Z. Alatwah

Submitted in partial fulfillment of the requirements for the degree of
Master of Science in Applied Earth Sciences
at the Delft University of Technology
to be defended publicly on August 21, 2023 at 01:00 PM

Thesis committee:

Supervisors:	Dr. Guillaume Rongier
	Dr. Femke Vossepoel
	Andrea Cuesta Cano, MSc
	Azin Karimzadanzabi, MSc
Place:	Faculty of Civil Engineering and Geosciences, Delft
Project Duration:	November 22, 2022 - August 21, 2023
Student number:	5469732

Faculty of Civil Engineering and Geosciences · Delft University of Technology

TU Delft

Preface

Having achieved this stage, I would want to convey my profound gratitude to everyone who contributed to my Master's degree success, including the TUDelft faculty, my family, my friends, and Saudi Aramco. Special thanks to my advisor, Dr. Guillaume Rongier, who participated in every element of the project from the beginning, when it came to selecting a study topic that would both fulfill my learning curiosity and expand my academic and practical expertise.

Guillaume made it possible to take pleasure in research and replace feelings of stress and obligation with a passion for the topic and a desire for better results. His unquestionable expertise and experience increased my learning curve at TUDelft, making the thesis project the most information-rich course of all.

Above all, he urged that I focus on themes that piqued my interest and directed me through the research with his unique, energetic, and optimistic personality. In addition, I would want to thank the other committee members, specifically Dr. Femke Vossepoel, for making herself available to answer any queries, providing assistance when necessary, and sharing instantly any new publications that can benefit my research. In addition, I would like to thank PhD candidates Andrea Cuesta Cano and Azin Karimzadanzabi for their continuous support.

Last but not least, I would want to convey my thanks to Saudi Aramco for financing my studies and being there for me throughout my time in the Netherlands, offering support and ensuring that the experience was worthwhile. Having a special place in my thoughts and heart, I would like to thank my NARCD family, in particular my manager Abdallah Al Shamsi and my mentor Khadem Al Nasser, for their innumerable efforts in offering all the available assistance to make me succeed not only as an employee but also as a person.

Saleh Z. Alatwah
Delft, August 2023

Abstract

Permeability, a key reservoir characteristic, governs the rate of fluid flow through reservoir rocks. Accurate permeability estimates are paramount for robust reservoir simulation, history matching, and production forecasting. Due to limited core data availability and intrinsic heterogeneity of permeability at different scales, establishing reliable permeability models can be challenging. This study aims to overcome these hurdles by predicting lab-measured core permeability from commonly acquired well logs, using various machine learning algorithms such as Support Vector Regression (SVR), Random Forest (RF), XGBoost, and LightGBM.

We examined two diverse datasets, representing a carbonate platform (Costa Field) and clastic formations (Volve Field). The Costa dataset, including 17 wells across a single reservoir, and the Volve dataset, comprising three wells across three different reservoirs, allowed for evaluating the robustness of our approach under different geological conditions.

A critical part of our methodology is feature engineering, particularly incorporating vertical variability. We integrated measurements from adjacent well log readings into our models, recognizing the importance of spatial context and the smoothing effect of well logs over small-scale heterogeneities. This improved prediction accuracy by accounting for shared geological history and depositional environments in proximity.

In Costa Field, blind tests showed R2 scores up to 0.64, and validation R2 scores reached up to 0.8 using a leave-one-well-out cross-validation method. For the Volve Field, blind test R2 scores were up to 0.84, 0.76, and 0.78 for Hugin, Sleipner, and Skagerrak formations, respectively. These results, while satisfactory, underscore the potential of machine learning methods in accurately predicting permeability and highlight the need for effective feature engineering.

This work advocates that while machine learning holds promise for automated feature engineering, human intervention, specifically to incorporate spatial context, can still significantly enhance predictions. Future advancements may seek to internalize this spatial awareness within the machine learning algorithms themselves.

Contents

1 Introduction

1.1. Problem Statement

Permeability is an integral parameter for reservoir characterization and management. Its significance is prevalent in various applications such as oil and gas extraction, CO2 or hydrogen storage, geothermal energy, among others. Permeability quantifies a fluid's ability to flow through porous media, thereby playing a fundamental role in comprehending and predicting reservoir performance.

Permeability is usually determined in a laboratory setting using core samples. This involves passing a fluid, with a known viscosity, through a core sample of fixed dimensions at a specific rate. The resulting pressure drop across the core is then measured. However, this procedure is both time-consuming and expensive, making it impractical to use on every well. Consequently, there is a shortage of direct core permeability measurements for an entire field. This scarcity of data significantly complicates the task of assessing reservoir performance and predicting future production behavior.

In light of the limitations, high costs, and impracticality associated with core sampling, this study explores the potential of machine learning algorithms as a more feasible alternative. Machine learning, a branch of artificial intelligence, uses statistical techniques to give computer systems the ability to learn from data, without being explicitly programmed. It aims to develop algorithms or models that can learn from and make decisions or predictions based on data. Here, we focus on several machine learning algorithms, which are Support Vector Regression (SVR), Random Forest (RF), XGBoost, and LightGBM. These algorithms are capable of learning patterns from wire-line log data, which is more readily available than core data, and use these patterns to make accurate and reliable predictions of permeability. This approach presents a potential solution that is both cost-effective and less data-intensive, addressing the challenges posed by the scarcity and uneven availability of core data.

1.2. Objectives and Method

The main research question this thesis aims to address is: "How can we leverage machine learning to predict permeability from well logs, achieving satisfactory predictions across datasets without the user having to perform in-depth tuning?"
The objectives of this study are:

- To benchmark the performance of selected machine learning algorithms in permeability prediction.
- To evaluate the impact of feature engineering techniques on these models.
- To assess the efficiency of hyperparameter tuning for each algorithm.
- To test the robustness and generalization potential of these algorithms using two specific datasets representing different geological scenarios.

To fulfill these objectives, we propose a methodological approach, also emphasizing robust validation. This validation, performed using 'leave-one-well-out' and 'k-fold' cross-validation methods, aims to mirror conditions as closely as possible to real applications.

Our approach includes:

- Assessing multiple machine learning techniques such as XGBoost, LightGBM, SVM, and Random Forest.
- Processing and using data from multiple geological datasets, mimicking real-life scenarios where core data is not uniformly available across the field, thus evaluating how different models trained on different sizes of core data generalize.
- Applying feature engineering techniques to enhance the predictive power of the models.
- Implementing hyperparameter tuning for each algorithm.
- Comparing the performance of the models both in their original form and after feature engineering and hyper-parameter tuning.

The detailed results of these processes will provide comprehensive insights into the potential of machine learning algorithms for permeability prediction in diverse geological contexts.

1.3. Document Structure

The structure of the thesis is as follows: Chapter 2 provides a literature review on permeability prediction, and the application of machine learning techniques in geo-energy engineering. Chapter 3 details the geological datasets used in the study and the preprocessing steps applied. It also introduces the machine learning algorithms used and the method for hyperparameter tuning. Chapter 4 presents a comprehensive benchmarking of the algorithms, with and without feature engineering and hyperparameter tuning, across the datasets. Chapter 5 provides a discussion of the findings, and Chapter 6 concludes with implications and potential future directions.

2

Background

The estimation of permeability, historically reliant on direct measurements and empirical petrophysical relationships, has evolved to embrace the use of machine learning techniques, offering potentially enhanced accuracy and predictive capability. This background section aims to provide a comprehensive overview of these aspects, beginning with an exploration of well-log data and its correlation with permeability, moving on to traditional petrophysical approaches for permeability estimation, and concluding with an examination of machine learning techniques used in modern permeability prediction.

2.0.1. Types of Well Logs and Their Relation to Permeability

Well logging, a fundamental tool in the geo-energy industries, provides a wealth of data that can be used to assess the characteristics of a reservoir. Different types of logs measure various physical properties of the formations and their contained fluids, many of which have a bearing on the permeability of the reservoir rock. (Table 2.1) summarizes the common types of well logs, their measurements, and the potential inferences for permeability

Well Log Type	**Measures**	**Units**	**General Link to Permeability**
Gamma Ray Log (GR)	Detects radioactivity, helps design lithological profiles	API	Useful for differentiating between reservoir and non-reservoir rocks, with indirect implications for permeability based on shale content (Carcione et al., 2020).
Resistivity Log (RT)	Measures resistivity, helps determine hydrocarbon versus water-bearing zones	Ohm.m	Resistivity is primarily used to determine fluid content, and can indirectly suggest permeability when used in conjunction with porosity and fluid saturation data (Saner et al., 1997).
Bulk Density Log (RHOB)	Measures bulk density, helps in porosity determination	g/cc	Can indirectly infer permeability through measures of porosity and differences between oil and gas-bearing zones (Asquith et al., 2004).
Neutron Porosity Log (NPHI)	Measures hydrogen concentration, helps in total porosity determination	V/V or %	High porosity can be linked with higher permeability, but this is dependent on pore interconnectivity (Asquith et al., 2004).
Sonic Log (DT)	Measures velocity of elastic waves, provides formation interval transit time	µs/ft	Faster travel times can suggest higher permeability. Often used in conjunction with other logs for a more accurate estimation (Asquith et al., 2004).

Table 2.1: Well Log Types and their Measures, Units, and Links to Permeability

2.0.2. Empirical Approaches

Several empirical models are used to estimate permeability from well logs. Among these, the Kozeny-Carman equation is a semi-empirical relationship that relates the permeability of a porous medium to its porosity and specific surface area (Gholanlo, 2021). Other empirical models, such as the Tixier, Timur, and Coates & Dumanoir methods, are also commonly employed (Aigbedion, 2007). These latter models are based on correlations between permeability, porosity, and irreducible water saturation. They primarily use resistivity measurements, assuming certain values for the cementation factor and/or saturation exponent (Mohaghegh et al., 1997). These methods are particularly applicable to clean sand formations where conditions of residual water saturation exist. However, it is important to note that all these empirical models may not perform well in heterogeneous formations due to the complex nature of these formations (Mohaghegh et al., 1997). A summarization of these methods is presented in (Table 2.2).

Method	Year	Equation
Kozeny-Carman	1937	$K = \frac{d^2\phi^3 C}{(1-\phi)^2 S^2}$
Tixier	1949	$\left(\frac{K}{20}\right)^{1/2} = \frac{2.3}{R_0(d_w - d_o)} \frac{\Delta R}{\Delta D}$
Timur	1968	$K = 0.136 \frac{\phi^{4.4}}{S_w^2}$
Coates & Dumanoir	1974	$K^{1/2} = \frac{C}{w^4} \frac{\phi^2}{R_w/R_{ti}}$
Coates	1981	$K^{1/2} = 100 \frac{\phi^2(1-S_{wirr})}{S_{wirr}}$

Table 2.2: Empirical Methods Used for Permeability Estiamtion adjusted after (Mohaghegh et al., 1997)

2.0.3. Machine Learning Approaches for Permeability Prediction

(Rezaee and Ekundayo, 2022) investigates the use of machine learning techniques to predict permeability based on well logs and core data from five boreholes in the Surat Basin, Australia. The authors found that the Artificial Neural Network with all seven available input logs performed the best, with a coefficient of determination (R2) of approximately 0.93 for the training data set and 0.87 for the blind testing data set. However, the study did not include data standardization or feature engineering, which may have limited the performance of other machine learning models and reduced their ability to accurately predict permeability.

(Otchere et al., 2021) presents a machine learning approach for predicting reservoir permeability and water saturation based on wireline logs from the Volve field in the North Sea. The authors utilized statistical methods and Extreme Gradient Boosting (for more detail on XGBoost, see section 3.4.3) regression to forecast these parameters, then proposed a novel ensemble model that combines the predictions of Random Forest and Lasso Regularization, and assigns importance to input features based on their relevance for permeability prediction. It is important to note, however, that the permeability used in this study is KLOGH, a calculated permeability from the Statoil model, not the actual permeability measurements obtained from core sampling. This consideration may affect the interpretation of the study's results and their direct comparability with others. Nonetheless, the ensemble model proposed was found to outperform the traditional XGBoost and hybrid PCA-XGBoost models in terms of precision, consistency, and accuracy. The strength of this study is likely rooted in the ensemble model, which leverages the predictions of multiple models to enhance overall prediction accuracy.

(Z. Zhang et al., 2021) aims to develop reliable methods for predicting permeability and porosity in the Lower Cambrian dolomite reservoir in the Tarim Basin in China. The authors collected 253 core samples with logging data and applied five typical permeability-porosity correlations and six machine learning methods to evaluate their performance. The authors discovered that the Flow Zone Indicator (FZI), combined with Discrete Rock Type (DRT) analysis, possessed a stronger ability to classify samples, and each DRT exhibited a unique permeability-porosity relationship. Based on these findings, an integrated indirect permeability prediction method was developed, combining petrophysical rock typing methods (FZI) with the Particle Swarm Optimization-Support Vector Machines (PSO-SVM) algorithm. The proposed integrated approach demonstrated improved permeability prediction

accuracy compared to direct machine learning prediction methods, reaching an R2 of 0.869 at its peak. One potential limitation of this study is that the authors performed a randomized holdout validation based on individual samples, rather than conducting a complete well blind test, to evaluate their proposed method's performance. Also, it is crucial to consider that the sample size of 253 core samples may not sufficiently capture the full spectrum of heterogeneity and complexity present in the reservoir.

(Elkatatny et al., 2017) aims to develop an artificial neural network (ANN) model for predicting permeability in heterogeneous carbonate reservoirs based on well logs. The authors used actual core data points with three logs (resistivity, bulk density, and neutron porosity) to train the model and test its performance on unseen data. They introduced a new term, the mobility index (MI), which is defined as follows:

$$MI = \frac{1}{\sqrt{R_{fz}}} - \frac{1}{\sqrt{R_{dp}}} \tag{2.1}$$

where R_{fz} is the normal value of the flushed zone resistivity and R_{dp} is the normal value of the deep resistivity. The mobility index was found to improve the permeability prediction compared to actual measured data. The ANN model showed an R2 of 0.95 and a root-mean-square error of 0.28 mD, and was comparable to support vector machine and adaptive neuro-fuzzy inference system models.

The bulk of petrophysical models used for traditional logging interpretation rely on linear or nonlinear relationships with well logs, and regression of logging sequences is commonly used to forecast parameters like as porosity, permeability, and water saturation, according to (S. Zhang et al., 2022). Nevertheless, ensemble learning approaches such as Light gradient boosting machine (for more detail on LightGBM, see section 3.4.4) have developed as a potentially effective answer to this issue. (S. Zhang et al., 2022) emphasized the efficacy of the suggested CRBM-Bayes-LightGBM hybrid predictor, which employs LightGBM as a cutting-edge ensemble learning model to forecast the target reservoir characteristics. (S. Zhang et al., 2022) conducted experiments using data from the Chang 8 member in the Jiyuan Oilfield, Ordos Basin, Northern China, to validate the performance of this model. The results demonstrate that the LightGBM-cored predictor is more accurate than three competing models, and that transfer learning may be used to make accurate predictions for small-volumetric datasets.

(G. Zhang et al., 2021) propose a novel approach for visualizing the predictive patterns learned by machine learning models for permeability prediction. They build machine learning models using support vector regression (for more detail on support vector regression, see section 3.4.1), random forest (for more detail on random forest, see section 3.4.2), and deep residual neural network (ResNet (He et al., 2016)) and use porosity and wireline logs as input features as well as ResNet features of thin sections. The results show that ResNet infused model performs the best in permeability prediction, with an R2 of 0.9 for training and 0.88 for testing. The use of ResNet is noteworthy, as it is a well-known pre-trained model for image classification. Overall, this study provides valuable insights into the use of machine learning for permeability prediction.

3

Material and methods

3.1. Datasets and Geological Setting

In this section, we introduce the two datasets that will be used to train and evaluate the machine learning models for predicting permeability using well logs and core data. These datasets are The COSTA field, and the Volve field. Each dataset consists of well logs including gamma ray (GR), density (RHOB), neutron porosity (NPHI), sonic (DT), and resistivity measurements, as well as stratigraphic well tops and core permeability data. These datasets will be used to train and test the performance of the machine learning algorithms developed in this study. We will also provide an overview of the characteristics of the datasets and any preprocessing steps that were taken prior to model training.

3.1.1. The COSTA field

The COSTA field dataset (Costa Gomes et al., 2022) represents the geological environment of the Upper Kharaib Member (Early Cretaceous) of the northeastern Rub Al Khali basin, a sub-basin of the larger Arabian Basin. This basin exhibits characteristics of a shelf-to-basin carbonate platform, one of the most productive aggradational carbonate parasequence formation sets globally in terms of oil and gas production (Costa Gomes et al., 2022).

The Kharaib Formation in the subsurface of this region is characterized by its distinct geological zones and is divided into two primary reservoir units, B-Upper and B-Lower. The B-Upper unit, subdivided into three main subzones, BI, BII, and BIIIU, is dominated by grain-supported limestone fabrics, including rudist-facies, bacinella-facies, and bioclast-grainstone facies (Strohmenger et al., 2004). These facies, originating from high and moderate-energy environments, are believed to have been reworked in shallower environments or during storm events, as depicted in (Figure 3.1).

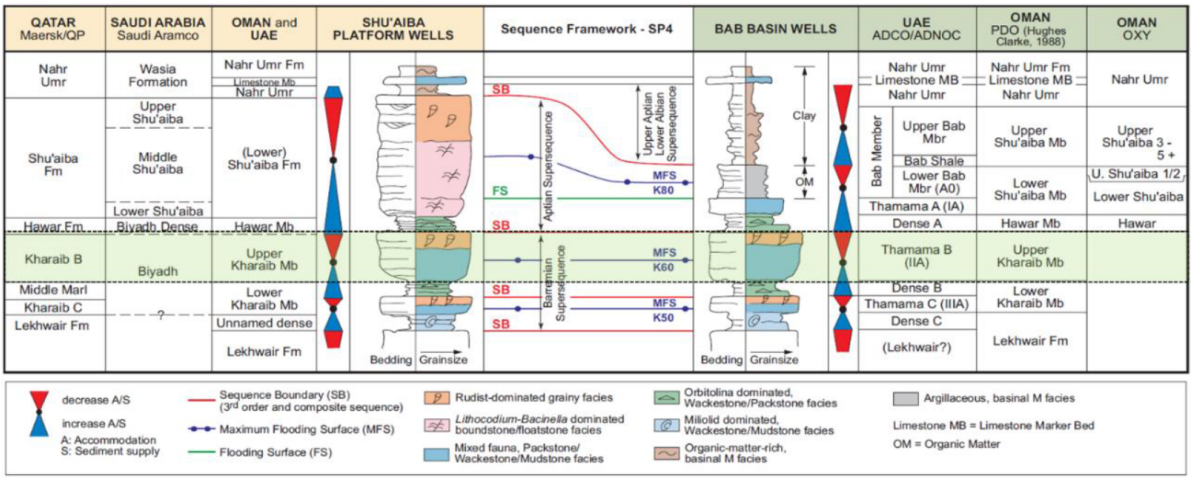


Figure 3.1: Regional lithostratigraphic representation of the Upper Kharaib Member (emphasized in green) spanning the southeast Arabian Plate from platform regions to basinal areas, showcasing its rudist-dominated grainy facies and the presence of packstone, wackstone, and mudstone facies (Costa Gomes et al., 2022).

In contrast, the B-Lower unit, further divided into two main subzones, BIIIL and BIV, is primarily associated with lime mudstone/wackestone facies, as illustrated in (Figure 3.2).

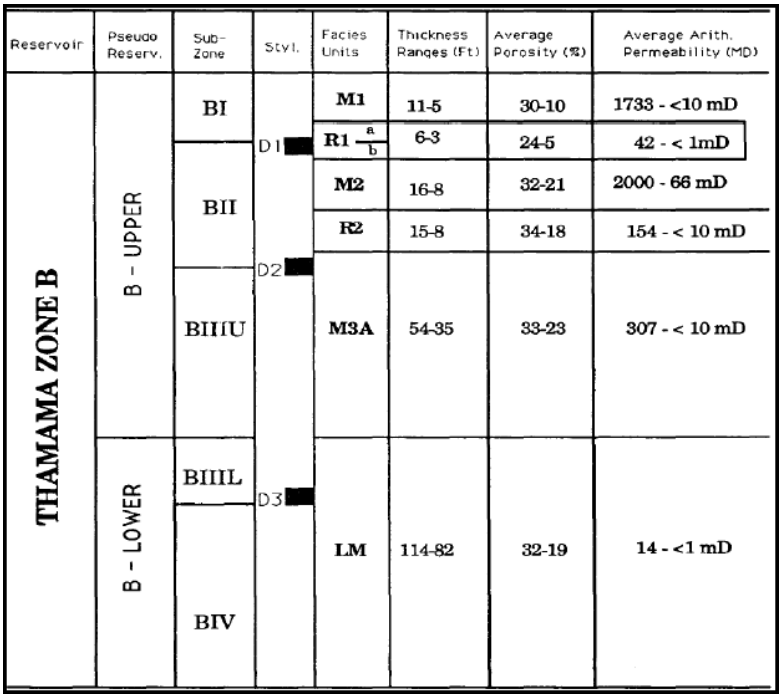


Figure 3.2: Upper Kharaib Member facies units from the Early Cretaceous period (within Thamama Zone B) and their associated reservoir properties (Costa Gomes et al., 2022).

The division between these reservoir units arises from their unique sedimentological and diagenetic characteristics. B-Upper has undergone early diagenetic dissolution, leading to highly permeable and porous reservoir zones, while B-Lower has experienced more extensive chemical and mechanical diagenesis. This differentiation provides a comprehensive understanding of the reservoir's potential and behavior, addressing the lithological and facies variations within the Kharaib Formation.
Within this carbonate platform, a distinctive pattern of interbedded porous and dense carbonate successions can be observed. The transitions between these zones can be attributed to successive transgressive (sea-level rise) and regressive (sea-level fall) conditions, which induced changes in the depositional environment(Costa Gomes et al., 2022).

The reservoir quality is closely tied to these sedimentary facies and their depositional dynamics, influencing critical reservoir properties such as porosity and permeability. Specifically, the Upper Kharaib reservoir exhibits an average porosity ranging from 18 to 27% and permeability values ranging from 25 to 70 mD. However, these averages may not accurately represent the complex heterogeneity of the reservoir, as there are thin layers with high permeability streaks exceeding 800 mD and areas with very low permeabilities below 1 mD (Torres et al., 2017). Hence, understanding these variations is crucial for permeability prediction.

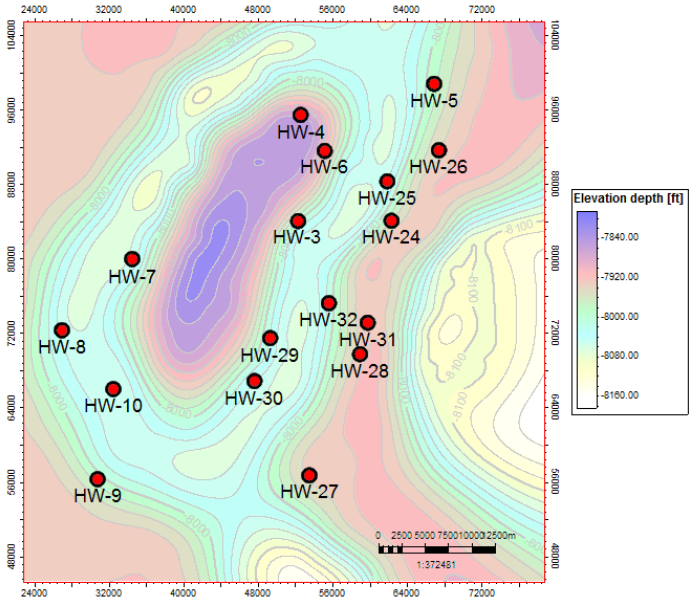


Figure 3.3: Contour map illustrating the top of the Upper Kharaib Member, with marked well locations, based on data from (Costa Gomes et al., 2022).

The dataset comprises a total of 6059 core permeability samples, obtained from 17 wells as depicted in (Figure 3.3). Each well intersects approximately a 49-meter thick section of the Upper Kharaib Member, which is characterized by 15 distinct geological zones. Among these zones, 8 are identified as reservoir zones (part of the two units mentioned before) while the remaining 7 are recognized as dense zones (lower porosity intervals that act as partial seals)(Costa Gomes et al., 2022). The dataset also includes corresponding well logs, which encompass gamma ray (GR), density (RHOB), neutron porosity (NPHI), sonic transit time (DT), resistivity (RT), and a calculated value for effective porosity (PHIF).

3.1.2. The Volve field

The Volve oil field(Equinor, 2022 Sanei et al., 2022), situated in the center portion of the North Sea, is located offshore Norway (Figure 3.4). With an approximate water depth of 80 meters, the field is situated in shallow waters.

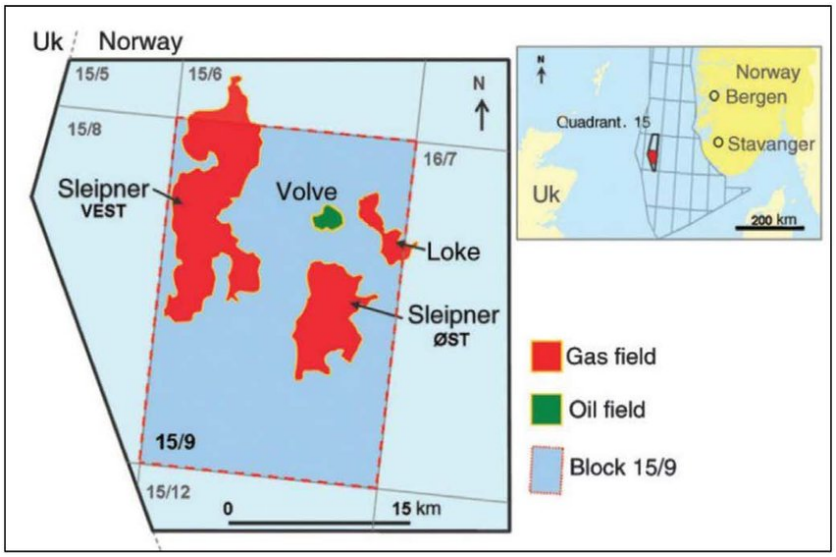


Figure 3.4: Location of Volve field in the North Sea (Sanei et al., 2022).

The dataset contains permeability core samples from 3 formations (Figure 3.5) in the Volve field, which are the Hugin Formation, Sleipner Formation and Skagerrak Formation.

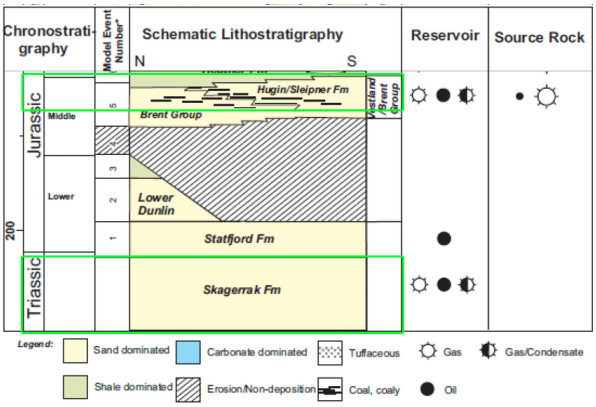


Figure 3.5: From the Triassic to the Jurassic, a generalized sequence of rock layers in the South Viking graben and the North Sea area. Highlighted in green are the formations of interest: Hugin Formation, Sleipner Formation, and Skagerrak Formation. Adapted from (Saha et al., 2022).

Hugin Formation

The Hugin Formation predominantly consists of light brown to yellow, very fine to medium-grained sandstones. These sandstones exhibit fair sorting, indicating a balanced mix of grain sizes, with grains that range from subangular to subrounded in shape. Intermittent layers of coarse-grained sediment, carbonaceous material, and laminated coal seams are present. Additionally, partings of shale and siltstone are evident, further contributing to the lithological diversity of the formation. Bioturbation is common in the sandstones, complemented by occasional cross-bedding.(Vollset and Dore, 1984).

The Hugin formation is composed of clastic sediments that were deposited in shallow and marginal marine environments. These sediments were deposited during a transgressive phase, which coincided with the southward migration of the Brent Delta systems (Kieft et al., 2010).

Figure 3.6: Core images form well 15/9-19 A displaying the Hugin Section of the core (NPD, 2023)).

Sleipner Formation

The formation is composed of coal-bearing sandstone and silty claystone lithology. The sandstones are non-calcareous, light to medium brown, fine to medium-grained, and contain occasional coarse and pebbly layers. The sandstones exhibit a moderate to poor degree of sorting, with granules that are sub-angular to sub-rounded. The silty claystones are medium to dark grey or greyish brown, micromicaceous, hard, and slightly fissile. Coal fragments, fossilized leaves, and root filaments are commonly discovered. The coals are mature, black, and massive, indicating a lack of internal layering. Thin laminations of silty claystone are frequently present (Vollset and Dore, 1984).

The Sleipner Formation can be interpreted as a clastic sedimentary sequence of continental origin, characterized by the deposition of fluvial and deltaic sediments, which are interbedded with coal layers (Vollset and Dore, 1984).

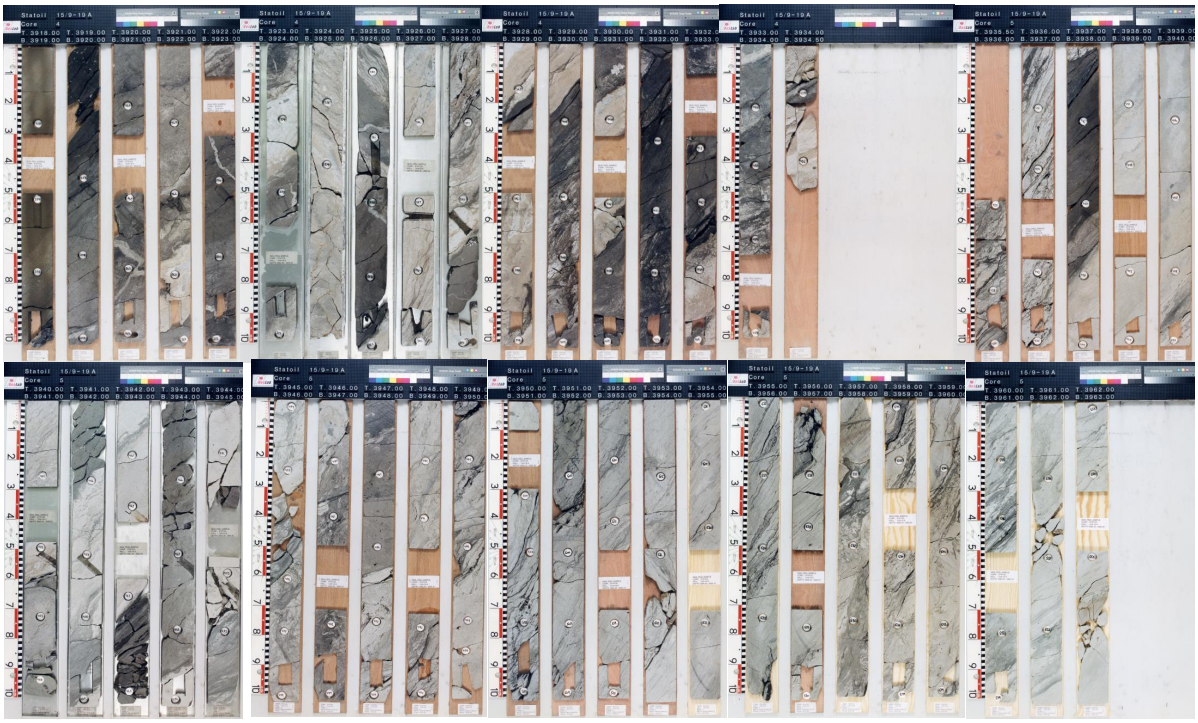


Figure 3.7: Core images form well 15/9-19 A displaying the Sleipner Section of the core (NPD, 2023)).

Skagerrak Formation

The formation is composed of alternating sequences of conglomerates, sandstones, siltstones, and shales. The sedimentary sequence exhibits a prevalence of reddish and brownish hues, with intercalations of grayish beds ranging from light to dark. The sandstones exhibit varying compositions, ranging from orthoquartzitic, arkosic to highly lithic. The lithologies of anhydrite, dolomite, and limestone are present in a subordinate manner (Deegan and Scull, 1977).

The majority of the Skagerrak Formation was probably deposited in a system of alluvial fans (Figure 3.8) that joined and were prograding along the eastern and southern edges of a basin that was under structural control. Some of the dark shale, carbonate, and anhydrite beds may have been deposited in lakes, based on the small area and badly preserved faunal components. Microfossils that are better preserved and other signs like glauconite show that some beds originated when small amounts of seawater got in between floods of continental clastics (Gray et al., 2020).

