FACIES CLASSIFICATION FROM WELL LOG DATA

NAME: MICHELLE KAVETZA

22ND February, 2025

INTRODUCTION

Business Problem

- The accurate classification of reservoir rock types is critical in the oil and gas industry for optimizing strategies, improving hydrocarbon recovery and reduce operational costs.

Objective

To develop a machine learning model to classify reservoir facies based on well log data. By predicting facies accurately, the goal is to support data driven decisions in exploration and procuction and maximizing resource extraction efficiency.

CONT….

Approach

1.Utilize well log data to build and evaluate various classification models.

Develop a machine learning model to classify facies with high accuracy.

2. Enhance the accuracy of geological interpretations for improved exploration and production decisions.

3. Identify key features from the well log data that influence facies classification.

Data Understanding

Overview

- Dataset contains well log data collected from multiple wells, including geophysical measurements that are essential for facies classification.
- Resistivity Logging(ILD_LOG10) - Logs the electrical resistivity of rock formations.
- Photoelectric Effect(PE) - Indicates mineral composition.
- Neutron-Density Porosity(DeltaPHI) - Difference between neutron and density porosity.
- Average Neutron-Density Pororsity(PHID) - Average neutron and density porosity.
- Nonmarine-Marine Indicator(NM_M) - Distinguishes between marine and nonmarine environments.
