**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 4, March 2026 Page: 2491−2506
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9302

# Comparison of Geothermal Well Productivity Using KNN, SVM and Gradient Boost Methods

**Mina Winawati Dwi Aryani*, Indra Nugraha Abdullah**

Faculty of Information Technology, Master of Computer Science Study Program, Budi Luhur University, Jakarta, Indonesia
Email: 1,*minadwiaryani@gmail.com, 2indra.nugraha@budiluhur.ac.id
Correspondence Author Email: minadwiaryani@gmail.com
Submitted: **28/01/2026**; Accepted: **19/03/2026**; Published**: 19/03/2026**

**Abstract**−Manual and conventional processing of geothermal well production data is computationally inefficient and requires several hours to days to generate productivity assessments, particularly when dealing with large-scale and non-linear operational datasets. The complexity of geothermal production parameters this paper such as wellhead pressure (WHP), enthalpy, steam flow, brine flow, total flow, and generated power this paper creates challenges for accurate and timely productivity classification at the well level. This study utilizes 74,912 daily production records collected from January 2018 to June 2024, comprising 13 operational and production-related attributes. The objective is to identify the most effective machine learning algorithm for classifying geothermal well productivity levels to support faster and more reliable operational decision-making. A comparative machine learning classification approach was conducted using K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), and Gradient Boosting. Model evaluation was performed using three train–test split ratios: 70:30, 80:20, and 90:10. Two modelling scenarios were implemented: with and without Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance. The results indicate that the K-NN model achieved the highest classification performance, reaching 94.22% accuracy using the 90:10 split ratio without SMOTE. Gradient Boosting demonstrated stable performance across all ratios, with its best accuracy of 91.39% at the 70:30 split without SMOTE. In contrast, SVM produced the lowest performance, with a maximum accuracy of 79.78% at the 90:10 ratio without SMOTE. The application of SMOTE improved minority class recall, particularly for SVM, but generally reduced overall model accuracy for K-NN and Gradient Boosting. These findings demonstrate that classical machine learning algorithms, particularly K-NN, provide an efficient and accurate solution for geothermal well productivity classification. The proposed approach significantly reduces processing time compared to conventional analytical methods and supports data-driven decision-making in geothermal production forecasting and development planning.

**Keywords:** Geothermal Well Productivity; Geothermal Production Data; Machine Learning; SMOTE; Support Vector Machine

## 1. INTRODUCTION

Daily geothermal well production data constitute primary operational data that directly reflect reservoir performance and well productivity behavior. Parameters such as wellhead pressure, enthalpy, steam flow, brine flow, total mass flow, and generated power form high-frequency time series datasets that accumulate rapidly over years of operation [1]. In large geothermal fields, these records reach tens of thousands of observations, creating a Big Data context characterized by high volume, multivariate structure, and nonlinear interdependencies among variables [2]. Accurate processing of these primary data is essential for estimating well productivity levels, forecasting future steam supply, planning pipeline capacity, and prioritizing well monitoring strategies [3]. However, conventional analytical approaches typically based on manual interpretation, feasibility study workflows, and deterministic reservoir simulations require extensive computational time and expert intervention [4]. Processing may take hours to days, limiting responsiveness in operational decision making, especially under increasing electricity demand and production optimization pressures. These limitations highlight the need for data driven methods capable of handling nonlinear production patterns efficiently and consistently [5]. To address these challenges, this study investigates machine learning–based classification for geothermal well productivity, with particular attention to two core issues: algorithm suitability for nonlinear daily production data and class imbalance within productivity categories (Low, Medium, High) [6].

Like Artificial Neural Networks (ANN), which often require large computational resources, extensive hyperparameter tuning, and risk overfitting on structured tabular data, classical machine learning algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Gradient Boosting offer theoretical and practical advantages for this case [7]. KNN is instance-based and effective for capturing local similarity patterns in multivariate operational data without complex model assumptions. SVM provides strong generalization capability in nonlinear classification through kernel functions, making it suitable for separating overlapping productivity classes [8]. Gradient Boosting, as an ensemble tree-based method, handles nonlinear interactions and feature importance effectively while maintaining computational efficiency on structured tabular datasets [9]. These three algorithms represent fundamentally different learning paradigms distance-based, margin-based, and ensemble boosting making them methodologically valuable to compare for geothermal productivity classification [10]. Furthermore, because geothermal productivity datasets often exhibit class imbalance (dominance of Medium class wells), this study explicitly evaluates the impact of the Synthetic Minority Over sampling Technique (SMOTE) to assess whether balancing strategies improve minority-class prediction without degrading overall model stability [11]. By integrating algorithm comparison with imbalance handling, this research aims to identify a computationally efficient and practically robust classification framework for well-level geothermal decision support [12].

Yehia (2024) also relied on ANN-based modelling to estimate production performance and optimize flow rate and power plant capacity [13]. Despite achieving satisfactory prediction performance, these studies concentrated on forecasting system outputs and did not evaluate alternative classification algorithms that may provide more efficient computation or improved interpretability. In addition, most prior works predominantly relied on single-model approaches, which restrict the ability to determine the most appropriate algorithm for different geothermal datasets. Overall, the existing literature reveals several key gaps. First, limited studies focus specifically on classification of geothermal well productivity levels at the operational decision-making scale. Second, comparative evaluations among multiple classification algorithms, particularly classical machine learning methods such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Gradient Boosting, remain scarce. Third, processing efficiency and practical implementation considerations, such as faster computation for real-time decision support, are rarely emphasized. Fourth, previous research tends to prioritize reservoir or system-level prediction rather than well-level productivity classification using integrated operational and production datasets [14].

Based on these considerations, this study analyzes geothermal well production data to classify well productivity levels using a comparative evaluation of classification algorithms. The objective is to identify the most suitable method to improve classification accuracy and enable faster, more efficient production prediction simulations [15]. This study contributes to geothermal data analytics by empirically comparing KNN, SVM, and Gradient Boosting to identify the most effective algorithm for classifying geothermal well productivity. It emphasizes well-level operational decision support and proposes a more efficient machine learning–based approach compared to conventional methods, enabling faster and more reliable production forecasting and strategic planning.

## 2. RESEARCH METHODOLOGY

### 2.1 Research Methodology Stages

This research applies the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework to ensure that the analytical process is systematic, measurable, and aligned with the research objectives. The methodology consists of the following stages:

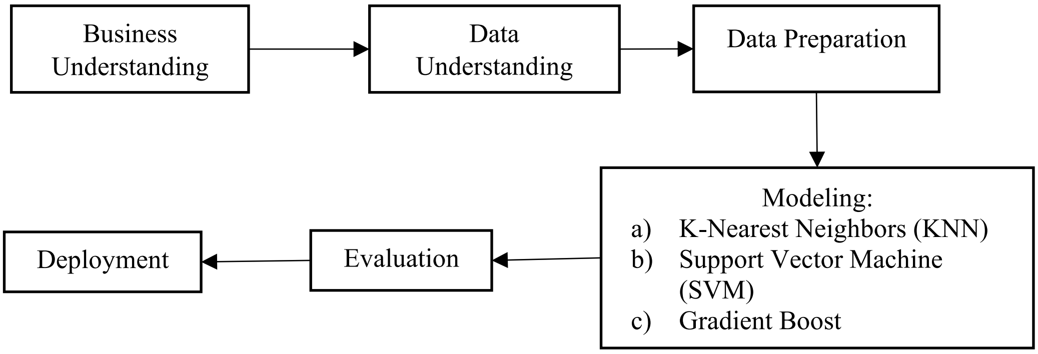


**Figure 1.** Method diagram

The Figure 1 illustrates the data model development process, which begins with business understanding, followed by data understanding and data preparation, then moves on to the modeling stage using the K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Gradient Boost methods, which are then evaluated before finally being implemented at the deployment stage. The study begins with business understanding by defining the objective of classifying geothermal well productivity and comparing KNN, SVM, and Gradient Boost algorithms based on accuracy. In the data understanding stage, historical production data such as WHP, flow rate, and enthalpy are collected and explored. Data preparation involves cleaning, transformation, and normalization to ensure quality. The modeling stage applies the three classification algorithms, followed by evaluation using accuracy metrics to determine the best-performing model. Finally, deployment implements the selected model for productivity prediction and operational decision support. Data were obtained through interviews, literature review, and observation.

This study applies the CRISP-DM framework, covering business understanding, data understanding, preparation, modeling, evaluation, and deployment. The objective is to classify geothermal well productivity levels (Low, Medium, High) using daily operational and production data to support well-level decision-making. The dataset consists of 74,912 daily records (January 2018–June 2024) with 13 features, including Wellhead Pressure (WHP), enthalpy, steam flow, brine flow, total flow, power (MW), and operational variables. Exploratory analysis identified class imbalance, with the Medium class dominating the dataset. Missing numerical values were handled using mean imputation, while categorical variables used mode imputation; records with excessive missing data were removed. Outliers were detected using the Interquartile Range (IQR) method and only invalid extreme values were excluded. Numerical features were normalized using Min-Max scaling. To address class imbalance, SMOTE was applied in a separate experimental scenario to generate synthetic samples for minority classes, enabling comparison between original and balanced data distributions.

Data were split using stratified sampling with three train–test ratios: 70:30, 80:20, and 90:10, to assess model stability under different training proportions. Three algorithms were implemented. K-Nearest Neighbors (KNN) used Euclidean distance and was optimized via 5-fold cross-validation, with K = 5 selected as optimal. Support Vector Machine (SVM) employed a Radial Basis Function (RBF) kernel, with hyperparameters tuned through grid search. Gradient Boosting was configured using optimized estimators, learning rate, and tree depth to balance accuracy and overfitting control. Model performance was evaluated using Accuracy, Precision, Recall, and F1-Score to ensure reliable assessment under imbalanced conditions, supported by confusion matrix analysis. The best-performing model was then deployed to classify new production data for operational monitoring and production forecasting support.

