Abstract—In this paper, we classify Kaizen (production process related best practices) from Volkswagen internal knowledge database into their adaptability status over other Volkswagen production plants with help of supervised machine learning. Different criteria’s like Return on Investment, implementation time-frame, impact on various other key performance indicators, plant-specific technical details, etc. are used to evaluate Kaizens and eventually cluster them into two categories. Empirical results show that the Decision tree model can predict the degree of adaptability (Success/Denied) with 85% accuracy.

Index Terms—Supervised Machine Learning, Kaizen, Best Practices Evaluation, Clustering, Production system

I. INTRODUCTION

Automotive industry with a global production network generally has knowledge repository of Kaizens with the aim of data transparency, accessibility, and transfer of productivity improvement among various manufacturing plants. The Kaizen (Japanese term for ‘improvement’, also known as Best practices) implemented at the production shop floor are documented as per organization standards and then stored in this knowledge repository. It contains information such as before and after situation, cost and other benefits, time frame for implementation as well as their impact on productivity-related key performance indicators (KPI). The adaptability of Kaizens in particular plants is determined by diverse parameters like the return of investment, manufacturing process similarity, topography, process-related technical restrictions, etc. The present process comprises of evaluating these Kaizens manually to see their adaptability, which is time enduring and requires expert knowledge in the related manufacturing process. The purpose of this paper is to investigate the adaptability of these Kaizens automatically in different Volkswagen manufacturing plants with classification based Machine learning models. This research serves two purposes for industrial application – minimizing risk for the production team by automatically evaluating new Kaizen beforehand and guiding the central team to identify and transfer plant-specific Kaizens. This research is an initial step towards designing a prescriptive analytics system working on KPI prediction along with Kaizen recommendation. This paper is organized as follows. In Section II, data cleaning and feature selection process is discussed. Section III focuses on machine learning model selection and optimization. Model performance comparison based on different sampling and optimization methods is reported in Section IV. Finally, Section V concludes the paper.

II. DATA PREPARATION

Paint shop data for 13 Volkswagen plants from the internal knowledge database for the last 12 years is studied. Each documented Kaizen in the database contains information about different KPI features along with their respective adaptability status depicting whether it was successfully implemented or denied in a particular Volkswagen plant. For training our Machine learning model, this Kaizen adaptability status (Success/Denied) is designated as a target indicator in the available labeled data set. Original Kaizen data from the knowledge repository consists of 64 initial features, with plant name as one of the feature columns. Besides this data set, another plant data set was prepared with technical specifications about each plant with paint shop details like the number of colors used, car segment manufactured, automation level, technology used, etc. Both data sets are merged to form 93 features, out of which 44 are eliminated due to knowledge redundancy and non-value addition to our training model after semi-structured interviews with related field experts. Furthermore features with unique values and with near-zero variance were excluded.

Only those Kaizen are considered for further analysis, which has finalized the status of implementation (Success/Denied). Initial data with 3473 data points is reduced to 2664 after data cleaning. Data cleaning constituted outlier detection, missing value replacement, duplicate data elimination, and constraints verification (Range, data-type, mandatory values). Outlier detection was carried out to remove erroneous observations present due to manual data recording. The graph of initial
features with numerical data was plotted against predicted label (Kaizen adaptability) and then analyzed to distinguish outliers. Machine learning analysis along with feature selection is done in RapidMiner software (V8.1) [1], which is an open-source data mining platform. It was preferred for this research, due to its compatible analytics visualization with other business stakeholders. For numerical features selection, filter type methods like Principal component analysis (PCA) and weight by information gain were used after normalizing the data. In addition, the correlation matrix was evaluated for features with binary data only to substantiate the absence of redundant features. The numerical features are normalized (Z-Transformation) before training and testing the data set.

III. METHODOLOGY

As the target indicator has more observations in one specific class (Success), two different oversampling methods are tested to abtain the model from susceptibility to majority class. The Bootstrapping method uses replacement by sampling repeatedly in the base data set to tackle data skewness. In SMOTE (Synthetic Minority Over-sampling Technique) sampling approach, the minority class is over-sampled by creating synthetic examples based on k-nearest neighbor [2].

ROC (Received Operator Characteristic) curve was plotted with a fraction of True positive versus the fraction of False positive for different clustering model comparisons. Model performance comparison for the same data set was done by analyzing Area Under the Curve (AUC) from ROC graph. ROC curves can exhibit the relation between sensitivity and specificity. For the classifier, Sensitivity (Class Recall) is the ability to select all the cases that need to be selected and Specificity is the ability to reject all the cases that need to be rejected. Model training is carried using cross-validation with Simple K-fold method (10 folds) and stratified sampling for building subsets. Model performance is done by evaluating accuracy from the confusion matrix, as it is an aggregate measure of classifier performance. Accuracy is calculated as below from True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

\[
\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)
\]

Performance comparison between grid search and evolutionary search (based on Genetic algorithm) is done for hyper tuning of the selected Machine learning model. Grid search moves the interval window in the defined range to capture the global minimum, but at times gets stuck at local optima. On the other hand, the Evolutionary algorithm covers large search space using mutation and crossover of successful models based on success criteria [3].

IV. FINDINGS

A. Model Selection

Decision tree, Random Forest, K-Nearest Neighbor & Naive Bayes was initially considered for classification based on feature characteristics present in data (binary, categorical, numerical type). The Decision tree was selected for hyper tuning due to the higher area under the curve as compared to other clustering models based on ROC graph. Overall model accuracy for decision tree ranged from 84% - 88% based on different sampling and optimization methods.

B. Model Performance

The selection of data sampling method played a key role in model accuracy as well as the Class recall ratio. With the application of bootstrap sampling and grid optimization, an accuracy of 88% was recorded. The class recall value of success was found to be higher (96%) as compared to Denied class (62%) indicating that the model is biased due to data imbalance. With SMOTE sampling, model accuracy was reduced to 85%. Nevertheless, class recall for the positive and negative class was nearly equal, which is critical for a balanced model (See Fig. 1 - Grid optimization). The similar results were observed for evolutionary optimization. The optimum model would be a decision tree that is hyper tuned using grid optimization along with SMOTE sampling with 85% accuracy for training our data set.

![Fig. 1. Performance comparison based on sampling and model optimization method](image)

V. CONCLUSION

The research depicts that Kaizen data along with plant-specific technical data can be used to predict their adaptability status in advance with good accuracy. Another practical application of this research can be to identify particular Kaizens from databases that have a higher likelihood of being successfully implemented at the maximum number of production plants. As further steps for this research, the current data along with text data needs to be tested using deep learning methods in conjunction with text mining. Besides, the optimized model can be created for other sections of plant like Body shop, Assembly line and Press shop.

REFERENCES