- Relative Position(RELPOS) - Represents the relative stratigraphic position within the reservoir.
- Facies – target variablre

 - Represents different resrvior rock types serving as target variable for classification

Observations

- The dataset contains multiple facies classes, which is a multi class classification problem.
- Presence of class imbalance which can impact performance was solved using SMOTE technique.
- Missing values and anomalies identified during EDA.
- Certain features like GR, ILD_log10 and PHIND showed strong correlations with specific facies.
- Outliers and missing data were handles to improve data quality and model accuracy.

Modeling

Model Selection

- Explored both Logistic Regression and Decision Trees.

Training

- Trained multiple models using the SMOTE balanced dataset.

Tuned hyperparameters for optimal features.

Evaluation Metrics

- Accuracy, Precision, Recall, F1 score
- Confusion matrix for class wise performance analysis.

Logistic Regression

- Attained an accuracy score of 53.25%

Performance Overview

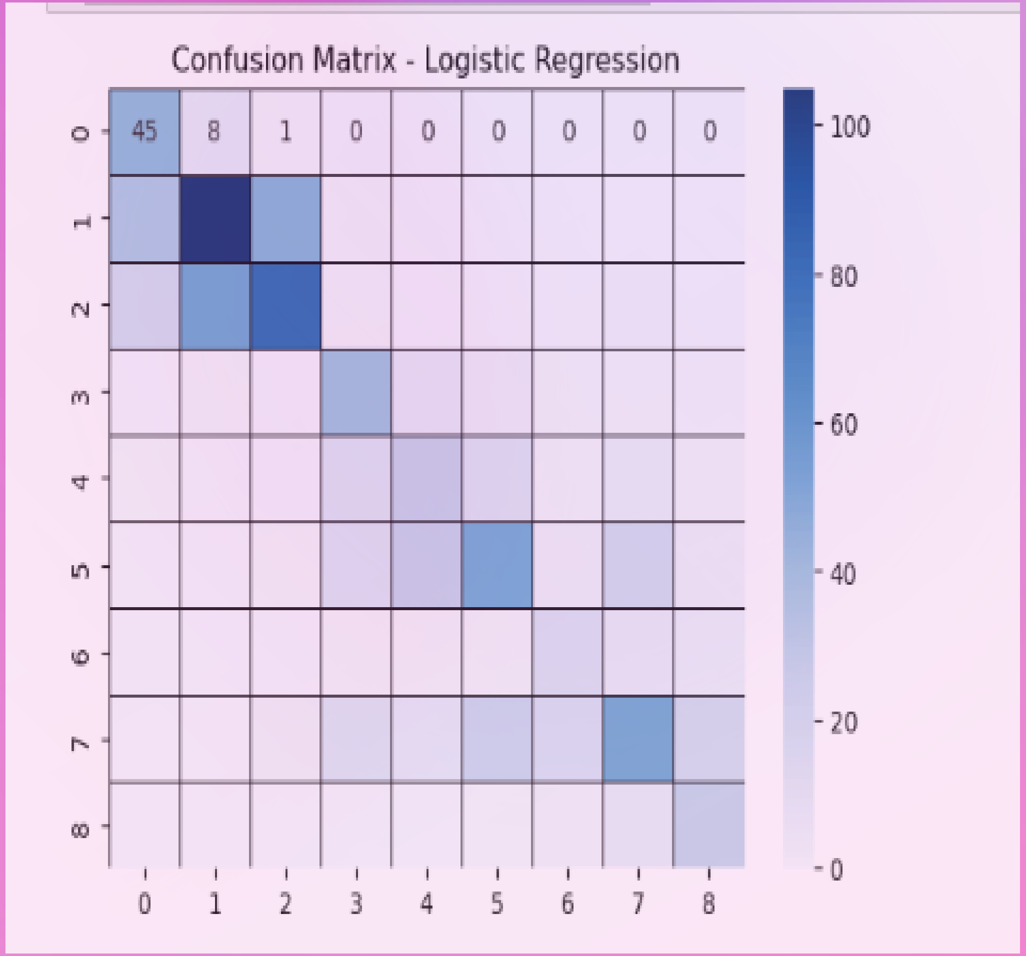
- The logistic Regression model achieved an accuracy of 53.25%
- High recall for class 0 (0.83) indicates effective identification of specific lithologies.
- Poor performance in Classes 5 and 8 could misclassify crucial reservoir rocks.
- The macro average fi score of 0.52 suggests moderate but even performance.
- Logistic Regression linearity limits its ability to capture complex geological patterns.