## 3. RESULT AND DISCUSSION

### 3.1 Business Understanding

The analysis of geothermal well production data processing indicates that the conventional methods currently used are not optimal for accurately and efficiently determining well productivity levels. These approaches are time-consuming and produce results that are insufficient for predictive simulations and performance evaluation. To address this limitation, this study applies data mining techniques by comparing K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), and Gradient Boost algorithms. These classification models utilize production parameters such as total flow, enthalpy, pressure, and other operational variables to determine geothermal well productivity levels. This approach is expected to improve accuracy and efficiency, support future production decision-making, and help geothermal energy companies optimize resource utilization [16].

### 3.2 Data Understanding

The data understanding phase focuses on comprehending the dataset through data exploration and variable identification, as outlined in Table 3.3. The study uses daily geothermal well production data categorized by operational function: PRD (Production), MON (Monitoring), COND (Condensate), and INJ (Injection). Operational status dataOpen (O), Shut-in (S), and Bleeding (B)are included to reflect well conditions over time [17]. Key operational parameters include flow control valve (FCV), wellhead pressure (WHP), enthalpy, and well identification (WELL_NAME). Productivity data consist of steam flow (STEAM_FLOW), brine flow (BRINE_FLOW), total flow (TOTAL_FLOW), and generated power (POWER_MW), which directly indicate well productivity.

The final dataset contains 74,912 rows and 13 columns. Initial inspection using Python functions such as head() and info() confirms that the data are well-structured with appropriate data types. Descriptive statistical analysis is then performed to examine distributions, central tendencies, and potential data issues such as outliers or missing values, This step ensures data quality and readiness for further predictive modeling [18].

### 3.3 Data Preparation

### 3.3.1 Data Cleaning

This stage involves correcting incomplete, inaccurate, or irrelevant data. This process begins by identifying missing data or data containing null values. Examples of incomplete or inaccurate data are shown in Table 1.

**Table 1.** Example of missing values

| HOURS ON | FCV | WHP | H | STEAM FLOW | BRINE FLOW | TOTAL FLOW | POWER MW | LEVEL |
|---|---|---|---|---|---|---|---|---|
| NULL | NULL | 17,71 | 2800 | 0 | NULL | 0 | 0 | Low |
| NULL | NULL | 17,71 | 2800 | 0 | NULL | 0 | 0 | Low |
| NULL | NULL | 17,71 | 2800 | 0 | NULL | 0 | 0 | Low |
| NULL | NULL | 17,71 | 2800 | 0 | NULL | 0 | 0 | Low |

Once the data has been identified, data containing empty or null values is filled with the value 0 according to the associated variable. This cleaning process ensures better and more consistent data quality for use in classification modeling. The data cleaning results are checked using the Python programming language. As an additional clarification, in geothermal production datasets the value **0** carries explicit physical meaning, such as indicating shut-in conditions, zero flow, or non-operational wells. Therefore, imputing all missing values with 0 may introduce bias if the missing entries actually represent unrecorded measurements rather than true zero production. To prevent distortion of the data distribution, the pattern of missing values was first examined. When missing data occurred during periods where wells were confirmed to be non-operational, zero imputation was retained because it accurately reflected real operational conditions. However, for numerical variables during active production periods this paper such as WHP, enthalpy, steam flow, brine flow, total flow, and power this paper statistical imputation methods were applied. Mean imputation was used for relatively symmetric distributions, while median imputation was selected for skewed variables to better preserve central tendency without artificially reducing variability. As an additional robustness check, a limited implementation of the KNN imputer was conducted to estimate missing values based on feature similarity. The results were consistent with the statistical imputation approach; therefore, mean/median imputation was maintained due to its computational efficiency and stability.

### 3.3.2 Data Formatting

At this stage, data formats are standardized to meet the requirements of the modeling process. Inconsistencies such as incorrect data types, varying date formats, and non-uniform categorical values are corrected. One-hot encoding is applied to convert categorical variables into binary representations (0 and 1), ensuring that all categorical features can be effectively processed by machine learning algorithms, which typically require numerical input. This standardization is essential for efficient analysis and classification in subsequent modeling stages.

### 3.3.3 Data Selection

Relevant data are selected based on the research objectives, focusing on daily geothermal well production variables, enthalpy, and fluid flow. The selection process is performed using SQL Server queries, as all data are stored in an SQL database. The dataset covers a five-year period from January 1, 2018, to June 30, 2024, and includes only wells with production status (PRD). This process results in 74,912 records ready for analysis. Variable relationships are then examined using a correlation matrix (Figure 4.4) to identify attributes strongly associated with well productivity. Only variables with a positive correlation coefficient above 0.5 are retained for modeling [19].

### 3.3.4 Data Transformation

After removing irrelevant or problematic data, transformation techniques are applied to improve data suitability for analysis and classification. This includes normalization to standardize the scale of numerical variables, ensuring uniformity across features. This step prevents bias toward variables with larger value ranges and allows machine learning algorithms to perform optimally during the modeling process.

### 3.4 Modeling

At the modeling stage, geothermal well productivity classification is performed using data that have undergone the preparation process. Two modeling approaches are applied: with SMOTE and without SMOTE. In the SMOTE-based approach, the Synthetic Minority Over-sampling Technique is used to balance the dataset by generating synthetic samples for minority classes, resulting in a more balanced class distribution. The performance of K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), and Gradient Boost models is then compared under both approaches to evaluate the effectiveness of SMOTE in handling class imbalance. Modeling is conducted using three different training and testing data split ratios: 70:30, 80:20, and 90:10. These variations are applied to analyze the impact of training and testing proportions on model performance and to assess the stability and robustness of each model across different data distributions.

### 3.4.1 K-NN Modeling

The K-Nearest Neighbors (K-NN) algorithm is employed to classify geothermal well productivity using the prepared dataset. The modeling process begins by inputting transformed data into the K-NN model. The optimal value of $K$ is determined iteratively through validation, selecting the value that yields the best classification performance. Based on evaluation results, the highest accuracy of approximately 93.6% is achieved at $K = 5$. Based on the evaluation results, the K-Nearest Neighbors (K-NN) model demonstrates strong and stable performance across all data split ratios. At the 70:30 ratio, K-NN achieves an accuracy of 93.31% without SMOTE, slightly decreasing to 92.71% with SMOTE. For the 80:20 split, the accuracy reaches 93.56% without SMOTE and decreases to 92.90% when SMOTE is applied. The highest accuracy is obtained at the 90:10 ratio, with 94.22% without SMOTE and 93.45% with SMOTE. These results indicate that the application of SMOTE does not consistently improve K-NN performance. Visualization of the accuracy results further confirms that K-NN generally performs better without SMOTE across all data ratios.

### 3.4.2 SVM Modeling

The Support Vector Machine (SVM) model is evaluated using the Radial Basis Function (RBF) kernel. Overall, SVM shows moderate performance, with accuracy remaining consistent across different data splits but lower compared to K-NN. At the 70:30 ratio, the model achieves 79.04% accuracy without SMOTE, which drops to 72.70% after SMOTE application. Similar patterns are observed at the 80:20 and 90:10 ratios, where accuracy without SMOTE reaches 79.24% and 79.78%, respectively, while accuracy with SMOTE remains lower. These findings suggest that SMOTE does not provide a significant benefit for SVM and may even reduce performance, particularly for datasets with non-linear class distributions.

### 3.4.3 Gradient Boost Modeling

Gradient Boosting (GB) exhibits consistently high accuracy across all train-test ratios. At the 70:30 split, the model achieves 91.39% accuracy without SMOTE, decreasing to 88.43% with SMOTE. For the 80:20 ratio, accuracy without SMOTE is 91.34%, while SMOTE reduces it to 88.57%. A similar trend is observed at the 90:10 ratio. The decline in performance after applying SMOTE indicates that Gradient Boosting inherently handles class imbalance effectively through its iterative, tree-based learning mechanism, making oversampling unnecessary.

Overall, the results demonstrate that all three modelsK-NN, SVM, and Gradient Boostingperform better without SMOTE. Among them, K-NN achieves the highest accuracy, while Gradient Boosting provides stable and reliable performance across all data splits. These findings highlight the importance of carefully considering the use of oversampling techniques such as SMOTE, as they do not always lead to improved model performance.

### 3.5 Model Evaluation

#### 3.5.1 Confusion Matrix

a. K-NN Confusion Matrix

The performance evaluation results of the K-Nearest Neighbors (K-NN) model with Non-SMOTE and SMOTE at train-to-test data ratios of 70:30, 80:20, and 90:10 can be analyzed using the following visualization.

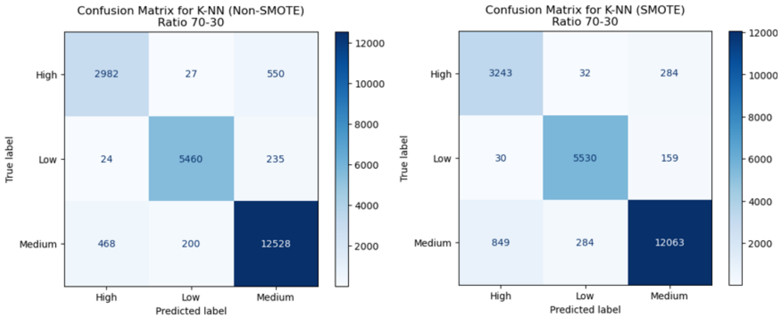


**Figure 1.** Confusion Matrix of K-NN with 70:30 Ratio

Figure 1 shows that using a 70:30 data ratio, the model without SMOTE performed well in predicting the Medium class, with the highest number of correct predictions at 12,528. However, the High and Low classes had several misclassifications, particularly the Medium class, which was frequently predicted as High. With SMOTE, the model showed improvement in handling the High and Low classes. However, there was a slight increase in misclassifications for the Medium class, with the number of correct predictions decreasing to 12,063.