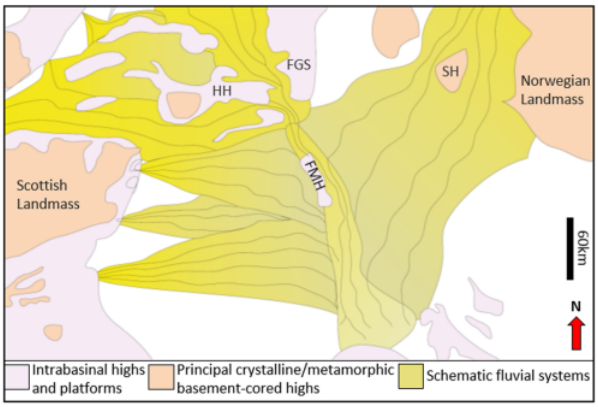


Figure 3.8: Reconstruction of the Skagerrak Deposition using a generalized schematic Gray et al., 2020.

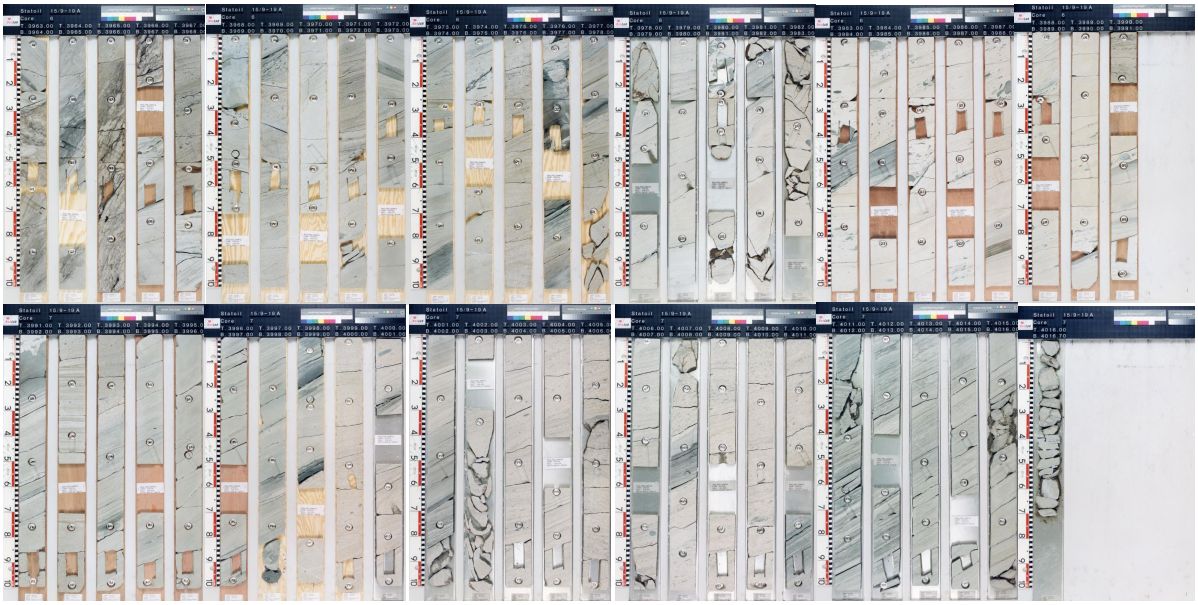


Figure 3.9: Core images form well 15/9-19 A displaying the Skagerrak Section of the core (NPD, 2023)).

In the Hugin section (Figure 3.6), dark intervals are observed, which are heavy oil stains in the high porosity and high permeability sandstones. This is consistent with the higher permeability values represented in the histogram within (Figure 3.10). The Hugin formation is also characterized by higher sand content as compared to Sleipner and Skagerrak, suggesting less shaly deposits.
Conversely, the Sleipner and Skagerrak sections (Figures 3.7, and 3.9) demonstrate a higher prevalence of alternating sequences of sand, silt, and shale. This variance in lithology is reflected in the distribution histogram (Figure 3.10), illustrating a greater number of samples with lower permeability from these formations. The observations from the core images are thus well-correlated with the permeability distribution seen in the histogram, further reinforcing the distinct geological and petrophysical characteristics of the Hugin, Sleipner, and Skagerrak formations.

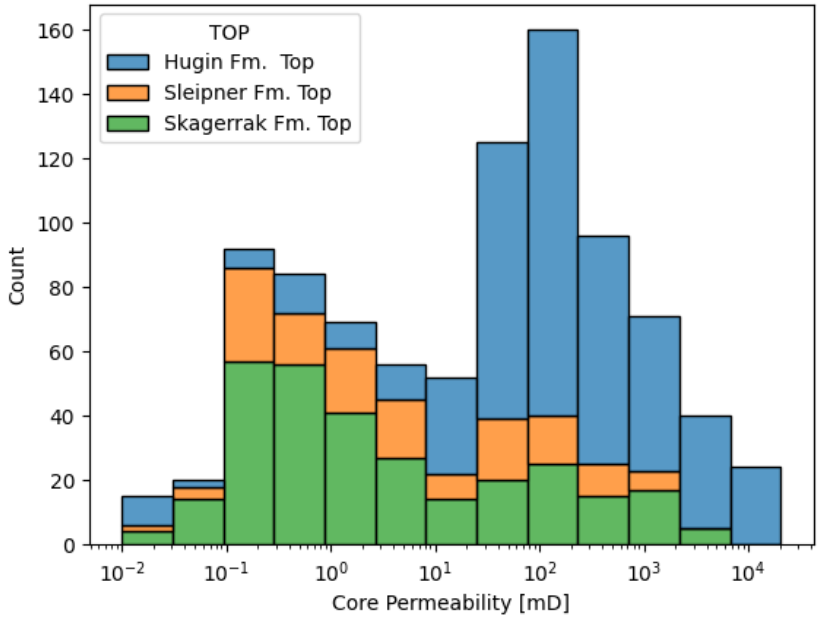


Figure 3.10: Distribution of Permeability, with color differentiation indicating the three respective intervals..

The dataset contains logs, core measurements, and formation tops from three wells (Figure 3.11)

(15/9-19 A, 15/9-19 BT2, and 15/9-19 SR). Gamma ray (GR), density (RHOB), neutron porosity (NPHI), sonic (DT), Deep (ILD) and Medium resistivity (ILM) are among the wireline logs accessible from the wells as well as computed petrophysical logs such as calculated effective porosity (PHIF) and volume of shale (VSH)

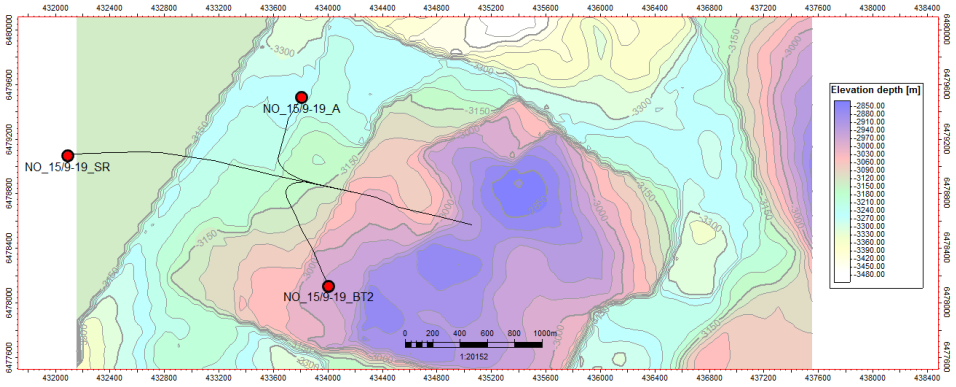


Figure 3.11: Contour map showing cored well locations in the Volve Oil Field generated from Sanei et al., 2022.

Statoil Modeled Permeability

Statoil, a Norwegian oil company currently named Equinor, conducted a comprehensive petrophysical study of the Volve field and developed a model based on a multivariable regression method (Equinor, 2022). This model uses effective porosity (PHIF) and volume of shale (VSH) (equation 3.1) to predict permeability (Sanei et al., 2022). In our research, we will compare the predictions from Statoil's model with the results from our machine learning methods .

$$\begin{aligned} KLOGH &= 10^{(2+8\cdot PHIF-9\cdot VSH)} \text{ Hugin Formation} \\ KLOGH &= 10^{(-3+32\cdot PHIF-2\cdot VSH)} \text{ Sleipner Formation} \\ KLOGH &= 10^{(-1.85+17.4\cdot PHIF-3\cdot VSH)} \text{ Skagerrak Formation} \end{aligned} \tag{3.1}$$

3.2. Research Process

Our research methodology (Figure 3.12) follows a systematic approach that integrates data processing, feature engineering, and the application of various machine learning techniques. A key aspect of this approach is the use of cross-validation strategies, including 'leave one well out' and k-fold cross-validation, which are essential for model validation and addressing the challenges of spatial data context inherent in geological datasets (Safaei-Farouji and Kadkhodaie, 2022). Detailed explanations of these strategies and their applications are provided in the subsequent sections.

The initial stage of the workflow involves ensuring the correct alignment of conventional core measurements, such as porosity and permeability, with the well depth as depicted in (Figure 18 , Table 11). These conventional core measurements provide the ground truth data necessary for training and validating the machine learning models.

The next step involves an in-depth examination of various well logs to understand the relationships between different log readings and their implications on permeability. This stage is crucial in identifying useful features for our machine learning models.

This is followed by the application of various feature engineering techniques. One-hot encoding, a method used to convert categorical data into a format that can be provided to machine learning algorithms, is applied on formation tops. Vertical variability features are computed to capture changes in well logs along the depth of the well. Further, neutron density separation calculations are performed as these features can provide valuable insights into porosity and fluid type (Glover, 2018).

After these steps, we apply min-max normalization. Min-max normalization is a data normalization technique that rescales the features to a specific range, typically 0 to 1, which can help improve the

performance of some machine learning algorithms (Kim, 2022).

The final stages of the process involve using machine learning algorithms to predict permeability as a continuous log, using the input features generated through the previous steps. The performance of the different machine learning models is then evaluated using various metrics to determine the most accurate model.

The methodologies described here are applied on two primary datasets used in this research: the Costa dataset and the Volve dataset (Section 3.1). These datasets provide a diverse range of geological conditions, which ensures that our machine learning models are robust and capable of handling different scenarios.

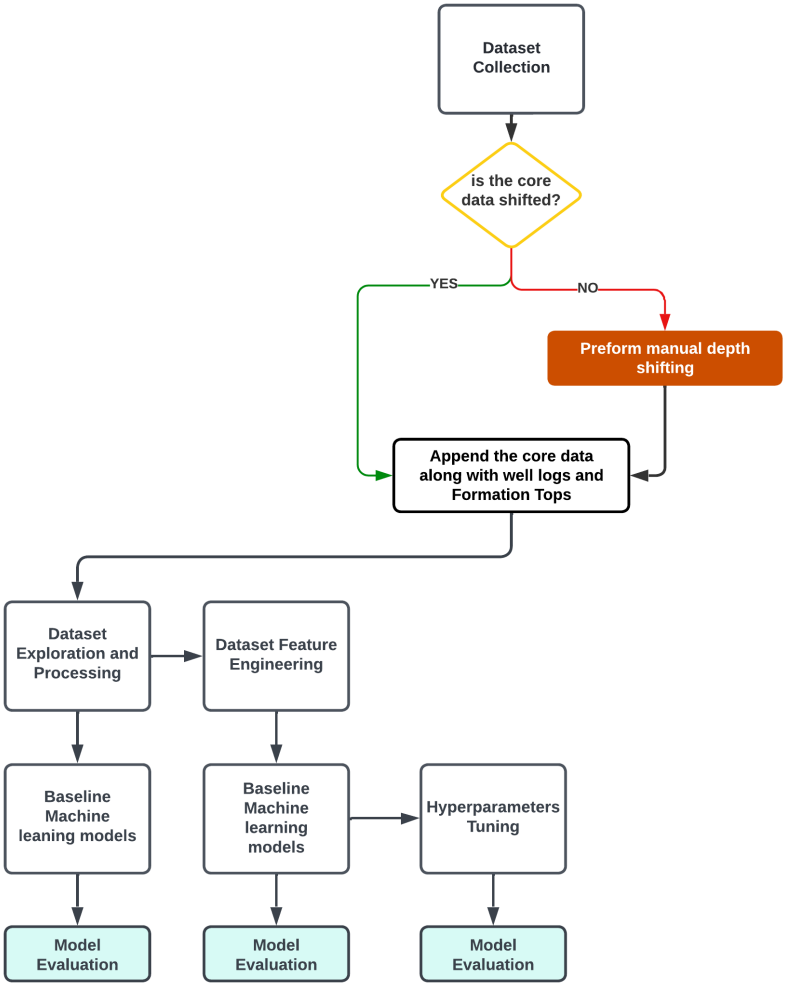


Figure 3.12: The overall process flow for this study.

3.3. Data Preparation

3.3.1. Feature Engineering

One Hot Encoding Formation Tops

One hot encoding, a popular method for encoding categorical data, turns each category into a binary vector. For formation tops, a binary vector is constructed for each top, with each top considered as a distinct category (Han and Kwon, 2021). Suppose there were three formation tops, namely A, B, and C. The binary vectors would be represented as [1, 0, 0], [0, 1, 0], and [0, 0, 1], respectively.

To illustrate the transformation, consider the hypothetical sample in Table 3.13, which includes the gamma ray (GR), bulk density (RHOB), neutron porosity (NPHI), and formation top (A, B, or C). After applying the one-hot encoding, each formation top (A, B, C) is represented by a separate binary vector, as shown in the table.

	Depth (m)	GR (API)	RHOB (g/cm3)	NPHI (v/v)	Formation	Permeability (mD)	Top A	Top B	Top C
0	2000	45.6	2.56	0.23	A	10.1	1	0	0
1	2001	48.2	2.60	0.21	A	9.8	1	0	0
2	2002	46.1	2.55	0.22	A	10.3	1	0	0
3	2003	47.8	2.57	0.20	A	9.5	1	0	0
4	2500	42.3	2.63	0.25	B	15.2	0	1	0
5	2501	43.6	2.62	0.24	B	14.8	0	1	0
6	2502	41.8	2.61	0.26	B	16.0	0	1	0
7	2503	42.7	2.62	0.25	B	15.3	0	1	0
8	3000	40.2	2.66	0.27	C	20.5	0	0	1
9	3001	40.8	2.67	0.26	C	19.8	0	0	1
10	3002	41.3	2.68	0.25	C	19.5	0	0	1
11	3003	40.6	2.66	0.27	C	20.2	0	0	1

Figure 3.13: Example of a data set before and after one-hot encoding. The original Formation variable has been expanded into three separate binary variables: Top A, Top B, and Top C.

One-hot encoding formation tops for the purpose of predicting permeability using well logs enables the machine learning model to factor in the influence of lithology changes on permeability. By converting formation tops into binary vectors, the model can begin to establish a correlation between specific lithology changes and shifts in permeability.

Moreover, the binary representation of formations can help the machine learning algorithm differentiate permeability populations per formation, given that abrupt transitions between formations often denote distinct lithological characteristics. This ultimately aids the machine learning algorithm in better understanding the geological environment.

A drawback of the one hot encoding is that it might considerably raise the dimensionality of the data, particularly if there are numerous categories. This can result in the curse of dimensionality, a phenomenon where the volume of the data grows exponentially with the dimensionality, making the data sparse. This sparsity can lead to overfitting as the model might start learning from the noise in the data rather than the actual patterns; (Verleysen and François, 2005). To address this issue, it is critical to carefully analyze the amount of categories and the potential impact on model performance.

Vertical Variability

Rock formations in proximity to one another are likely to exhibit similar sedimentological and petrophysical characteristics due to shared geological history and depositional environments. Additionally, well logs capture a volume, so local heterogeneity is smoothened. This smoothing effect is much smaller on core samples, so their measurements tend to show more variability at smaller scale. Hence, introducing log measurements above and below could help to better capture the true variability of a formation. Therefore, one feature engineering strategy utilized in this study involves the integration of features from measurements taken immediately above and below a given observation, done iteratively up to 13 nearest neighbors which is 3 meters sampled at a rate of 0.15m.

Specifically, well log readings from above and below each observation were included, considering a predefined sampling rate. For instance, if a reading was taken at 3000 meters, the well log values from the observations at 2999.85 meters and 3000.15 meters were added as features. This approach can aid the model in better predicting permeability by capturing the dependencies and commonalities amongst neighboring observations.

Additionally, quantifying vertical variability can be achieved through another feature engineering technique: calculating forward and backward differences of well log values. The difference between a

reading at one depth and the immediate depth above or below can provide an understanding of the rate of change in petrophysical properties across the rock formation. For example, the forward difference between the well log values from observations at depths of 2999.85 and 3000 meters can reveal variations in the well log values over a unit depth interval. Similarly, the backward difference can give insight into changes in well log values over the previous depth interval. Such variations can be indicative of changes in lithology, pore structure, and consequently, permeability.

Neutron and density Sepration (NDS)

In petrophysical analysis, neutron and density logs are frequently used to calculate rock parameters including porosity and lithology. However, because of the compaction effect, using bulk density logs alone might occasionally produce estimates of these attributes that are imprecise (Crampin, 2008). This problem can be fixed using the petrophysical method of neutron-density separation (NDS), which is frequently included as a feature in permeability prediction models(Fadairo and Awuyo, 2019).

$$\text{NDS} = \frac{(\text{RHOB} - 1.95)}{0.05} - \frac{(0.45 - \text{NPHI})}{0.03} \tag{3.2}$$

Neutron-Density Separation (NDS) is a graphical representation of the distance between the neutron and density curves, calculated using a limestone matrix 3.2 calibration, as used in the COSTA dataset, and visualized using a corresponding limestone scale (Fadairo and Awuyo, 2019). However, this methodology is not directly applicable in the Volve case as it was specifically tailored for carbonate environments. The neutron and density logs are utilized to compute the NDS log, a parameter that cannot be measured directly. Given its ability to capture lithological differences in the formation while remaining unaffected by compaction, NDS calculation proves to be a valuable tool in facies and permeability modeling(Fadairo and Awuyo, 2019).

3.3.2. Normalization

Normalization, while not always necessary for every machine learning algorithm such as random forest, is a prevalent preprocessing step that can help in scaling the input data to a specific range, often between 0 and 1 (Kim, 2022), or sometimes even applying standardization for statistical or optimization reasons resulting in unbounded values . This process helps in ensuring that the model can learn equitably from all features, irrespective of their original scale. In the context of permeability prediction utilizing well logs, the input data comprise gamma ray, resistivity, density, and neutron log measurements. These measurements can significantly vary in scale, making normalization a valuable step to ensure the model can appropriately learn from all features.

The MinMax scaler, a well-known normalizing technique, scales the range of each feature to between 0 and 1 (Bisong, 2019). This normalization method is implemented by subtracting the feature's minimum value and then dividing by its range (i.e., the difference between the maximum and minimum values of the feature). Below is the formula for the MinMax scaler (Pedregosa et al., 2011):

$$x_{norm} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{3.3}$$

However, it is vital to be aware that the MinMax scaler is sensitive to data outliers (Choudhury, 2020). Outliers can significantly impact the scaling process, leading to a skewed normalization of other data points. Therefore, appropriate handling of outliers is critical when applying the MinMax scaler. This could involve replacing outliers with the closest non-outlier value or removing them entirely. It should be noted, though, that identifying outliers can be a complex process and must be approached with caution.

3.4. Machine Learning

3.4.1. Support Vector Regression

Support Vector Machine (SVM) regression is a machine learning algorithm that constructs an optimal hyperplane to predict output variables. This hyperplane is defined by the data points, known as

support vectors, that lie closest to the decision boundary (Kramer, 2016). The SVM regression algorithm aims to both maximize the margin between these support vectors and minimize the error between predicted and actual values. The concept of a margin here is a range around the hyperplane where errors are tolerated, and the balance between minimizing the error of data points falling outside this range and maximizing the number of data points falling inside the range is represented by the epsilon-insensitive loss function in SVM regression (G. Zhang et al., 2021) as illustrated in (Figure 3.14)

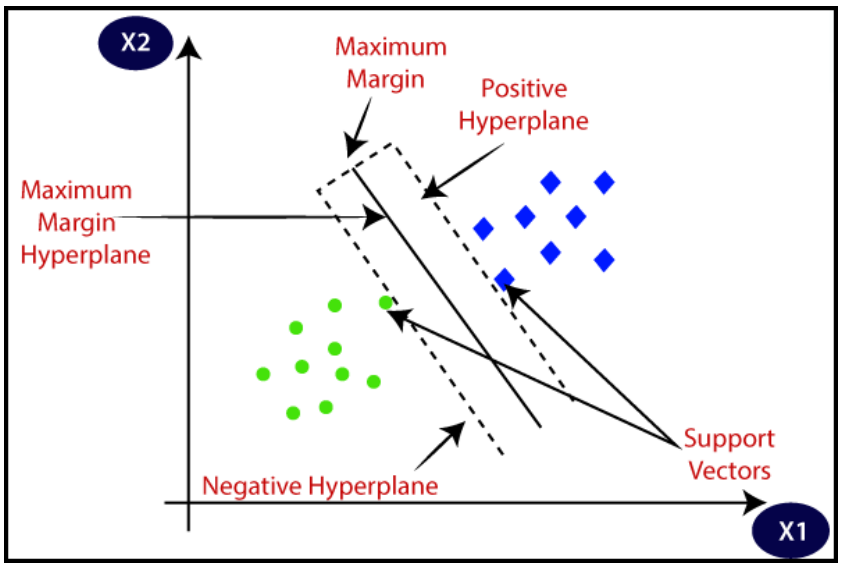


Figure 3.14: The figure depicts the SVM algorithm's objective, which is to establish an optimal decision boundary, known as a hyperplane, to efficiently classify n-dimensional space. SVM determines this hyperplane using select extreme data points termed as support vectors, leading to the name Support Vector Machine. The illustration showcases two distinct categories separated by such a hyperplane. (Javapoint, 2020)).

The SVM regression algorithm employs a kernel function to transform the input data into a higher-dimensional feature space, allowing the algorithm to create a linear decision boundary in the new feature space. The most commonly used kernel functions are linear, polynomial, and radial basis function (RBF) (Vapnik, 1998).

Once the data is transformed, the algorithm selects a subset of the training data, the support vectors, and identifies the coefficients of the hyperplane that most effectively separates the data. This is achieved by reducing the sum of squared errors between the anticipated and actual values and maximizing the margin between the support vectors. The hyperplane's coefficients are optimized using a technique such as gradient descent (G. Zhang et al., 2021).

The SVM regression algorithm employs a regularization parameter known as C during the training process. This parameter is used to balance between maximizing the margin and decreasing the error. A smaller value of C results in a wider margin but may increase mistakes on the training data, whereas a larger value of C results in a narrower margin but may overfit the data (Brereton and Lloyd, 2010).

After training, the SVM regression model can be applied to new, unknown data to predict the output variable. The kernel function is used to translate the input data into the same feature space, and the anticipated output is generated by assessing the input data's position relative to the hyperplane (G. Zhang et al., 2021).

3.4.2. Random Forest Regression

the random forest regression algorithm works by building a decision tree based on randomly selecting a subset of features from the training data. The process of recursively dividing the data into smaller subsets depending on the values of the selected features continues until a stopping requirement,

such as a maximum tree depth or a minimum amount of data points in a leaf node, is met (Feng et al., 2021) . This process is repeated to generate several decision trees, and the final prediction is obtained by averaging the predictions of the entire forest as illustrated in (Figure).

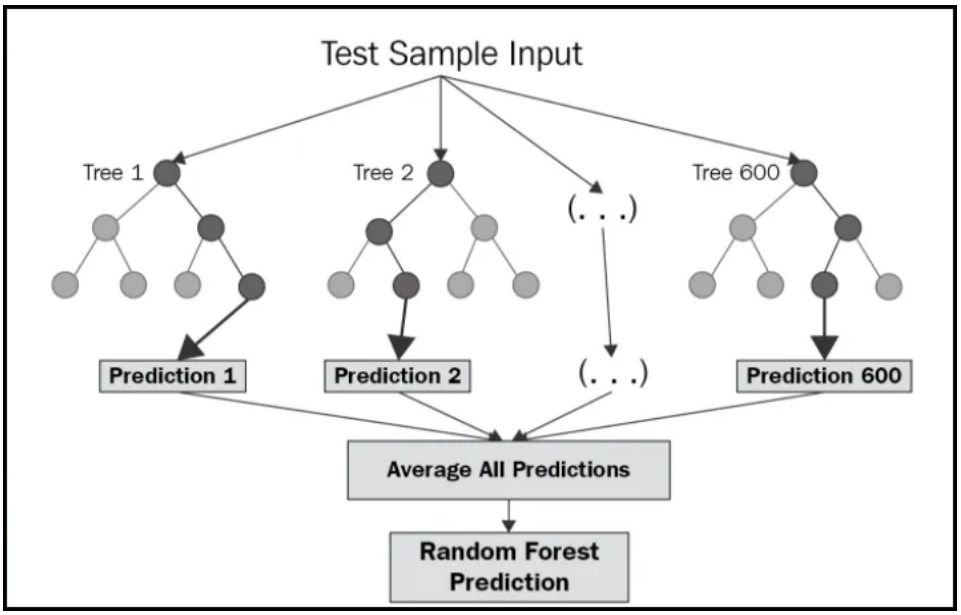


Figure 3.15: The figure illustrates the architecture of a Random Forest, highlighting that the trees function concurrently without interplay. A Random Forest builds multiple decision trees during training and predicts based on the average output value from these trees for regression tasks. (Chaya, 2022)).

Its capacity to handle complex and nonlinear interactions between input variables and output variables, as well as noisy and missing data, is the algorithm's greatest strength. In addition, the method is capable of identifying the most crucial input factors for the prediction, so providing vital insight into the underlying geologic processes (Feng et al., 2021).

To optimize the performance of a random forest regression model, its hyperparameters must be properly tuned. Hyperparameters are parameters that are established prior to the learning process and have a substantial impact on the algorithm's behavior and performance (Feng et al., 2021) .

The following hyperparameters are crucial for random forest regression (Pedregosa et al., 2011):

- n_estimators is the number of decision trees there are in the forest. Although adding more trees can raise the model's accuracy, doing so also lengthens the computing process.
- max_depth: represents the maximum number of levels a decision tree can contain. Raising the maximum depth can make the model more complex and possibly overfit the data, whilst decreasing it can make it too simple and underfit the data.
- min_samples_split: The minimum amount of data points required to split a decision tree node. Raising this parameter prevents overfitting, whilst decreasing it makes the model more flexible and may result in overfitting the data.
- max_features: maximum amount of features that can be used to split a node in the decision tree. This parameter can be increased to make the model more diverse, while decreasing it can make the model more similar and possibly overfit the data.

The appropriate values for these hyperparameters are dependent on the data set and problem being addressed. Thus, a detailed process of hyperparameter tuning is required to identify the ideal values for a given dataset.

3.4.3. Extreme Gradient Boosting (XGBoost) Regression

Extreme Gradient Boosting (XGBoost) is a powerful machine learning algorithm predominantly used in regression tasks. XGBoost operates by constructing an ensemble of decision trees and iteratively refining the model by training each tree to correct the errors made by the preceding trees (Chen and Guestrin, 2016) as illustrated in (Figure 3.16).

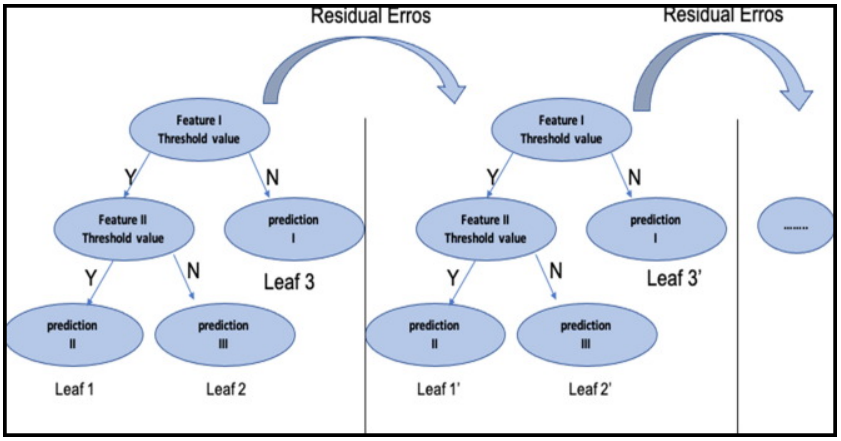


Figure 3.16: The figure demonstrates the XGBoost mechanism, an implementation of gradient boosting machines. It showcases the formation of multiple decision trees, acting as weak learners, with each tree iteratively focusing on the residuals of the preceding tree to optimize the model's accuracy. This ensemble approach seeks to minimize both loss and regularization objectives. (Ibrahem Ahmed Osman et al., 2021)).

The mean squared error (MSE) between the predicted and actual values of the target variable is the objective function that is minimized during the training process.

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \Omega(f) \tag{3.4}$$

Where $\hat{y}_i$ represents the predicted value of the i-th sample, y_i represents the actual value, and $\Omega(f)$ is a regularization term that governs the model's complexity. Similar to the XGBoost algorithm, XGBoost Regressor provides a number of advanced capabilities, including the ability to handle missing values, sparse data, and parallel processing.

In addition, XGBoost Regressor permits the specification of a custom objective function, allowing the model to be further optimized for specific challenges. Overall, XGBoost Regressor is a potent and efficient technique for regression tasks that is frequently utilized in real-world applications due to its scalability and speed (Wade, 2020).

There are numerous significant hyper-parameters in XGBoost that can be adjusted to increase the model's performance (Wade, 2020). Among the most significant are the following:

- learning_rate controls the magnitude of the algorithm's parameter updates. The process will converge more slowly, but the resulting model will be more accurate if the learning rate is decreased.
- n_estimators regulate the amount of trees in an ensemble. Increasing the number of trees will enhance the complexity and accuracy of the model, but will also raise the computing cost.
- colsample_bytree: This variable controls the percentage of characteristics that are utilized to train each tree. A smaller number can reduce overfitting, but at the expense of the model's capacity to capture key traits.
- gamma: This variable determines the minimum loss reduction necessary for a split. Higher values make the model more conservative and can help to prevent overfitting.

These are some of the most essential hyper-parameters in XGBoost, however there are more that can be tweaked based on the task at hand. Using techniques such as as grid search or random search

to explore a variety of hyper-parameter settings and evaluate the performance of the model on a validation set is the most effective method for determining the ideal hyper-parameters.

3.4.4. Light Gradient Boosting Machine (LightGBM) Regression

LightGBM is an open-source gradient boosting framework that is used extensively for regression analysis in various domains, including geoscience and petrophysics(S. Zhang et al., 2022). It is specifically designed to efficiently handle datasets with a large number of instances, potentially reaching up to millions, although the concept of ”large” may vary depending on the context and task. However, LightGBM maintains accuracy and computational efficiency across a wide range of dataset sizes. The LightGBM regressor constructs a series of decision trees iteratively, where each tree attempts to fit the residual errors of the previous tree (Ke et al., 2017).

Both LightGBM and XGBoost utilize the method of fitting the residual error in their gradient boosting algorithms. However, a distinguishing feature of LightGBM is its adoption of the leaf-wise growth strategy, as opposed to the level-wise approach commonly used in XGBoost (Saha, 2023). This difference in tree growth strategies can significantly impact the model's performance and complexity as illustrated in (Figure 3.17).

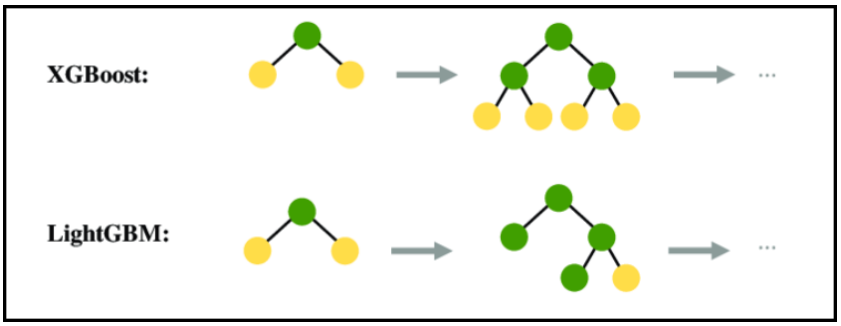


Figure 3.17: Comparative illustration of tree growth strategies between LightGBM (leaf-wise growth) and XGBoost (level-wise growth). This distinction highlights the unique approach of each algorithm in handling data splits and optimizing model performance. (Kumar, 2022)).

