

**SHRI VISHNU ENGINEERING COLLEGE FOR
WOMEN(AUTONOMOUS)::BHIMAVARAM**
**Department of Computer Science and
Engineering IV B.Tech II SEM - Project
Status Report**

Section – B

A.Y: 2020-2021

Project Batch No:	7
Project Title	Chronic Kidney Disease Prediction
Project Guide	Mr. K. Bhadrachalam
Project Aim (in 250 words)	The aim of the project is to detect CKD using machine learning algorithms by considering the least number of tests or features. We approach this by applying machine learning classifiers on a small dataset of 400 records. To remove the null values, we perform Data Cleaning. In order to increase the size of the dataset, we perform data augmentation. In order to reduce the number of features, the association between variables has been studied. Feature selection methods has been applied to the attributes and found that there are some of the attributes which impact mostly to predict the CKD.
Literature survey (list of books, online material studied)	https://www.kidney.org/atoz/content/about-chronic-kidney-disease https://www.kaggle.com/akshayksingh/kidney-disease-dataset https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease https://machinelearningmastery.com/types-of-classification-in-machine-learning/ https://www.analytixlabs.co.in/blog/classification-in-machine-learning/
Tasks / Modules completed so far(List the modules and percentage)	Data Cleaning (100%) Visualization (100%) Data Augmentation (100%) Feature Selection (100%) Applying Models (100%) Result Analysis and User Interface (100%)

<p>Role of each Project member</p>	<p>K.L.N. Pravallika – Python Programming(Data Cleaning, Feature Selection, Data Augmentation) and Designing User Interface</p> <p>M. Dedeepya - Python Programming(Applying Algorithms to dataset, Data Augmentation)</p> <p>D. Srilakshmi - Python Programming(Visualization) , Data Augmentation , Designing User Interface.</p> <p>N.S.L. Meghana - Python Programming(Visualization), Data Augmentation</p>
<p>Questions asked in previous reviews</p>	<p>Review1</p> <ol style="list-style-type: none"> 1. Justify the title of your project? A. Our project is to predict whether a person is suffering from CKD or not. 2. What is the Existing System? A. The existing system of diagnosis is based on the examination of urine with the help of serum creatinine level. 3. Drawbacks of Existing System? A. Loss of money. Delay in time. Needs more human effort. <p>Review2</p> <ol style="list-style-type: none"> 1. Abstract of the project? A. We will consider kidney disease dataset, first we clean the dataset which means filling the null values. Later we apply feature selection techniques. We apply algorithms on the dataset which contain only selected features. This predicts whether a person is "ckd" or "not ckd" . 2. Techniques Filling missing values? A. You can use the attribute mean to fill in the missing value. Use the most probable value to fill in the missing value. You can fill in the missing value manually. 3. How can you predict a person ckd or not? A. We apply feature selection techniques. We apply algorithms on the dataset which contain only selected features. This predicts whether a person effected with

	<p>kidney disease or not.</p> <p>4. Objective of the project?</p> <p>A. To predict whether a person effected with kidney disease or not.</p> <p>5. How you calculate accuracy, precision and recall?</p> <p>A. Accuracy= $(TP + TN) / (TP+FP+FN+TN)$ Precision = $TP / (TP+FP)$ Recall = $TP / TP+FN$</p> <p>6. Type 1, 2 errors?</p> <p>A. Type I error is a false positive conclusion Type II error is a false negative conclusion.</p> <p>Review 3</p> <p>1. What is the System Design?</p> <p>Firstly, we will be considering a dataset that contains 25 attributes and performing data cleaning methods. Later, we divide the dataset into a training set and testing set(80-20; 70-30; 60-40). Further, we carry out feature selection on the training data by implementing filter methods and wrapper methods. After this, we will apply machine learning algorithms such as Random forests, Support Vector Machineand then will achieve a training model, using this we will be predicting the disease either it is CKD or non-CKD by considering the accuracy measure of the algorithms.</p> <p>2. What are the algorithms used?</p> <p>SVM, Random Forest</p> <p>3. What is percentage represented in Distplot?</p> <p>Percentage of missing values in that column</p> <p>Reviw4</p> <p>1. What is the toughest work in your project?</p> <p>Feature Selection</p> <p>2. How did you perform models?</p> <p>Using the in-built functions</p> <p>3. How did you calculate accuracies?</p> <p>Using the confusion matrix.</p> <p>4. What does the scoring parameter represent in Feature Selection?</p> <p>It denotes on which basis the parameters are considerd.</p>

Project Guide Signature:

Project Teams Members:

Name	Redg No	Signature
K. L. N. Pravallika	17B01A0567	K. L. N. Pravallika
M. Dedeepya	17B01A0585	M. Dedeepya
D. SriLakshmi	17B01A0591	D. SriLakshmi
N.S.L. Meghana	17B01A05A4	N. S. L. Meghana

Project Coordinator Signature

Head of the Department Signature