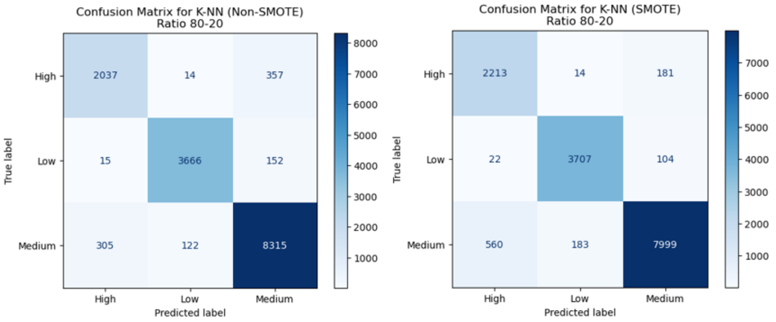


**Figure 2.** Confusion Matrix of K-NN with 80:20 Ratio

Figure 2, using an 80:20 data ratio, without SMOTE, still showed dominant performance in predicting the Medium class, with 8,315 correct predictions. However, there were more misclassifications for the High class, which was predicted as either Medium or Low. With SMOTE, the model improved predictions for the High class, with the number of correct predictions increasing to 2,213. However, misclassifications still occurred for the Medium class, albeit at a slightly reduced rate.

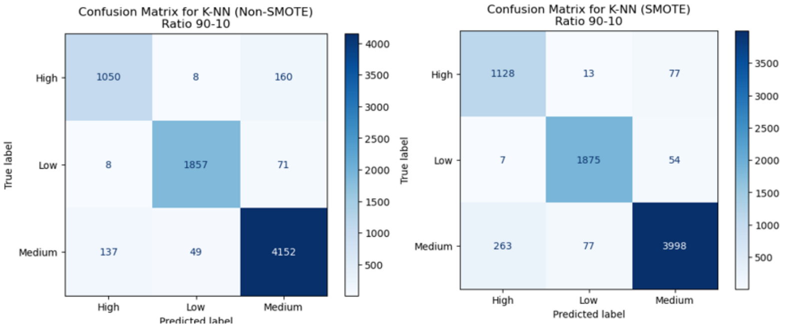


**Figure 3.** Confusion Matrix for K-NN with 90:10 Ratio

Figure 3 shows that using a 90:10 data ratio, without SMOTE, accuracy decreased, especially for the High and Medium classes, with more prediction errors. The number of correct predictions for the Medium class was 4,152, while the High class showed significant errors. With SMOTE, model performance improved for the High and Low classes, with the number of correct predictions for the High class reaching 1,128. However, the Medium class still showed classification errors, with the number of correct predictions decreasing slightly to 3,998. Overall, applying SMOTE to the K-NN model helped improve predictions for the minority High and Low classes, but tended to decrease accuracy for the majority Medium class. This suggests that balancing the data using SMOTE has a positive impact on infrequent classes, but may compromise accuracy for the dominant class [20].

b. SVM Confusion Matrix

The performance evaluation results of the SVM models with Non-SMOTE and SMOTE at train-to-test data ratios of 70:30, 80:20, and 90:10 can be analyzed using the following visualization.

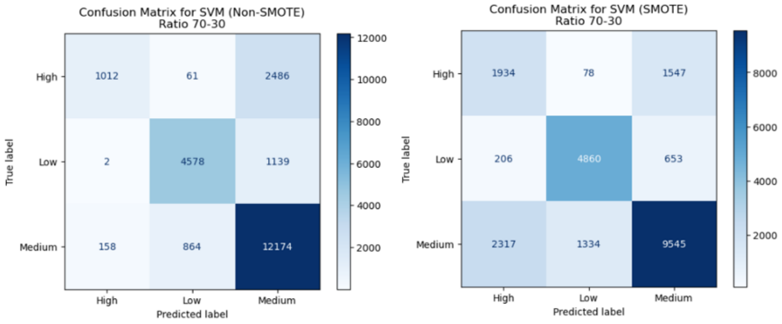


**Figure 4.** Confusion Matrix SVM with 70:30 Ratio

Figure 4, using a 70:30 train-to-test data ratio, without SMOTE, demonstrated good performance in predicting the Medium class, with 12,174 correct predictions. However, the High and Low classes experienced significant misclassifications, particularly with the high class being predicted as Medium. With SMOTE, the model's performance improved in the High class, where correct predictions increased to 1,934. However, misclassifications in the Medium class increased, with the number of correct predictions dropping to 9,545.

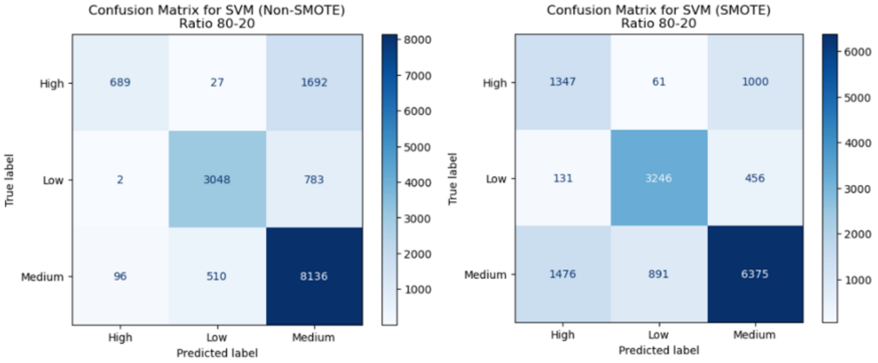


**Figure 5.** Confusion Matrix SVM with 80:20 Ratio

Figure 5 shows that using an 80:20 train-to-test data ratio, the model without SMOTE performed best for the Medium class, with 8,136 correct predictions. However, the High and Low classes still had relatively high error rates, especially the High class, which is often predicted as Medium. With SMOTE, the model showed significant improvement in the High class, with the number of correct predictions increasing to 1,347. However, there was an increase in misclassifications for the Medium class, with the number of correct predictions dropping to 6,375.

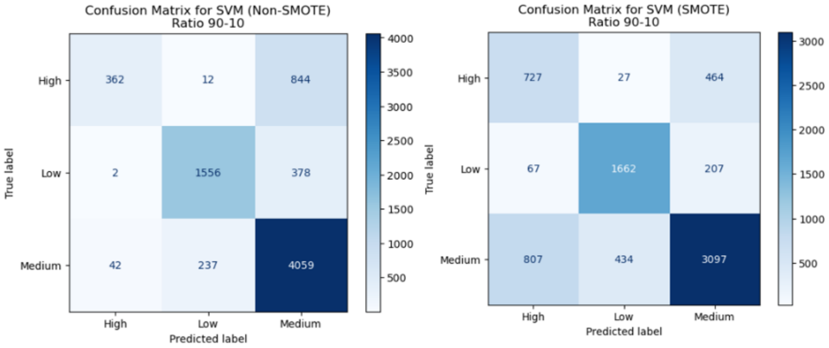


**Figure 6.** Confusion Matrix GB with 90:10 Ratio

Figure 6 shows a decrease in performance using a 90:10 train-to-test data ratio without SMOTE, with the Medium class having 4,059 correct predictions. The High and Low classes exhibited higher misclassifications, especially the High class, which was mostly predicted as Medium. With SMOTE, the model showed improved predictions for the High and Low classes, with the number of correct predictions for the High class increasing to 727. However, the Medium class experienced an increase in errors, with only 3,097 correct predictions.

Overall, applying SMOTE to the SVM model helped improve the model's performance in predicting the High class, which is the minority class. However, this often resulted in decreased prediction accuracy for the majority Medium class, which is a trade-off when the SMOTE oversampling technique is applied [21].

c. Confusion Matrix Gradient Boosting

The performance evaluation results of the SVM models with Non-SMOTE and SMOTE at train-to-test data ratios of 70:30, 80:20, and 90:10 can be analyzed using the following visualization.

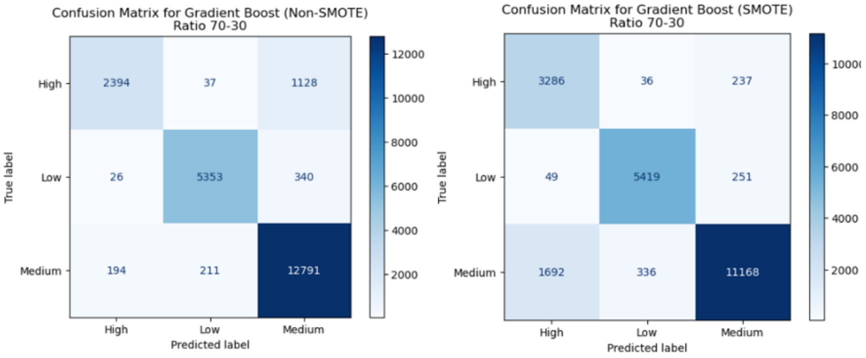


**Figure 7.** Confusion Matrix GB with 70:30 ratio

Figure 7, using a 70:30 train-to-test data ratio, shows good performance without SMOTE in predicting the Medium class, with 12,174 correct predictions. However, the High and Low classes experienced significant misclassifications, particularly with the high class being predicted as Medium. With SMOTE, model performance improved for the High class, where correct predictions increased to 1,934. However, misclassifications for the Medium class increased, with the number of correct predictions dropping to 9,545.

In Figure 8, using an 80:20 train-to-test data ratio, the model's performance without SMOTE remains stable, especially for the Medium class, although the High and Low classes have a higher number of incorrect predictions. Applying SMOTE at this ratio appears to improve predictions for the High and Low classes, but at the expense of a slight decrease in performance for the Medium class.