Geological impact

- Effective identification of distinct lithologies, Class 0 benefits exploration.
- Misclassification risks in Class 5 & 8 could impact drilling and reservoir patterns.
- Geological data often exhibits non linear patterns.

Confusion Matrix

- Visualises how well the L.R model classifies each facies.
- Diagonal cells show correct classification with Class 0 having the highest count(45)
- Off diagonal cells represent misclassification indicating where the model struggled.
- Classes 2,5,7 show several misclassifications.



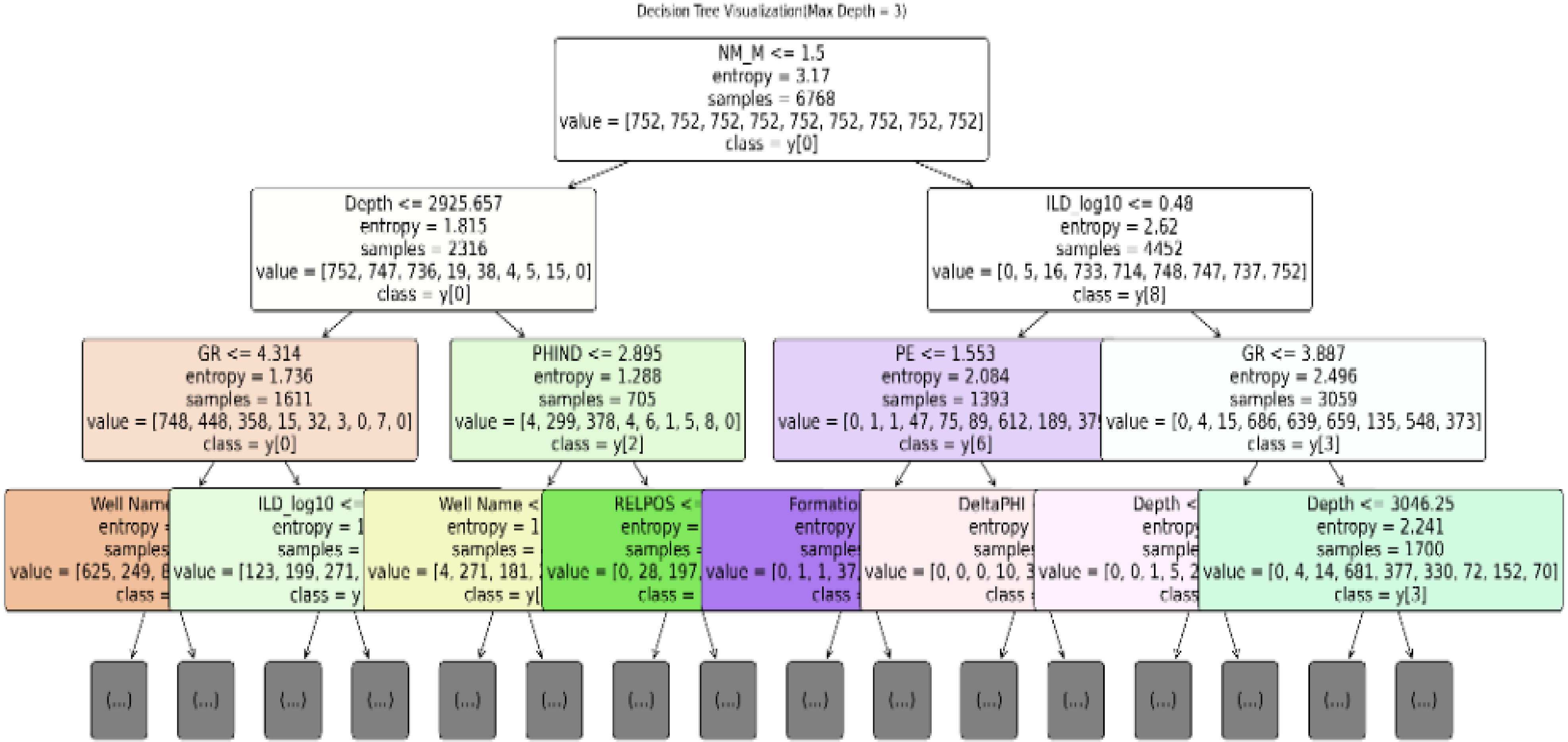
Model Evaluation

- The Decision tree achieved strong performance with an accuracy of 73.13% outperforming L.R.
- High f1 score for Key classes 1,2,3,8 suggest balanced precision and recall.
- Macro weighted average of 0.73 highlight consistent handling of class imbalance.

Geological impact

- Accurate identification of dominant rock types, Class2,3,8 supports better reservoir modeling.
- Misclassification risks in Class 5 could affect drilling and evaluation.
- D.T captures complex, non linear interactions in geological data.
- Achieves higher accuracy and better feature interpretability compared to logistic regression.

Decision Tree



Hyperparameter Tuning

- Optimised Decision Tree after hyper tuning attained an accuracy of 73.61%.
- Improved accuracy post tuning.
- Strong performance for Classes 1,2,3, 8.
- Class 5 showed improvement but remains the least accurate.
- Enhanced ability to predict complex lithological variations.
- High recall in Class 8 supports identification of distinct geological formations.
- This improved model precision and recall.

Random Forest

- Achieved an accuracy of 81.33%.
- Significant improvement over Logistic Regression(53.25%) and Decision Tree(73.61%).
- High F1 score for classes 1(0.89), 2(0.87), 3(0.79) and 8(0.92).
- Class 5 improved f1score (0.72) but remains a moderate performer.

