LAB Logbook

Lab 1:

**File\_read:**

Reading the file in csv format

df=pd.read\_csv("telecom\_churn.csv")

**change data type of column:**

it changes the data type of the particular column

df["Churn"] = df["Churn"].astype("int64")

**Sorting:**

It sorts the value in ascending or descending order, sometimes depends on booleen condition which we give

df.sort\_values(by="Total day charge", ascending=False).head()

**Sort by slicing:**

This is used to sort and slice in single line of code, where we can able to order it and get the needed range of data

df.sort\_values(by=["Churn","Total day charge"], ascending=[True,False]).head(-5)-

**Functions:**

Functions have predefined functions, which is also called as build in functions. We can able to perform max, min, average with mean and count.

df.apply(np.max)

**Lambda functions:**

Lambda functions are very convenient in such scenarios.

For example, if we need to select all states starting with W, we can do it like this:

df[df["State"].apply(lambda state: state[0]=="W")].head()

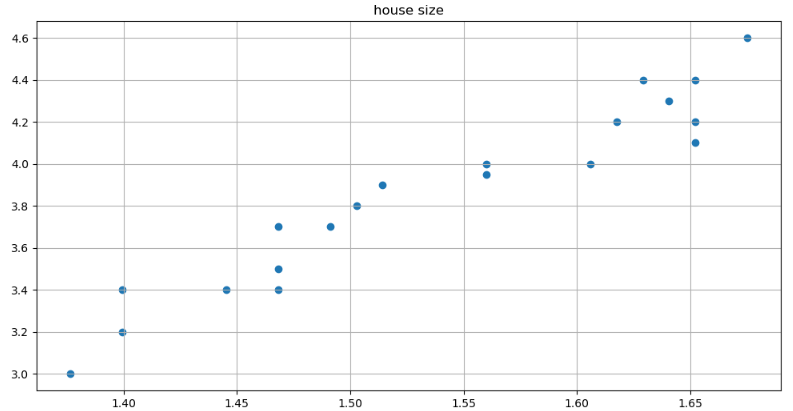
**Groupby:**

 Grouping data in Pandas works as follows: df.groupby(by=grouping\_columns)[columns\_to\_show].function()

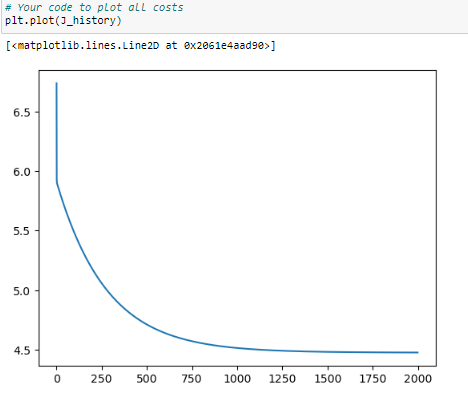
columns\_to\_show = ["Total day minutes","Total eve minutes", "Total night minutes"]

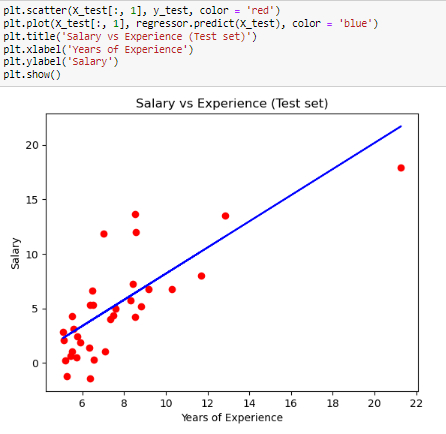
df.groupby(["Churn"])[columns\_to\_show].describe(percentiles=[])