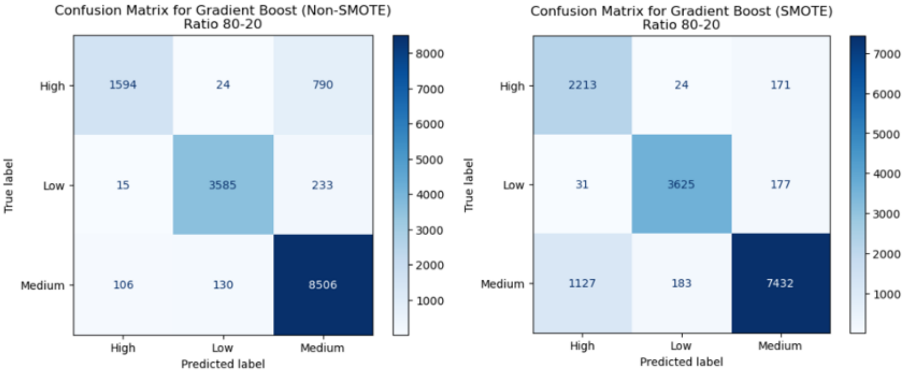


**Figure 8.** Confusion Matrix GB with 80:20 Ratio

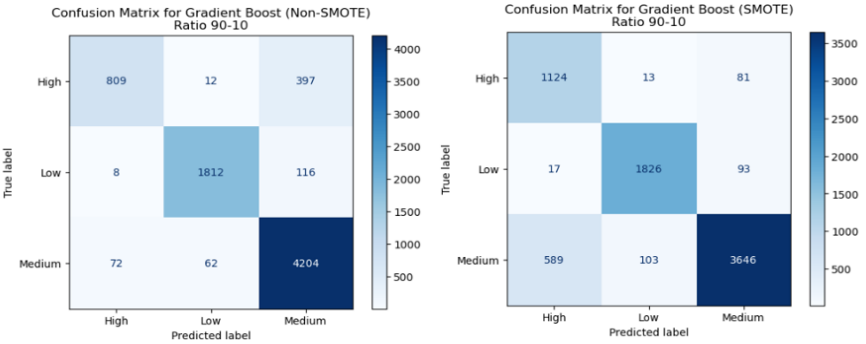
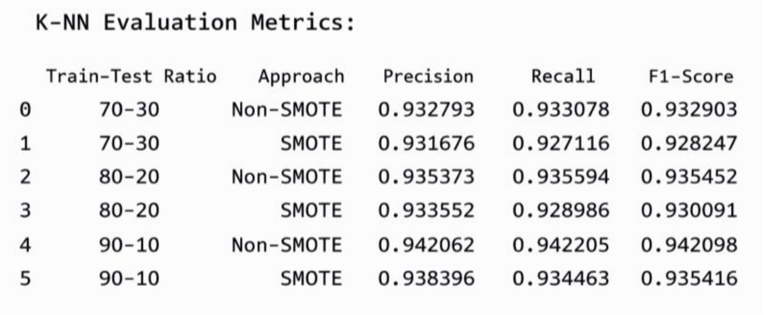


**Figure 9.** Confusion Matrix GB with a 90:10 ratio

In Figure 9, using a 90:10 train-to-test data ratio, the model without SMOTE tends to experience decreased performance, especially for the High class, which has a higher number of mispredictions. Using SMOTE improves the prediction distribution for the High class, but the prediction accuracy for the Medium and Low classes still shows some misclassifications[22]. Overall, applying SMOTE helps the Gradient Boosting model improve predictions for the High and Low minority classes, but often results in increased misclassifications for the Medium majority class. This indicates that balancing data with SMOTE produces varying results depending on the train-to-test distribution used [23].

### 3.6 Evaluation Metrics

Evaluation metrics are used to evaluate the performance of the applied classification model. The goal is to understand the extent to which the model can correctly predict the data and to comprehensively measure its performance. Precision and recall are used to evaluate the model's performance in predicting specific classes, especially on imbalanced data. The F1-score combines precision and recall to provide a harmonic mean, making it suitable for use in situations with uneven class distribution.



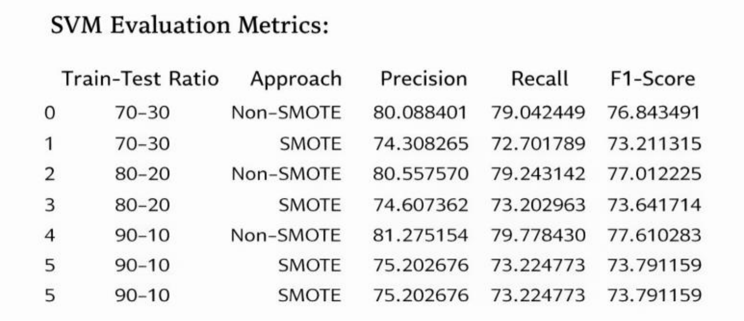
```
K-NN Evaluation Metrics:

  Train-Test Ratio   Approach  Precision    Recall  F1-Score
0            70-30  Non-SMOTE   0.932793  0.933078  0.932903
1            70-30      SMOTE   0.931676  0.927116  0.928247
2            80-20  Non-SMOTE   0.935373  0.935594  0.935452
3            80-20      SMOTE   0.933552  0.928986  0.930091
4            90-10  Non-SMOTE   0.942062  0.942205  0.942098
5            90-10      SMOTE   0.938396  0.934463  0.935416
```

**Figure 10.** K-NN Evaluation Metrics

In the metric evaluation results for the K-NN model in Figure 10, model performance was measured using train-to-test data ratios of 70:30, 80:20, and 90:10, with and without SMOTE. The following is an explanation:

a. With a 70:30 data ratio, the non-SMOTE model had a Precision of 0.932793 and a Recall of 0.933078, while applying SMOTE slightly decreased the metric values to 0.931676 and 0.927116.
b. With an 80:20 data ratio, the non-SMOTE model produced the highest Precision and Recall values of 0.935373 and 0.935594, respectively, but these values decreased slightly when using SMOTE.
c. At a 90:10 data ratio, the non-SMOTE model performed best with a Precision of 0.942062 and a Recall of 0.942205, while SMOTE produced slightly lower values.

Overall, applying SMOTE did not consistently improve the performance of the K-NN model, particularly for certain train-to-test data ratios, where the non-SMOTE model demonstrated better and more stable metric value.



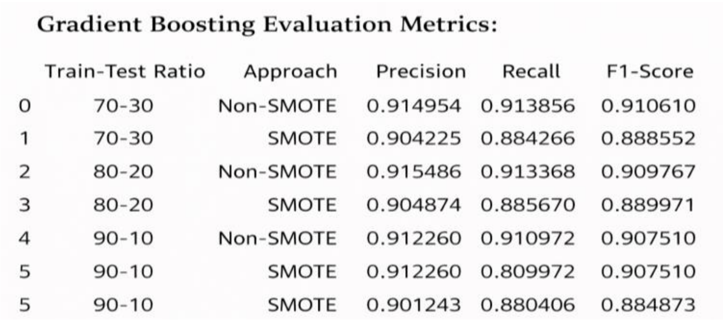
```
SVM Evaluation Metrics:

  Train-Test Ratio   Approach  Precision     Recall   F1-Score
0            70-30  Non-SMOTE  80.088401  79.042449  76.843491
1            70-30      SMOTE  74.308265  72.701789  73.211315
2            80-20  Non-SMOTE  80.557570  79.243142  77.012225
3            80-20      SMOTE  74.607362  73.202963  73.641714
4            90-10  Non-SMOTE  81.275154  79.778430  77.610283
5            90-10      SMOTE  75.202676  73.224773  73.791159
5            90-10      SMOTE  75.202676  73.224773  73.791159
```

**Figure 11.** SVM Evaluation Metrics

In the metric evaluation results for the SVM model in Figure 11, model performance was measured using various train-to-test data ratios (70:30, 80:20, and 90:10) with and without SMOTE. The following is an explanation:

a. At a 70:30 data ratio, the non-SMOTE model performed better with a Precision of 80.08%, a Recall of 79.04%, and an F1-Score of 76.84%. However, applying SMOTE decreased performance to a Precision of 74.31% and a Recall of 72.70%.
b. At a 80:20 data ratio, the best performance was still achieved by non-SMOTE with a Precision of 80.56% and a Recall of 79.24%, while SMOTE produced lower performance with a Precision of 74.61% and a Recall of 73.20%.
c. With a 90:10 data ratio, Non-SMOTE again produced the highest performance with 81.28% Precision, 79.78% Recall, and 77.61% F1-Score. Meanwhile, SMOTE yielded lower results, with 75.20% Precision and 73.22% Recall.

Overall, the SVM model performed better when using Non-SMOTE at all train-to-test data ratios. Applying SMOTE tended to decrease model performance, especially in Precision, indicating that the SVM is less than optimal for handling data processed with oversampling techniques.



Gradient Boosting Evaluation Metrics:

| | Train-Test Ratio | Approach | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| 0 | 70-30 | Non-SMOTE | 0.914954 | 0.913856 | 0.910610 |
| 1 | 70-30 | SMOTE | 0.904225 | 0.884266 | 0.888552 |
| 2 | 80-20 | Non-SMOTE | 0.915486 | 0.913368 | 0.909767 |
| 3 | 80-20 | SMOTE | 0.904874 | 0.885670 | 0.889971 |
| 4 | 90-10 | Non-SMOTE | 0.912260 | 0.910972 | 0.907510 |
| 5 | 90-10 | SMOTE | 0.912260 | 0.809972 | 0.907510 |
| 5 | 90-10 | SMOTE | 0.901243 | 0.880406 | 0.884873 |

**Figure 12.** GB Evaluation Metric

In the metric evaluation results for the Gradient Boost model in Figure 12, model performance was measured using various train-to-test data ratios (70:30, 80:20, and 90:10) with and without SMOTE. The following is an explanation:

a. At a 70:30 data ratio, the non-SMOTE model performed best with a Precision of 91.49%, a Recall of 91.39%, and an F1-Score of 91.06%. Meanwhile, applying SMOTE significantly decreased the metrics, with Precision at 90.42%, Recall at 88.43%, and an F1-Score at 88.86%.
b. At a 80:20 data ratio, the non-SMOTE model still performed superiorly with a Precision of 91.55%, a Recall at 91.34%, and an F1-Score at 90.98%. Applying SMOTE yielded lower values, with a Precision of 90.49% and a Recall of 88.57%.
c. At a 90:10 data ratio, Non-SMOTE still performed best, with a Precision of 91.23%, a Recall of 91.10%, and an F1-Score of 90.75%. However, performance decreased when using SMOTE, with Precision dropping to 90.12% and a Recall of 88.04%.

Overall, the Gradient Boost model performed better with Non-SMOTE compared to SMOTE across all train-to-test data ratios. Applying SMOTE tended to decrease Precision, Recall, and F1-Score values, indicating that Gradient Boost performs better without oversampling on naturally balanced data.

