**CAT -3**

**BIG DATA ANALYTICS LAB - 19MAM67**

**TEAM MEMBERS:**

HARINI SRI S - 1934009

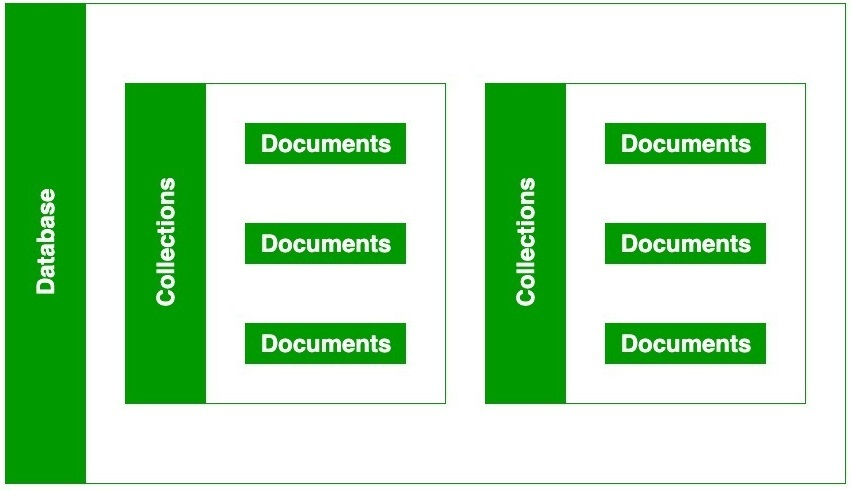
RITHIKA SREE K S - 1934031

SHAMRUTHI R - 1934041

**MONGODB :**

The MongoDB database is developed and managed by MongoDB.Inc under SSPL(Server Side Public License) and initially released in February 2009. It also provides official driver support for all the popular languages like C, C++, C#, and .Net, Go, Java, Node.js, Perl, PHP, Python, Motor, Ruby, Scala, Swift, Mongoid. So, that you can create an application using any of these languages. Nowadays there are so many companies that use MongoDB like Facebook, Nokia, eBay, Adobe, Google, etc. to store their large amounts of data.

The MongoDB environment gives you a server that you can start and then create multiple databases on it using MongoDB. Because of its NoSQL database, the data is stored in the collections and documents. Hence the database, collection, and documents are related to each.



**HOW MONGODB HELPS IN BIG DATA:**

MongoDB is an open-source document-oriented database that is **designed to store a large scale of data and also allows you to work with that data very efficiently**. It is categorized under the NoSQL (Not only SQL) database because the storage and retrieval of data in the MongoDB are not in the form of tables.