There are numerous significant hyper-parameters in LightGBM that can be adjusted to increase the model's performance . Among the most significant are the following:

- num_leaves: This hyperparameter regulates the maximum number of leaves per decision tree. A greater value of num leaves can result in a more complex model that may overfit the data, whereas a smaller value can lead to underfitting.
- learning_rate: This hyperparameter determines the step size for each iteration of the boosting procedure. A higher value of learning rate may result in faster convergence but overshoot the optimal solution, whereas a lower value may result in slower convergence but greater accuracy.
- max_depth: represents the maximum number of levels a decision tree can contain. Raising the maximum depth can make the model more complex and possibly overfit the data, whilst decreasing it can make it too simple and underfit the data.
- min_child_samples: This hyperparameter determines the minimum amount of samples necessary to create a new leaf node. A higher value may lead to underfitting, whereas a lower value may result in overfitting.
- feature_fraction: This hyperparameter determines the proportion of features selected at random at each decision tree split. A greater value can result in a more sophisticated model that may overfit the data, whereas a smaller value can lead to underfitting.
- min_gain_to_split: This hyperparameter determines the minimal gain required to divide each decision tree. A higher value may lead to underfitting, whereas a lower value may result in overfitting.

- bagging_fraction: This hyperparameter determines the proportion of samples taken at random throughout each iteration of the boosting procedure. A greater value can result in a more sophisticated model that may overfit the data, whereas a smaller value can lead to underfitting.
- bagging_freq: This hyperparameter affects the frequency of bagging throughout each boost iteration. A greater value can result in a more sophisticated model that may overfit the data, whereas a smaller value can lead to underfitting.
- n_estimators regulate the amount of trees in an ensemble. Increasing the number of trees will enhance the complexity and accuracy of the model, but will also raise the computing cost.

3.4.5. Hyperparameter Tuning

The selection of optimal hyperparameters for machine learning models, crucial for improving model performance, was carried out using random grid search. This method randomly selects hyperparameters from a predefined range and evaluates them over a specified number of iterations(Kavzoglu and Teke, 2022). The ranges specified in (Table 3.1) correspond to the minimum and maximum values of uniform distributions used for hyperparameter tuning, with some distributions on a log scale (e.g., for reg_lambda). Random grid search is preferred over exhaustive grid search due to its efficiency in computation time, while still ensuring a broad exploration of possible parameters. The ranges of the hyperparameters explored for each algorithm are available in (Table 3.1) below. The final chosen hyperparameters after the tuning process can be found in (Appendix 6).

Algorithm	**Hyperparameter**	**Range**
XGBRegressor	colsample_bytree	0.8 to 1.0
	gamma	0.2 to 0.6
	learning_rate	0.1 to 0.5
	max_depth	10 to 110
	min_child_weight	10 to 30
	n_estimators	50 to 200
LGBMRegressor	bagging_fraction	0.8 to 1.0
	bagging_freq	1 to 5
	feature_fraction	0.2 to 0.6
	learning_rate	0.1 to 0.3
	min_child_samples	5 to 20
	min_split_gain	0.1 to 0.4
	n_estimators	50 to 150
	num_leaves	2 to 10
	reg_alpha	0.1 to 1
	reg_lambda	1×10^{-7} to 1×10^{-5}
RandomForestRegressor	max_depth	10 to 110
	max_features	'sqrt', 'log2'
	min_samples_leaf	1 to 5
	min_samples_split	5 to 15
	n_estimators	200 to 600
	bootstrap	True, False

Table 3.1: Hyperparameter Ranges for Each Algorithm

3.5. Cross Validation Techniques

Cross-validation is a fundamental technique used to evaluate a model's performance and prevent overfitting. It includes partitioning the dataset into subsets, training the model on X-1 subsets, and predicting on the remaining subset to estimate the generalization error. This process helps in understanding how well the model will perform on unseen data. Common types of cross-validation include k-fold cross-validation and leave-one-well-out cross-validation.

3.5.1. K-fold cross-validation

K-fold cross-validation (Pedregosa et al., 2011) divides the dataset into k equal-sized or almost equal-sized folds, depending on the distribution of the data. This process is repeated k times, with each fold serving as the test set just once. The results of each fold are then averaged to determine the model's performance. This strategy ensures that each sample is used for both training and testing, and reduces the possibility of overfitting by utilizing several folds for validation.

3.5.2. Leave-one-well-out cross-validation

Leave-one-well-out Cross-validation (LeaveOneGroupOut) (Pedregosa et al., 2011) is a form of k-fold cross-validation that is frequently used in geoscience applications where spatial autocorrelation may be present in the data. The data is divided based on a grouping variable, such as the well name or location, and the model is trained on all groups but one before being evaluated on the remaining group.

This method is repeated for each group in the dataset, and the results are averaged to provide an estimate of the model's performance. This ensures that the model does not exhibit bias towards any specific well or location and is capable of more accurately capturing spatial variance in the data.

The choice between k-fold and leave-one-well-out cross-validation is not just about their respective advantages and disadvantages, but also about the practical application of the model. If an algorithm is developed to interpolate between many permeability samples along a well, then k-fold cross-validation may be a better fit as it effectively handles randomly scattered large datasets with computational efficiency. Conversely, in an exploration context where the goal might be to predict values for a new well, the Leave-one-well-out method provides a more accurate performance estimate, especially when spatial autocorrelation is present in the data. Thus, the selection between these validation methods depends on both the characteristics of the dataset and the specific research issue at hand. In this study, both methods will be used to evaluate the generalizability of these models across different configurations.

3.6. Model Evaluation

3.6.1. R^2 Score and Mean Absolute Error

The performance evaluation of our machine learning models was based on two fundamental metrics, namely the coefficient of determination (R^2 score) and the mean absolute error (MAE).
The R^2 score is a quantitative metric used to evaluate the degree to which the variability in the dependent variable can be explained by the independent variable(s) (Lewinson, 2023).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3.5}$$

where SS_{res} is the sum of squares of residuals and SS_{tot} is the total sum of squares. An R^2 score of 1 indicates that the model perfectly predicts the target variable, while an R^2 score of 0 suggests that the model does not explain any of the variation in the target.
The MAE, on the other hand, is a measure of how close predictions are to the actual outcomes. It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. MAE is computed as (Lewinson, 2023):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.6}$$

where y_i represents the actual values, $\hat{y}_i$ the predicted values, and n is the number of observations. Lower MAE values indicate better model performance as they signify smaller differences between predicted and actual values.
Together, the R^2 score and MAE provide a comprehensive overview of model accuracy and predictive power.

3.6.2. Shapley Additive Explanations for Model Interpretation

In our research, we employed a tool called Shapley Additive Explanations (SHAP)(Mitchell et al., 2020) to help us understand our machine learning model's predictions. This tool assigns each feature in our dataset a SHAP value, which can be considered as the contribution of that feature towards the prediction.
The SHAP values are calculated in a way inspired by game theory. The model's prediction for a particular instance is expressed as an additive combination of the SHAP values. This is mathematically represented by:

$$f(x_i) = \varphi_0 + \sum_{j=1}^{n}\varphi_{i,j}z'_{i,j} \tag{3.7}$$

Here, $x_{i,j}$ is the j^{th} feature for the i^{th} instance. The symbol n is the total number of instances. The function $f(x_i)$ is the predicted value for instance x_i. The term φ_0 is the base value of the model, representing the average prediction across the entire dataset without considering specific features. It serves as a baseline prediction, obtained by averaging the model's output for all possible feature

combinations (Lundberg, 2020). The term $\varphi_{i,j}$ is the SHAP value for feature $x_{i,j}$, and z'_{ij} is a binary vector that represents the presence or absence of feature combinations.
The SHAP value for a feature $x_{i,j}$ is calculated as follows:

$$\varphi_{i,j} = \sum_{S \subseteq N\, x_{i,j}} \frac{|S|!(n-|S|-1)!}{n!}[f_{x_i}(S \cup x_{i,j}) - f_{x_i}(S)], j \geq 1 \tag{3.8}$$

In this equation, N is the set of all features, and S is a subset of N. The function $f_{x_i}(S)$ gives the predicted value using the instances in subset S.
Using SHAP, we were able to better understand our model's behavior. It helped us see how different features contribute to the predictions, and also gave us insights into which features were the most significant. This made our machine learning model more interpretable and reliable.

4

Results

4.1. The Costa field

4.1.1. Train/Test Splitting Strategy

We utilized two different methods to divide our data into training and testing wells. The first approach involved randomly picking 3 wells for testing and the remaining 14 wells for training, as indicated in (Table 4.1).

Training Wells	Testing Wells
HW-10	HW-26
HW-24	HW-29
HW-25	HW-7
HW-27	
HW-28	
HW-3	
HW-30	
HW-31	
HW-32	
HW-4	
HW-5	
HW-6	
HW-8	
HW-9	

Table 4.1: Wells used for Training and Testing

The second approach was a more manual process where we split the data into three clusters based on their geographical locations, as demonstrated in (Figure 4.1). These location-based clusters were then used as criteria for training and testing. In this method, wells from two designated clusters were used for training, while the remaining cluster was used for testing purposes.Furthermore, to evaluate the model's performance within each individual cluster, we adopted the 'leave one well out' cross-validation technique. This method facilitated the evaluation of the model's robustness within a geographically constrained section using less data. In other words, each well within a cluster was alternately treated as a blind well, thereby enabling us to benchmark the performance of the model in scenarios with limited data availability.

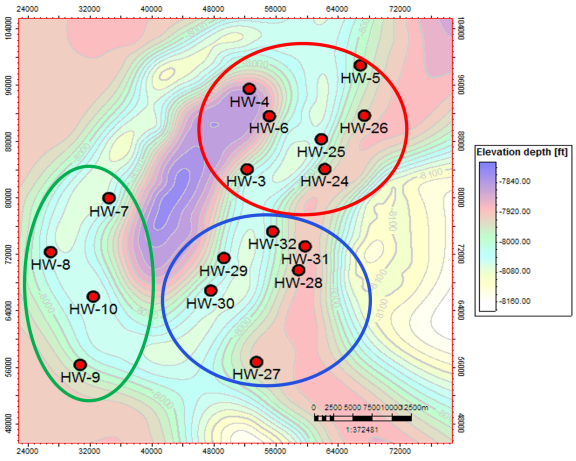


Figure 4.1: Contour map showcasing the top of the Upper Kharaib Member with marked well locations, based on data from (Costa Gomes et al., 2022). The map also highlights the geographical clustering of wells into three distinct regions: northeast, southeast, and southwest, illustrating the spatial distribution of data utilized for model validation.

4.1.2. Data Exploration

Exploratory Data Analysis (EDA) is a crucial step where we investigate and break down our data sets to highlight their main statistical characteristics, often using data visualization methods. EDA helps us figure out how to handle our data to get the answers we need, which makes it easier to spot patterns, identify irregularities, verify our hypotheses, or test our assumptions.

To aid in this, we compiled the statistical characteristics of the data points used in our machine learning models in (Figure 4.2). This summary provides a snapshot of the data distribution across all datasets used in our study. Importantly, for the sake of comprehensive analysis and reproducibility, all features available in the datasets were utilized.

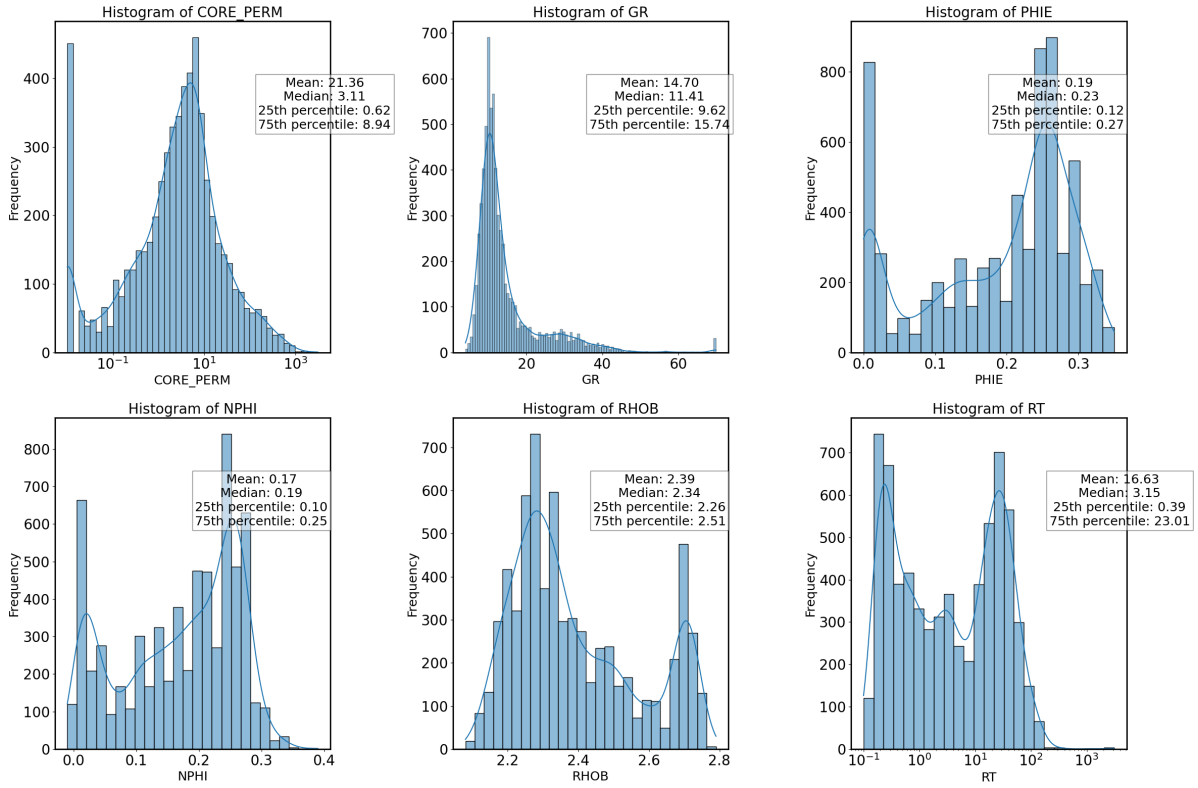


Figure 4.2: Summary statistics of the Costa dataset.

Moreover, this step is critical as it provides insights that inform the next steps of feature engineering and model building. The dataset comprises a total of 6059 data points gathered from seventeen wells.

4.1.3. Features Evaluation

In general, how a model performs largely depends on the input features we use for its training and validation (Janiesch et al., 2021). It is critical to pinpoint the most relevant features to predict permeability. The quality of features is important because if they are not up to the mark, they can cause redundancy and make the model harder to interpret. In our study, we evaluated the correlation between different features and log-transformed permeability using the Pearson correlation coefficient (p), which specifically measures linear relationships (Cohen et al., 2009). It's worth noting that this method may not capture non-linear relationships, but we chose to use all features in our model to ensure a comprehensive approach.

Features with p values around or greater than ±0.6 demonstrated a moderately strong relationship with log-transformed permeability, serving as our threshold for relevance. Looking at specific features, PHIE, NPHI, and RHOB exhibited the strongest correlations with p values of approximately 0.72, 0.68, and -0.68, respectively. GR displayed a moderate correlation with a p value of around -0.43. Despite their importance in identifying hydrocarbon pay zones, NDS and log10RT demonstrated a weak correlation with log-transformed permeability, with p values of around -0.23 and -0.06 respectively. Nevertheless, all features were used in the final model, as shown in (Figure 4.3).

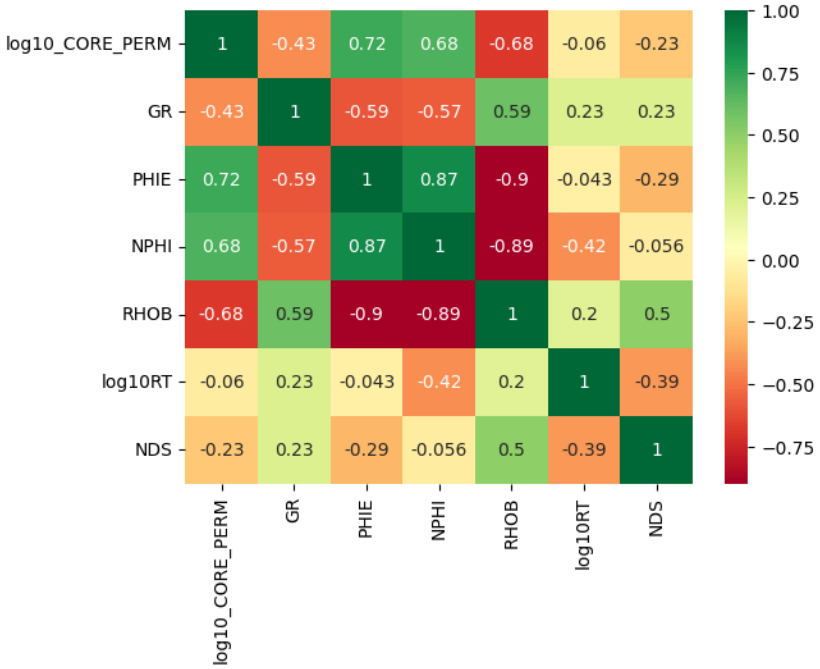


Figure 4.3: Pearson correlation matrix for the Costa dataset. The color scheme indicates the nature of correlations: green represents positive correlations, while red signifies negative correlations. The intensity of the color indicates the strength of the correlation, with darker shades representing stronger relationships.

4.1.4. Leave One Well Out Validation Results

In this section, the model results acquired from the Costa dataset are discussed, revealing a noticeable progression of enhancement through the stages of the study (Figure 4.4). Beginning with the base model performance, where no feature engineering is applied, it is clear that the SVM model provides the most accurate predictions, demonstrated by a mean R^2 score of 0.368. This indicates that, among the base models, SVM was most efficient in capturing the underlying relationships in the Costa dataset.

Significant improvements in all the model performances were observed with the implementation of feature engineering. The models which benefited the most from this were the Random Forest and

SVM (Figure 4.4a). In particular, Random Forest demonstrated the most significant increase in mean R^2, improving from 0.281 to 0.432. This indicates that the engineered features were effectively capturing the underlying data structures, subsequently enhancing the Random Forest's ability to predict core permeability from well logs.(Figure 4.4 b).
Further improvements in the models' predictive capabilities were observed upon tuning the hyperparameters of the feature-engineered models (Figure 4.4c). Remarkable improvements were seen in the mean R^2 scores of the tuned LightGBM and XGBoost models. In particular, the tuned LightGBM model yielded the highest mean mean R^2 of 0.510 among all the models, suggesting that with the appropriate hyperparameters and feature engineering, the LightGBM model provided the most accurate permeability predictions for the Costa dataset. This highlights the importance of model optimization in machine learning-based geology tasks.

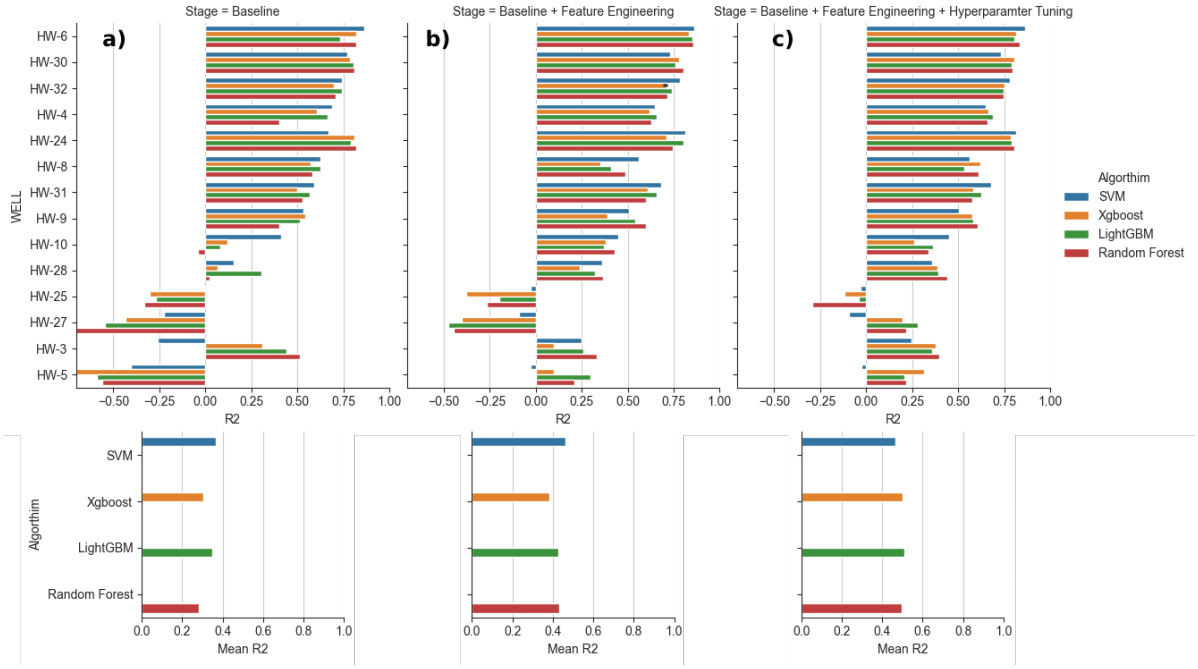


Figure 4.4: Comparative Validation Results at Various Processing Levels.
a) Displays R^2 scores for individual wells using baseline models without feature engineering. The bottom section provides the mean R^2 for each algorithm.
b) Illustrates R^2 scores for each well when employing baseline models that incorporate feature engineering. The corresponding mean R^2 per algorithm is denoted below.
c) Represents R^2 scores per well, obtained from baseline models that both implement feature engineering and undergo hyperparameter tuning. Notably, the mean R^2 for each algorithm is indicated at the base. It's worth noting that in this figure, a single well exhibits a negative performance.

Throughout the study, certain wells exhibited a significant transformation in their performance, particularly when evaluated using specific algorithms. For instance, HW-10, when assessed using the Random Forest model, demonstrated a marked improvement. It began with an R^2 score of -0.0422 during the baseline stage, but after the integration of feature engineering and hyperparameter tuning, its score surged to 0.3377. Similarly, HW-27, when evaluated with the LightGBM model, transitioned from an initial R^2 score of -0.5447 to a commendable 0.2820 post-optimization. Lastly, HW-5, under the Xgboost model, showcased a remarkable shift from an R^2 score of -0.8357 to 0.3160 after the application of the aforementioned methodologies. These substantial improvements emphasize the profound impact of feature engineering and model optimization in enhancing the predictive capabilities of machine learning models.
These significant improvements in model performance through feature engineering and hyperparameter tuning demonstrate the effectiveness of these specific methods in refining predictive models in geology. It becomes clear from this study that when feature engineered and hyperparameter tuned, machine learning models can assist in predicting core permeability from well logs in carbonate rocks, like those found in the Costa dataset. This further emphasizes the potential of machine learning as a

useful tool in petrophysics, and future studies could explore other optimization techniques to further enhance prediction accuracy. The validation wells plots can be found in (Appendix 6)

4.1.5. Sensitivity Analysis of Vertical Variability

The sensitivity analysis of the vertical variability feature engineering technique, as shown in (Figure 4.5), displays a significant influence on the performance of all four machine learning models. The practice of integrating values from above and below the data points into the model seems to enhance the model's predictive power. This implies that spatial continuity and stratigraphic consistency are critical factors when predicting core permeability from well logs.

From the results, it is observed that as vertical variability increases (from 0 to 3 meters), the mean R^2 value also rises, indicating a better model performance. Of particular note, the Support Vector Regression (SVR) model benefited the most from this feature engineering technique, as evidenced by a significant rise in mean R^2 from 0.3891 (with no vertical variability) to 0.4941 (with 3 meters of vertical variability). This improvement suggests that by incorporating vertical variability, the SVR is better able to capture the stratigraphic dependencies in the permeability data.

In addition, the Random Forest Regressor and the Light Gradient Boosting Machine showed improved results with the incorporation of vertical variability, further validating its effectiveness. However, the Extreme Gradient Boosting model seems to benefit less from this, possibly due to the model's inherent configuration not being able to fully utilize the added features. This highlights that while vertical variability improves model performance in general, the extent of improvement is model-specific.

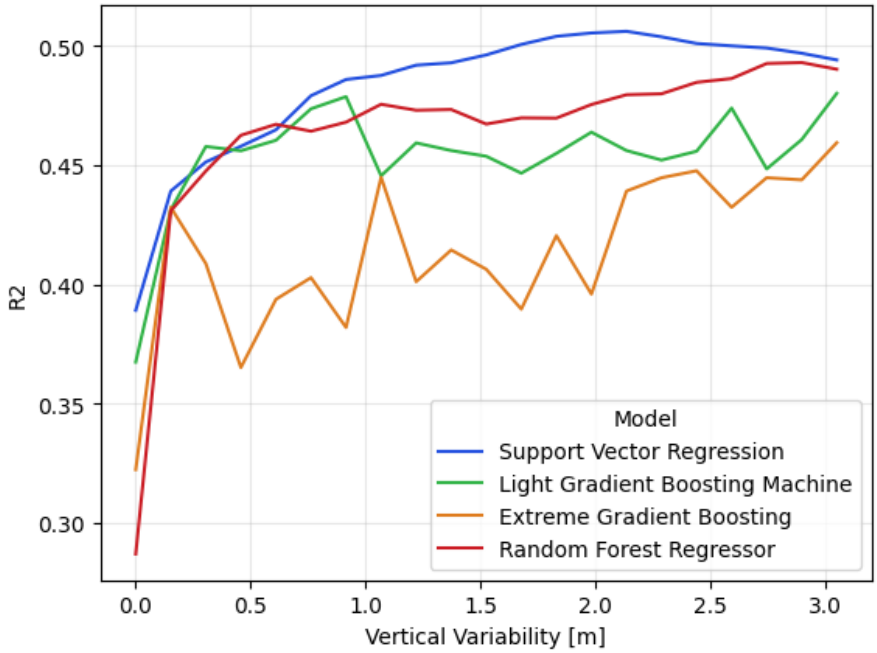


Figure 4.5: Vertical Variability Sensitivity Analysis. The analysis incorporates well log values from up to 3 meters above and below, equivalent to approximately 13 measurements given a 0.15-meter sampling rate. Each line is colored based on a specific ML algorithm, illustrating the enhancement in model performance as more vertical variability is introduced into the dataset.

In conclusion, the incorporation of vertical variability as a feature not only improves prediction accuracy, but also makes these models more physically meaningful by reflecting the inherent geological principle that rock properties do not vary abruptly, but rather gradually along depth. Therefore, vertical variability, as a feature engineering technique, proves to be an effective approach to capture stratigraphic continuity in petrophysical prediction tasks.

4.1.6. Blind testing Results on The Test Wells

Blind tests play a crucial role in evaluating the generalizability of machine learning models. In this study, blind tests refer to wells that were not part of the training dataset. By testing on these wells, we

aim to understand how well the feature-engineered and hyperparameter-tuned models can predict unseen data. The results of these tests are visualized as well logs in Figures 4.6, 4.7, and 4.8. An encouraging observation is the average R^2 value, which is above 0.6 as depicted in Figure 4.9. This suggests that our refined models are not only fitting the training data but are also generalizing well to new, unseen data. Each result is color-coded by the specific ML model used, emphasizing that all results presented are from the feature-engineered and hyperparameter-tuned models.

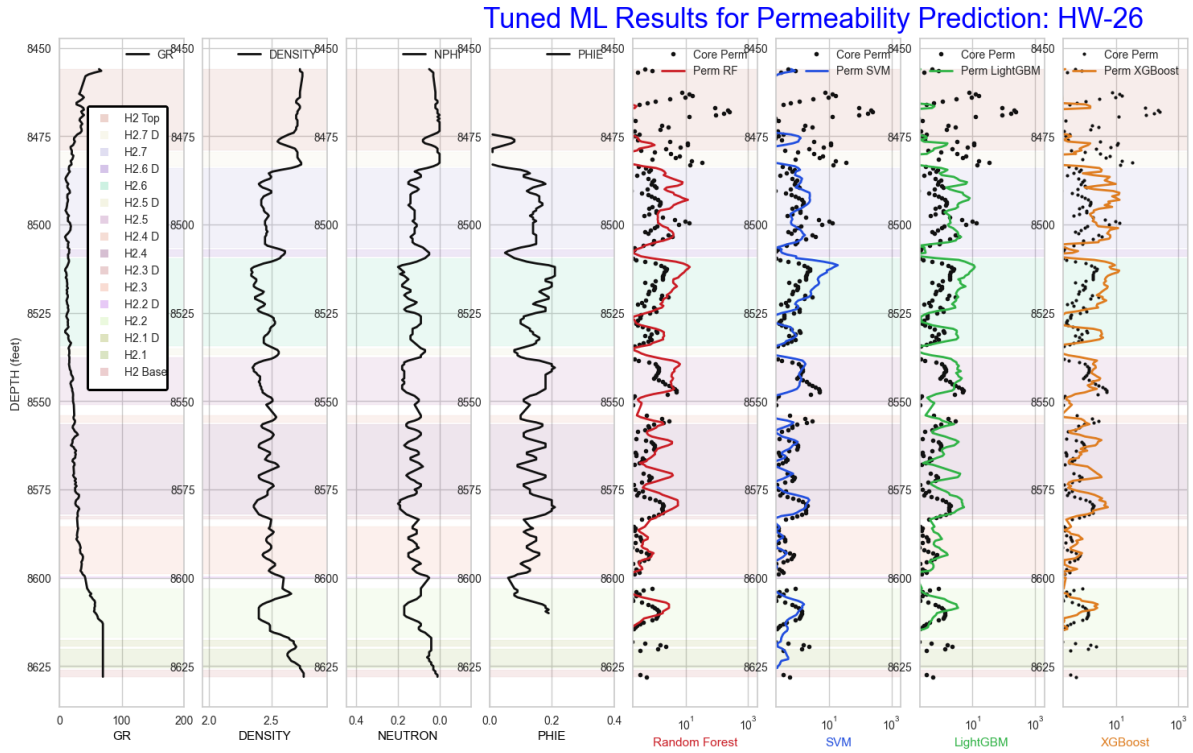


Figure 4.6: Prediction Results for Blind WELL: HW-26. The well log view displays GR, DENSITY, NEUTRON, and calculated porosity (PHIE). Predicted permeability is shown in varying colors, with each color representing a specific ML algorithm. Dots signify the actual core-measured permeability, and the background represent the reservoir zone.

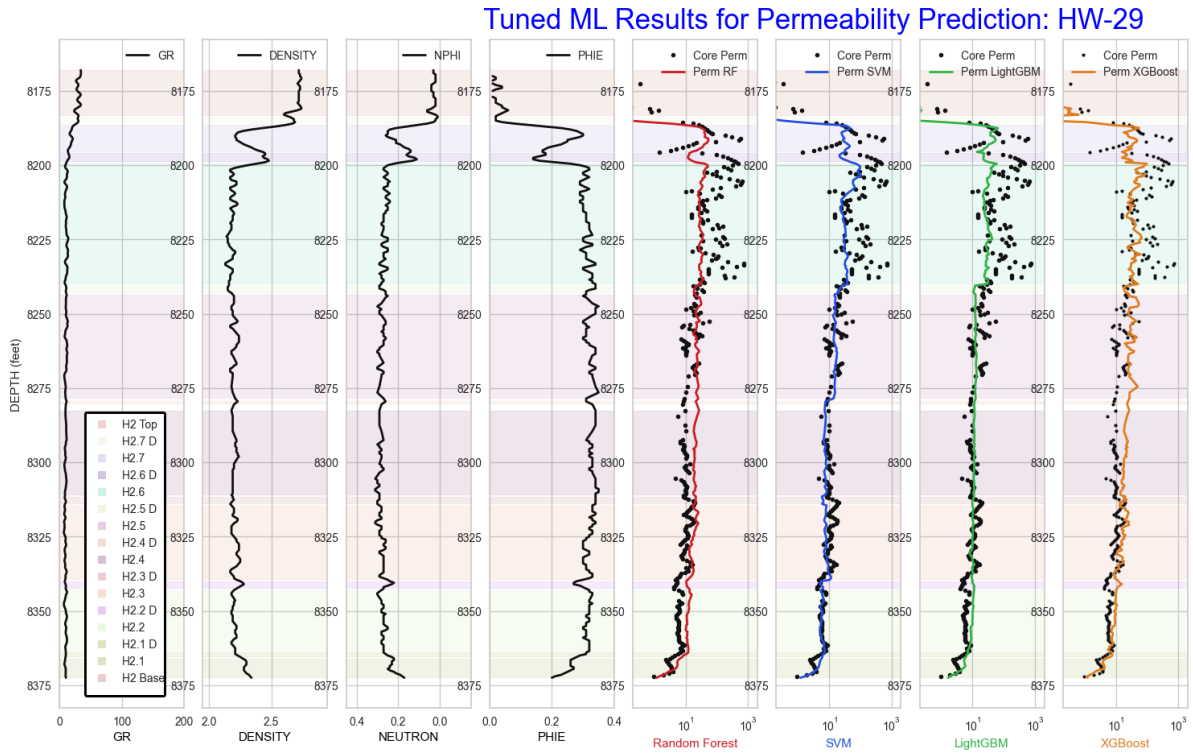


Figure 4.7: Prediction Results for Blind WELL: HW-29. The well log view showcases GR, DENSITY, NEUTRON, and calculated porosity (PHIE). Predicted permeability values are color-coded by the ML algorithm used, while dots highlight the core-measured permeability, and the background represent the reservoir zone.