### 3.7 Classification Report

The Classification Report is used to analyze the performance of the classification model in more depth. The confusion matrix presents a comparison between the true labels and the model's predicted labels in matrix form, making it easier to identify correct and incorrect predictions for each class. The goal is to calculate evaluation metrics for a more detailed understanding of model performance, particularly in handling majority and minority classes. The Classification Report visualization allows viewing the distribution of misclassifications for each target class, allowing for clearer identification of the model's strengths and weaknesses [24].

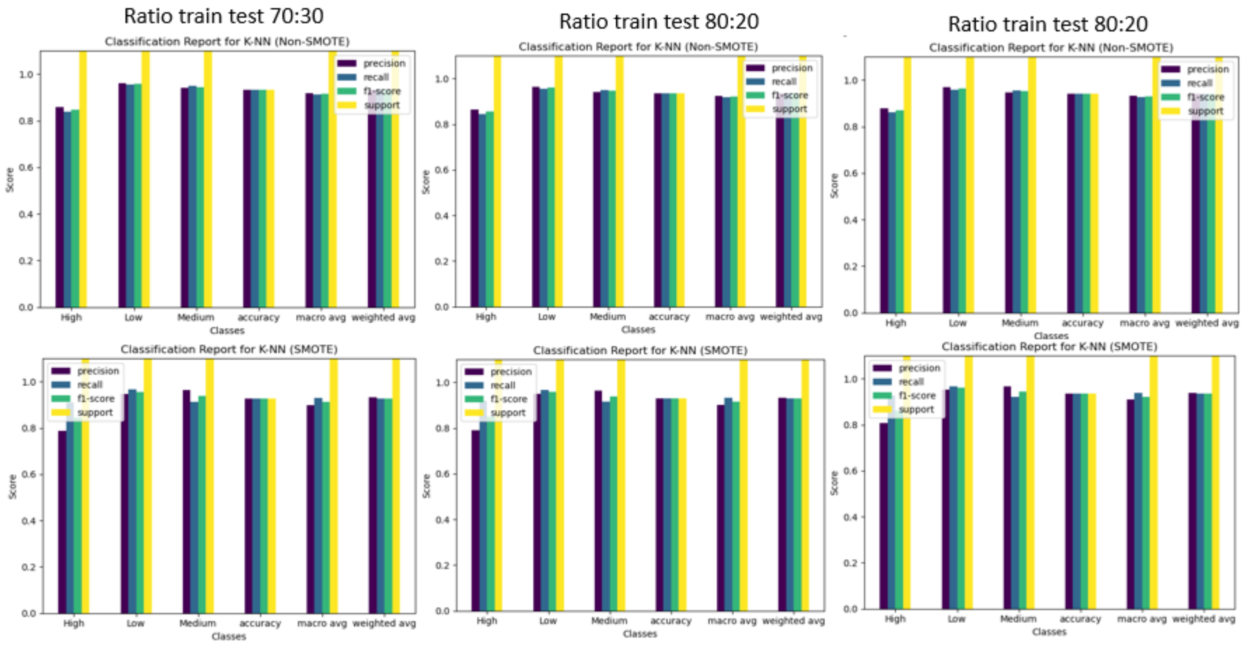


**Figure 13.** K-NN Classification Report

Based on the classification report visualization for the K-NN model in Figure 13, model performance was evaluated using Precision, Recall, F1-Score, and Support metrics at train-to-test data ratios of 70:30, 80:20, and 90:10, respectively, using both the Non-SMOTE and SMOTE approaches.

a. 70:30 Ratio: In the Non-SMOTE approach, the model's performance for the Medium class had the highest scores across all metrics compared to the High and Low classes. With SMOTE, performance improved for the High and Low classes, but performance decreased slightly for the Medium class. This indicates a compromise in the prediction of the majority class when SMOTE was applied.
b. 80:20 Ratio: In the Non-SMOTE approach, model performance remained the highest for the Medium class, while the High and Low classes had lower scores, particularly in the Recall metric. With SMOTE, the scores for the High and Low classes improved slightly, but overall performance decreased slightly, especially for the Medium class.
c. 90:10 Ratio: With the non-SMOTE approach, the model maintained good performance for the Medium class, but accuracy for the High and Low classes decreased due to the imbalanced data distribution. With SMOTE, scores for the minority classes, High and Low, improved, but performance for the majority class, Medium, decreased slightly.

Overall, applying SMOTE helped improve the performance of the K-NN model for the High and Low classes, which are the minority classes. However, this improvement was often accompanied by a decrease in scores for the Medium class, particularly in the Recall and F1-Score metrics. This indicates that using SMOTE is effective in addressing data imbalance, but consideration should be given to balancing the performance of the majority class.

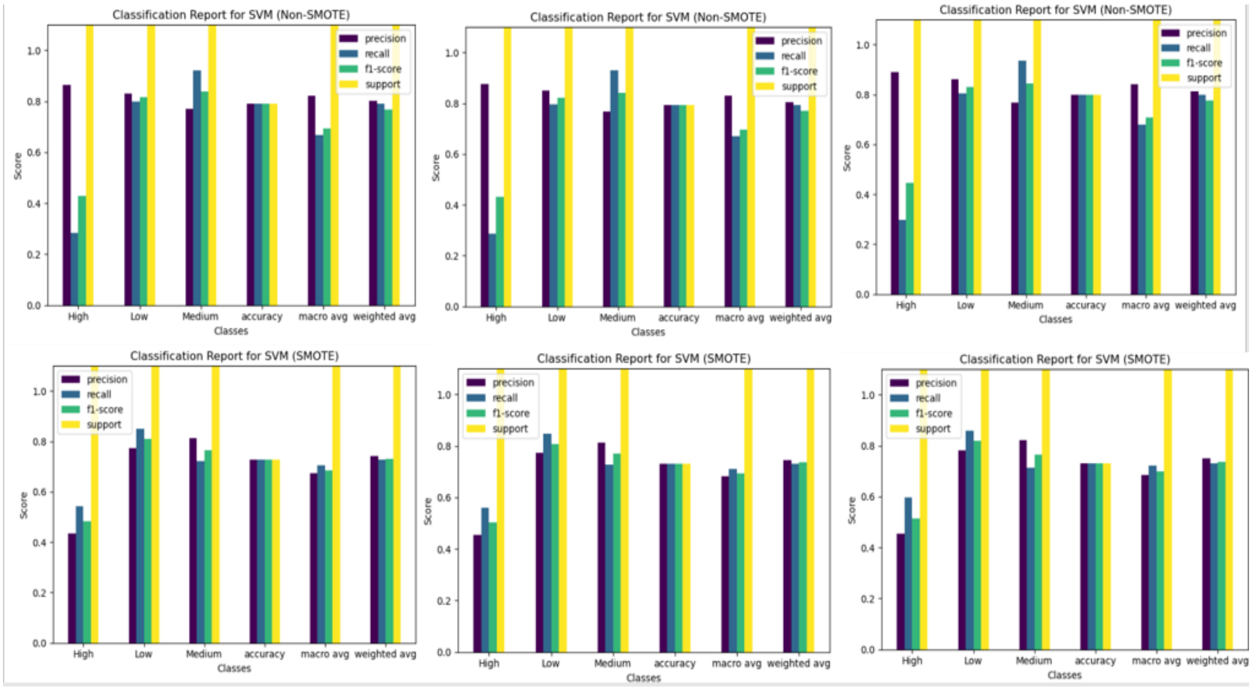


**Figure 14.** SVM Classification Report

Based on the classification report for the SVM model in Figure 14, model performance was analyzed using Precision, Recall, F1-Score, and Support metrics at train-to-test data ratios of 70:30, 80:20, and 90:10, respectively, using the Non-SMOTE and SMOTE approaches.

a. 70:30 Ratio: With the Non-SMOTE approach, the model performed well for the Medium class, but the scores for the High and Low classes were lower, especially in the Recall metric. With SMOTE, the model's performance for the High class improved in Precision and Recall, but the score for the Medium class decreased slightly. This indicates that SMOTE successfully improved minority class predictions.
b. 80:20 Ratio: Non-SMOTE showed higher Precision and Recall scores for the Medium class compared to the other classes. However, the High and Low classes had relatively low scores. With SMOTE, the scores for the High and Low classes increased, but performance on the Medium class decreased slightly, as evidenced by the lower F1-score.
c. 90:10 Ratio: With the non-SMOTE approach, model performance decreased as the training data portion decreased, particularly for the High and Low classes, which have low Recall due to data imbalance. With SMOTE, the Precision and Recall scores for the High and Low classes increased significantly, although this was accompanied by a decrease in the score for the Medium class.

Overall, applying SMOTE to the SVM model helped improve predictive performance on the High and Low classes, which are minority classes. However, applying SMOTE also tended to decrease the score for the Medium

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 4, March 2026 Page: 2491−2506
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9302

class, particularly for Recall and F1-score, as the model focused more on the previously infrequent class. This suggests a trade-off between improving performance on the minority class and decreasing performance on the majority class.