Lab 2

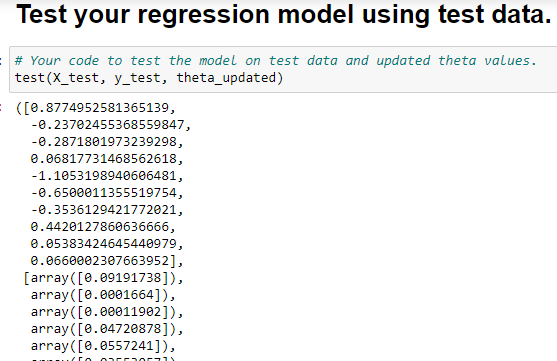


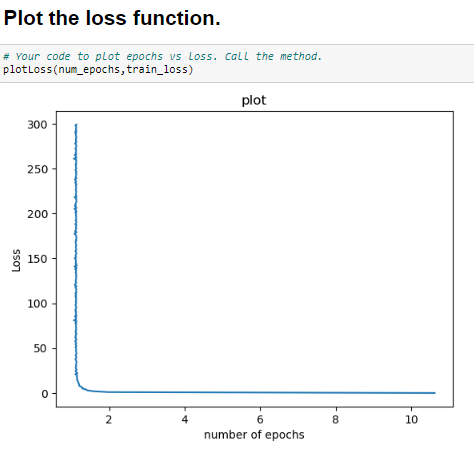
Lab 3

a



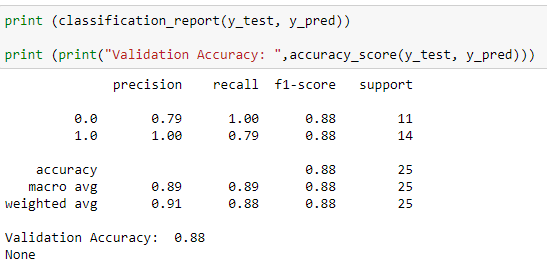
Lab 4

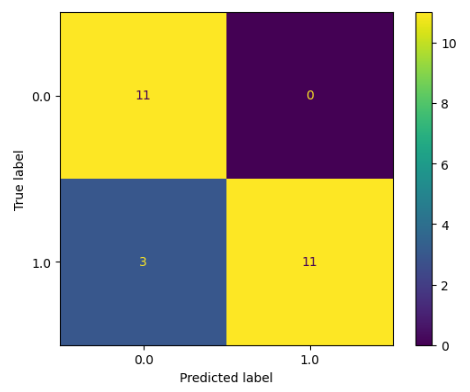




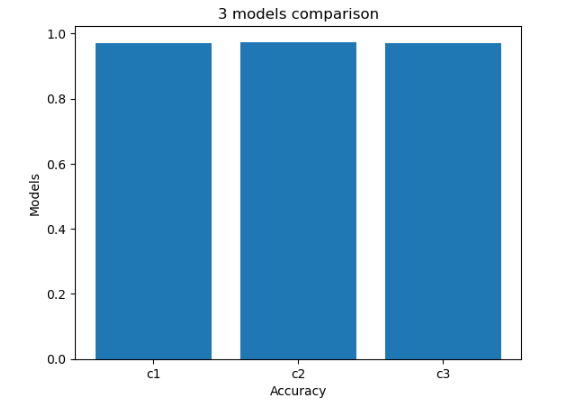


Lab 5

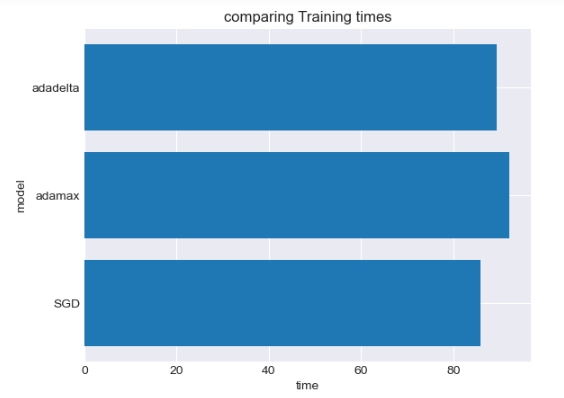




Lab 6



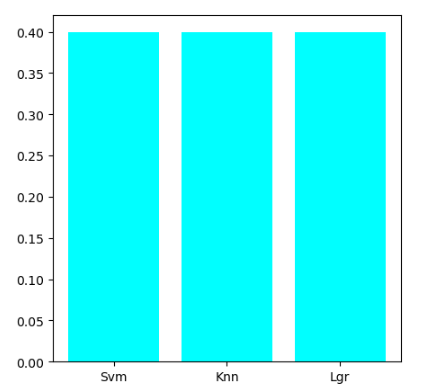
Lab 7



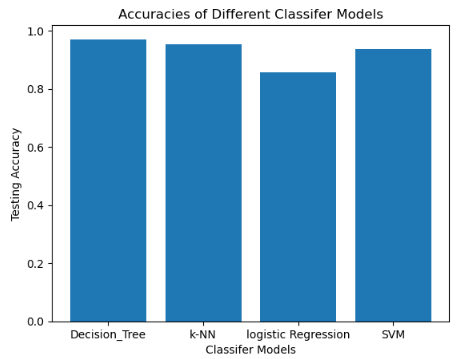
Lab 8

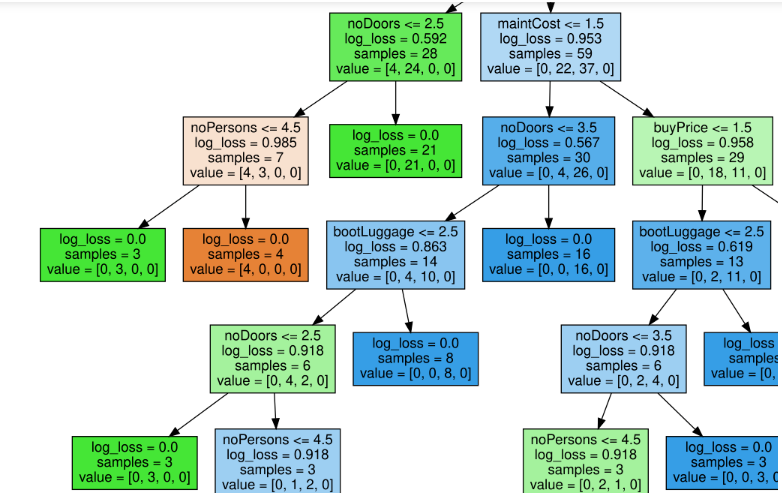


Lab 9



Lab 10





1. What is the decision criterion at each node of the decision tree?

* The decision criterion at each node of the decision tree is typically based on measure of impurity called entropy.
* It has two major types called classification tree and regression tree.

1. How is entropy/gini index/loss change at each node?

* The concept of entropy, gini index, and loss are related to the decision tree algorithm. It determines the best split at each node.

1. Is entropy/gini index/loss change decrease or increase in as we move down in tree?

* Entropy is a measure of disorder at each node, gini index is another measure of impurity which is commonly used in decision tree algorithm and loss is a regression trees. The loss function mostly used to measure the error or variance.

4. How many samples are left at each node? Can you see any pattern?

* The number of samples left at each node in a decision tree can vary based on the dataset and the splitting criteria applied to each node.
* There are certain types of nodes such as, root node, internal node, leaf node and pattern.

5. What information is contained at leaf node?

* A leaf node contains information which includes, majority class or average, class distribution, impurity, and regression value range.