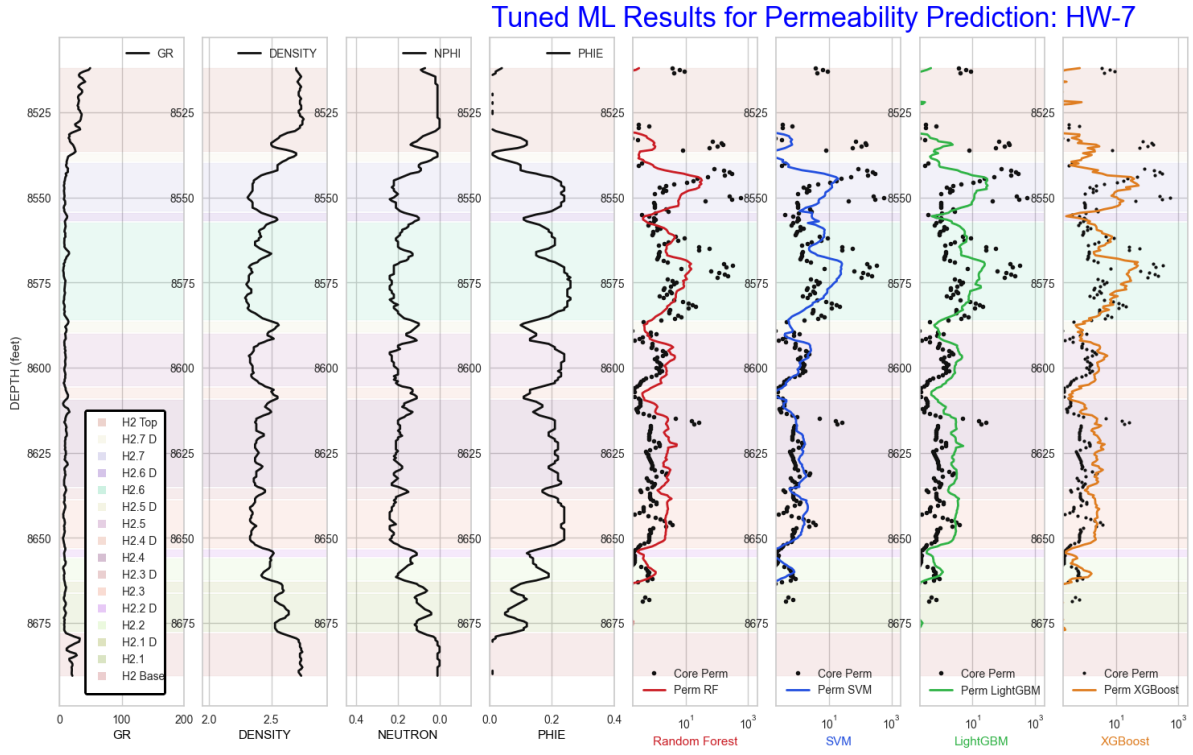


Figure 4.8: Prediction Results for Blind WELL: HW-7. Displayed are the well log views for GR, DENSITY, NEUTRON, and calculated porosity (PHIE). Predicted permeability is color-coded based on the ML algorithm, with dots signifying the actual core-measured permeability, and the background represent the reservoir zone.

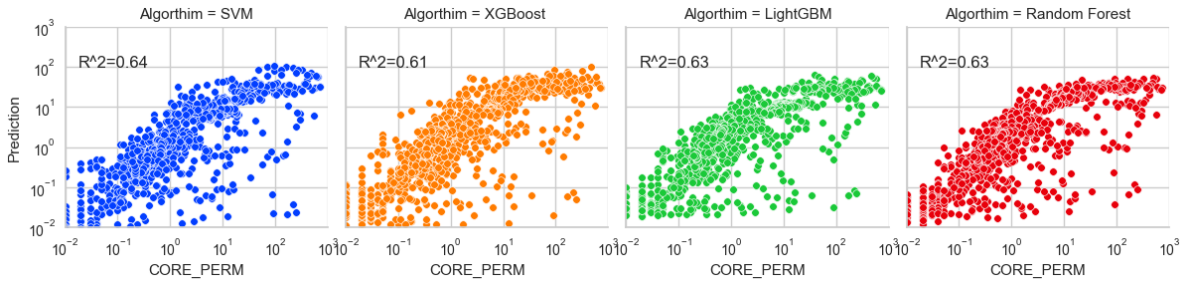


Figure 4.9: Scatter plots of Predicted Permeability vs. Measured Permeability for the Blind Test wells, color-coded by the ML algorithm. The plot provides a direct comparison between model predictions and actual measurements, emphasizing the model's accuracy and generalizability.

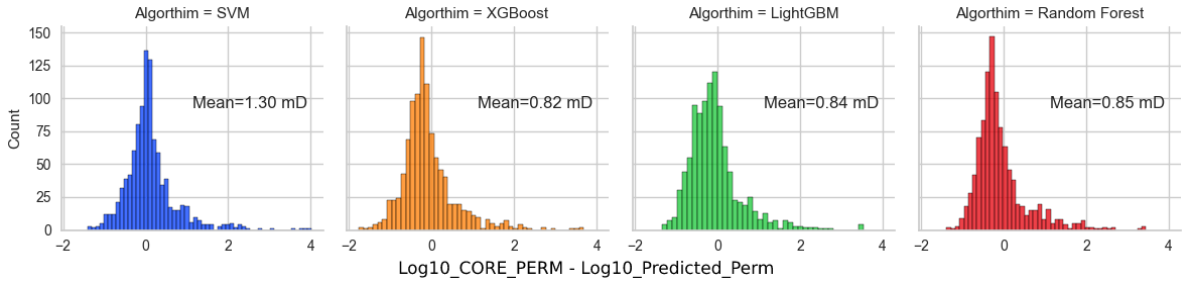


Figure 4.10: Histograms displaying the absolute error between Predicted Permeability and Measured Permeability for the Blind Test wells. The results are color-coded based on the ML algorithm used

4.1.7. Geographical Location Clustering Results

This study employed a validation technique that divided the carbonate platform into three distinct geographical clusters: northeast (Cluster1), southeast (Cluster2), and southwest (Cluster3). These clusters provide a significant testing ground for evaluating the generalization capabilities of machine learning models under geological variations, as illustrated in (Figure 4.11). The findings have substantial implications for the predictive reliability of these models.

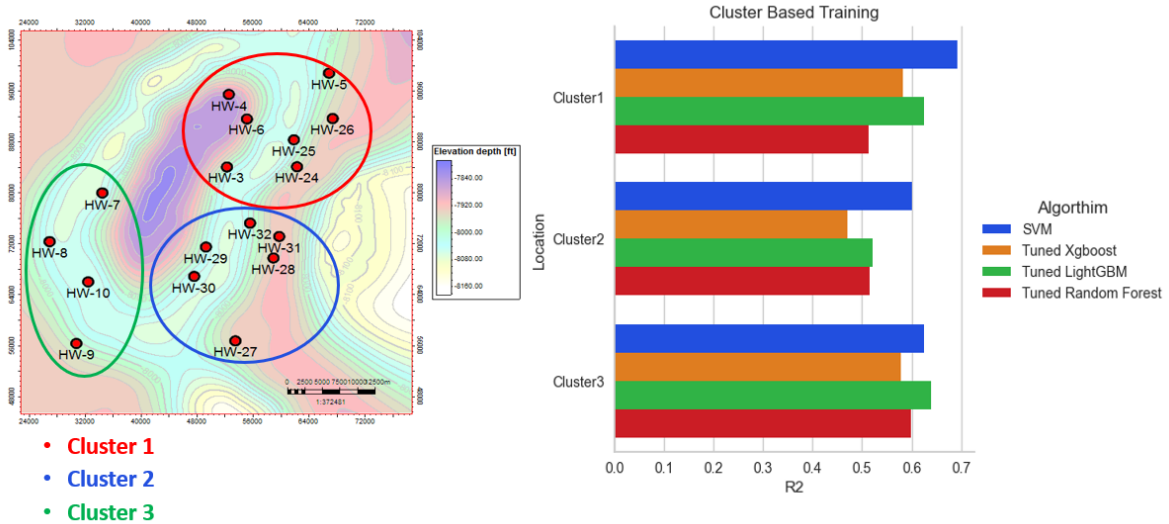


Figure 4.11: Tuned Prediction Results Using Clusters of Wells to Predict a Cluster. The contour map showcases the top of the Upper Kharaib Member. The plot illustrates the performance of each ML algorithm when two clusters of wells are utilized to predict one distinct cluster.

In this experiment, the Support Vector Machine (SVM) model consistently demonstrated the most ro-

bust performance across all three geographically diverse clusters. It recorded the highest R^2 values in each scenario (Figure 4.11). The steady performance of SVM suggests its robustness towards geological changes, thereby establishing it as a powerful tool for permeability prediction across varied geologic settings within the carbonate platform. The superior performance of SVM, showcased in (Figure 4.11), is likely due to its inherent effectiveness in handling smaller datasets, which are often encountered in geological explorations.
By contrast, the performances of Tuned LightGBM, Tuned Xgboost, and Tuned Random Forest models displayed a high degree of variability across the three clusters, as evident in (Figure 4.11). Tuned LightGBM displayed optimal performance in Cluster3 and Cluster 1, which likely share similar geological characteristics. However, its performance dropped when applied to Cluster2, suggesting potential limitations in generalizing across disparate geological formations. Similarly, the Tuned Xgboost model exhibited subpar performance in Cluster2 and Cluster3 compared to Cluster1, hinting at possible restrictions in its adaptability to varied geological settings.
The Tuned Random Forest model showcased the least consistency across the three clusters, yielding the lowest R^2 values. As depicted in (Figure 4.11), this inconsistency suggests that while Random Forest is a potent tool for handling complex non-linear relationships and preventing overfitting, it might struggle to generalize effectively when trained on smaller datasets (**RF_small_Sharma_2023)**. This cross-validation approach, with each distinct geological cluster serving as a test set, emphasizes the importance of evaluating machine learning models under varying geological scenarios to ensure accurate permeability predictions. Furthermore, it highlights that different machine learning algorithms may yield optimal performance under varying conditions, such as dataset size and the geological heterogeneities captured in the data. Therefore, it is crucial to explore multiple models across different scenarios and select the one that provides the highest mean metric across these different scenarios. The results and analysis represented in (Figure 4.11) affirm this strategy.
Incorporating a 'leave-one-well-out' cross-validation within each cluster offers unique insights into each model's performance. In this method, all wells within a single cluster were used for training, except one, which was then assessed for predictive accuracy using the R^2 metric.

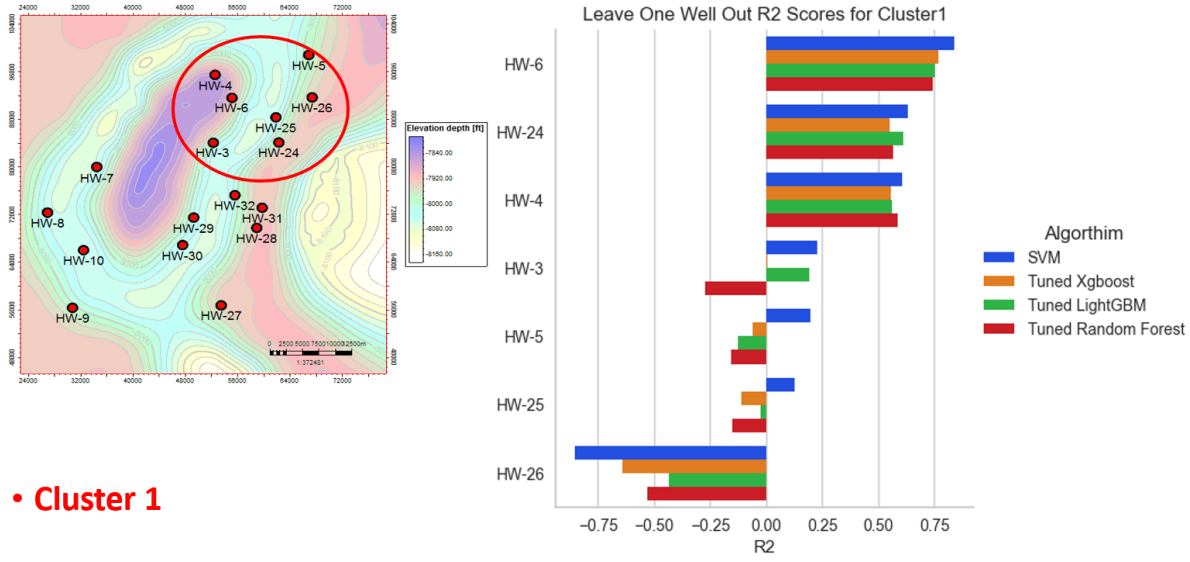


Figure 4.12: Tuned Prediction within Cluster1.

The data reveals that SVM, Tuned Xgboost, Tuned LightGBM, and Tuned Random Forest all display varying performance based on individual wells within the clusters. This variability underscores the impact of spatial correlation and geological heterogeneities on permeability prediction.
For example, within Cluster 1 (Figure 4.12), wells like HW-25, HW-26, HW-5, and HW-3 consistently underperformed across all models, as reflected by low R^2 values. This consistent underperformance suggests the presence of local heterogeneities that the models may struggle to capture due to smaller

training data. These heterogeneities could include abrupt shifts in rock properties, changes in hydrocarbon saturation, or other geological features that influence permeability predictions.

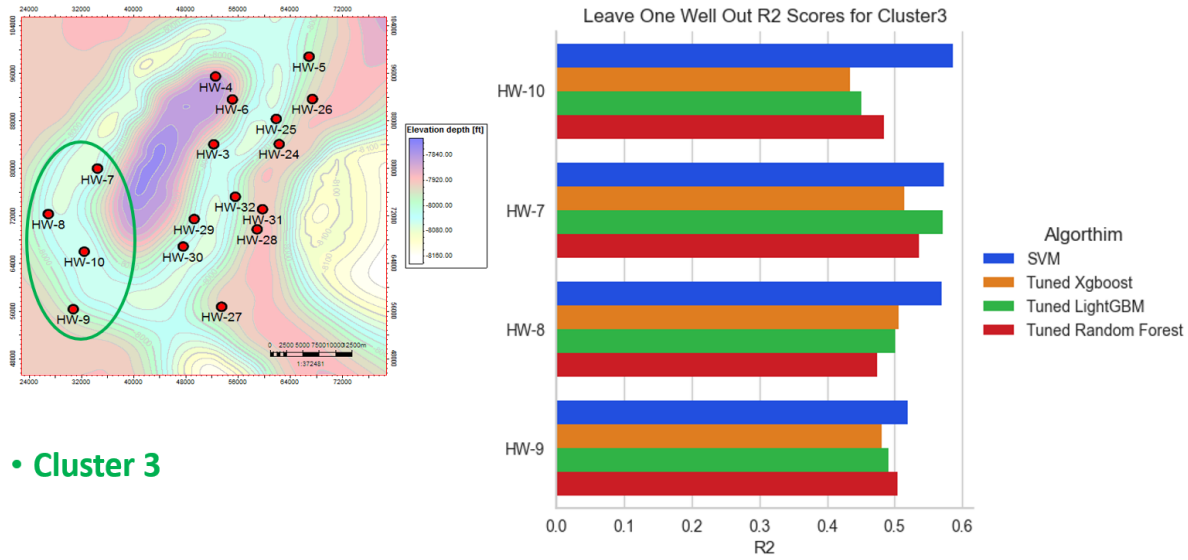


Figure 4.13: Tuned Prediction within Cluster3.

In contrast, Cluster 3 performed well overall (Figure 4.13), particularly for wells not mentioned above, indicating a high degree of spatial correlation among these wells. The consistent performance across most wells in this cluster underscores the effectiveness of machine learning models when spatial correlation is high.
Similarly, within Cluster 2 (Figure 4.14), well HW-27 also exhibited poor performance across all models, indicating the presence of local heterogeneities. This observation underlines how individual outliers or wells with unique geological features can significantly influence the predictive ability of machine learning models.

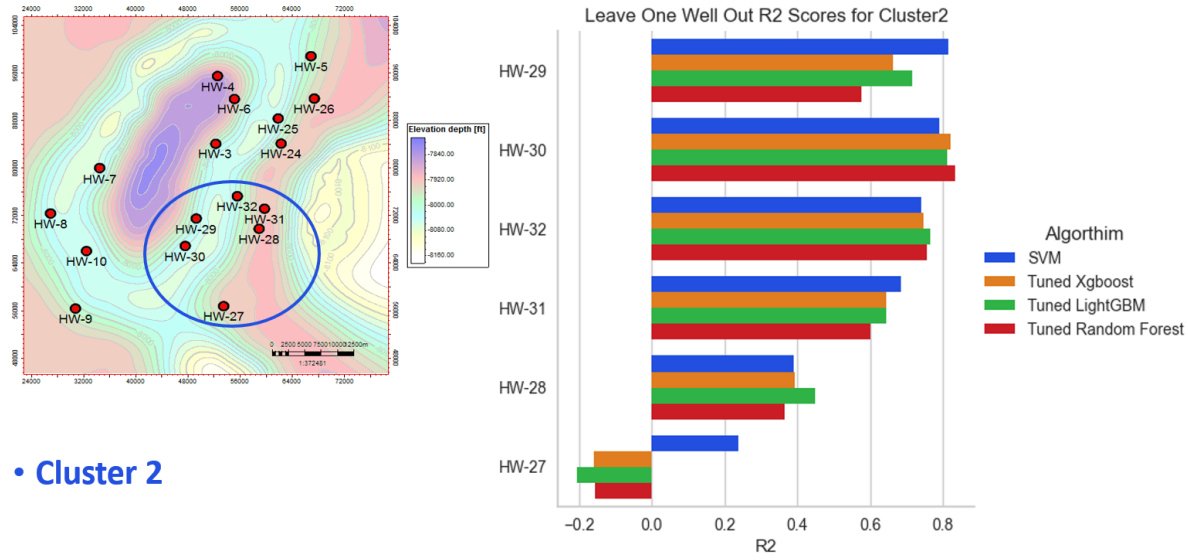


Figure 4.14: Tuned Prediction within Cluster2.

These findings emphasize the significance of combining geological understanding with machine learning for permeability predictions. They also highlight the need for comprehensive and varied training

data that encapsulates a wide range of geological variations, thus enhancing the models' generalization capability. Consequently, integrating domain knowledge with data-driven models can lead to improved reservoir prediction and management.

4.1.8. SHAP (SHapley Additive exPlanations)

In this study, the XGBoost machine learning model was applied to predict permeability, providing significant insights into the importance of different features, as quantified through Shapley Additive exPlanations (SHAP) values. SHAP values measure the average marginal contribution of a feature to the model's prediction across all possible feature combinations.

The Gamma Ray (GR) feature played a significant role in the XGBoost model's permeability predictions, as indicated by its high mean SHAP value of 1.065. The GR log, which provides information on lithology and shale content, could be accounting for the influence of shaly interbeds or the stylolites concentration in the Upper Kharaib Member.

Effective Porosity (PHIE) demonstrated substantial influence with a high mean SHAP value of 1.395. The influence of porosity on permeability, coupled with the variability introduced by different degrees of pore system interconnectivity, is well captured by the XGBoost model.

The DEPTH feature was shown by the XGBoost model to impact permeability predictions considerably, as reflected by its mean SHAP value of 1.199. The model identified a correlation between depth and permeability, which is consistent with our understanding of the effects of depth on compaction and diagenesis.

X_LOG and Y_LOG, representing spatial coordinates of well logs, exhibited a moderate influence on the permeability predictions, suggesting spatial variability in depositional and diagenetic events within the reservoir.

As shown in (Figures 4.15,4.16), A SHAP contribution plot visualizes the impact of each feature on the model's prediction for each instance, with the x-axis representing the SHAP value (contribution to total permeability) and each point representing a specific instance, allowing us to understand how the presence or value of a feature influences the prediction.

In conclusion, while the two models largely agree on the importance of several features, there are significant differences. These discrepancies are likely due to the models' unique ways of interpreting the relationships in the data.

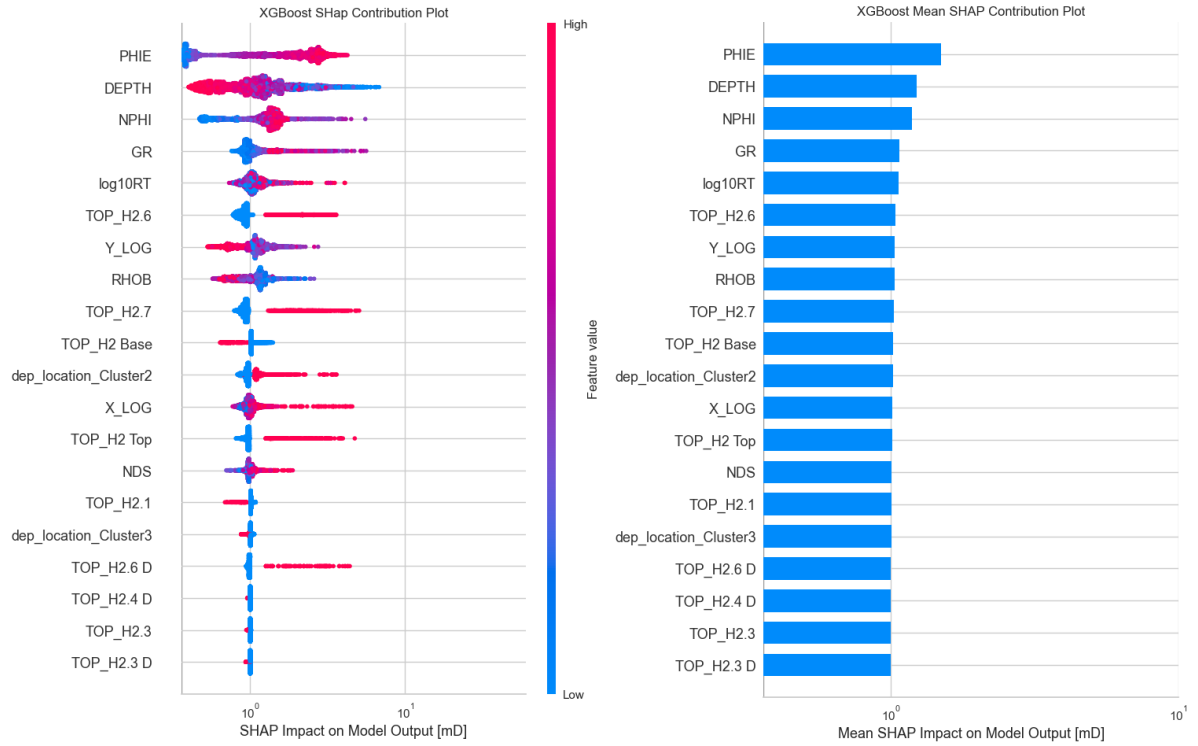


Figure 4.15: SHAP Summary Plots for the XGBoost Model Using the Costa Dataset. SHAP values measure the average marginal contribution of a feature to the model's prediction across all possible feature combinations. In the plot, each dot represents a data point. The x-axis indicates the SHAP value, which shows the impact of the feature on the model's prediction. Features are listed on the y-axis, with the most influential feature at the top. The color of the dots represents the feature's value, with red indicating higher values and blue indicating lower values. A dot's position on the x-axis shows whether the feature pushes the prediction higher (to the right) or lower (to the left).

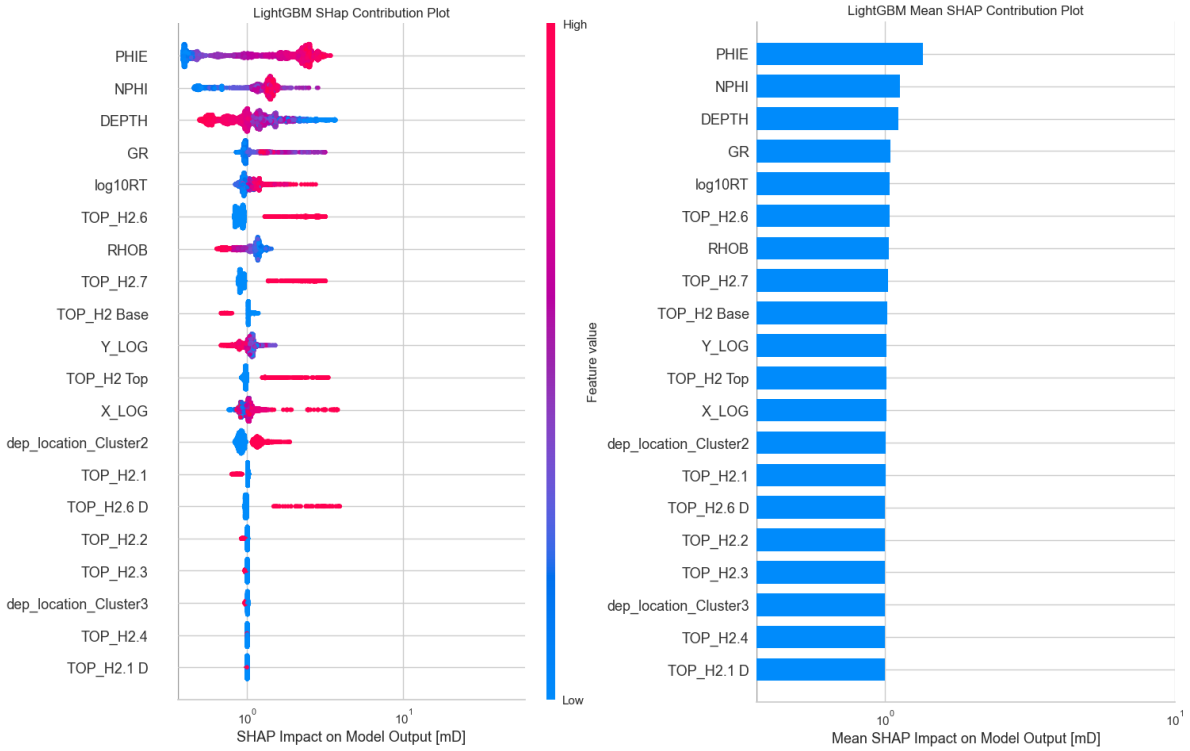


Figure 4.16: SHAP Summary Plots for the Lightgbm Model Using the Costa Dataset. SHAP values measure the average marginal contribution of a feature to the model's prediction across all possible feature combinations. In the plot, each dot represents a data point. The x-axis indicates the SHAP value, which shows the impact of the feature on the model's prediction. Features are listed on the y-axis, with the most influential feature at the top. The color of the dots represents the feature's value, with red indicating higher values and blue indicating lower values. A dot's position on the x-axis shows whether the feature pushes the prediction higher (to the right) or lower (to the left).

4.2. The Volve field

4.2.1. Train/Test Splitting Strategy

In this study of the Volve dataset, we adopted two strategies to segregate our data for model training and testing. The initial strategy was the 'leave one well out' method, where we trained on two wells and tested on the third, rotating through all three wells in turn (Table 4.2).

Training Wells	Testing Well
15/9-19_BT2, 15/9-19_SR	15/9-19_A
15/9-19_A, 15/9-19_SR	15/9-19_BT2
15/9-19_A, 15/9-19_BT2	15/9-19_SR

Table 4.2: Wells used for Training and Testing

The second strategy employed a formation-centric split, with each formation's data divided into an 80% training set and a 20% test set. This methodology was particularly relevant because not all formations were present in all wells as depicted in (Table 4.3). By splitting the data this way, we accounted for the unique petrophysical properties and potential location-based heterogeneities within each formation. Consequently, this approach emphasized understanding the formation characteristics, in contrast to the first strategy which provided a well-centric understanding. These two methods offered comprehensive and complementary insights into the data.

WELL	FORMATION	COUNT
15/9-19_A	Hugin Fm. Top	279
	Skagerrak Fm. Top	190
	Sleipner Fm. Top	147
15/9-19_BT2	Hugin Fm. Top	129
15/9-19_SR	Skagerrak Fm. Top	105
	Hugin Fm. Top	54

Table 4.3: Permeability measurements available per well, Formation and their count

4.2.2. Data Exploration

Exploratory Data Analysis (EDA) is a crucial step where we investigate and break down our data sets to highlight their main statistical characteristics, often using data visualization methods. EDA helps us figure out how to handle our data to get the answers we need, which makes it easier to spot patterns, identify irregularities, verify our hypotheses, or test our assumptions.
we have laid out the statistical characteristics of the data points that we fed into our machine learning models in (Figures 4.17,4.18,4.19). This gives a snapshot of the data distribution across all cored formations within the Volve Oil Field dataset.

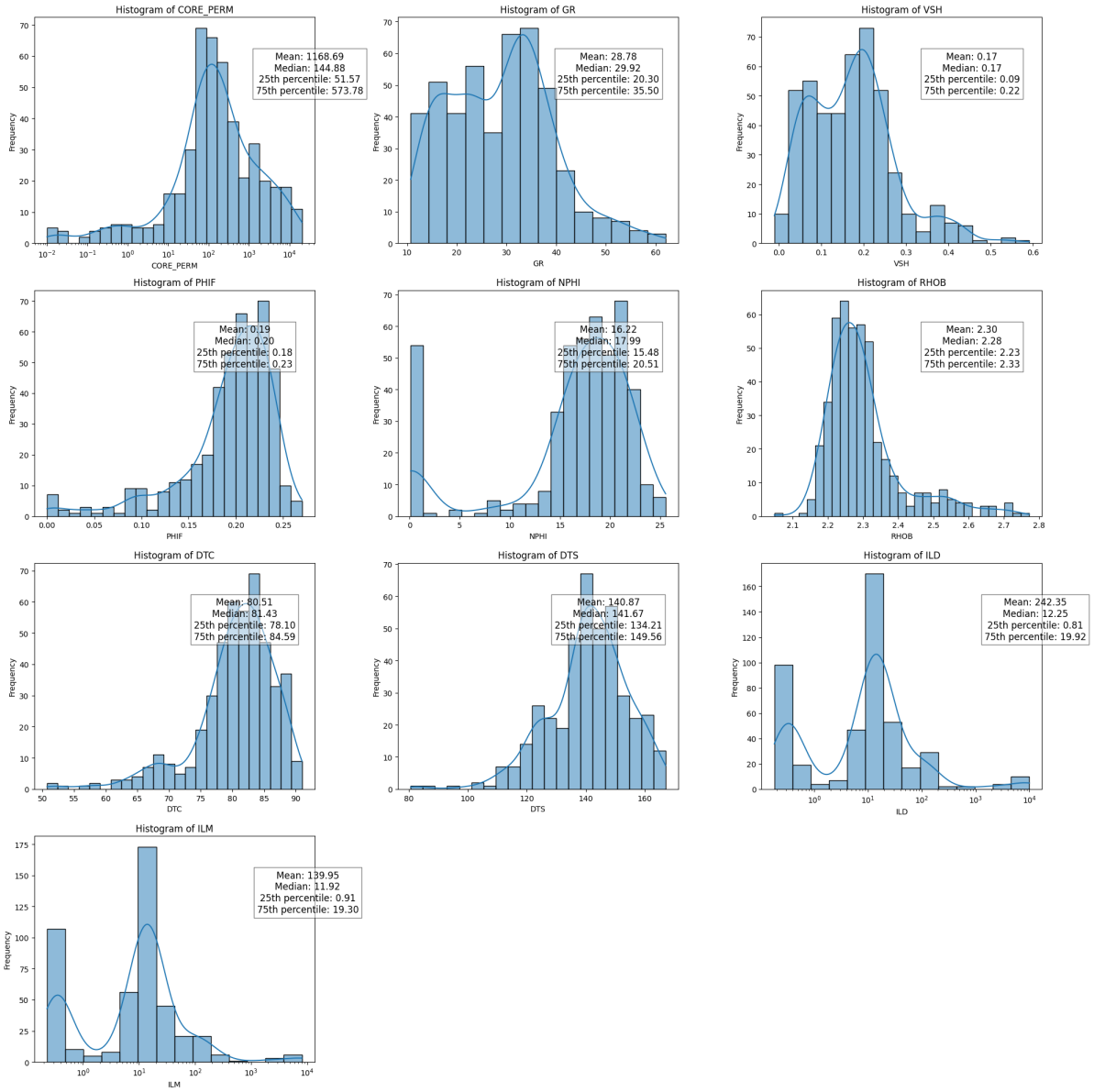


Figure 4.17: Summary statistics of the Hugin Formation.

