



**R V College
Of
Engineering**

Go, change the world

**MACHINE LEARNING ASSIGNMENT
PHASE -01**

**TITLE –DIABETES PREDICTION USING MACHINE
LEARNING**

Group Members:

1RV20MC026 HARSHA VARDHANA P D
1RV20MC108 VIDYA ARADHYA G R

1. PROBLEM DEFINITION

- Diabetes Mellitus is among critical diseases and lots of people are suffering from this disease.
- Age, obesity, lack of exercise, hereditary diabetes, living style, bad diet, high blood pressure, etc. can cause Diabetes Mellitus.
- People having diabetes have high risk of diseases like heart disease, kidney disease, stroke, eye problem, nerve damage, etc.
- Current practice in hospital is to collect required information for diabetes diagnosis through various tests and appropriate treatment is provided based on diagnosis.
- In this project we are going to predict whether a person has diabetes or not.
- So we will be using one of the important machine learning algorithms support vector machine.
- So we will do our coding in python and first of all let's try to understand about this problem statement and we understand about this support vector machine algorithm.

2.DATA SET AND FEATURES CONSIDERED

About the Dataset:

- Pregnancies- Count of pregnant women.
- Glucose- Glucose level of the women.
- Blood Pressure- Rate of blood pressure of particular women.
- Skin Thickness- Deals with the fat content.
- Insulin -Insulin level of a particular women.
- BMI -obtained by dividing the weight by eight squared.
- Diabetes Pedigree Function- It's a number which indicates the diabetic value.
- Age- Age group of the women.
- Outcome:
 - ✓ 0 - for non-diabetic.
 - ✓ 1- for diabetic.

3.MACHINE LEARNING ALGORITHM USED

- With respect to this problem we are using Support Vector Machine Algorithm for prediction.
- This is one of the important supervised learning algorithm.
- We are going to feed the training data into the support vector machine and later it tries to plot the data in the form of graph.
- Later it tries to find out the hyper plane which separates the two data i.e., diabetic and non diabetic.
- A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems.
- After giving an SVM model sets of labelled training data for each category, they're able to categorize new text.
- Two main advantages:
 - ✓ Higher speed and better performance with a limited number of samples.
 - ✓ This makes the algorithm very suitable for text classification problems, where it's common to have access to a dataset of at most a couple of thousands of tagged samples.

4. SAMPLE DATASETS

The data was collected and made available by “National Institute of Diabetes and Digestive and Kidney Diseases”

	A	B	C	D	E	F	G	H	I
1	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
2	6	148	72	35	0	33.6	0.627	50	1
3	1	85	66	29	0	26.6	0.351	31	0
4	8	183	64	0	0	23.3	0.672	32	1
5	1	89	66	23	94	28.1	0.167	21	0
6	0	137	40	35	168	43.1	2.288	33	1
7	5	116	74	0	0	25.6	0.201	30	0
8	3	78	50	32	88	31	0.248	26	1
9	10	115	0	0	0	35.3	0.134	29	0
10	2	197	70	45	543	30.5	0.158	53	1



THANK YOU