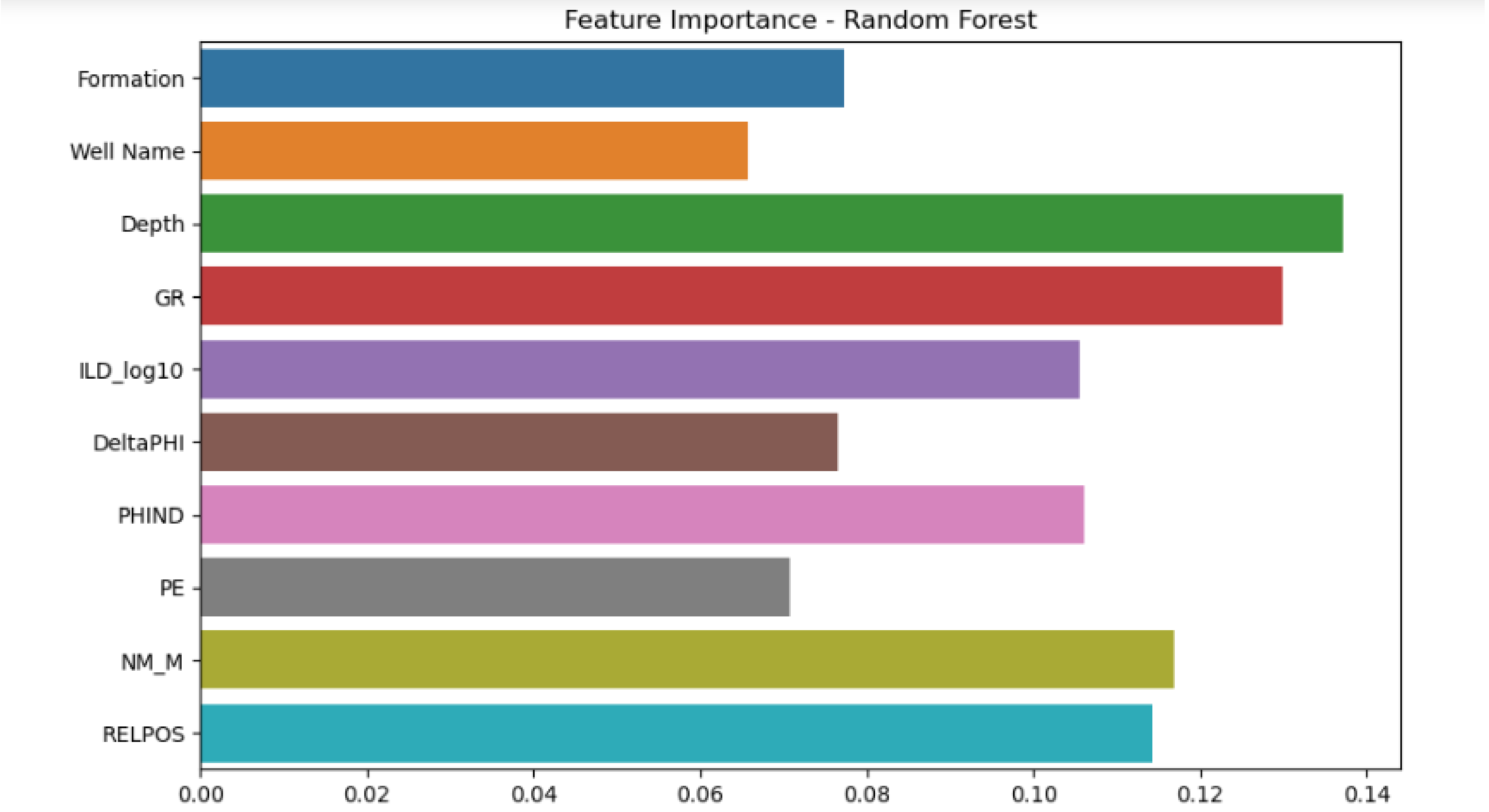
Geological Impact

- Enhanced consistency and accuracy in classifying distinct reservoir rock types.
- Strong results for Classes 1 and 8 indicate clear lithological boundaries.

Justification

- Best performing model with robust handling of nonlinear relationships.
- Suitable for complex geological datasets, offering superior generalization and reduced overfitting.

Feature Importance



Interperation

- Depth and GR are the most influential features indicating their strong role in facies differentiation.
- PHIND, NM_M and RELPOS also show significant importance reflecting their relevance in geological formations.
- Less impactful features like Well Name suggest limited variability across wells.
- The model's focus on Depth and GR highlights their geological significance in classifying reservoir facies.
- Enhanced accuracy supports more reliable reservoir characterization.
- Random Forest's ensemble approach effectively handles complex, nonlinear relationships.
- Its superior accuracy and balanced performance across classes make it the best choice for this dataset.

Recommendation

Adopt Random Forest Model:

Due to its high accuracy and strong generalization capabilities, Random Forest is recommended for operational deployment.

Focus on Key Features:

Prioritise Depth, GR and PHIND in future data collection and analysis to enhance model accuracy.

Address Misclassified Facies:

Investigate geological factors causing misclassification in Class 5 and 6 for improved model precision.

Next Step

Data Enrichment:

Incorporate additional geological and geophysical data to further refine the model.

Model Monitoring:

Implement continuous monitoring to ensure model stability and update as new data becomes available.

Explore Advanced Techniques:

Evaluate deep learning models and ensemble strategies for potential performance gains.

THANK YOU!