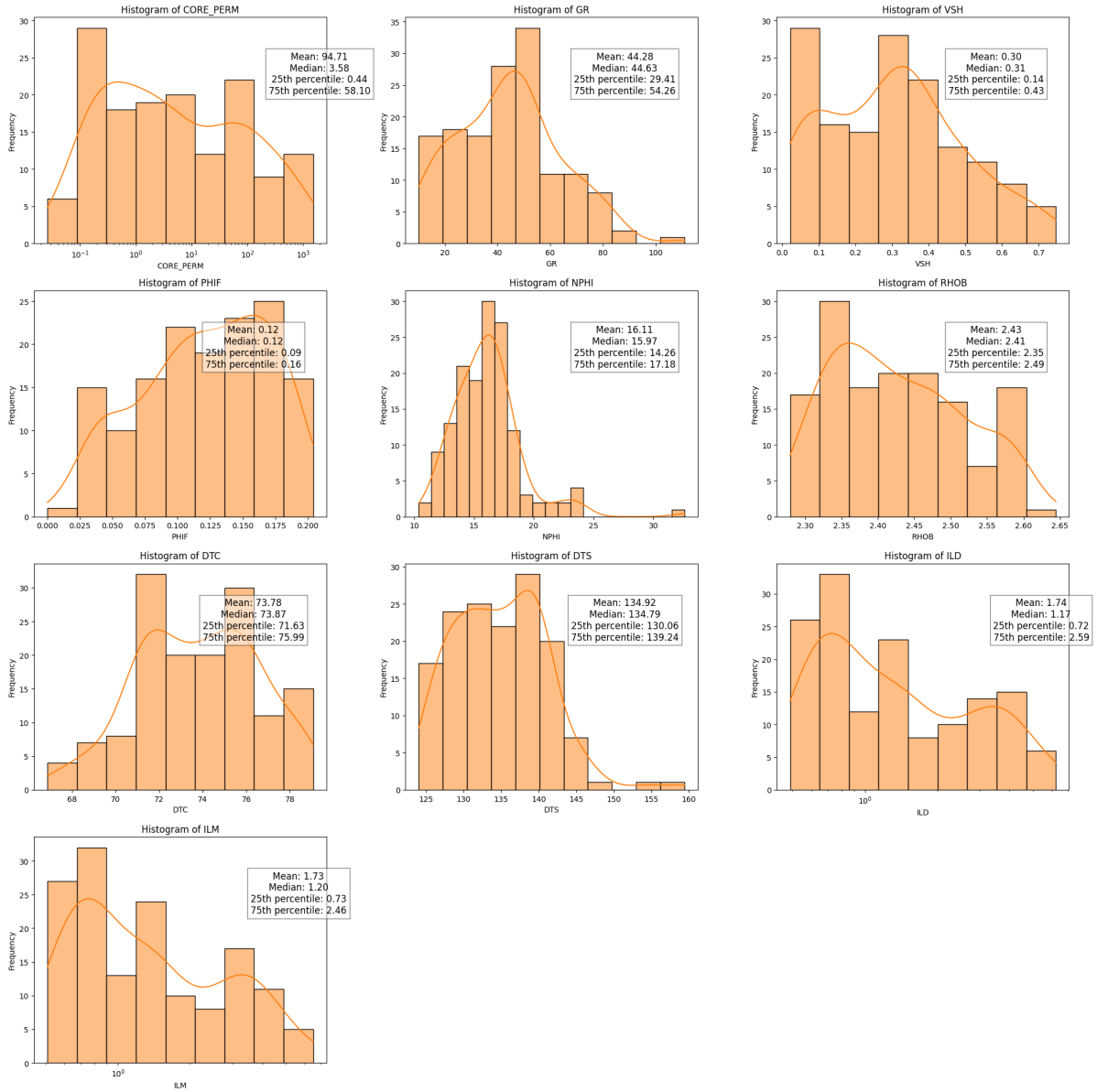


Figure 4.18: Summary statistics of the Sleipner Formation.

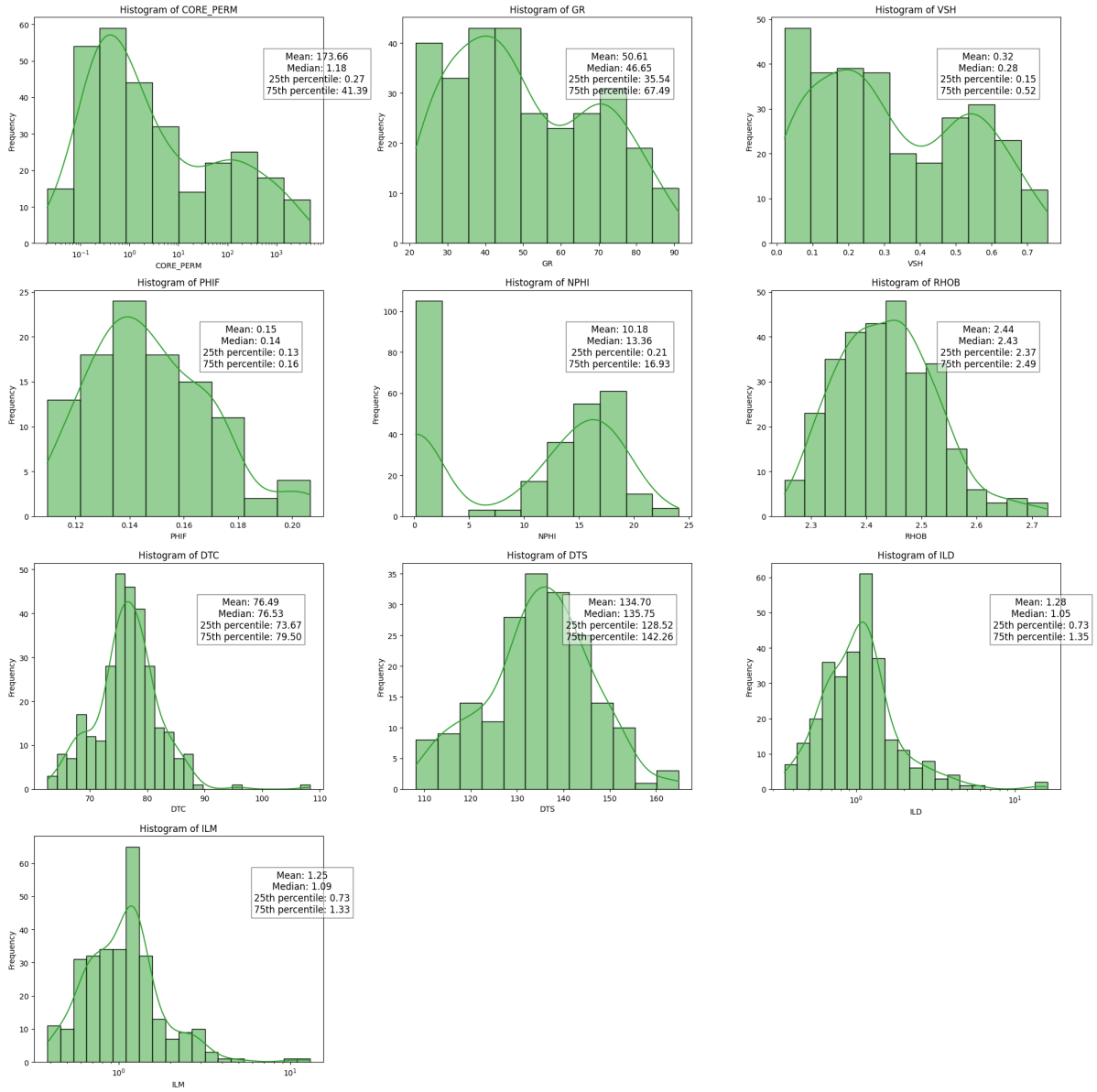


Figure 4.19: Summary statistics of the Skagerrak Formation.

Moreover, this step is critical as it provides insights that inform the next steps of feature engineering and model building. The dataset comprises a total of 916 data points gathered from three wells.

4.2.3. Features Evaluation

In the case of the Hugin Formation, as shown in (Figure 4.21) there appears to be a noteworthy negative correlation between the permeability and several petrophysical parameters: Gamma Ray (GR), Bulk Density (RHOB), deep and medium resistivity (log10ILD and log10ILM), and Volumetric Shale (VSH). In particular, RHOB showed a strong negative correlation of -0.68. This suggests that an increase in these parameters may correspond to a decrease in permeability in the Hugin Formation. On the other hand, there is a strong positive correlation (0.66) between permeability and the effective porosity (PHIF). This implies that an increase in PHIF may enhance permeability within this formation. Compressional and shear transit times (DTC and DTS) showed moderate positive correlations with permeability.

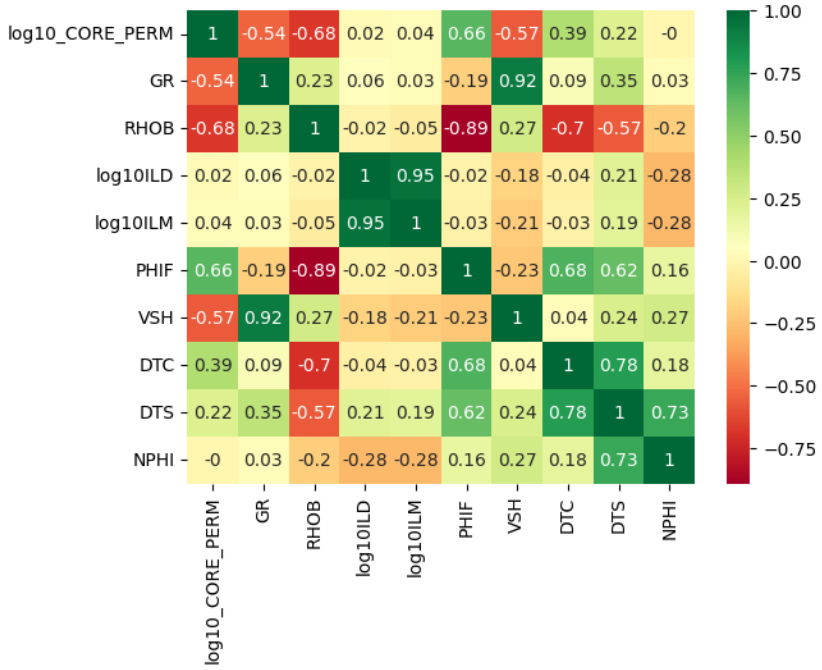


Figure 4.20: Pearson Correlation Matrix for the Volve Oil Field Dataset - Hugin Formation. The color scheme indicates the nature of correlations: green represents positive correlations, while red signifies negative correlations. The intensity of the color indicates the strength of the correlation, with darker shades representing stronger relationships.

In the Sleipner Formation, as shown in (Figure 4.21) similar to the Hugin Formation, there are negative correlations between the permeability and RHOB, log10ILD, log10ILM, and VSH. The strongest negative correlation was observed with RHOB (-0.64). This indicates a potentially inverse relationship between these properties and the permeability in the Sleipner Formation.
Conversely, there is a strong positive correlation (0.58) between permeability and PHIF, suggesting that increased effective porosity could potentially contribute to an increase in permeability within this formation. DTC and DTS also showed positive correlations, with DTC being moderately correlated (0.39).

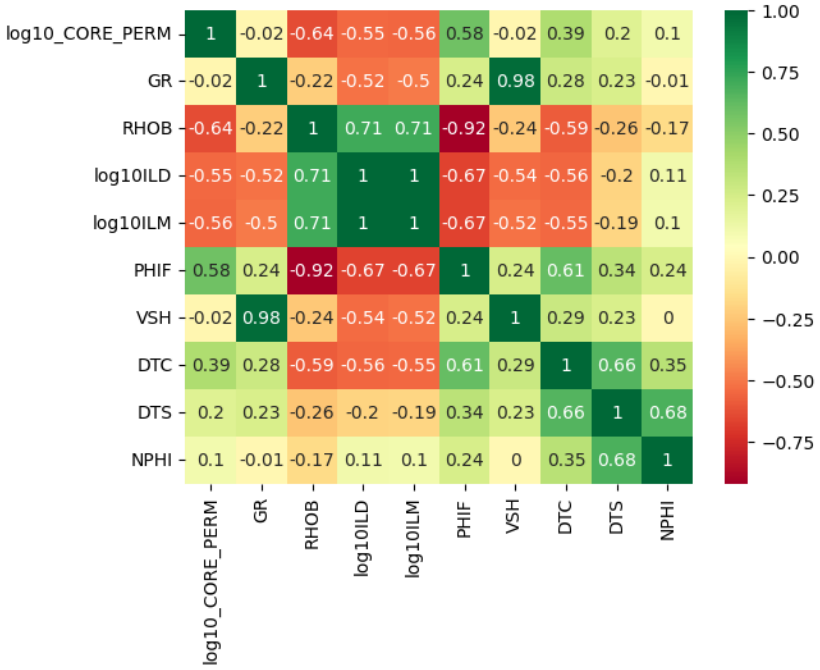


Figure 4.21: Pearson Correlation Matrix for the Volve Oil Field Dataset - Sleipner Formation. The color scheme indicates the nature of correlations: green represents positive correlations, while red signifies negative correlations. The intensity of the color indicates the strength of the correlation, with darker shades representing stronger relationships.

For the Skagerrak Formation, as shown in (Figure 4.22) permeability negatively correlates with GR, RHOB, log10ILD, log10ILM, and VSH. RHOB showed the strongest negative correlation (-0.55), similar to the previous formations.
On the contrary, PHIF, DTS and Neutron Porosity (NPHI) show positive correlation with permeability. PHIF and NPHI have moderate positive correlations (0.60 and 0.54, respectively), suggesting a potential influence on permeability.

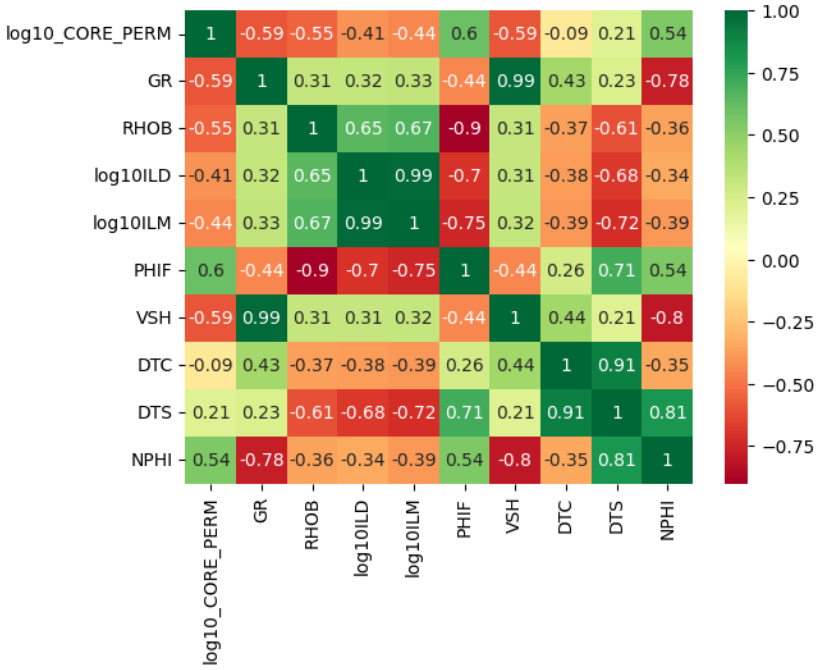


Figure 4.22: Pearson Correlation Matrix for the Volve Oil Field Dataset - Skagerrak Formation. The color scheme indicates the nature of correlations: green represents positive correlations, while red signifies negative correlations. The intensity of the color indicates the strength of the correlation, with darker shades representing stronger relationships.

4.2.4. Leave One Well Out Validation Results

Initially, the baseline models were presented with the raw data, not considering the geological context explicitly. The varied geological characteristics, such as the unique sediment composition in the Hugin Formation, likely contributed to the subpar performance of the models at this stage. More specifically, models seemed to struggle particularly with the 15/9-19 _A well, which includes Hugin, Sleipner and Skagerrak formations. This struggle could potentially be a result of the Sleipner Formation's fluvial and deltaic deposits and the mixed lithologies present in Skagerrak Formation (Vollset and Dore, 1984), which might have introduced greater prediction difficulties. In the second stage, feature engineering was introduced to equip the models with more context-rich information. By integrating one hot encoding of formation tops and vertical variability measures, the models were given a taste of the subsurface geological narrative. This new contextual understanding of the formations led to a noteworthy improvement across all models such as for wells 15/9-19 _A and 15/9-19 _SR the SVM preformance doubled and quadrupled respectively. The performance increase was particularly pronounced for the 15/9-19 _A well which has the highest number of samples, suggesting that the richer feature set enabled the models to better handle the mixed lithologies and depositional environment complexities.

The third stage, which involved model tuning, witnessed an even further improvement in the models' R^2 metric . The improvements in accuracy across all three stages of the model development process - initial model, feature-engineered model, and hyperparameter-tuned model - are visualized in (Figure 4.23).

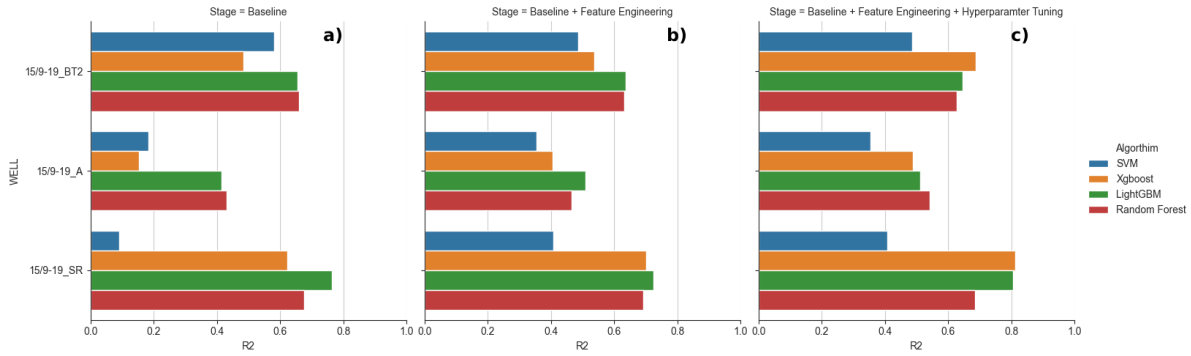


Figure 4.23: Comparative Validation Results at Various Processing Levels.
a) Displays R^2 scores for individual wells using baseline models without feature engineering.
b) Illustrates R^2 scores for each well when employing baseline models that incorporate feature engineering.
c) Represents R^2 scores per well, obtained from baseline models that both implement feature engineering and undergo hyperparameter tuning.

4.2.5. Validation Results on every well

Validation on individual wells plays a pivotal role in assessing the generalizability of machine learning models. In this context, each well is treated as its own distinct test, similar to a 'blind test' approach. This means that the models are evaluated on wells they haven't been trained on, offering a rigorous test of their predictive capabilities. The outcomes of these validations are captured in the well log plots, as illustrated in Figures 4.24, 4.25, and 4.26.

A noteworthy observation from the results is the performance metrics presented in Figures 4.27 and 4.28. These metrics, particularly the R^2 scores and Mean Absolute Error, provide insights into how closely the model's predictions align with actual measured values.

Furthermore, the color-coding in the plots signifies the specific ML model used, emphasizing that the showcased results are derived from the meticulously feature-engineered and hyperparameter-tuned models. This highlights the models' ability not just to fit the training data but also to generalize effectively to new, unseen data.

the well log plots depicted in (Figures 4.24, 4.25, 4.26) illustrate the outcomes of permeability predictions when each well is individually considered as a blind test. Moreover, (Figures 4.27,4.28) refer to comparing validation wells results vs real values measuring R^2 scores and Mean Absolute Error.

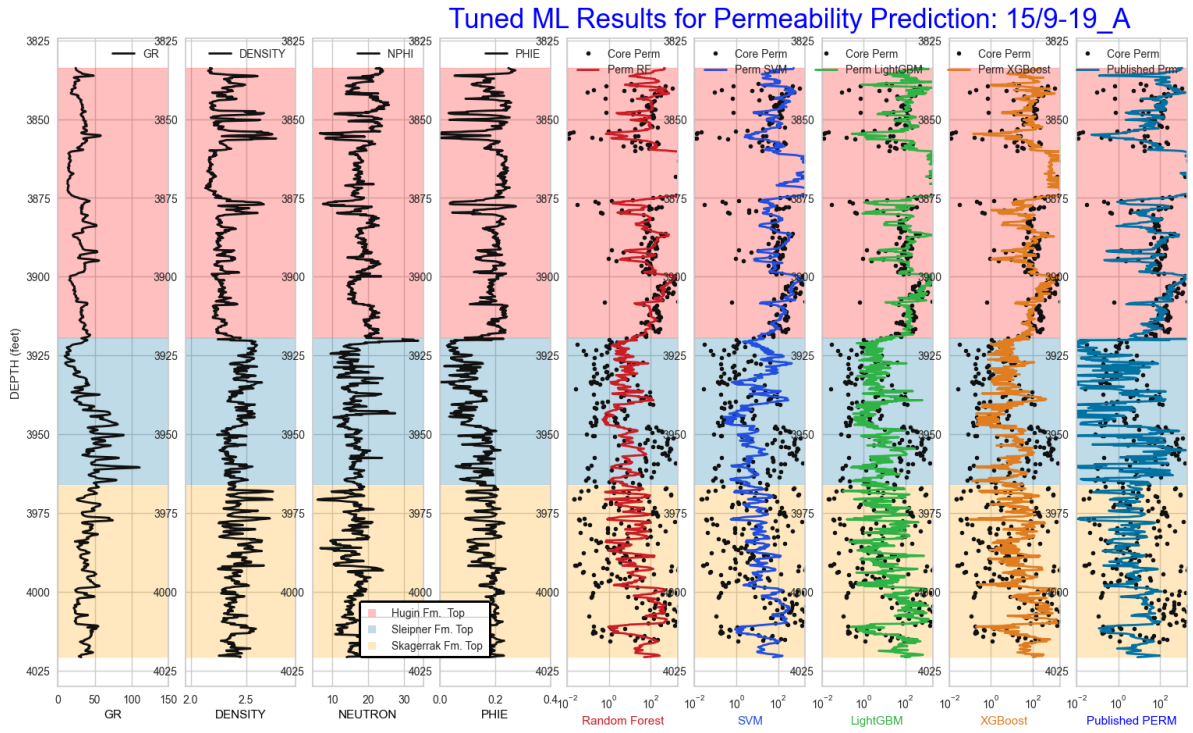


Figure 4.24: Prediction Results for Blind WELL: 15/9-19_A. Displayed are the well log views for GR, DENSITY, NEUTRON, and calculated porosity (PHIE). Predicted permeability is color-coded based on the ML algorithm, with dots signifying the actual core-measured permeability. The background represents the formations, covering Hugin, Sleipner, and Skagerrak.

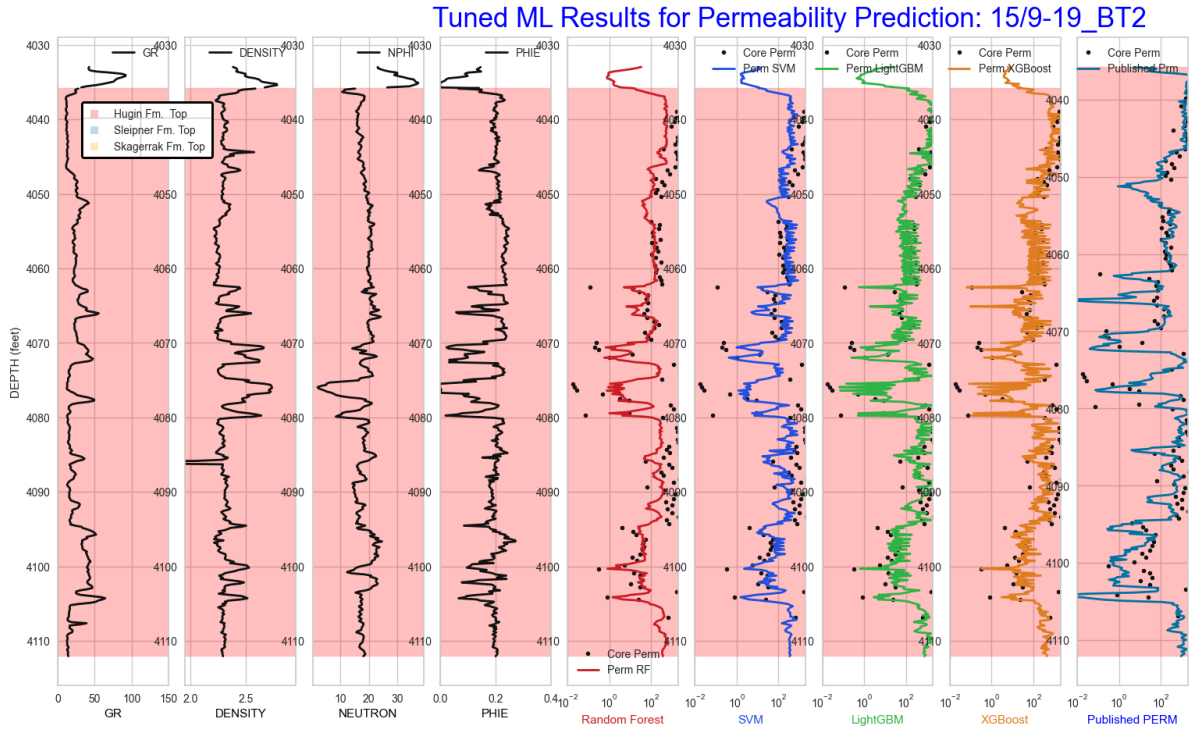


Figure 4.25: Prediction Results for Blind WELL: 15/9-19_BT2. Displayed are the well log views for GR, DENSITY, NEUTRON, and calculated porosity (PHIE). Predicted permeability is color-coded based on the ML algorithm, with dots signifying the actual core-measured permeability. The background represents the Hugin formation.

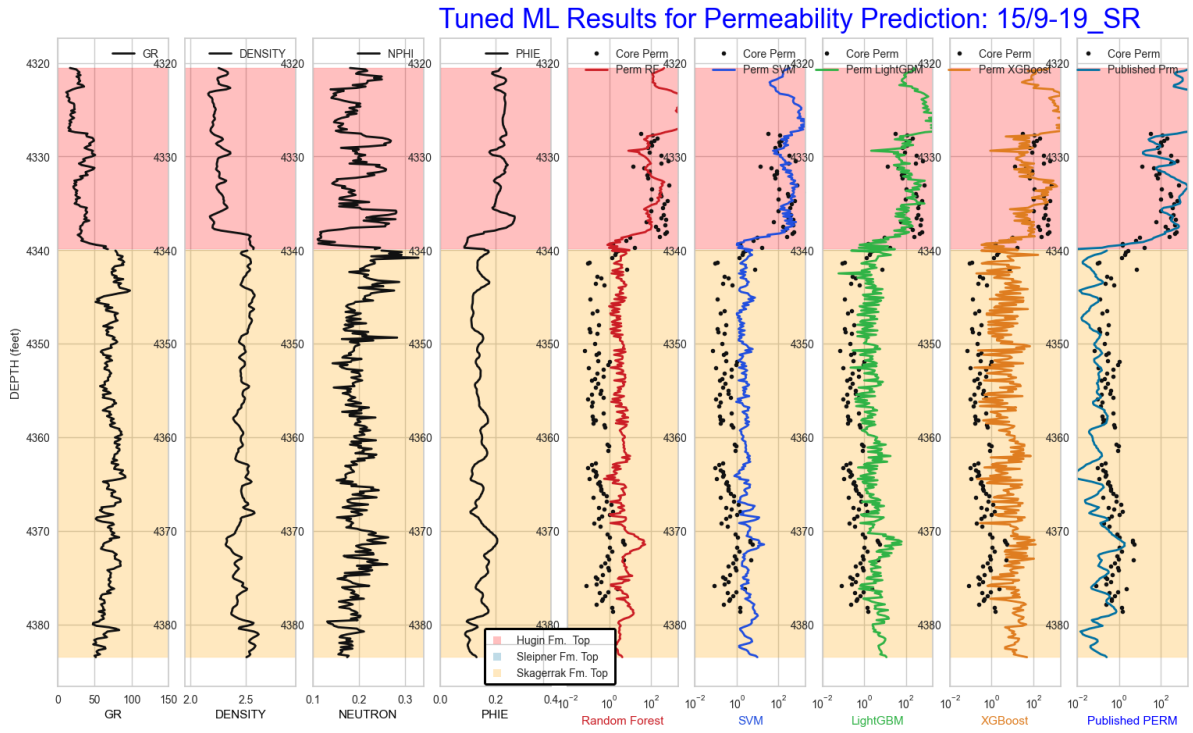


Figure 4.26: Prediction Results for Blind WELL: 15/9-19_SR. Displayed are the well log views for GR, DENSITY, NEUTRON, and calculated porosity (PHIE). Predicted permeability is color-coded based on the ML algorithm, with dots signifying the actual core-measured permeability. The background represents both the Hugin and Sleipner formations.

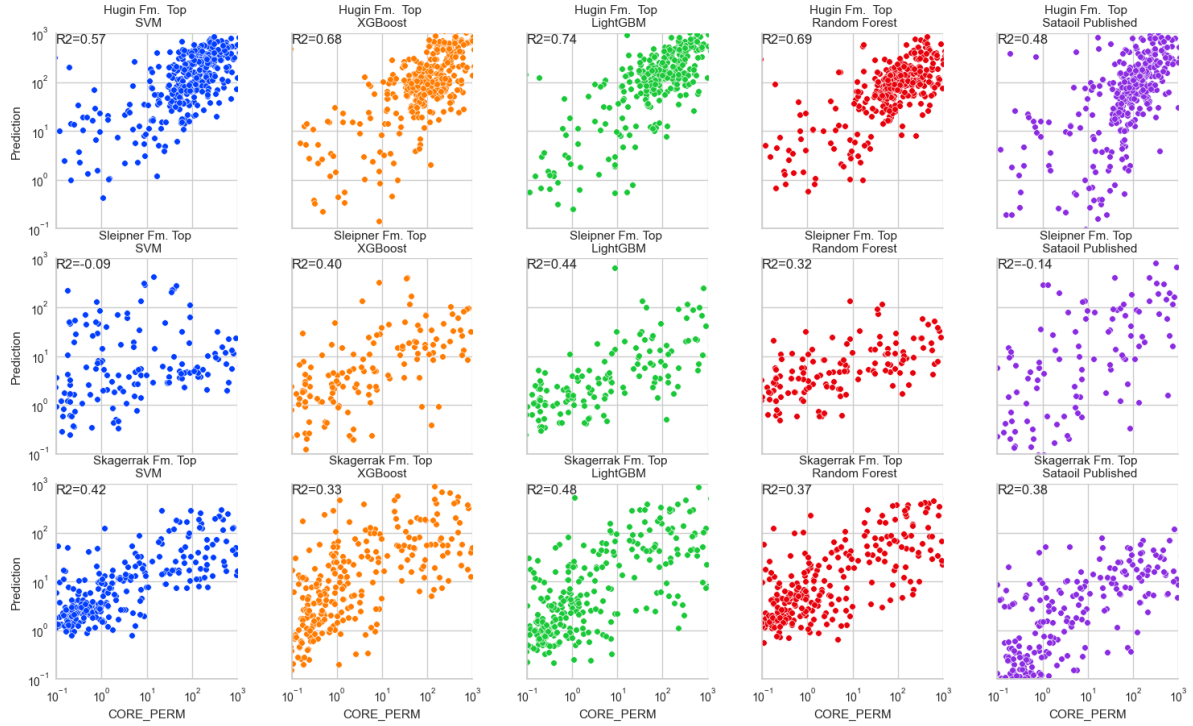


Figure 4.27: Scatter plots of blind prediction per formation top versus actual core permeability values, excluding the published Sataoil permeability. When using the complete well as a blind test, the values represent all the wells when treated as blind tests. Each scatter plot is color-coded based on the specific ML algorithm used. Notably, the scatter plots colored in purple represent the predictions from the published Statoil model.

