PROJECT TOPIC: Diagnosing Chronic Kidney Disease Using Machine Learning Algorithms

Project Group Members:

1) Abhishek Gupta (Sec- G Univ Roll No. - 181500016)
2) Naman Shukla (Sec- G Univ Roll No. - 181500312)

Project Supervisor: Mr Rahul Pradhan, Assistant Professor

About the Project:

Chronic kidney disease (CKD) is a global health problem with high mortality rate, and it induces other diseases. Since there are no obvious symptoms during the early stages of CKD, patients often fail to notice the disease. Early detection of CKD enables patients to receive timely treatment to ameliorate the progression of this disease. The new analysis suggests that in 2017, the global prevalence of CKD was 9.1% (697.5 million cases). The age-standardized global prevalence of CKD was higher in women and girls (9.5%) than in men and boys (7.3%). In this study, we propose a machine learning methodology for diagnosing CKD. In this study, we will be using six Machine Learning algorithms (logistic regression, random forest, support vector machine, k-nearest neighbour, naïve Bayes classifier and feed forward neural network) to establish the model.

Motivation:

As mentioned above how the CKD is a global problem. CKD does not show obvious symptoms in its early stages. Therefore, the disease may not be detected until the kidney loses about 25% of its function. In addition, CKD has high morbidity and mortality, with a global impact on the human body. It can induce the occurrence of cardiovascular disease. Hence, the prediction and diagnosis of CKD in its early stages is quite essential.

Project Planning:

- The CKD data set is obtained from the University of California Irvine (UCI) machine learning repository, which has a large number of missing values.
- KNN imputation will be used to fill in the missing values, which selects several complete samples with the most similar measurements to process the missing data for each incomplete sample.
- Logistic regression (LOG), RF, SVM, KNN, naive Bayes classifier (NB) and feed forward neural network (FNN) will be used to establish CKD diagnostic models on the complete CKD data sets.

- The models with better performance will be extracted for misjudgement analysis.

