Aim: To understand and implement the Ensemble learning technique
(bagging/boosting).

Theory:

Bagging: It reduces variance error and help to avoid overfitting. Uses sampling with replacement to generate multiple sample of given size. Sample may contain repeat data points.

Ada Boosting: It decrease the bias error and build strong predictive models. The algorithm allocates weight to each resulting model.Weights are re-assigned to each instance, with higher weights to incorrectly classified instance.

Library used:

- **SimpleImputer:** It help in handling missing data in predictive model. It replace N$_a$N value with specified placeholder. (here we used mean value).

- **AdaBoost Classifier:** It begin fitting a classifier and fits additional copies of classifier on original Data set .The weight of incorrectly classified are adjusted such that classified focus more difficult cases.

- **Bagging Classifier:** It fits base classifier each on random subsets of Data set and then aggregate their individual prediction.

- **Classification_report:** It used to measure quality of prediction, evaluation metric to show precision, recall, F1 score and support score of model.

- **Data set:** "Diabetes.csv"

The objective of Data set is to diagnostically predict whether or not patient has diabetes, based on certain diagnostic measurements.

Conclusion: Hence, we have successfully implemented ensemble technique like bagging and boosting.

Aim: To understand and implement the linear regression algorithm.

Theory: Linear regression is machine learning algorithm based on supervised learning. A LR model predicts values based on independent variable it was initially trained on via a line of best fit that can be used to extrapolate new values based on dependant variables.
It is used for finding out relationship between variable and forecasting. Fits a line minimizing the sum of mean-squared error for each data point.

General form:

$$Y=m_1x_1 + m_2x_2 + m_3x_3 +........+ m_nx_n + c + e$$

Where,
Y= dependent variable.
Xi= Independent variable.
E= random/stochastic error term.

Library used:

- **Pandas**: It is derived from the word panel data.It can perform five significant step required for processing and analysis of data i.e load, manipulate, prepare, model and analyze.

- **Numpy:** It stand for 'Numerical python'. It consist of multidimensional array objects and collection of routines for processing of array.

- **Linear Regression:** It uses relationship between data-points to draw a straight line through all them.

- **Matplotlib:** It uses to create 2D graphs and plots by using python scripts.

- **SK learn.metrics:** It implement several loss, score and utility function to measure classification performance.

- **Data set:** "Salary_Data.csv".

Conclusion: hence, we successfully implemented linear regression algorithm.