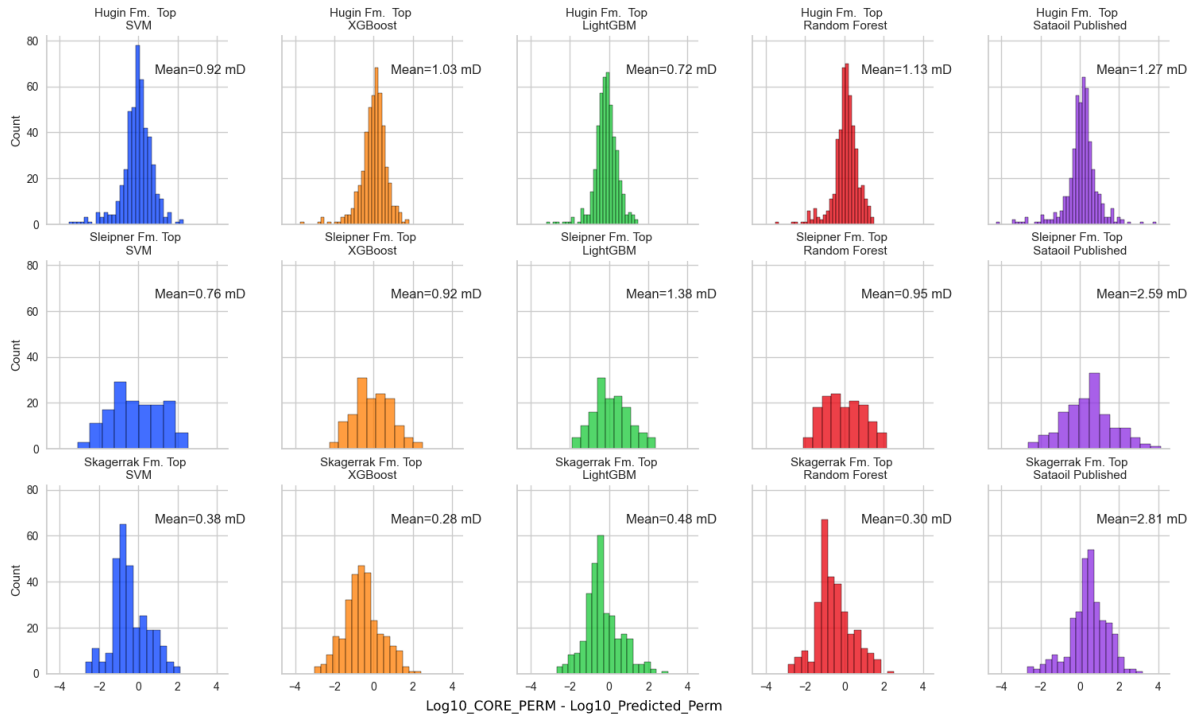


Figure 4.28: Histograms showcasing the Log10 Error between Predicted Permeability and Measured Permeability for each well when treated as a blind test. This excludes the published Statoil permeability values. The histograms provide a distribution view of the prediction errors

4.2.6. Tuned Models Results on 20% Testing and 80% Training

We assessed the machine learning models against the Statoil model, which is a non-linear transformation of a linear model, by splitting the data into 20% for testing and 80% for training. The results revealed a clear advantage of machine learning models, such as SVM, Random Forest, and gradient boosting techniques, over the Statoil's transformed linear approach, with a marked difference in both training and testing phases as depicted in (Figures 4.27 4.29 4.31). The predictive performance of each model seems to be influenced by the unique geological characteristics of each formation.

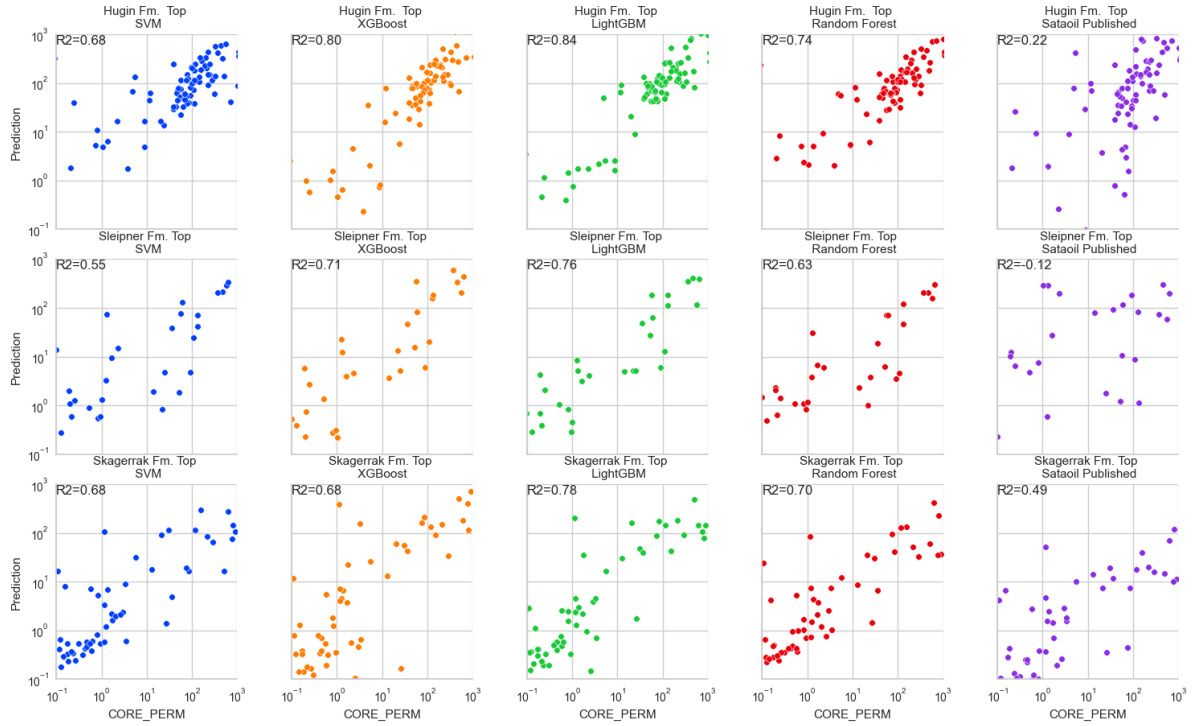


Figure 4.29: Scatter plots comparing the predicted permeability values for the 20% testing set against the actual core permeability values for each formation top. This visualization excludes the published Sataoil permeability values. The plots provide a direct comparison between model predictions and real measurements, emphasizing the model's performance on the testing subset.

For the Hugin Formation, characterized by its fine to medium grained sandstones with intermittent layers of coarse-grained sediment, the complexity inherent in its granular structure seems to favor the LightGBM model. With an R^2 score of 0.838, LightGBM has markedly outperformed other algorithms, indicating its strong ability to handle intricate non-linear patterns embedded within the data. Given the Hugin formation's various granular structures, frequent bioturbation, and occasional cross-bedding, alongside the calcareous and glauconitic characteristics of the sandstones, it can be inferred that the patterns in the formation's permeability are complex, non-linear, and intricate. Compared to the Statoil's transformed linear model's R^2 score of 0.223, the robustness of LightGBM in tackling non-linear patterns clearly manifests.

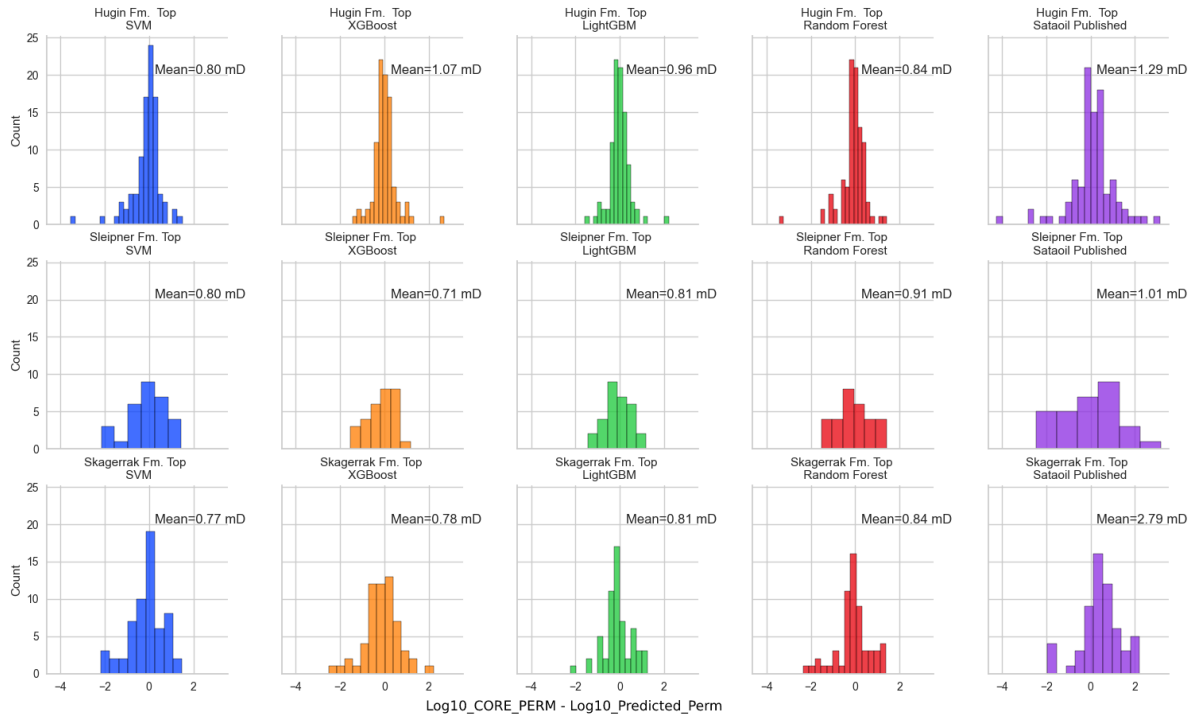


Figure 4.30: Histograms illustrating the Log10 Error between Predicted Permeability and Measured Permeability for the 20% testing values. Notably, while the published Statoil permeability values are excluded from this representation, they were not blind and were included in the fitting of their respective model. The histograms offer a distributional view of the prediction errors, emphasizing the model's accuracy on the designated testing subset

When considering the Skagerrak Formation, the geological characteristics and subsequent permeability distributions appear to be highly heterogeneous. This is indicated by the diverse composition of conglomerates, sandstones, siltstones, and shales as well as the lithologies of anhydrite, dolomite, and limestone. This heterogeneity likely contributes to the superior performance of all machine learning models (LightGBM at 0.785, Random Forest at 0.696, XGBoost at 0.677 and SVM at 0.685) over the Statoil's transformed linear model (0.492). As machine learning models are known for their capacity to capture complex, non-linear patterns, the irregular permeability distribution probably associated with the alluvial fan and interludes of lacustrine deposition system might have rendered these models more effective.

Finally, the Sleipner Formation, with an R^2 score for the LightGBM model of 0.764, Random Forest model of 0.629, SVM model of 0.55, and XGBoost model of 0.709, shows a significant performance difference over the Statoil model, which reported a negative R^2 score (-0.121). It suggests the likely existence of intricate non-linear associations in the data from this formation, characterized by the deposition of fluvial and deltaic sediments interbedded with coal layers. Models like the Statoil's transformed linear approach might struggle with such complex patterns, underscoring the value of advanced machine learning techniques in this context.

In conclusion, the results underscore the significance of integrating machine learning with geological insights, such as one-hot encoding formation tops or modeling each formation separately, for effective permeability prediction. The models' performance improved significantly with the introduction of feature engineering and tuning.

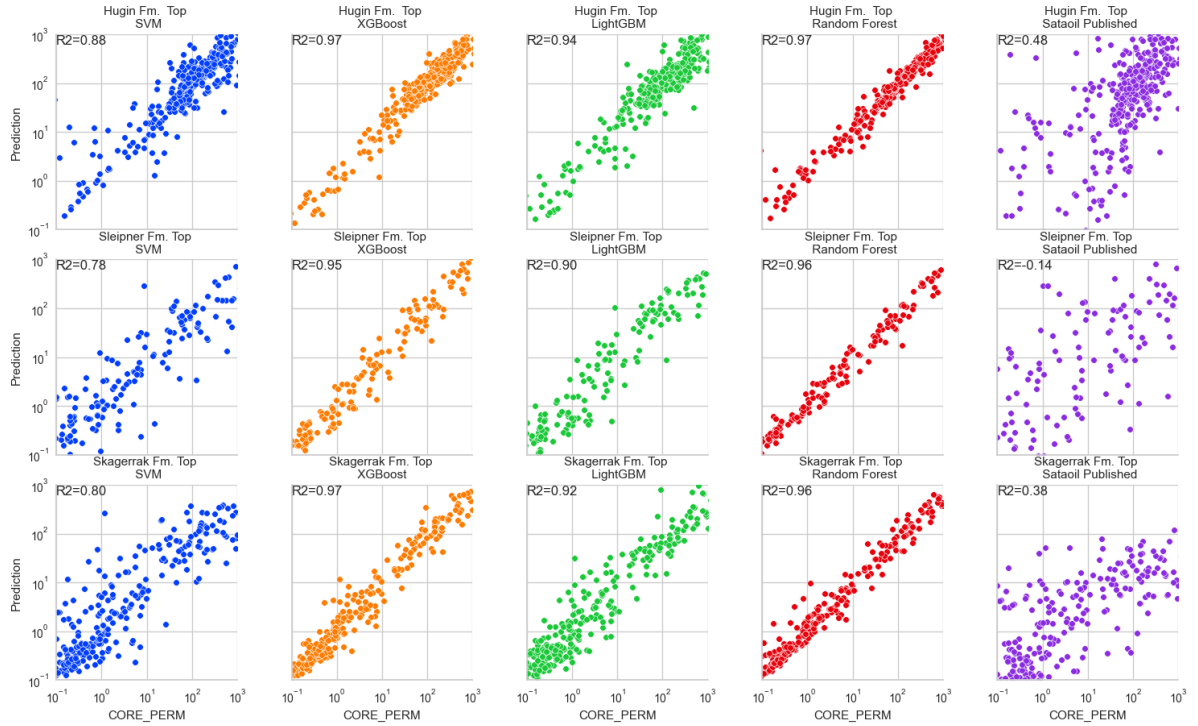


Figure 4.31: Scatter plots showcasing the R^2 scores from training against the actual core permeability values for each formation top. The plots provide a direct comparison between the model's training performance and real measurements, emphasizing the model's ability to capture the underlying patterns in the training data for each formation

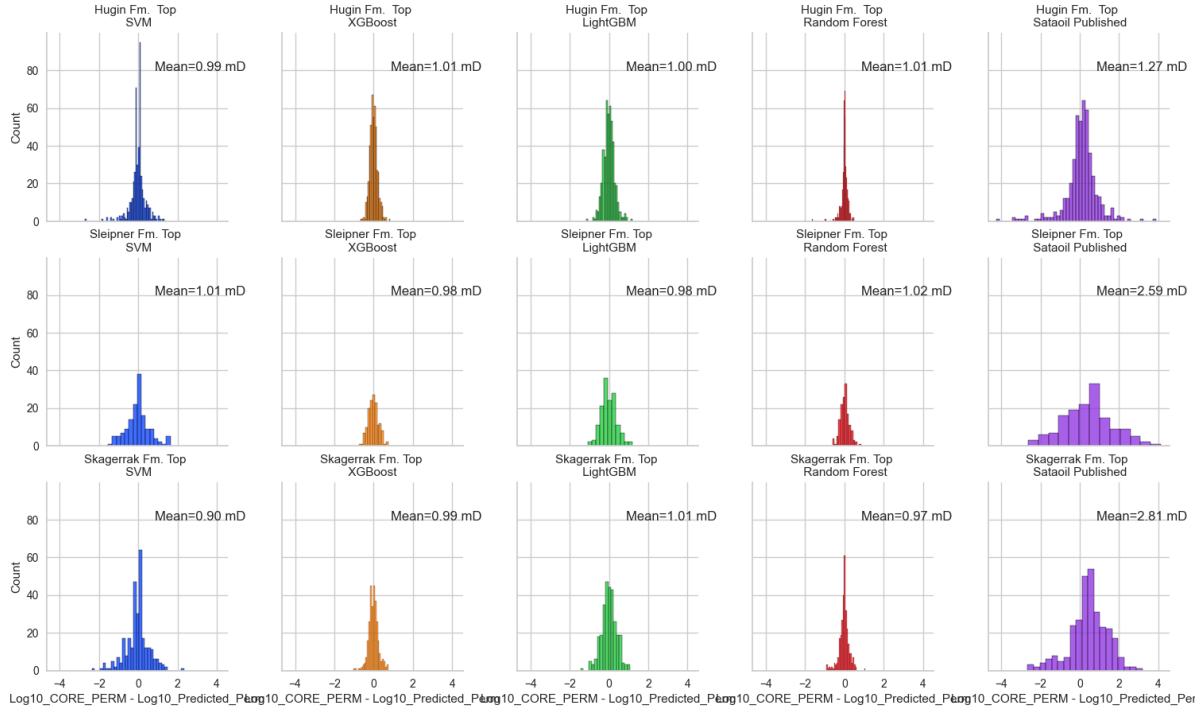


Figure 4.32: Histograms depicting the Log10 Error between Predicted Permeability and Measured Permeability for the 80% training set. The histograms provide a distributional perspective on the prediction errors

4.2.7. SHAP (SHapley Additive exPlanations)

The top contributors according to SHAP values in the XGBoost model (Figure 4.33) are Bulk Density (RHOB),the Gamma Ray (GR), and Effective Porosity (PHIF), while in the LightGBM model, they are Volume of Shale (VSH), PHIF, and RHOB.

GR typically indicates the presence of clay or shale (Katahara, 2008). In the context of the three formations, GR may highlight the abundance of shales in the Sleipner and Skagerrak formations and the alternating shale and siltstone layers within the Hugin formation. High GR values would suggest a higher clay or shale content, which are typically associated with low permeability. Therefore, it is logical that XGBoost uses GR as a significant predictor for permeability.

RHOB often reflects the rock's composition and compaction state, influencing permeability. Higher RHOB values could indicate denser rocks such as cemented sandstones or shales, which typically have lower permeability. Considering the sedimentary context of the three formations, which contain varying grades of sandstone, claystone, siltstone, and shale, RHOB serves as an essential factor for both models in differentiating between lithologies and their corresponding permeabilities.

PHIF represents the effective porosity, or the interconnected pore volume or void space that contributes to fluid flow or permeability in a reservoir. In all three formations, the degree of sorting, grain shape, and abundance of carbonaceous material can affect PHIF, with better-sorted, rounded grains and less carbonaceous material typically resulting in higher porosities and, consequently, higher permeabilities (Shepherd, 1989). Hence, it is consistent to observe PHIF as a top contributing factor in both models.

In the LightGBM model, VSH becomes a top contributor. The Sleipner formation with its coal-bearing sandstones and claystones, the Hugin formation with its intermittent shale and siltstone partings, and the Skagerrak formation with its sequences of siltstones and shales could exhibit significant VSH. Higher shale volumes usually correspond to lower permeabilities due to the low permeability of shale itself, making VSH an essential factor in permeability predictions.

In the context of the three formations, the Hugin Formation appears to have a larger influence in the LightGBM model than in the XGBoost model, based on the mean SHAP values. This could be attributed to the unique lithological features of the Hugin Formation, such as its calcareous and glauconitic sandstones and frequent bioturbation, which could lead to a greater range of permeabilities. The standard deviations indicate how much the contribution of each feature varies. Larger standard deviations for GR, RHOB, and PHIF in the XGBoost model may reflect the wide range of lithologies and sedimentary environments encountered within the Hugin, Sleipner, and Skagerrak Formations.

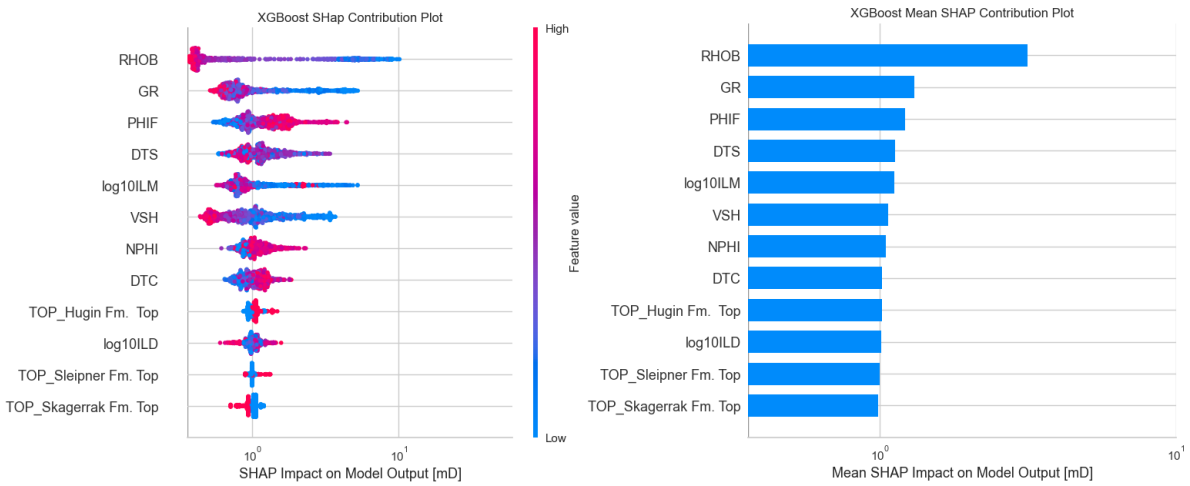


Figure 4.33: SHAP Summary Plots for the XGBoost Model Using the Volve Oil Field Dataset. SHAP values measure the average marginal contribution of a feature to the model's prediction across all possible feature combinations. In the plot, each dot represents a data point. The x-axis indicates the SHAP value, which shows the impact of the feature on the model's prediction. Features are listed on the y-axis, with the most influential feature at the top. The color of the dots represents the feature's value, with red indicating higher values and blue indicating lower values. A dot's position on the x-axis shows whether the feature pushes the prediction higher (to the right) or lower (to the left).

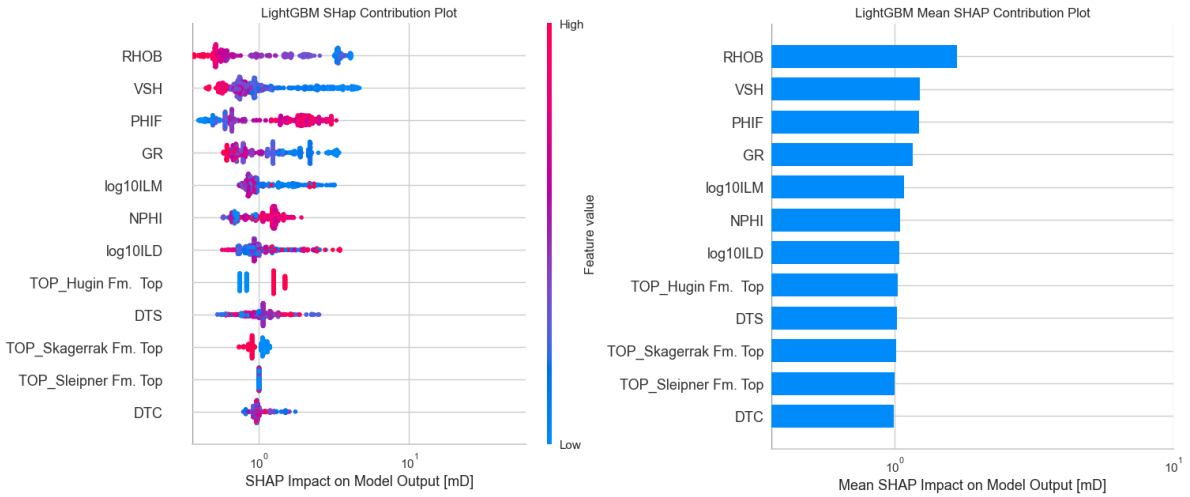


Figure 4.34: SHAP Summary Plots for the LightGBM Model Using the Volve Oil Field Dataset. SHAP values measure the average marginal contribution of a feature to the model's prediction across all possible feature combinations. In the plot, each dot represents a data point. The x-axis indicates the SHAP value, which shows the impact of the feature on the model's prediction. Features are listed on the y-axis, with the most influential feature at the top. The color of the dots represents the feature's value, with red indicating higher values and blue indicating lower values. A dot's position on the x-axis shows whether the feature pushes the prediction higher (to the right) or lower (to the left).

5

Discussion

Measuring core permeability in the laboratory is fraught with uncertainties stemming from various sources (McPhee et al., 2015). The inherent heterogeneity of core samples , combined with potential alterations from sample handling and preparation, can introduce variability in measurements (McPhee et al., 2015). The precision of instruments, the methodologies employed, and the specific laboratory conditions, including stress and temperature, further compound these uncertainties (McPhee et al., 2015). Additionally, factors like pore connectivity, tortuosity, and potential alterations in rock fabric due to chemical reactions or physical changes during drilling can skew results (McPhee et al., 2015). A significant challenge also arises from the error in aligning core measurements to the correct depth in the wellbore, which can lead to misinterpretations and inaccuracies (McPhee et al., 2015). When applying machine learning methods to propagate permeability as a well log, it is essential to recognize that these laboratory uncertainties, including depth alignment errors, are inherently carried over. This can influence the reliability of predictions in terms of aligning closely with in-situ values.

In permeability prediction, both traditional petrophysical methods like the Tixier, Timur, and Coates models and machine learning techniques often prioritize porosity-based features (Figure 15). Traditional methods rely on well log data such as porosity and resistivity , along with core-derived parameters like the cementation factor and saturation exponent (Mohaghegh et al., 1997). However, their application can be limited in heterogeneous reservoirs, often requiring zone-specific adjustments. Machine learning, on the other hand, not only considers these conventional parameters but also integrates a diverse range of features, offering refined predictions for each unique geological setting.

Feature engineering, as underscored in our work and the literature, plays a pivotal role in the development of robust and efficient models in petrophysics. The process of creating new features from existing ones can significantly enhance the performance of machine learning models. In our work, we have given considerable attention to feature engineering, and our choices have been guided by the nature of the data and the specific requirements of the problem at hand. A key technique we used was introducing Vertical Variability to account for the inherent characteristics of well logs and core samples. Well logs capture a volume as depicted in (Figure 5.1), so local heterogeneity is smoothened. This smoothing effect is much smaller on core samples, so their measurements tend to show more variability at a smaller scale. By introducing Vertical Variability, we were able to improve the prediction results significantly, as evidenced by the performance metrics of our models.

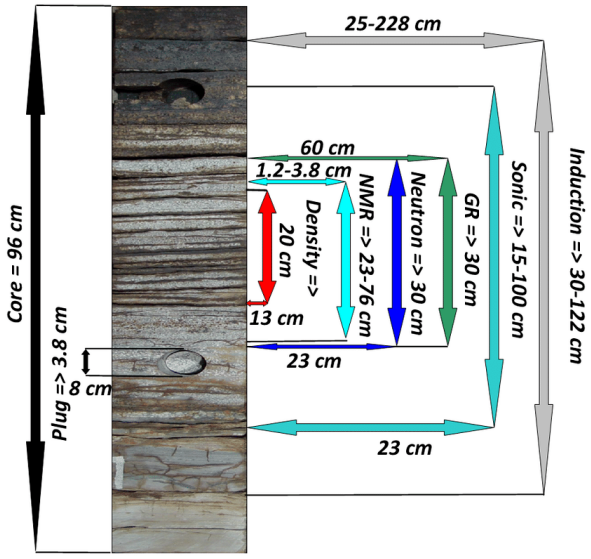


Figure 5.1: This graphic compares the vertical resolution and radial depth of investigation between core, plug, and well logs. The photo shows a core segment from a laminated reservoir zone. Well logs lack the vertical resolution to fully reproduce these thin laminations. Vertical and horizontal descriptions estimate the resolution and radial investigation length of the well logs, respectively (Diniz Ferreira and Torres-Verdin, 2012).

The inclusion of spatial context, such as vertical variability features as additional inputs in our machine learning models is not a novel concept in machine learning (Wood, 2022), but its application in petrophysical analysis has shown significant improvements in prediction accuracy. This is not surprising as spatial context often plays a crucial role in geological formations and processes. However, there is always room for improvement and exploration of new methods. For instance, the study by (Talebi et al., 2021) introduces a spatial random forests technique that captures spatial context within the algorithm itself, rather than through the inputs. This approach, which uses local spatial-spectral information to learn intrinsic heterogeneity, spatial dependencies, and complex spatial patterns, has shown to yield better results. It also handles multi-resolution data and missing values more effectively. Therefore, while our current approach has demonstrated promising results, future work could explore the integration of such spatial-aware algorithms to further enhance the predictive performance of our models.

The choice to utilize Neutron-Density Separation (NDS) in our petrophysical analysis is influenced by its unique advantages. NDS, a graphical representation of the distance between Neutron and Density curves plotted on a well-matched limestone scale, is not affected by compaction, making it a reliable tool for comparing similar facies at different depths (Fadairo and Awuyo, 2019). It also accounts for porosity loss due to compaction, a crucial aspect in petrophysical analysis (Hansen, 1996). Furthermore, NDS is adept at identifying distinctive lithologies, as certain minerals exhibit unusual combinations of density and neutron porosity, making them identifiable on a combination log (Glover, 2018). However, the choice of NDS or any other formula should be guided by the specific requirements of the analysis and the characteristics of the data. For instance, in formations where shale is present, other formulas that account for the effect of bound water might be more appropriate. Therefore, while there are numerous petrophysical formulas available, the use of NDS is justified by its unique advantages in handling compaction effects and identifying distinctive lithologies.

Our decision to use tree-based methods over neural networks was informed by several factors and is supported by the machine-learning literature. As highlighted by (Grinsztajn et al., 2022), tree-based models like XGBoost and Random Forests often outperform deep learning methods on medium-sized tabular data, due to their ability to handle irregular patterns and uninformative features more effectively. Moreover, tree-based models offer greater interpretability, which is crucial in petrophysics applications where understanding the model's decision-making process is important. This aligns with the arguments presented by (Rudin, 2019), who emphasizes the importance of using interpretable

models over black box models for high stakes decisions. Our work reflects these insights, demonstrating the effectiveness of tree-based models and feature engineering techniques in enhancing model performance, and underscores the importance of model choice and transparency in the field of machine learning in petrophysics.

Our study found that machine learning models generally outperformed non-linear transformations of linear models (Statoil Model) , even in blind tests where the Statoil Model had used the blind data in the fitting process, as shown in (Figure 4.29). However, each approach has its merits. For instance, the Statoil model for the Volve field, a non-linear transformation of a linear model, was specifically designed to fit physical parameters, offering a tailored solution for permeability prediction. Furthermore, interpretability techniques like SHAP can provide insights into the relationships used by the machine learnin models for predictions. Therefore, while machine learning models demonstrated superior performance, the choice between custom and more generic models should be guided by the specific problem requirements such as modeling in 1D or 3D domain.

6

Conclusion

This thesis sought to investigate the application and performance of machine learning algorithms in permeability prediction across a carbonate reservoir and a clastic reservoir with varying petrophysical characteristics. The study focused on four machine learning algorithms, namely Support Vector Regression (SVR), Random Forest (RF), XGBoost, and LightGBM, all of which were trained on two datasets, Costa Field and Volve Field, with diverse geological conditions.

The application of machine learning algorithms showed promising results, yielding good accuracy in predicting reservoir permeability. In the Costa Field dataset, blind test R2 scores reached as high as 0.64. However, when utilizing a leave-one-well-out cross-validation approach, the validation R2 scores reached up to 0.8 for individual wells, signifying a high degree of precision in predicting permeability. In the Volve Field dataset, the algorithms also yielded high R2 scores during blind tests, ranging from 0.76 to 0.84 across the Hugin Formation, Sleipner Formation, and Skagerrak Formation.

These results highlight the significant impact of feature engineering and hyperparameter tuning on the performance of machine learning models. After applying these techniques, the prediction accuracy improved significantly from negative R2 up to 0.3 . This underscores the potential of machine learning techniques as a cost-effective and efficient alternative for permeability prediction, especially in scenarios where core data may be scarce or unavailable.

However, while these results are encouraging, the effective application of machine learning models in permeability prediction still requires careful considerations, especially regarding the dataset's quality and representativeness, model interpretability, and the choice of appropriate model complexity. It's also essential to acknowledge that the achieved predictive accuracy may vary across different geological conditions and formations. Consequently, it is crucial to adapt these models and the associated preprocessing steps to the specifics of the geological dataset in question. One recommendation for future researchers is to test out stacking machine learning models and experimenting with different variations. Another suggestion is to incorporate additional sedimentological categorical features, such as depositional systems and geometries based on well log trends as shown in (Figures, 19,20), and evaluate its effect model accuracy.

Overall, this research has provided valuable insights into the potential of machine learning algorithms for permeability prediction. It has also highlighted areas that require further exploration, such as the impact of feature selection, other forms of preprocessing, and the investigation of other machine learning techniques. The results, however, have paved the way for future work in this direction. The high prediction accuracy achieved points towards the feasibility of developing a robust, reliable, and versatile machine learning model for permeability prediction, that can help enhance reservoir characterization and management processes. Future research may also explore the application of these techniques in predicting other crucial subsurface properties (eg. Porosity , TOC , Young's Modulus).

References

Aigbedion, I. (2007). A case study of permeability modeling and reservoir performancenbsp; in the absence of core data in the niger delta, nigeria. *Journal of Applied Sciences*, *7*(5), 772–776. https://doi.org/10.3923/jas.2007.772.776

Asquith, G. B., Krygowski, D., Gibson, C. R., & Asquith, G. B. (2004). *Basic well log analysis*. American Association of Petroleum Geologists.

Bisong, E. (2019). Introduction to scikit-learn. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, 215–229. https://doi.org/10.1007/978-1-4842-4470-8_18

Brereton, R. G., & Lloyd, G. R. (2010). Support vector machines for classification and regression. *The Analyst*, *135*(2), 230–267. https://doi.org/10.1039/b918972f

Carcione, J. M., Gei, D., Picotti, S., Misnan, M. S., Rashidi, M. R., Bakar, Z. A., Harith, Z. Z., Bahri, N. H., & Hashim, N. (2020). Porosity and permeability of the overburden from wireline logs: A case study from offshore malaysia. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, *6*(3). https://doi.org/10.1007/s40948-020-00172-y

Chaya. (2022). Random forest regression. https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939785

Choudhury, K. (2020). Feature scaling-effect of different scikit-learn scalers: Deep dive. https://towardsdatascience.com/feature-scaling-effect-of-different-scikit-learn-scalers-deep-dive-8dec775d4946

Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. *Noise reduction in speech processing*, 1–4.

