### SYNOPSIS

**Project Team No:** 21SOCU1122

### Register No: Name:

1. 122003073 Gandrothu Pavan Sai Kishore
2. 122003138 Maddu Bhaskar Vamsi Krishna
3. 122014053 Tamalampudi Venkata Rama Reddy

**Project Title:** CHRONIC KIDNEY DISEASE PREDICTION USING MACHINE LEARNING

**Name of Guide:** Dr. Renuga Devi. T, Asst. Professor - II, School of Computing

**Abstract**

Since curing a disease is possible only after detection, disease detection has become an essential factor in the medical sector. Chronic kidney disease (CKD), one among the major diseases that needs to be detected at early stages. Nowadays, machine learning (ML) is playing a crucial role in the detection of diseases. So, in this paper, we mainly focus on finding the best ML model for prediction of CKD. For this perspective, a data set was collected from UCI repository. We applied 7 ML classifier algorithms, namely, K-Nearest Neighbor (KNN), Logistic Regression (LR), Linear Support Vector Machine (LSVM) with penalties L1 and L2, Random Forest (RF), CHAID, C 4.5, and Artificial Neural Network (ANN) for prediction. We apply Feature Selection (FS) methods such as full feature selection, correlation-based Feature Selection (CFS), wrapper FS, LASSO FS for selecting the features. A data set for the SMOTE, a data balancing technique is applied to the data set for balancing the target class. Classifier performance is measured using accuracy, precision, recall, error, F-measure, AUC, and GINI index.

In this paper, the results are computed for all classifier algorithms applied to each feature selection method. From the results, the best-performing algorithms with the best feature selection methods are selected, and their performances are computed with the application of SMOTE in each case. From the results, we observed that LR, LSVM L1, LSVM L2, CHAID, C 4.5, and RF perform better in measure of accuracy and other evaluation metrics than Full Features and LASSO Feature Selection in measure of accuracy and other evaluation metrics. After applying SMOTE with LASSO Feature Selection, Random Forest performs best with an accuracy of 98.49%, and LSVM with penalty L2 has an accuracy of 97.73%.

### Specific Contribution

* + Data Preprocessing.
  + Implemented Feature Selection techniques and ML algorithms.

### Specific Learning

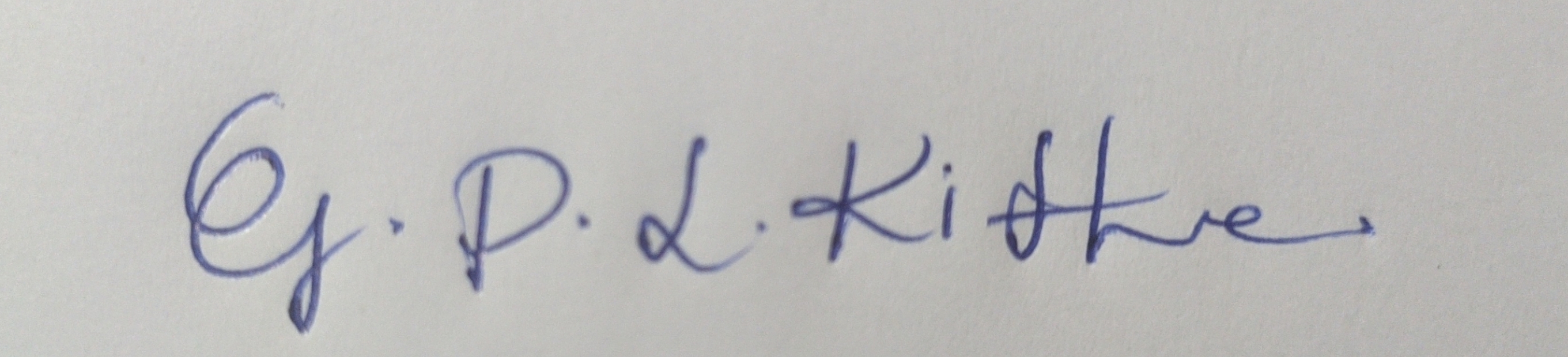
* + Implementing various ML algorithms in Python.
  + Learned about the different feature selection techniques.

### Technical Limitations & Ethical Challenges faced

* + Data imbalance in the data set.
  + Implementing all 3 feature selection techniques.

***Keywords:*** *Feature Selection techniques,Data preprocessing, ML algorithms.*

**Gandrothu Pavan Sai Kishore**



**Name & Signature of the Student Signature of Guide Date:** 09/01/2022