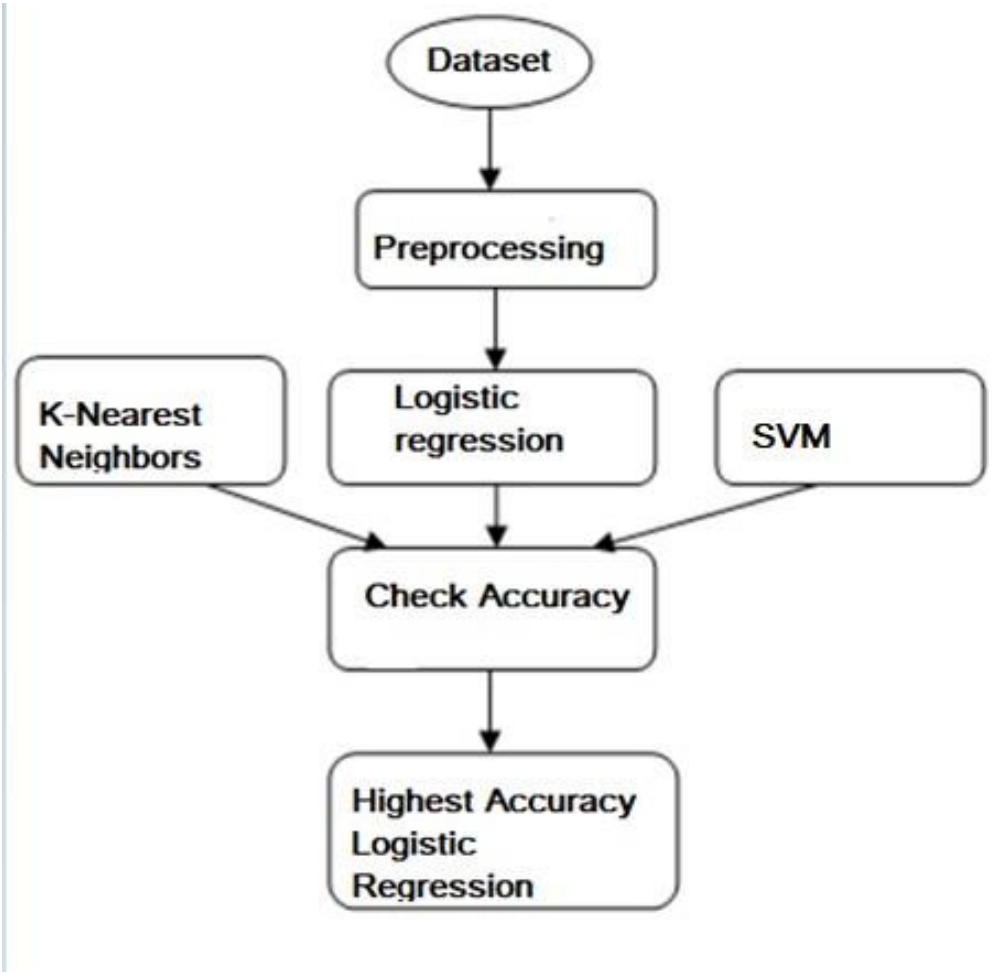

Fig:2 Data Flow Diagram

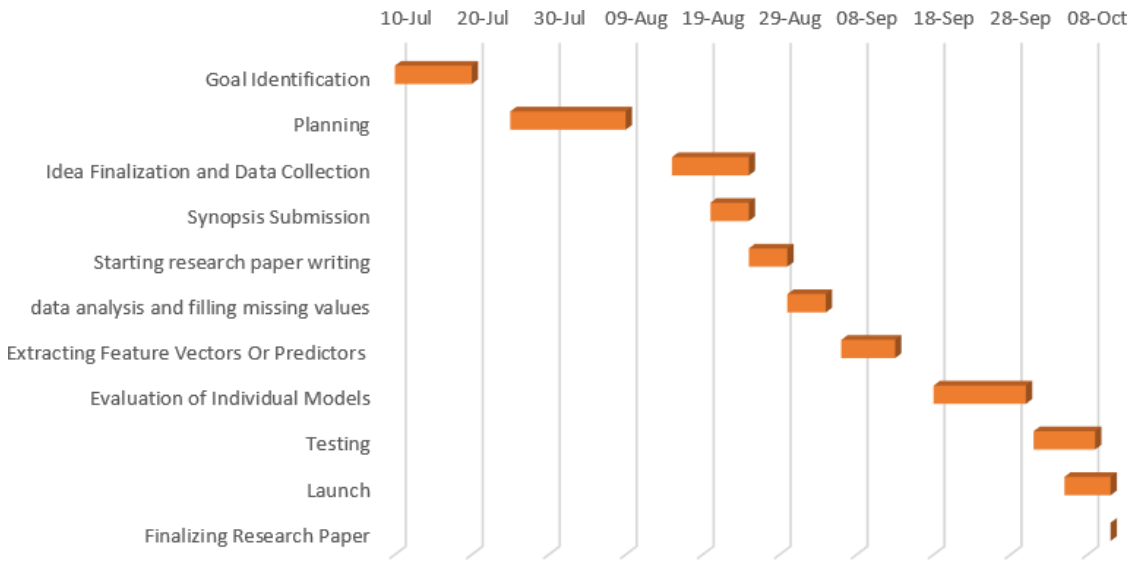

Fig 2: Gantt Chart

Tools required:

- **Hardware Requirements:**

  I. Laptop with i3 5th Generation
  II. Min. 4GB Ram

- **Software Requirements:**
  Python 3
  Libraries:
  - Numpy
  - Pandas
  - Matplotlib
  - Scikit-learn
  - Scipy

References:

- Chronic Kidney Disease Prediction using Machine Learning Models S.Revathy, B.Bharathi, P.Jeyanthi, M.Ramesh
- A Machine Learning Methodology for Diagnosing Chronic Kidney Disease ,Bin Chen
- Kaggle (For dataset Collection)
- https://ieeexplore.ieee.org