Costa Gomes, J., Geiger, S., & Arnold, D. (2022). The design of an open-source carbonate reservoir model. *Petroleum Geoscience*, *28*(3). https://doi.org/10.1144/petgeo2021-067

Crampin, T. (2008). Well log facies classification for improved regional exploration. *Exploration Geophysics*, *39*(2), 115–123. https://doi.org/10.1071/eg08012

Deegan, C. E., & Scull, B. J. ((1977). A standard lithostratigraphic nomenclature for the central and northern north sea. *UK Institute of Geological Sciences, Report*, *1*, 36.

Diniz Ferreira, E. L., & Torres-Verdin, C. (2012). Improved estimation of pore connectivity and permeability in deepwater carbonates with the construction of multi-layer static and dynamic petrophysical models.

Elkatatny, S., Mahmoud, M., Tariq, Z., & Abdulraheem, A. (2017). New insights into the prediction of heterogeneous carbonate reservoir permeability from well logs using artificial intelligence network. *Neural Computing and Applications*, *30*(9), 2673–2683. https://doi.org/10.1007/s00521-017-2850-x

Equinor. (2022). Volve field data set. https://www.equinor.com/energy/volve-data-sharing

Fadairo, A., & Awuyo, O. (2019). Achieving best practices in log pre-processing for facies and permeability modeling. *International Journal of Mechanical Engineering and Technology*, *10*.

Feng, R., Grana, D., & Balling, N. (2021). Imputation of missing well log data by random forest and its uncertainty analysis. *Computers amp;amp; Geosciences*, *152*, 104763. https://doi.org/10.1016/j.cageo.2021.104763

Gholanlo, H. H. (2021). Analysis of permeability based on petrophysical logs: Comparison between heuristic numerical and analytical methods. *Journal of Petroleum Exploration and Production Technology*, *11*(5), 2097–2111. https://doi.org/10.1007/s13202-021-01163-9

Glover, P. W. (2018). Petrophysics msc course notes. https://homepages.see.leeds.ac.uk/~earpwjg/PG_EN/CD%20Contents/GGL-66565%20Petrophysics%20English/Chapter%2015.PDF

Gray, E., Hartley, A., & Howell, J. (2020). The influence of stratigraphy and facies distribution on reservoir quality and production performance in the triassic skagerrak formation of the uk and norwegian central north sea. *Geological Society, London, Special Publications*, *494*(1), 379–409. https://doi.org/10.1144/sp494-2019-68

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data?

Han, D., & Kwon, S. (2021). Application of machine learning method of data-driven deep learning model to predict well production rate in the shale gas reservoirs. *Energies*, *14*(12), 3629. https://doi.org/10.3390/en14123629

Hansen, S. (1996). A compaction trend for cretaceous and tertiary shales on the norwegian shelf based on sonic transit times. *Petroleum Geoscience*, *2*(2), 159–166.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Ibrahem Ahmed Osman, A., Najah Ahmed, A., Chow, M. F., Feng Huang, Y., & El-Shafie, A. (2021). Extreme gradient boosting (xgboost) model to predict the groundwater levels in selangor malaysia. *Ain Shams Engineering Journal*, *12*(2), 1545–1556. https://doi.org/https://doi.org/10.1016/j.asej.2020.11.011

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Javapoint. (2020). Support vector machine (svm) algorithm - javatpoint. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

Katahara, K. (2008). What is shale to a petrophysicist? *The Leading Edge*, *27*(6), 738–741.

Kavzoglu, T., & Teke, A. (2022). Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (xgboost). *Bulletin of Engineering Geology and the Environment*, *81*(5). https://doi.org/10.1007/s10064-022-02708-w

Ke, G., Meng, Q., Finely, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems 30 (NIP 2017)*.

Kieft, R. L., Jackson, C. A.-L., Hampson, G. J., & Larsen, E. (2010). Sedimentology and sequence stratigraphy of the hugin formation, quadrant 15, norwegian sector, south viking graben. *Geological Society, London, Petroleum Geology Conference Series*, *7*(1), 157–176. https://doi.org/10.1144/0070157

Kim, J. (2022). Synthetic shear sonic log generation utilizing hybrid machine learning techniques. *Artificial Intelligence in Geosciences*, *3*, 53–70. https://doi.org/https://doi.org/10.1016/j.aiig.2022.09.001

Kramer, O. (2016). Scikit-learn. In *Machine learning for evolution strategies* (pp. 45–53). Springer International Publishing. https://doi.org/10.1007/978-3-319-33383-0_5

Kumar, A. (2022). Xgboost vs lightgbm. https://www.linkedin.com/pulse/xgboost-vs-lightgbm-ashik-kumar/

Lewinson, E. (2023). A comprehensive overview of regression evaluation metrics. https://developer.nvidia.com/blog/a-comprehensive-overview-of-regression-evaluation-metrics/

Lundberg, S. (2020). The science behind interpretml: Shap. Retrieved July 8, 2023, from https://www.youtube.com/watch?v=VB9uV-x0gtg

McPhee, C., Reed, J., & Zubizarreta, I. (2015). Routine core analysis. *Developments in Petroleum Science*, 181–268. https://doi.org/10.1016/b978-0-444-63533-4.00005-6

Mitchell, R., Frank, E., & Holmes, G. (2020). Gputreeshap: Fast parallel tree interpretability. *CoRR*, *abs/2010.13972*. https://arxiv.org/abs/2010.13972

Mohaghegh, S., Balan, B., & Ameri, S. (1997). Permeability determination from well log data. *SPE Formation Evaluation*, *12*(03), 170–174. https://doi.org/10.2118/30978-pa

NPD. (2023). Norwegian petroleum directorate fact pages. wellbore 15/9-19 a. https://factpages.npd.no/en/wellbore/PageView/Exploration/All/3145

Otchere, D. A., Ganat, T. O., Gholami, R., & Lawal, M. (2021). A novel custom ensemble learning model for an improved reservoir permeability and water saturation prediction. *Journal of Natural Gas Science and Engineering*, *91*, 103962. https://doi.org/10.1016/j.jngse.2021.103962

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Radwan, A. E. (2020). Modeling the depositional environment of the sandstone reservoir in the middle miocene sidri member, badri field, gulf of suez basin, egypt: Integration of gamma-ray log patterns and petrographic characteristics of lithology. *Natural Resources Research*, *30*(1), 431–449. https://doi.org/10.1007/s11053-020-09757-6

Rezaee, R., & Ekundayo, J. (2022). Permeability prediction using machine learning methods for the co2 injectivity of the precipice sandstone in surat basin, australia. *Energies*, *15*(6), 2053. https://doi.org/10.3390/en15062053

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Safaei-Farouji, M., & Kadkhodaie, A. (2022). A comparative study of individual and hybrid machine learning methods for estimation of vitrinite reflectance (ro) from petrophysical well logs. *Modeling Earth Systems and Environment*, *8*(4), 4867–4881. https://doi.org/10.1007/s40808-022-01381-y

Saha, S., Vishal, V., Mahanta, B., & Pradhan, S. P. (2022). Geomechanical model construction to resolve field stress profile and reservoir rock properties of jurassic hugin formation, volve field, north sea. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, *8*(2). https://doi.org/10.1007/s40948-022-00359-5

Saha, S. (2023). Xgboost vs lightgbm: How are they different. https://neptune.ai/blog/xgboost-vs-lightgbm#:~:text=In%20contrast%20to%20the%20level,higher%20accuracy%20while%20being%20faster.

Sanei, M., Ramezanzadeh, A., & Asgari, A. (2022). Building 1d and 3d static reservoir geomechanical properties models in the oil field - journal of petroleum exploration and production technology. https://link.springer.com/article/10.1007/s13202-022-01553-7

Saner, S., Kissami, M., & Nufaili, S. A. (1997). Estimation of permeability from well logs using resistivity and saturation data. *SPE Formation Evaluation*, *12*(01), 27–31. https://doi.org/10.2118/26277-pa

Shepherd, R. G. (1989). Correlations of permeability and grain size. *Groundwater*, *27*(5), 633–638.

Strohmenger, C. J., Weber, L. J., Ghani, A., Rebelle, M., Al-Mehsin, K., Al-Jeelani, O., Al-Mansoori, A., & Suwaina, O. (2004). High-resolution sequence stratigraphy of the kharaib formation (lower cretaceous, u.a.e.) *All Days*. https://doi.org/10.2118/88729-ms

Talebi, H., Peeters, L. J., Otto, A., & Tolosana-Delgado, R. (2021). A truly spatial random forests algorithm for geoscience data analysis and modelling. *Mathematical Geosciences*, *54*(1), 1–22. https://doi.org/10.1007/s11004-021-09946-w

Torres, K. M., Ugonoh, M. S., & Al Hashmi, N. F. (2017). High-resolution sequence stratigraphy analysis and diagenesis evolution of a barremian carbonate platform kharaib formation, onshore abu dhabi. *Day 1 Mon, November 13, 2017*. https://doi.org/10.2118/188875-ms

Vapnik, V. (1998). The support vector method of function estimation. In J. A. K. Suykens & J. Vandewalle (Eds.), *Nonlinear modeling: Advanced black-box techniques* (pp. 55–85). Springer US. https://doi.org/10.1007/978-1-4615-5703-6_3

Verleysen, M., & François, D. (2005). The curse of dimensionality in data mining and time series prediction. *Computational Intelligence and Bioinspired Systems*, 758–770. https://doi.org/10.1007/11494669_93

Vollset, J., & Dore, G. (1984). A revised triassic and jurassic lithostratigraphic nomenclature for the norwegian north sea. *NPD-Bulletin No*, *3*, 53.

Wade, C. (2020). *Hands-on gradient boosting with xgboost and scikit-learn: Perform accessible machine learning and extreme gradient boosting with python*. Packt Publishing.

Wood, D. A. (2022). Gamma-ray log derivative and volatility attributes assist facies characterization in clastic sedimentary sequences for formulaic and machine learning analysis. *Advances in Geo-Energy Research*, *6*(1), 69–85.

Zhang, G., Wang, Z., Mohaghegh, S., Lin, C., Sun, Y., & Pei, S. (2021). Pattern visualization and understanding of machine learning models for permeability prediction in tight sandstone reservoirs. *Journal of Petroleum Science and Engineering*, *200*, 108142. https://doi.org/10.1016/j.petrol.2020.108142

Zhang, S., Gu, Y., Gao, Y., Wang, X., Zhang, D., & Zhou, L. (2022). Petrophysical regression regarding porosity, permeability, and water saturation driven by logging-based ensemble and trans-

fer learnings: A case study of sandy-mud reservoirs. *Geofluids*, *2022*, 1–31. https://doi.org/10.1155/2022/9443955

Zhang, Z., Zhang, H., Li, J., & Cai, Z. (2021). Permeability and porosity prediction using logging data in a heterogeneous dolomite reservoir: An integrated approach. *Journal of Natural Gas Science and Engineering*, *86*, 103743. https://doi.org/10.1016/j.jngse.2020.103743