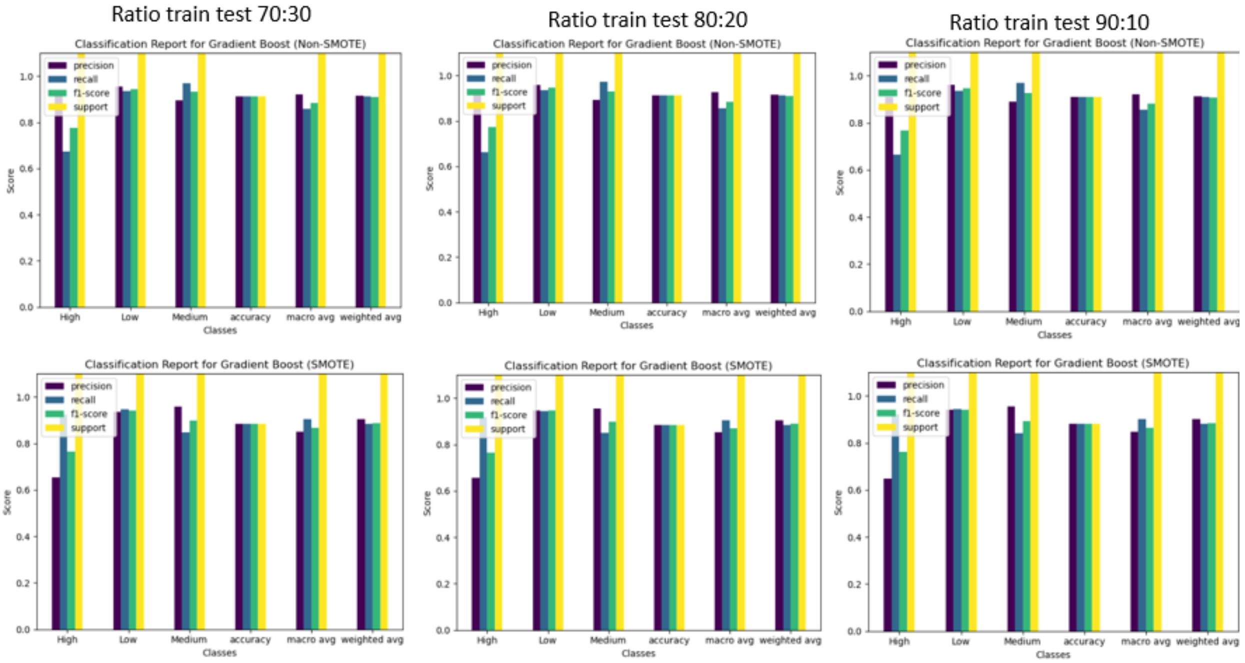


**Figure 15.** Classification Report GB

Based on the classification report for the Gradient Boost model in Figure 15, model performance was evaluated using Precision, Recall, F1-Score, and Support metrics at train-to-test data ratios of 70:30, 80:20, and 90:10 with the Non-SMOTE and SMOTE approaches.

a. 70:30 ratio: Non-SMOTE demonstrated excellent performance for the Medium class, with Precision, Recall, and F1-Score scores approaching the maximum. However, the scores for the High and Low classes remained lower. With SMOTE, model performance for the High and Low classes improved, particularly for the Recall metric, indicating an improved model capability in handling minority classes.
b. 80:20 ratio: With the Non-SMOTE approach, the model still demonstrated high performance for the Medium class, while the High and Low classes had lower scores, particularly for Recall. With SMOTE, performance for the High and Low classes improved, although it resulted in a slight decrease in the F1-score for the Medium class. This indicates that SMOTE helps balance model performance on the minority class.
c. 90:10 ratio: Non-SMOTE performed best for the Medium class, but scores for the High and Low classes were lower due to the increasingly imbalanced data distribution. With SMOTE, score improvements were seen for the High and Low classes, especially for Recall, although performance for the Medium class decreased slightly.

Overall, the Gradient Boost model demonstrated stable performance across all train-to-test data ratios, especially for the majority Medium class. Applying SMOTE helped improve performance on the minority classes, namely High and Low, especially for the Recall metric, but sometimes decreased performance for the majority class. This indicates that SMOTE is effective in handling imbalanced data, although there was a balance in overall model accuracy.

### 3.8 Model Comparison

Based on the results of the classification model performance comparison conducted in this study, comparisons were conducted based on accuracy. This was to evaluate the ability of each model to handle data with different distributions, including imbalanced data. Additionally, analyses were conducted on three train-to-test data ratios: 70:30, 80:20, and 90:10, using both the Non-SMOTE and SMOTE approaches, to understand the impact of oversampling on the performance of each model.

**Table 2.** Comparison of Model Accuracy

| Train-Test Ratio | K-NN | | SVM | | Gradient Boost | |
|---|---|---|---|---|---|---|
| | Non-SMOTE | SMOTE | Non-SMOTE | SMOTE | Non-SMOTE | SMOTE |
| 70:30 | 93.31% | 92.71% | 79.04% | 72.70% | 91.39% | 88.43% |
| 80:20 | 93.56% | 92.90% | 79.24% | 73.20% | 91.34% | 88.57% |
| 90:10 | 94.22% | 93.45% | 79.78% | 73.22% | 91.10% | 88.04% |

Table 2 shows a comparison of model accuracy. The K-NN model had the highest accuracy compared to the SVM and Gradient Boost models across all data ratios. The K-NN model demonstrated more consistent performance across both the Non-SMOTE and SMOTE approaches. The Gradient Boost model demonstrated stable performance with high accuracy across all data ratios, although it tended to be more optimal on the Non-SMOTE data. Meanwhile, the SVM model had lower accuracy than the other two models, and its performance was more significantly affected by the use of SMOTE. In all cases, applying SMOTE to the data often decreased model accuracy, especially for the majority class, although it improved performance for the minority class. Overall, the K-NN model performed best in terms of accuracy, followed by Gradient Boost and SVM, with the Non-SMOTE approach providing the best performance for most models [16].

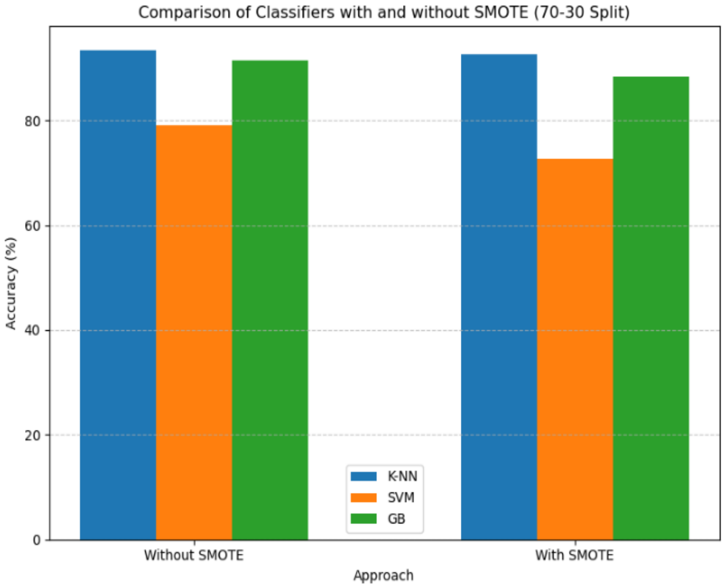


**Figure 16.** Comparison of 70:30 Data Ratio Models

The model accuracy comparison in Figure 16, using a 70:30 train-to-test data ratio, yielded the highest accuracy for both the K-NN model with and without SMOTE, with a slight decrease when SMOTE was applied. The Gradient Boost model performed consistently but tended to experience a decrease in accuracy after SMOTE was applied. On the other hand, the SVM model showed the lowest performance compared to the other two models with both approaches, with a more significant decrease when SMOTE was applied.

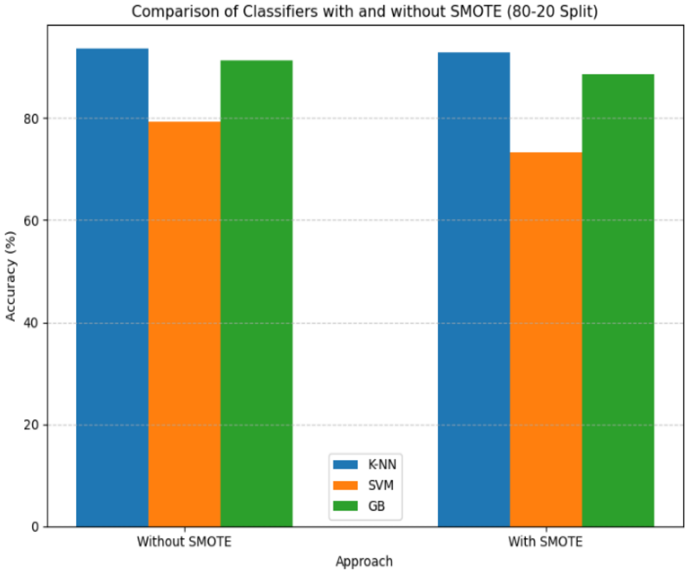


**Figure 17.** Comparison of 80:20 Data Ratio Models

The model accuracy comparison in Figure 17 uses an 80:20 train-to-test data ratio, using both the SMOTE and SMOTE approaches. The K-NN model again demonstrated the best performance with the highest accuracy, both without SMOTE and with SMOTE. The application of SMOTE slightly decreased the accuracy of the K-NN model, although not significantly. The Gradient Boost model performed consistently and closely matched the K-NN model, but also showed a slight decrease in accuracy after the application of SMOTE. Meanwhile, the SVM model performed the lowest compared to the other two models. Furthermore, the application of SMOTE to the SVM model had a less than optimal impact, with accuracy remaining lower than without SMOTE.

