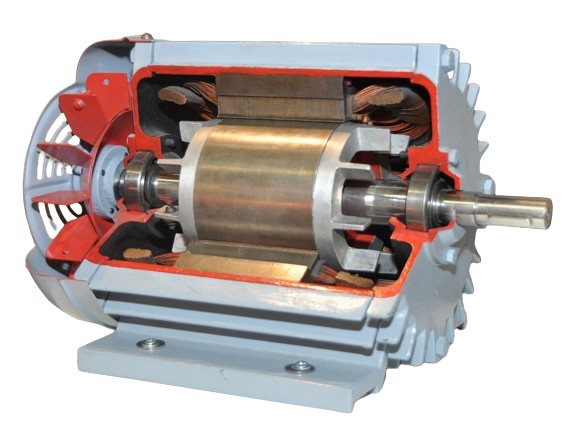
**DS-323-Machine Learning for Data Science (Fall 2023)**

------------------------------------------------------------------------------

**Assignment**

---------------------------------

|Final Report

* **Three-Phase** Current

Signature Data of loaded **Induction Motor**

**Objective**

The goal of this project was to build and evaluate a k-NN (k-Nearest Neighbors) classification model on a given dataset. The analysis includes applying the holdout method, 10-fold cross-validation, and incorporating PCA (Principal Component Analysis) for dimensionality reduction.

**Dataset**

The preprocessed dataset used for this project consists of features and labels. The features represent various attributes, and the labels indicate the class to which each instance belongs.

**Data Preprocessing**

Label Encoding: The categorical labels were encoded into numerical format using the LabelEncoder.

Train-Test Split: The dataset was split into training and testing sets using the holdout method and 10-fold cross-validation.

Modeling and Evaluation:

Baseline **k-NN Model**

The initial k-NN model was implemented with a specified value for k.

The holdout method was used to evaluate the model's performance, considering metrics such as accuracy, precision, recall, and F1 score.

**Holdout Method** with 10-Fold Cross-Validation

The holdout method was applied to split the data into training and testing sets.

Additionally, 10-fold cross-validation was performed to assess the model's robustness.

PCA Integration:

**PCA** was applied to reduce the dimensionality of the dataset.

The k-NN model was then trained and evaluated on the PCA-transformed data.

Both holdout and 10-fold cross-validation were conducted for PCA-integrated models.

Results:

The results were compared across **different scenarios:**

**Without PCA (Holdout Method):**

Accuracy: 0.2209

Precision: 0.2191

Recall: 0.2209

F1 Score: 0.2151

**With PCA (Holdout Method):**

Accuracy: 0.1977

Precision: 0.2191

Recall: 0.2074

F1 Score: 0.2014

Specificity: 0.5238

Sensitivity: 0.9801

**Average Metrics with PCA (10-Fold Cross-Validation):**

Average Accuracy: 0.2089

Average Precision: 0.2411

Average Recall: 0.2309

Average F1 Score: 0.2108

Average Specificity: 0.6298

Average Sensitivity: 0.9758

**Conclusion:**

* The model without PCA achieved slightly better accuracy in both the holdout method

and 10-fold cross-validation.

* The PCA-integrated model demonstrated potential benefits in terms of feature reduction, but it showed lower accuracy.
* The choice between models should consider the trade-offs between accuracy, interpretability, and computational efficiency based on specific application requirements.

**Recommendations:**

Further exploration of hyperparameter tuning for k in k-NN models may be beneficial.

Consideration of other dimensionality reduction techniques and their impact on model performance.

Evaluation of additional machine learning algorithms for comparison and model selection.

This project provides insights into the performance of k-NN models under different scenarios and lays the groundwork for future enhancements and optimizations.