Appendix

Costa Dataset

Validation wells plots for Costa Dataset

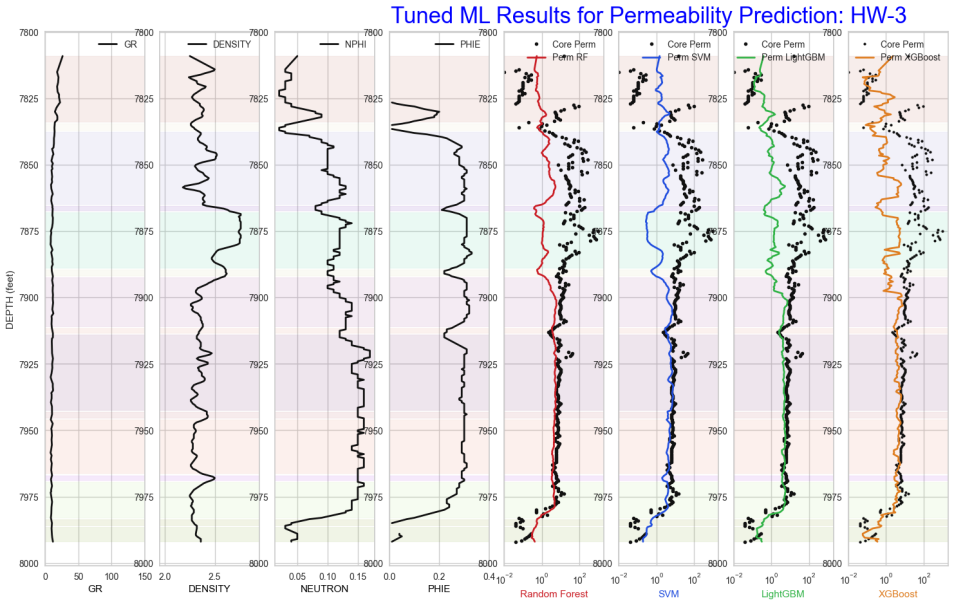


Figure 1: Prediction Results for Validation WELL : HW-3

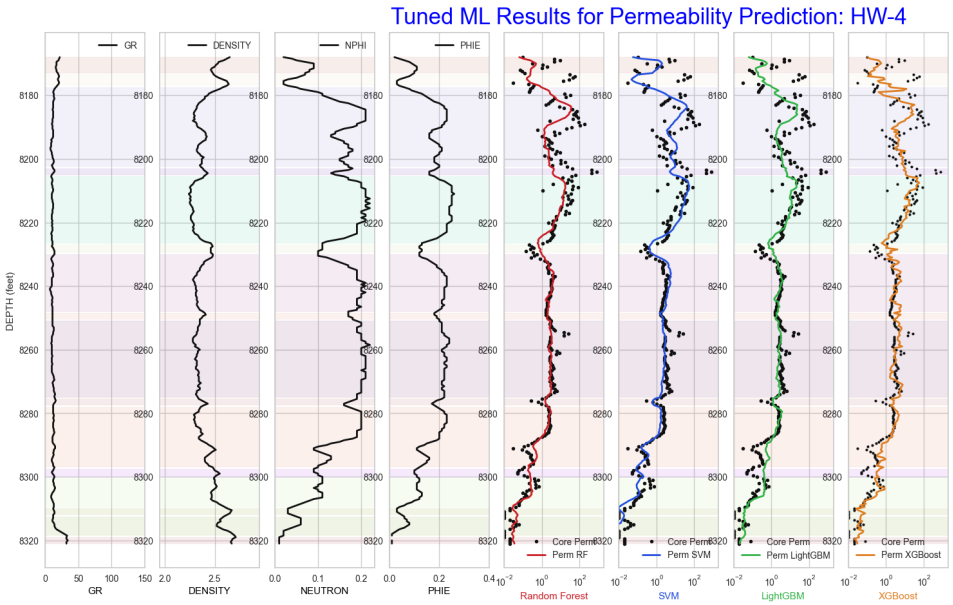


Figure 2: Prediction Results for Validation WELL : HW-4

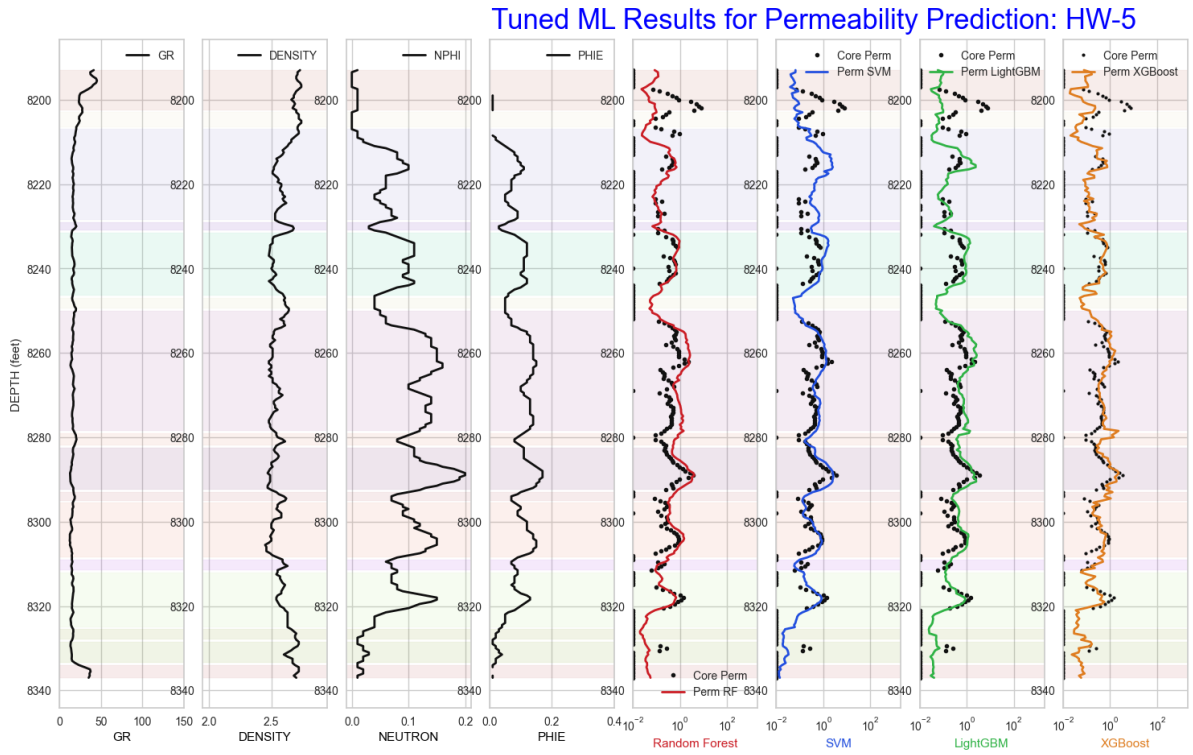


Figure 3: Prediction Results for Validation WELL : HW-5

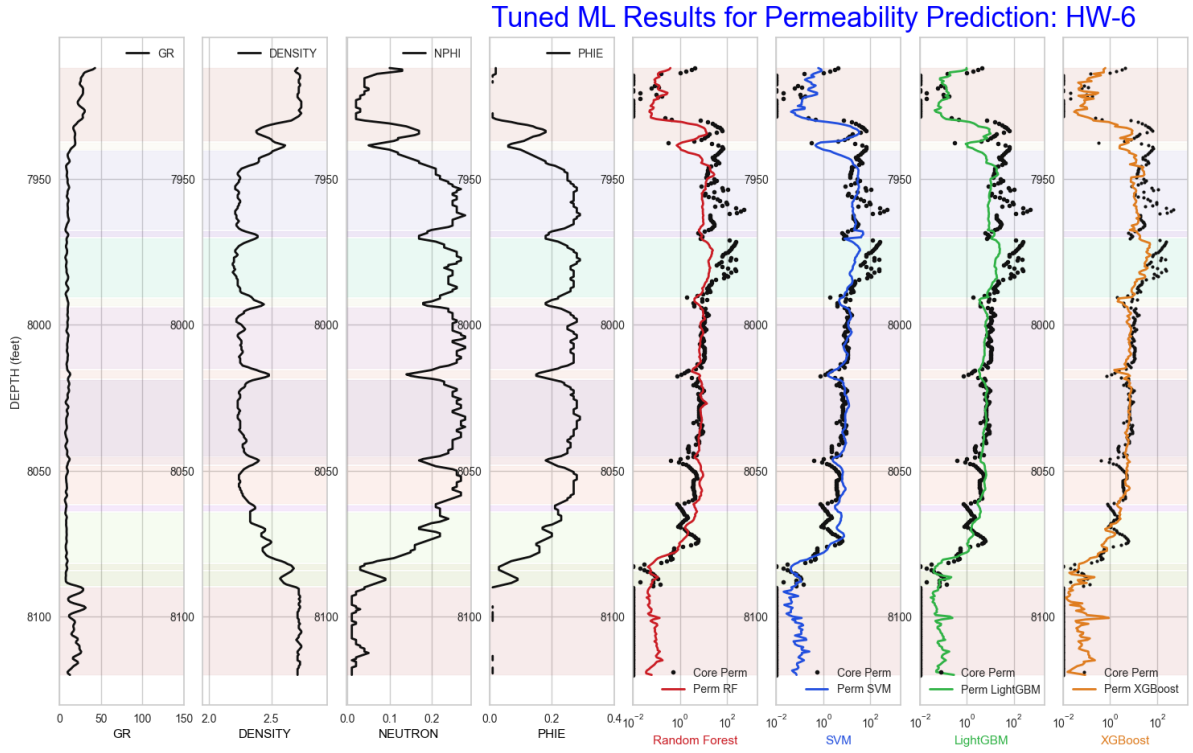


Figure 4: Prediction Results for Validation WELL : HW-6

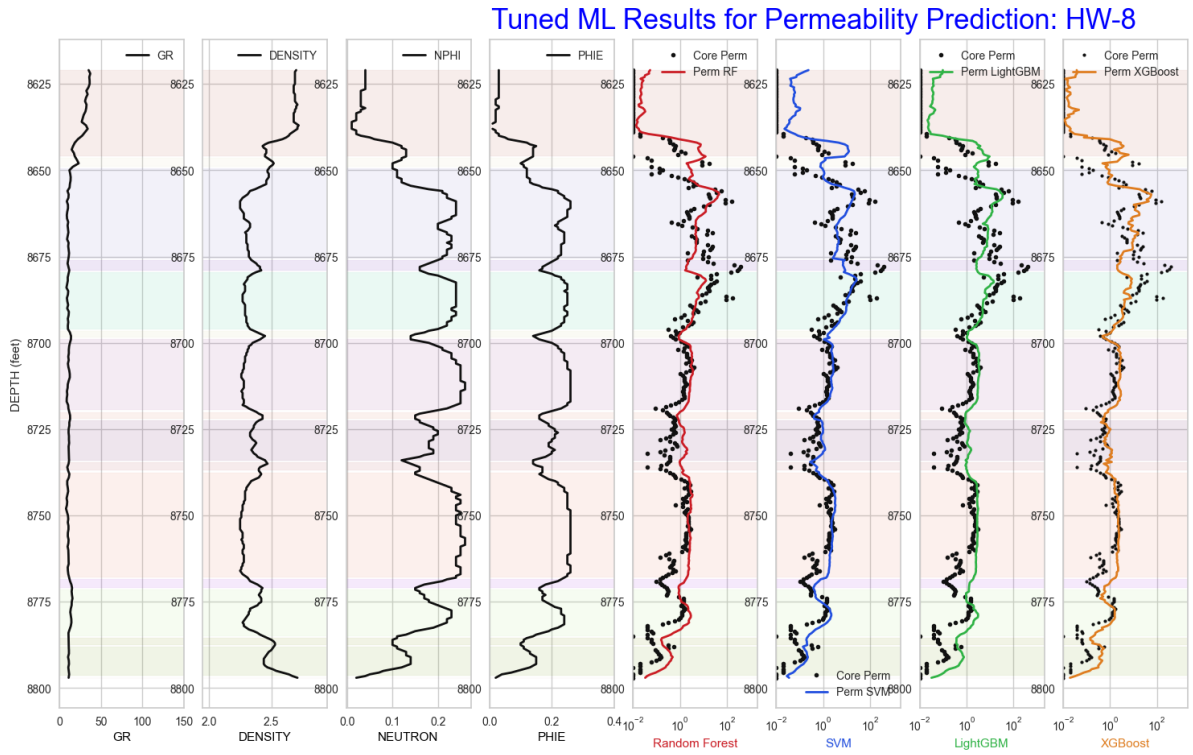


Figure 5: Prediction Results for Validation WELL : HW-8

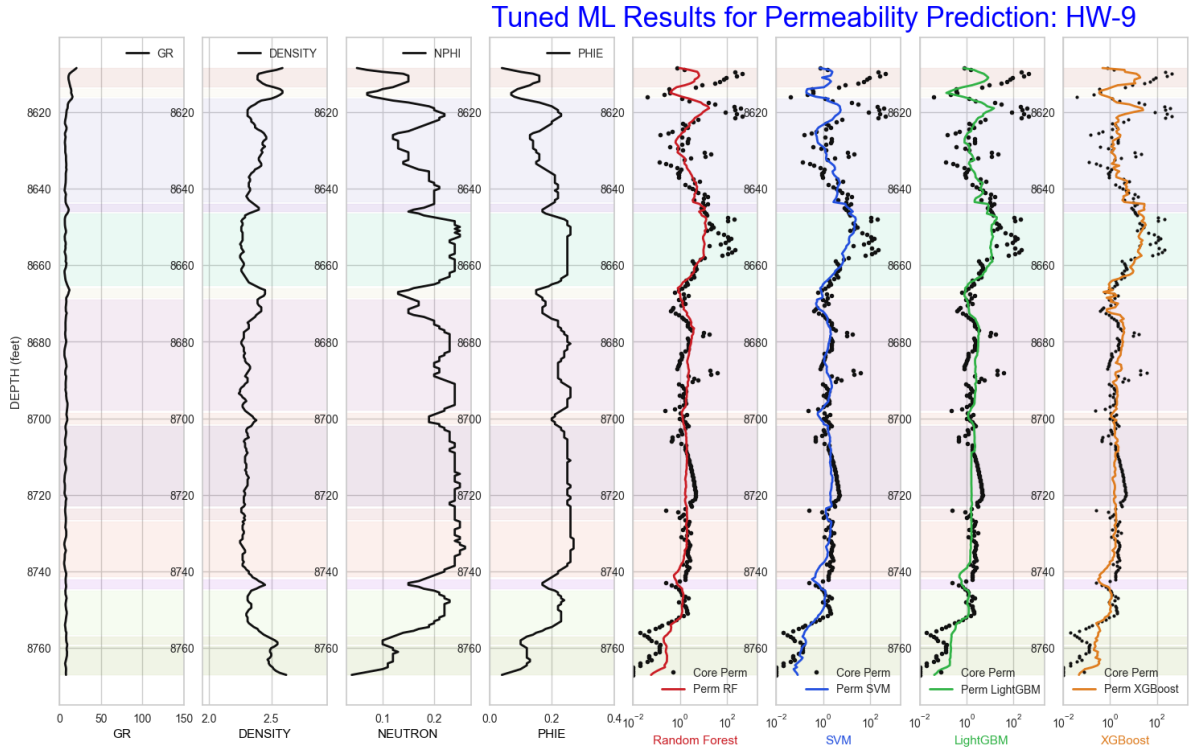


Figure 6: Prediction Results for Validation WELL : HW-9

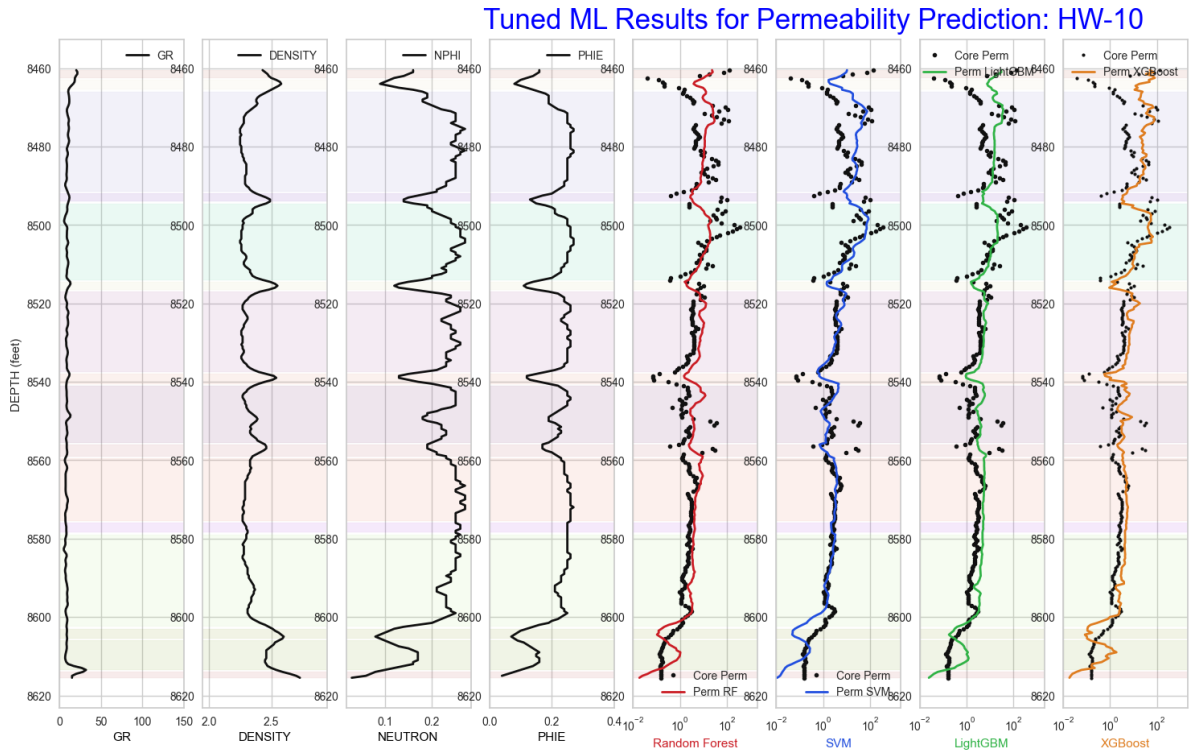


Figure 7: Prediction Results for Validation WELL : HW-10

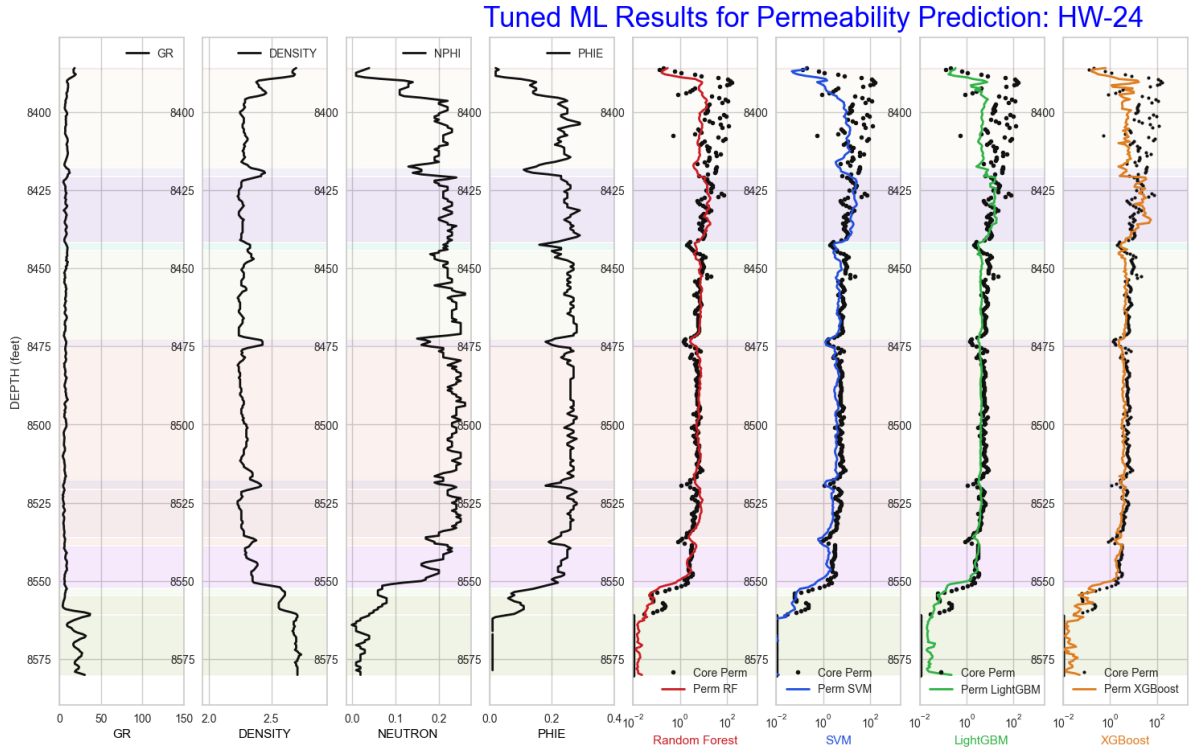


Figure 8: Prediction Results for Validation WELL : HW-24

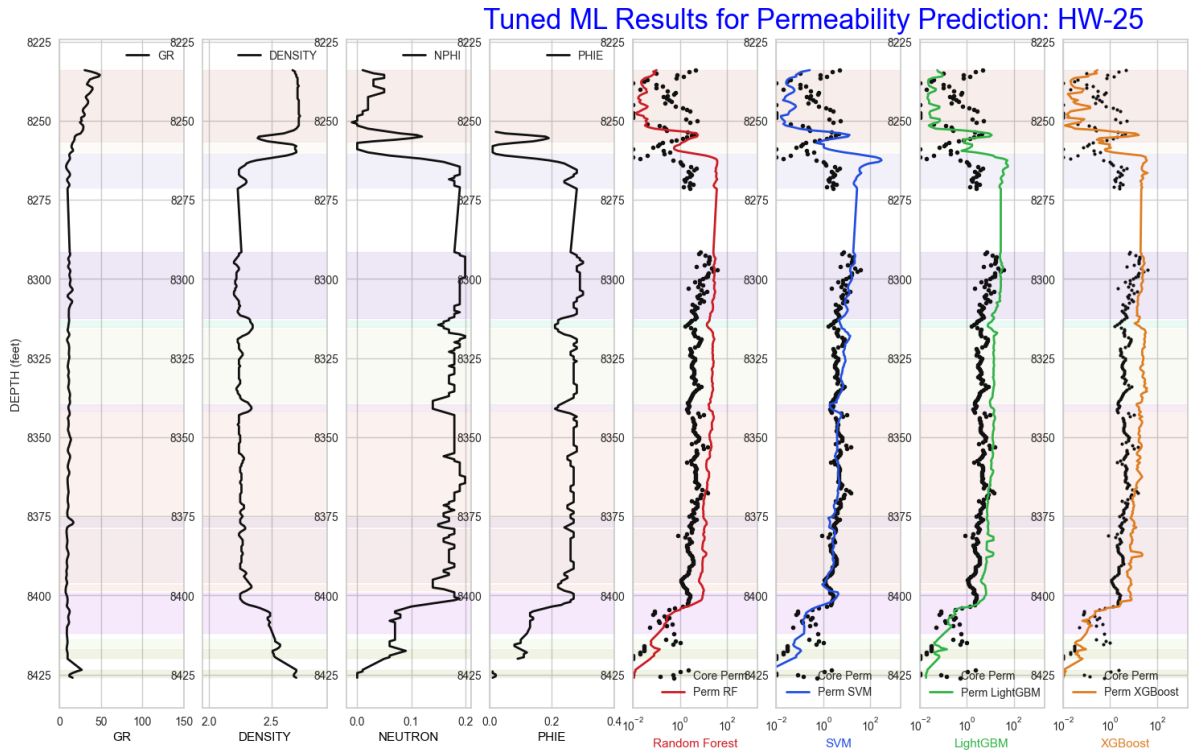


Figure 9: Prediction Results for Validation WELL : HW-25

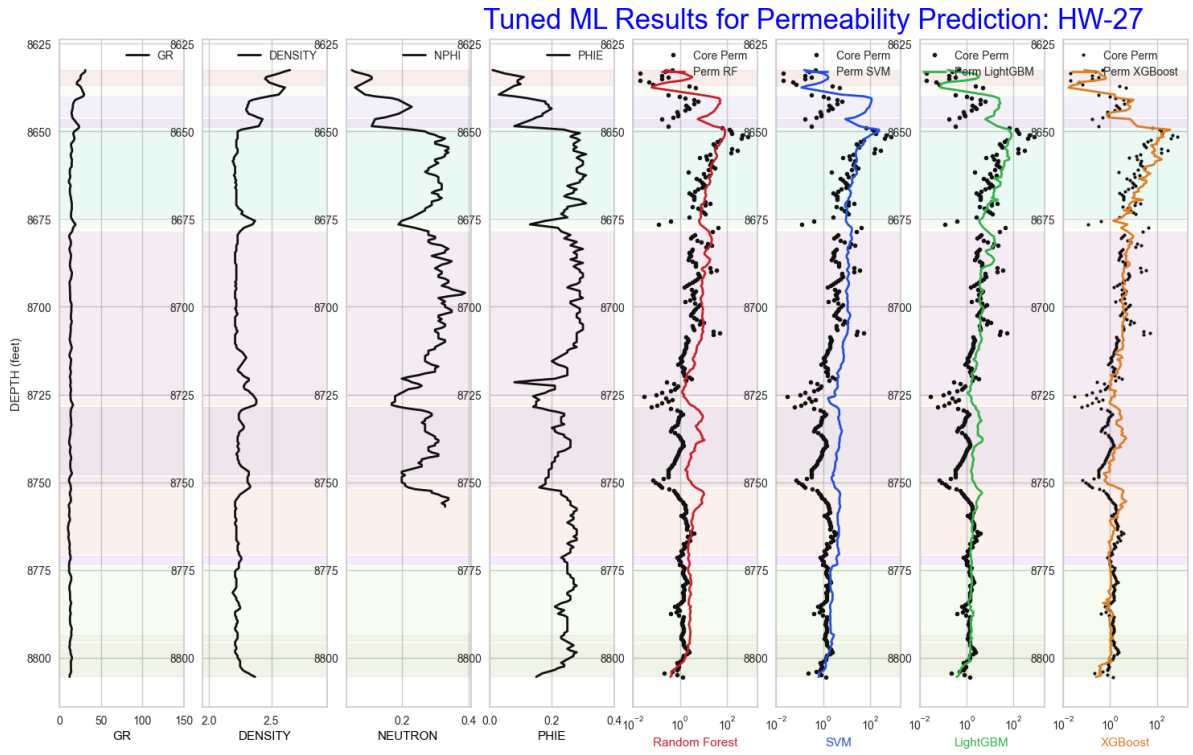


Figure 10: Prediction Results for Validation WELL : HW-27

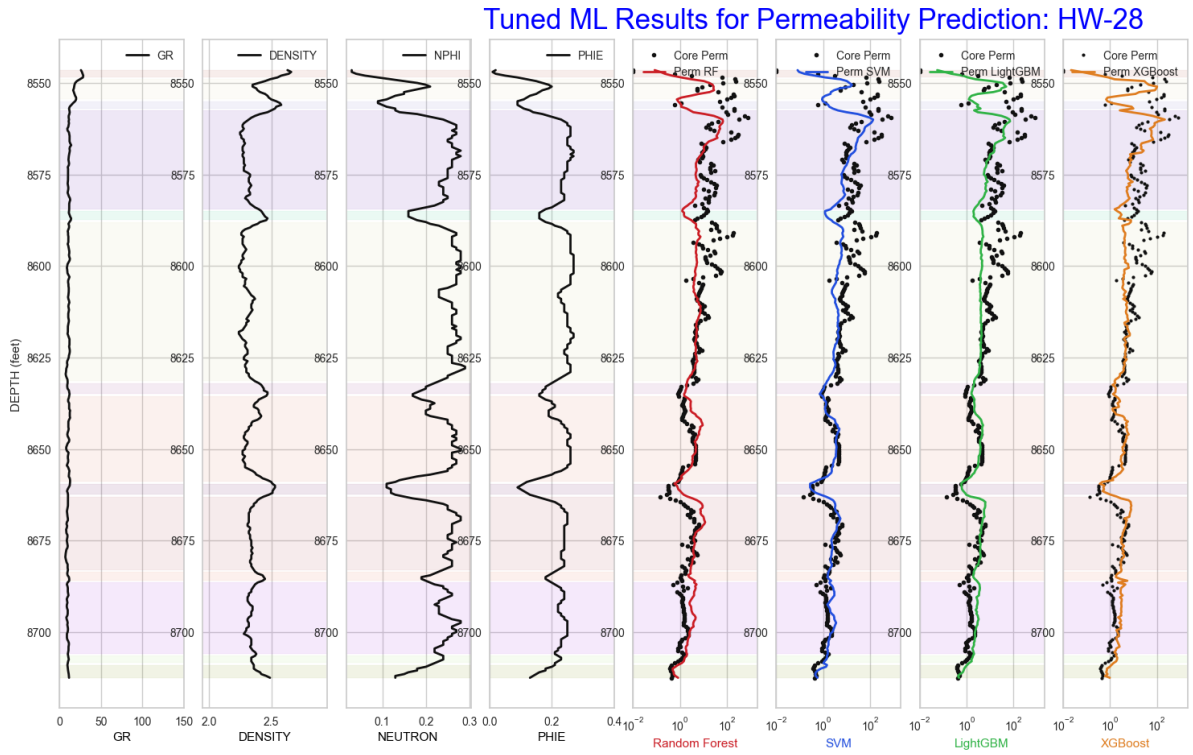


Figure 11: Prediction Results for Validation WELL : HW-28

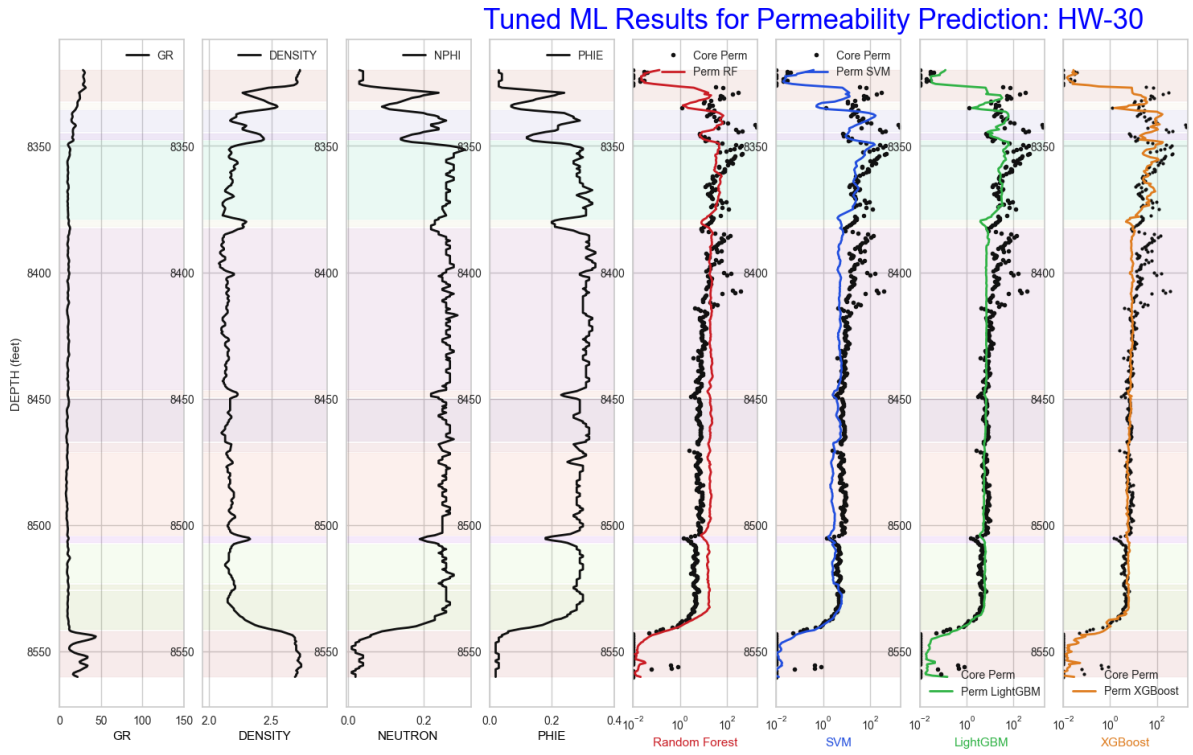


Figure 12: Prediction Results for Validation WELL : HW-30

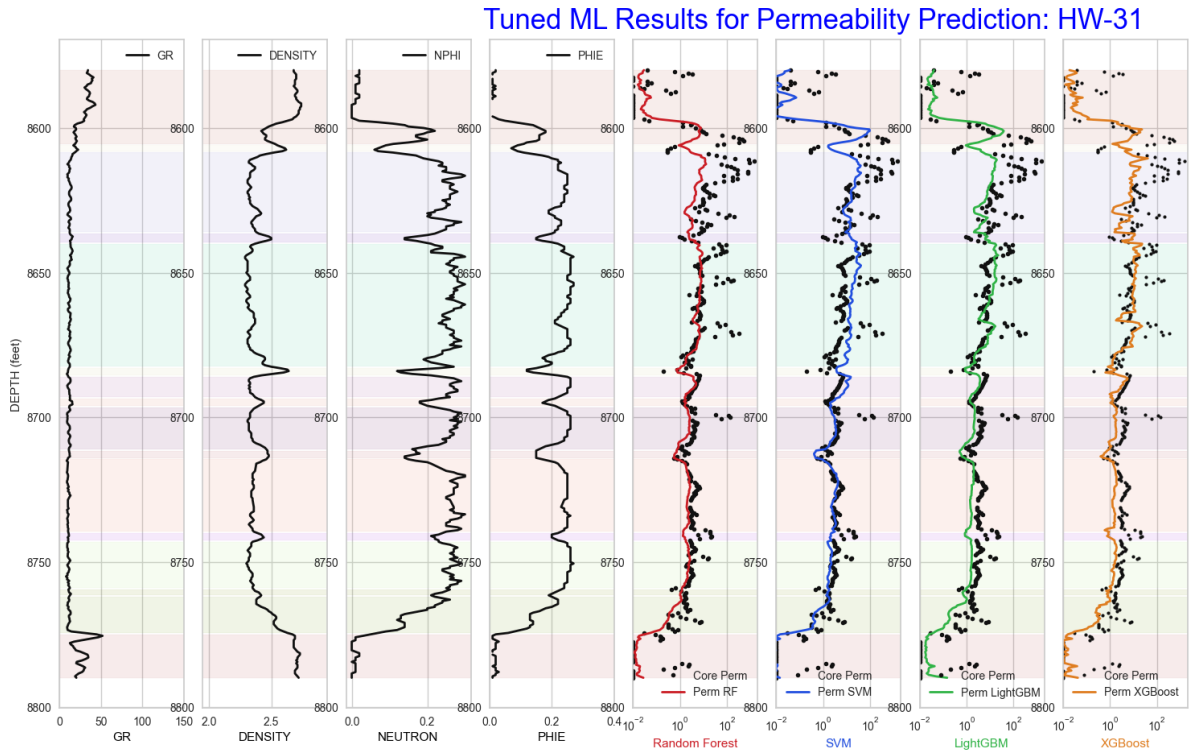


Figure 13: Prediction Results for Validation WELL : HW-31

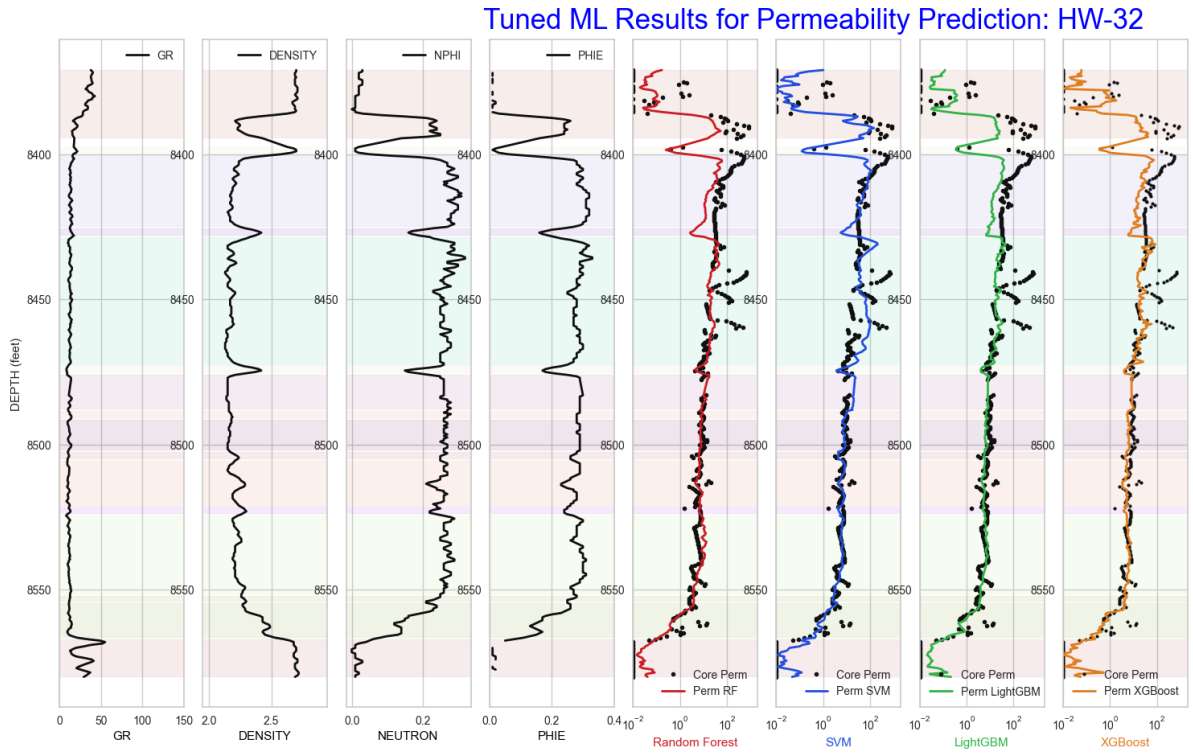


Figure 14: Prediction Results for Validation WELL : HW-32

Hyperparameters For Costa Dataset

Hyperparameter	Value
colsample_bytree	0.8888888888888888
gamma	0.6666666666666666
learning_rate	0.09
min_child_weight	12
n_estimators	100

Table 1: Hyperparameters for XGBRegressor

Hyperparameter	Value
bagging_fraction	0.7
bagging_freq	5
feature_fraction	0.7
learning_rate	0.01
min_child_samples	11
min_split_gain	0.3
n_estimators	260
num_leaves	20
reg_alpha	0.0001
reg_lambda	1

Table 2: Hyperparameters for LGBMRegressor

Hyperparameter	Value
bootstrap	False
max_depth	70
max_features	'sqrt'
min_samples_leaf	4
min_samples_split	10
n_estimators	400

Table 3: Hyperparameters for RandomForestRegressor

Hyperparameter	Value
kernel	'rbf'
degree	3
gamma	'scale'
coef0	0.0
tol	0.001
C	1.0
epsilon	0.1
shrinking	True
cache_size	200

Table 4: Hyperparameters for SVR

Additional Plots

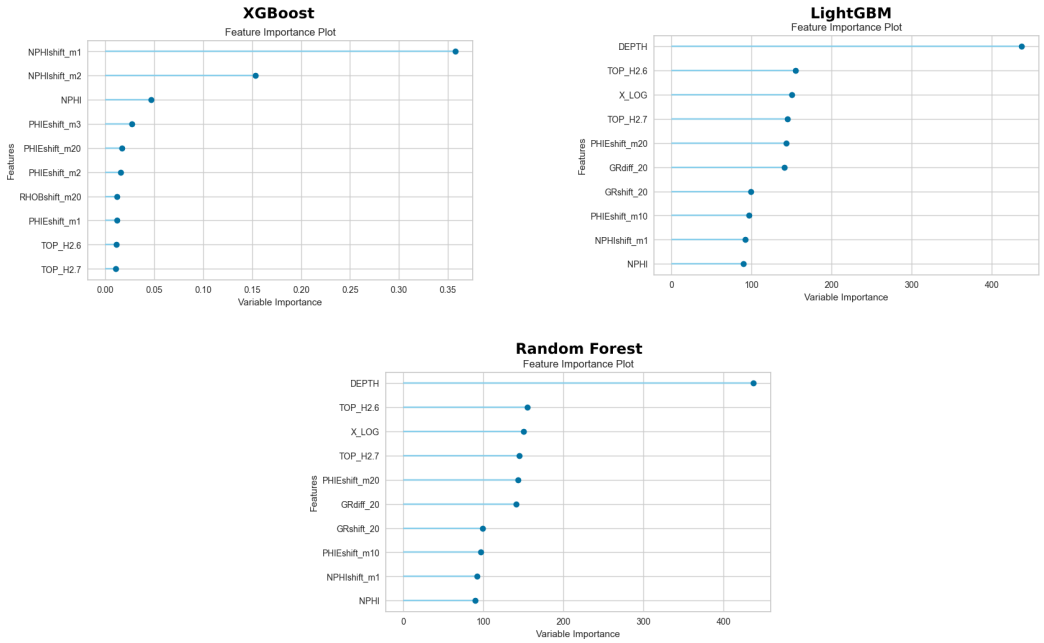


Figure 15: Feature Importances for Tree-Based Algorthims

SHAP statstics

	PHIE	DEPTH	NPHI	GR	log10RT	TOP_H2.6	Y_LOG	RHOB	TOP_H2.7	TOP_H2 Base	dep_location_Cluster2	X_LOG	TOP_H2 Top	NDS	TOP_H2.1	dep_location_Cluster3	TOP_H2.6 D	TOP_H2.4 D	TOP_H2.3	TOP_H2.3 D
mean	1.39	1.20	1.16	1.07	1.06	1.03	1.03	1.03	1.02	1.02	1.02	1.01	1.01	1.01	1.01	1.00	1.00	1.00	1.00	1.00
std	0.83	0.74	0.37	0.25	0.15	0.38	0.23	0.21	0.32	0.08	0.13	0.17	0.18	0.07	0.04	0.02	0.13	0.00	0.01	0.00
min	0.02	0.10	0.26	0.73	0.69	0.75	0.38	0.45	0.77	0.56	0.81	0.74	0.79	0.65	0.63	0.87	0.94	0.96	0.94	0.93
25%	0.76	0.72	1.12	0.94	0.98	0.86	0.91	0.88	0.91	1.01	0.95	0.95	0.96	0.97	1.01	1.00	0.99	1.00	1.00	1.00
50%	1.70	1.12	1.28	0.98	1.04	0.91	1.07	1.10	0.93	1.02	0.97	0.99	0.98	0.99	1.01	1.00	0.99	1.00	1.00	1.00
75%	2.02	1.44	1.36	1.09	1.13	0.95	1.17	1.17	0.95	1.02	1.08	1.03	0.98	1.03	1.01	1.01	0.99	1.00	1.00	1.00
max	3.42	6.17	4.77	4.85	3.27	2.78	1.99	1.93	4.25	1.33	2.79	3.74	3.92	1.62	1.08	1.06	3.57	1.00	1.01	1.00

Figure 16: SHAP statistics for Xgboost model using Costa Dataset.

	PHIE	NPHI	DEPTH	GR	log10RT	TOP_H2.6	RHOB	TOP_H2.7	TOP_H2 Base	Y_LOG	TOP_H2 Top	X_LOG	dep_location_Cluster2	TOP_H2.1	TOP_H2.6 D	TOP_H2.2	TOP_H2.3	dep_location_Cluster3	TOP_H2.4	TOP_H2.1 D
mean	1.30	1.11	1.10	1.04	1.04	1.03	1.03	1.02	1.02	1.01	1.01	1.01	1.01	1.00	1.00	1.00	1.00	1.00	1.0	1.00
std	0.73	0.36	0.44	0.15	0.12	0.37	0.17	0.31	0.07	0.10	0.14	0.17	0.14	0.03	0.15	0.02	0.01	0.01	0.0	0.00
min	0.07	0.23	0.31	0.84	0.83	0.81	0.56	0.87	0.62	0.62	0.91	0.73	0.82	0.78	0.96	0.92	0.96	0.96	1.0	0.99
25%	0.56	1.10	0.80	0.96	0.95	0.86	0.92	0.90	1.02	0.98	0.97	0.93	0.90	1.01	0.98	1.00	1.00	1.00	1.0	1.00
50%	1.59	1.28	1.08	0.98	1.01	0.90	1.10	0.92	1.02	1.04	0.98	1.01	0.94	1.01	0.99	1.00	1.00	1.00	1.0	1.00
75%	1.89	1.34	1.32	1.03	1.11	0.94	1.16	0.95	1.03	1.08	0.99	1.04	1.14	1.01	0.99	1.01	1.00	1.01	1.0	1.00
max	2.61	2.07	2.87	2.36	1.99	2.37	1.35	2.38	1.16	1.41	2.51	2.97	1.62	1.03	3.10	1.01	1.01	1.01	1.0	1.00

Figure 17: SHAP statistics for LightGBM model using Costa Dataset.

The Volve Dataset

Hyperparameters For The Volve Dataset

Table 5: Hyperparameters for XGBRegressor

Hyperparameter	Value
booster	'gbtree'
colsample_bytree	0.8888888888888888
gamma	0.4444444444444444
learning_rate	0.36
max_depth	80
min_child_weight	23
n_estimators	100
n_jobs	-1

Table 6: Hyperparameters for LGBMRegressor

Hyperparameter	Value
bagging_fraction	1.0
bagging_freq	2
device	'gpu'
feature_fraction	0.4
learning_rate	0.2
min_child_samples	11
min_split_gain	0.2
n_estimators	70
num_leaves	4
random_state	123
reg_alpha	0.5
reg_lambda	1e-06

Table 7: Hyperparameters for RandomForestRegressor

Hyperparameter	Value
bootstrap	False
max_depth	70
max_features	'sqrt'
min_samples_leaf	4
min_samples_split	10
n_estimators	400

Hyperparameter	Value
kernel	'rbf'
degree	3
gamma	'scale'
coef0	0.0
tol	0.001
C	1.0
epsilon	0.1
shrinking	True
cache_size	200

Table 8: Hyperparameters for SVR

Volve Field

	GR	RHOB	log10 ILD	log10 ILM	PHIF	VSH	DTC	DTS	NPHI	TOP Skagerrak Fm. Top	TOP Hugin Fm. Top	TOP Sleipner Fm. Top
mean	1.40	1.64	1.03	1.16	1.35	1.00	1.01	1.07	1.00	1.01	1.00	1.02
std	1.36	1.24	0.14	0.62	0.89	0.41	0.15	0.33	0.16	0.11	0.00	0.05
min	0.20	0.06	0.65	0.61	0.09	0.32	0.63	0.38	0.54	0.76	1.00	0.88
25%	0.51	0.37	0.91	0.77	0.46	0.70	0.91	0.84	0.90	0.95	1.00	1.01
50%	0.82	1.19	1.02	0.95	1.57	0.93	0.98	0.99	0.99	1.03	1.00	1.01
75%	1.96	2.64	1.11	1.37	2.05	1.27	1.08	1.23	1.09	1.05	1.00	1.04
max	6.20	5.09	1.49	4.68	3.12	2.19	1.55	2.64	1.76	1.18	1.00	1.17

Table 9: SHAP values for the XGBoost model

	GR	RHOB	log10 ILD	log10 ILM	PHIF	VSH	DTC	DTS	NPHI	TOP Sk-ager-rak Fm. Top	TOP Hugin Fm. Top	TOP Sleip-ner Fm. Top
mean	1.09	1.40	1.09	1.03	1.22	1.24	1.00	1.03	1.04	1.00	1.14	1.00
std	0.51	0.93	0.34	0.26	0.65	0.82	0.10	0.25	0.20	0.00	0.56	0.08
min	0.41	0.02	0.57	0.71	0.10	0.23	0.83	0.20	0.64	1.00	0.51	0.94
25%	0.62	0.45	0.86	0.81	0.50	0.68	0.95	0.89	0.95	1.00	0.59	0.97
50%	0.97	1.20	0.92	0.90	1.54	0.89	0.97	0.97	1.02	1.00	1.25	0.99
75%	1.64	2.47	1.30	1.21	1.73	1.81	0.98	1.22	1.07	1.00	1.68	0.99
max	2.34	2.69	2.70	1.77	2.27	3.92	1.44	1.77	2.01	1.00	1.95	1.38

Table 10: SHAP values for the LightGBM model

Depth Shifing Volve Dataset

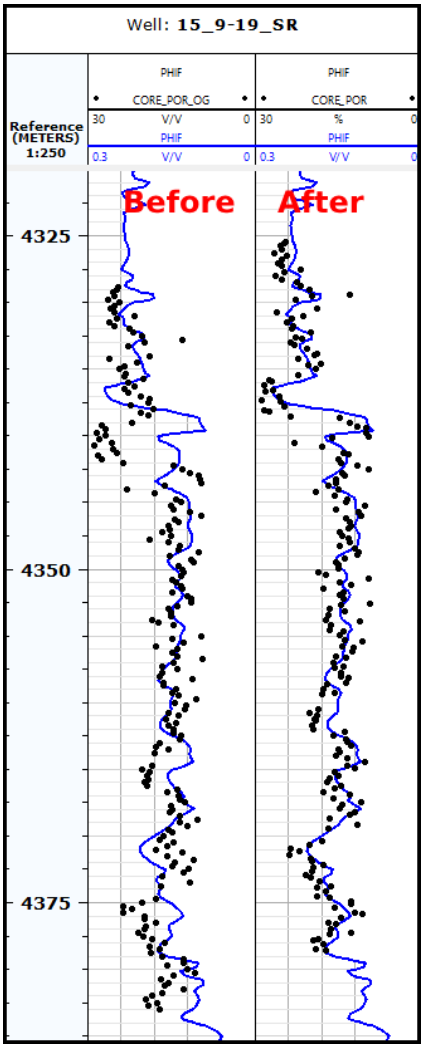


Figure 18: Example of well 15/9-19_SR after before and after Depth shifting Core porosity

WELL	Core Id	Top Depth (meters)	Base Depth (meters)	Top Shift Depth (meters)	Base Shift Depth (meters)
15/9-19 SR	1	3643.0	3647.5	3643.0	3647.5
15/9-19 SR	2	4328.85	4354.2	4325.384	4350.734
15/9-19 SR	3	4355.0	4383.0	4350.654	4378.654
15/9-19 BT2	1	4037.0	4049.77	4037.6	4050.37
15/9-19 BT2	2	4054.5	4071.8	4053.4	4070.7
15/9-19 BT2	3	4072.0	4089.95	4070.6	4088.55
15/9-19 BT2	4	4090.5	4109.2	4088.5	4107.2
15/9-19 A	1	3837.0	3852.2	3838.6	3853.8
15/9-19 A	2	3854.0	3881.45	3854.2	3881.65
15/9-19 A	3	3881.5	3908.25	3882.1	3908.85
15/9-19 A	4	3908.5	3931.92	3909.1	3932.52
15/9-19 A	5	3935.5	3963.0	3935.3	3962.8
15/9-19 A	6	3963.0	3991.0	3963.0	3991.0
15/9-19 A	7	3991.0	4016.7	3991.2	4016.9

Table 11: Depth Shift Table for Volve Field Wells (Equinor, 2022)

Additional Plots

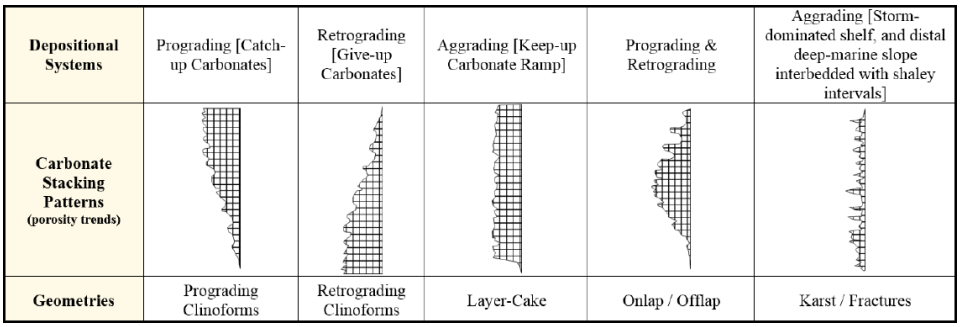


Figure 19: Additional Feature Engineering Suggested for Carbonate Data-sets (Costa Gomes et al., 2022)

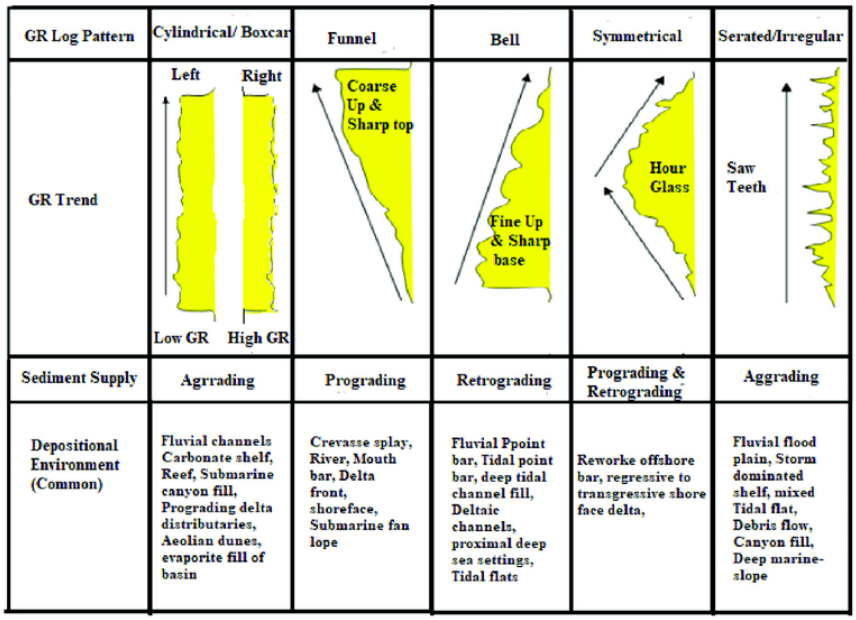


Figure 20: Additional Feature Engineering Suggested for Clastic Data-sets (Radwan, 2020)