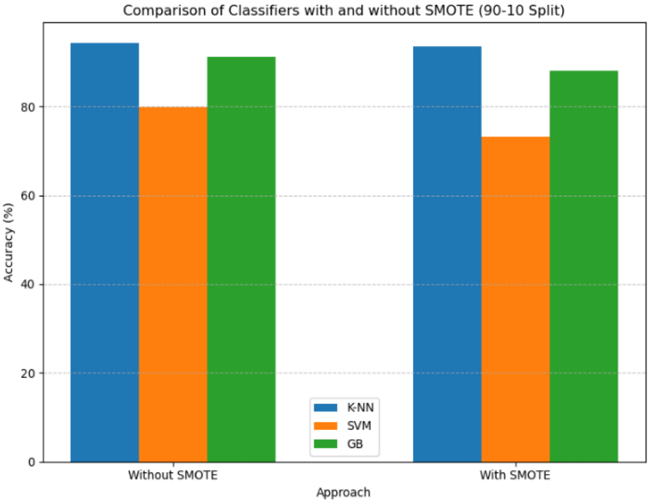


**Figure 18.** 90:10 Model Ratio Comparison

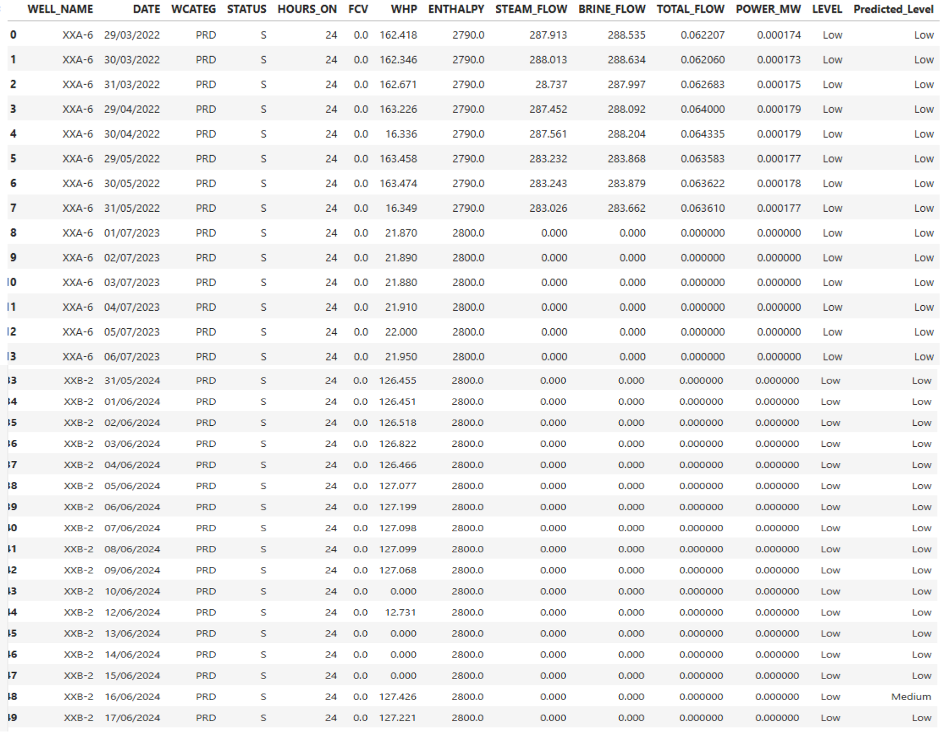
The model accuracy comparison in Figure 18 uses a 90:10 train-to-test data ratio, using both the SMOTE and SMOTE approaches. The K-NN model still demonstrated the best performance with the highest accuracy among the three models, both without SMOTE and with SMOTE. The application of SMOTE slightly decreased the accuracy of the K-NN model, although its performance remained superior compared to the other models. The Gradient Boost model maintained stable accuracy and was the second-best performing model after K-NN. However, the accuracy of Gradient Boost also decreased after the application of SMOTE, although the decrease was relatively small. On the other hand, the SVM model demonstrated the lowest performance, both without SMOTE and with SMOTE. The application of SMOTE to the SVM model did not provide significant improvement and even tended to decrease its accuracy.

From these comparisons, the best model can be identified, which not only has high accuracy but also provides good performance in both the majority and minority classes.

## 3.9 Deployment

### 3.9.1 Data Selection

The data used for data visualization deployment is the data with the highest accuracy, namely the classification results using the K-NN method without SMOTE, using a 90:10 train-to-test data ratio, resulting in an accuracy of 94.22%.



| | WELL_NAME | DATE | WCATEG | STATUS | HOURS_ON | FCV | WHP | ENTHALPY | STEAM_FLOW | BRINE_FLOW | TOTAL_FLOW | POWER_MW | LEVEL | Predicted_Level |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | XXA-6 | 29/03/2022 | PRD | S | 24 | 0.0 | 162.418 | 2790.0 | 287.913 | 288.535 | 0.062207 | 0.000174 | Low | Low |
| 1 | XXA-6 | 30/03/2022 | PRD | S | 24 | 0.0 | 162.346 | 2790.0 | 288.013 | 288.634 | 0.062060 | 0.000173 | Low | Low |
| 2 | XXA-6 | 31/03/2022 | PRD | S | 24 | 0.0 | 162.671 | 2790.0 | 28.737 | 287.997 | 0.062683 | 0.000175 | Low | Low |
| 3 | XXA-6 | 29/04/2022 | PRD | S | 24 | 0.0 | 163.226 | 2790.0 | 287.452 | 288.092 | 0.064000 | 0.000179 | Low | Low |
| 4 | XXA-6 | 30/04/2022 | PRD | S | 24 | 0.0 | 16.336 | 2790.0 | 287.561 | 288.204 | 0.064335 | 0.000179 | Low | Low |
| 5 | XXA-6 | 29/05/2022 | PRD | S | 24 | 0.0 | 163.458 | 2790.0 | 283.232 | 283.868 | 0.063583 | 0.000177 | Low | Low |
| 6 | XXA-6 | 30/05/2022 | PRD | S | 24 | 0.0 | 163.474 | 2790.0 | 283.243 | 283.879 | 0.063622 | 0.000178 | Low | Low |
| 7 | XXA-6 | 31/05/2022 | PRD | S | 24 | 0.0 | 16.349 | 2790.0 | 283.026 | 283.662 | 0.063610 | 0.000177 | Low | Low |
| 8 | XXA-6 | 01/07/2023 | PRD | S | 24 | 0.0 | 21.870 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 9 | XXA-6 | 02/07/2023 | PRD | S | 24 | 0.0 | 21.890 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 10 | XXA-6 | 03/07/2023 | PRD | S | 24 | 0.0 | 21.880 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 11 | XXA-6 | 04/07/2023 | PRD | S | 24 | 0.0 | 21.910 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 12 | XXA-6 | 05/07/2023 | PRD | S | 24 | 0.0 | 22.000 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 13 | XXA-6 | 06/07/2023 | PRD | S | 24 | 0.0 | 21.950 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 3 | XXB-2 | 31/05/2024 | PRD | S | 24 | 0.0 | 126.455 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 4 | XXB-2 | 01/06/2024 | PRD | S | 24 | 0.0 | 126.451 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 5 | XXB-2 | 02/06/2024 | PRD | S | 24 | 0.0 | 126.518 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 6 | XXB-2 | 03/06/2024 | PRD | S | 24 | 0.0 | 126.822 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 7 | XXB-2 | 04/06/2024 | PRD | S | 24 | 0.0 | 126.466 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 8 | XXB-2 | 05/06/2024 | PRD | S | 24 | 0.0 | 127.077 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 9 | XXB-2 | 06/06/2024 | PRD | S | 24 | 0.0 | 127.199 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 0 | XXB-2 | 07/06/2024 | PRD | S | 24 | 0.0 | 127.098 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 1 | XXB-2 | 08/06/2024 | PRD | S | 24 | 0.0 | 127.099 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 2 | XXB-2 | 09/06/2024 | PRD | S | 24 | 0.0 | 127.068 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 3 | XXB-2 | 10/06/2024 | PRD | S | 24 | 0.0 | 0.000 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 4 | XXB-2 | 12/06/2024 | PRD | S | 24 | 0.0 | 12.731 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 5 | XXB-2 | 13/06/2024 | PRD | S | 24 | 0.0 | 0.000 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 6 | XXB-2 | 14/06/2024 | PRD | S | 24 | 0.0 | 0.000 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 7 | XXB-2 | 15/06/2024 | PRD | S | 24 | 0.0 | 0.000 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |
| 8 | XXB-2 | 16/06/2024 | PRD | S | 24 | 0.0 | 127.426 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Medium |
| 9 | XXB-2 | 17/06/2024 | PRD | S | 24 | 0.0 | 127.221 | 2800.0 | 0.000 | 0.000 | 0.000000 | 0.000000 | Low | Low |

**Figure 19.** Best Accuracy Classification Results

### 3.9.2 Dashboard Visualization

The dashboard was created to make it easier for users to understand and analyze the results of the well productivity classification (Low, Medium, and High) based on attributes selected based on needs, enabling dynamic data integration and informative visual presentation.

WELL NAME
XXA-3
1/1/2022 12/31/2022

XXA-3
01/01/2022 - 31/12/2022

| Highest Steam (ton/hour) | Highest Brine (ton/hour) | Highest Power (MW) | Highest Total Flow (ton/hour) |
|---|---|---|---|
| 298.3870 | 0.0009 | 0.0835 | 302.1120 |

**Figure 20.** Production Data Dashboard

Figure 20 shows the design of an interactive dashboard to display geothermal well production data based on the well name and selected date range. This dashboard presents various important information in a structured manner to support well productivity analysis.

The dashboard also displays the distribution of well operating status data in the form of a pie chart showing the number of wells with an operating status of "O" and those with an operating status of "S." This dashboard was designed to provide comprehensive insight into well performance based on historical data, assisting users in more targeted analysis and better decision-making.

The superior performance of K-Nearest Neighbors (KNN) can be explained by the structural characteristics of the geothermal production dataset. The main predictive variables this paper such as Wellhead Pressure (WHP), steam flow, brine flow, total flow, and generated power this paperare continuous operational parameters that exhibit relatively consistent local patterns within each productivity class. In daily production data, wells with similar thermodynamic and flow characteristics tend to cluster naturally. Because KNN is a distance based algorithm that relies on local similarity rather than global decision boundaries, it performs effectively when class separation occurs in localized regions of the feature space. The normalization process further enhanced this behavior by ensuring proportional distance measurement across variables. As a result, the nearest neighbor principle was well aligned with the intrinsic distribution of geothermal productivity data, leading to the highest observed accuracy of 94.22%.

In contrast, the Support Vector Machine (SVM) model produced the lowest performance. Although SVM is theoretically strong in handling nonlinear classification through the Radial Basis Function (RBF) kernel, its effectiveness depends on the existence of a clear margin between classes. In this study, overlapping productivity characteristics were observed between Medium and High classes, particularly in transitional operational conditions. Daily geothermal data are also subject to operational fluctuations, measurement noise, and temporary instabilities, which reduce margin separability. Under such conditions, SVM may struggle to construct an optimal hyperplane that generalizes well, leading to reduced classification accuracy compared to instance-based and ensemble methods.

The impact of SMOTE further clarifies the dataset behavior. While SMOTE improved Recall for minority classes, it generally reduced overall accuracy for KNN and Gradient Boosting. This phenomenon can be explained by over-generalization. SMOTE generates synthetic minority samples by interpolating between neighboring minority instances. However, when class boundaries are not strictly separated, synthetic samples may be created in regions that overlap with majority-class distributions. This artificially expands minority-class regions beyond their natural density, potentially distorting the true decision boundary. For distance-based methods such as KNN and tree-based ensembles such as Gradient Boosting, which already captured the natural structure of the data effectively, the introduction of synthetic samples reduced boundary precision. Therefore, although SMOTE mitigated class imbalance, it simultaneously introduced noise in boundary regions, explaining the observed decrease in overall model accuracy.

## 4. CONCLUSION

Based on the research background, methodology, and experimental results, several conclusions can be drawn. Dataset preparation aligned with operational business objectives plays a decisive role in model performance, particularly in defining relevant production attributes and productivity class labels. The dataset consisted of imbalanced classes, with the Medium productivity category dominating the observations, which significantly influenced classification behavior. Among the evaluated algorithms, K-Nearest Neighbors (KNN) achieved the best overall performance, reaching the highest accuracy of 94.22% using the 90:10 train–test split without SMOTE. Gradient Boosting followed as the second-best model, with its highest accuracy of 91.39% at the 70:30 split without SMOTE, demonstrating stable performance across all scenarios. Support Vector Machine (SVM) produced the lowest results, with a maximum accuracy of 79.78% at the 90:10 split without SMOTE. These results confirm the performance ranking of KNN > Gradient Boosting > SVM for daily geothermal production data classification. The application of SMOTE improved minority-class Recall, particularly for SVM, with noticeable gains in detecting High and Low productivity wells. However, for KNN and Gradient Boosting, SMOTE reduced overall accuracy by approximately 0.7–3.3 percentage points depending on the data split, indicating that oversampling did not consistently enhance global performance. This finding suggests that the original data distribution was already sufficiently informative for distance-based and ensemble methods, while margin-based SVM benefited more from class balancing. Practically, the improvement from conventional manual processingthis paperrequiring hours to daysthis paperto an automated classification model achieving over 94% accuracy represents a substantial operational impact. The selected KNN model enables faster productivity classification, supports near-real-time monitoring, and provides more reliable input for geothermal production forecasting and development planning. Thus, this study offers a quantitatively validated and operationally implementable framework for well-level geothermal productivity assessment.

## REFERENCES

[1] A. Al-Fakih, A. Al-khudafi, A. Koeshidayatullah, S. Kaka, and A. Al-Gathe, “Forecasting geothermal temperature in western Yemen with Bayesian-optimized machine learning regression models,” *Geothermal Energy*, vol. 13, no. 1, p. 4, Jan. 2025, doi: 10.1186/s40517-024-00324-3.

[2] N. Saxena *et al.*, “Hybrid KNN-SVM machine learning approach for solar power forecasting,” *Environmental Challenges*, vol. 14, p. 100838, Jan. 2024, doi: 10.1016/j.envc.2024.100838.

[3] A.-M. Koray, E. Gyimah, M. Metwally, H. Rahnema, and O. Tomomewo, “Leveraging machine learning for enhanced reservoir permeability estimation in geothermal hotspots: a case study of the Williston Basin,” *Geothermal Energy*, vol. 13, no. 1, p. 8, Jan. 2025, doi: 10.1186/s40517-024-00323-4.

[4] B. Mukherjee, S. Kar, and K. Sain, “Machine Learning Assisted State-of-the-Art-of Petrographic Classification From Geophysical Logs,” *Pure Appl. Geophys.*, vol. 181, no. 9, pp. 2839–2871, Sep. 2024, doi: 10.1007/s00024-024-03563-4.

[5] R. Dwivedi *et al.*, “Granite porosity prediction under varied thermal conditions using machine learning models,” *Earth Sci. Inform.*, vol. 18, no. 2, p. 211, Jun. 2025, doi: 10.1007/s12145-025-01726-y.

[6] J. He, K. Li, X. Wang, N. Gao, X. Mao, and L. Jia, “A Machine Learning Methodology for Predicting Geothermal Heat Flow in the Bohai Bay Basin, China,” *Natural Resources Research*, vol. 31, no. 1, pp. 237–260, Feb. 2022, doi: 10.1007/s11053-021-10002-x.

[7] Mahmoud. M. AlGaiar, “Unlocking Geothermal Potential: Advancing Exploration with Artificial Intelligence for Sustainable Energy Solutions,” in *ADIPEC*, SPE, Nov. 2024, p. 25. doi: 10.2118/222269-MS.

[8] M. Mahetaji and J. Brahma, “Enhancing wellbore stability through machine learning for sustainable hydrocarbon exploitation,” *Sci. Rep.*, vol. 15, no. 1, p. 35268, Oct. 2025, doi: 10.1038/s41598-025-17588-9.

[9] A. Saad *et al.*, “Enhancing Geothermal Drilling Performance: Predicting Rate of Penetration with Machine Learning Utilizing Geomechanical and Petrophysical Data,” in *SPE Offshore Europe Conference & Exhibition*, SPE, Sep. 2025. doi: 10.2118/226760-MS.

[10] Z. Guo, K. Li, H. Zhang, C. Gu, and J. He, “Predicting Geothermal Heat Flow in the Bohai Bay Basin Based on Machine Learning Methods,” *Math. Geosci.*, vol. 57, no. 5, pp. 925–949, Jul. 2025, doi: 10.1007/s11004-025-10182-9.

[11] Y. Xiong, M. Zhu, Y. Li, K. Huang, Y. Chen, and J. Liao, “Recognition of Geothermal Surface Manifestations: A Comparison of Machine Learning and Deep Learning,” *Energies (Basel).*, vol. 15, no. 8, p. 2913, Apr. 2022, doi: 10.3390/en15082913.

[12] E. R. Ugarte and S. Salehi, “A Robust Screening Tool to Repurpose Hydrocarbon Wells to Geothermal Wells in Oklahoma,” in *SPE Oklahoma City Oil and Gas Symposium*, Oklahoma: SPE, Apr. 2023, pp. 11–21. doi: 10.2118/213068-MS.

[13] T. yehia, M. Gasser, H. Ebaid, E. Okoroafor, and N. Meehan, “A Comparative Analysis of Machine Learning Techniques for Geothermal Wells’ Drilling Rate of Penetration (Rop) Prediction,” in *Unconventional Resources Technology Conference*, Texas: SPE/AAPG/SEG, Jun. 2024, pp. 5–8. doi: https://doi.org/10.15530/urtec-2024-4044244.

[14] H. Ebaid, E. Rita Okoroafor, N. Meehan, T. Yehia, and M. Gasser, “A Comparative Analysis of Machine Learning Techniques for Geothermal Wells’ Drilling Rate of Penetration (ROP) Prediction,” in *The Unconventional Resources Technology Conference*, Tulsa, OK USA: American Association of Petroleum Geologists, Jun. 2024, pp. 8–16. doi: 10.15530/urtec-2024-4044244.

[15] A. Al-Fakih, A. Abdulraheem, and S. Kaka, “Application of machine learning and deep learning in geothermal resource development: Trends and perspectives,” *Deep Underground Science and Engineering*, vol. 3, no. 3, pp. 286–301, Sep. 2024, doi: 10.1002/dug2.12098.

[16] M. Ahmadi, “Interpretable Machine Learning for High-Accuracy Reservoir Temperature Prediction in Geothermal Energy Systems,” *Energies (Basel).*, vol. 18, no. 13, p. 3366, Jun. 2025, doi: 10.3390/en18133366.

[17] A. Shahdi, S. Lee, A. Karpatne, and B. Nojabaei, “Exploratory analysis of machine learning methods in predicting subsurface temperature and geothermal gradient of Northeastern United States,” *Geothermal Energy*, vol. 9, no. 1, p. 18, Dec. 2021, doi: 10.1186/s40517-021-00200-4.

[18] T. Yehia, M. Gasser, H. Ebaid, N. Meehan, and E. R. Okoroafor, “Comparative analysis of machine learning techniques for predicting drilling rate of penetration (ROP) in geothermal wells: A case study of FORGE site,” *Geothermics*, vol. 121, p. 103028, Jul. 2024, doi: 10.1016/j.geothermics.2024.103028.

[19] S. I. Mohammad *et al.*, “Expedited and dependable geothermal rock characterization and absolute permeability modeling using advanced data-driven techniques,” *Geothermal Energy*, vol. 13, no. 1, p. 45, Dec. 2025, doi: 10.1186/s40517-025-00371-4.

[20] Y. Chen, K. Li, H. Zhang, C. Gu, J. He, and W. Lu, “Enhanced Prediction of Geothermal Heat Flow Using an Improved GBRT Model with Genetic Algorithm,” *Natural Resources Research*, vol. 34, no. 5, pp. 2509–2535, Oct. 2025, doi: 10.1007/s11053-025-10501-1.

[21] L. Wang, Z. Yu, Y. Zhang, and P. Yao, “Review of machine learning methods applied to enhanced geothermal systems,” *Environ. Earth Sci.*, vol. 82, no. 3, p. 69, Feb. 2023, doi: 10.1007/s12665-023-10749-x.

[22] Y. Duan, Y. Liang, Q. Ji, and Z. Wang, “A Machine Learning Approach for the Clustering and Classification of Geothermal Reservoirs in the Ying-Qiong Basin,” *J. Mar. Sci. Eng.*, vol. 13, no. 3, p. 415, Feb. 2025, doi: 10.3390/jmse13030415.

[23] S. Novianti, T. Nurkholifa, M. Suryana, E. Susanto Program Studi Usaha Perjalanan Wisata, and J. Administrasi Niaga Politeknik Negeri Bandung, “Penggunaan Geographical Information System (GIS) untuk Visualisasi Analisis Perilaku Spasial Wisatawan,” vol. 4, no. 2, pp. 2654–3894, 2021, doi: 10.17509/jithor.v4i2,%20October.37168.

[24] P. U. Ekeopara, C. J. Nwosu, F. M. Kelechi, C. P. Nwadiaro, and K. K. ThankGod, “Prediction of Thermal Conductivity of Rocks in Geothermal Field Using Machine Learning Methods: a Comparative Approach,” in *SPE Nigeria Annual International Conference and Exhibition*, Lagos: SPE, Jul. 2023, pp. 11–14. doi: 10.2118/217217-MS.