**FEATURES OF MONGODB:**

## Indexing appropriately for better query executions

## Replication for better data availability and stability

## Sharding

## Load balancing

* Schema-less Database
* Scalability
* Aggregation
* High Performance

**ADVANTAGES OF MONGODB:**

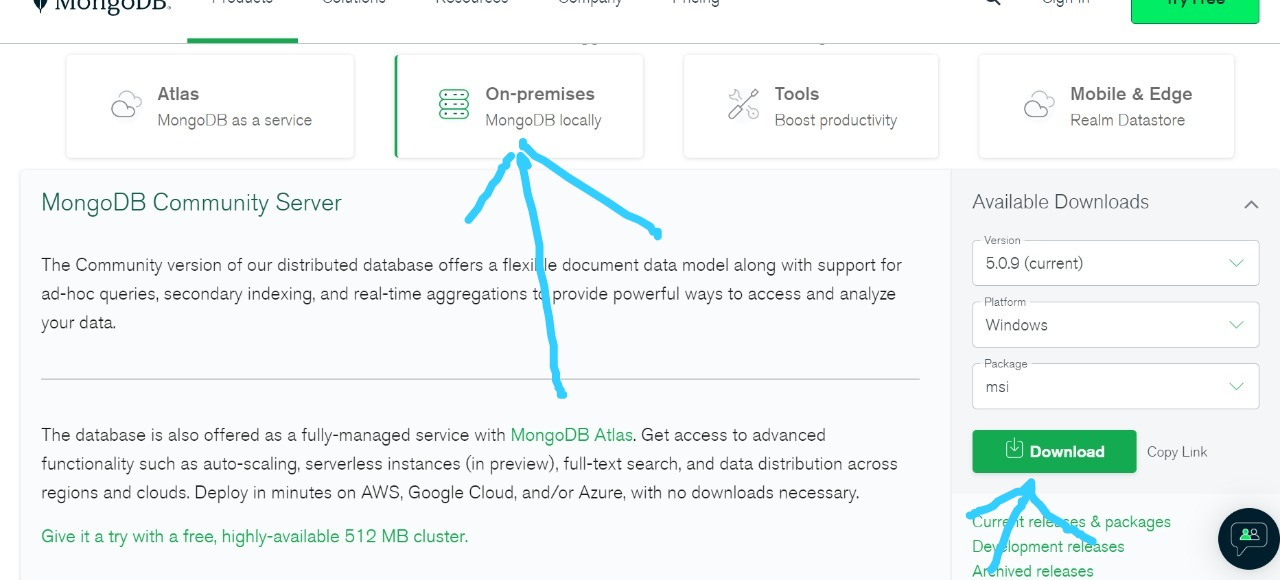
* It is a schemaless NoSQL database. You need not to design the schema of the database when you are working with MongoDB.
* It does not support join operation.
* It provides great flexibility to the fields in the documents.
* It contains heterogeneous data.
* It provides high performance, availability, and scalability.
* It supports Geospatial efficiently.
* It is a document oriented database and the data is stored in BSON documents.
* It also supports multiple document ACID transitions(string from MongoDB 4.0).
* It is easily integrated with Big Data Hadoop

**DISADVANTAGES OF MONGODB:**

* It uses high memory for data storage.
* You are not allowed to store more than 16MB data in the documents.
* The nesting of data in BSON is also limited; you are not allowed to nest data more than 100 levels.

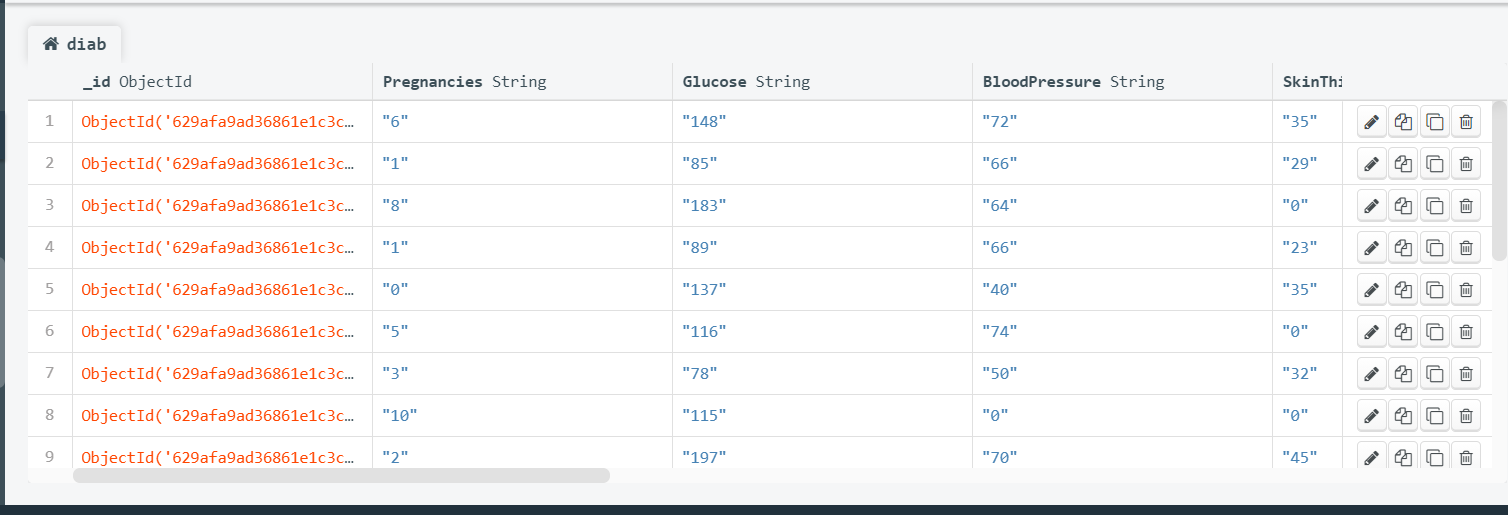
**Installation of mongodb:**

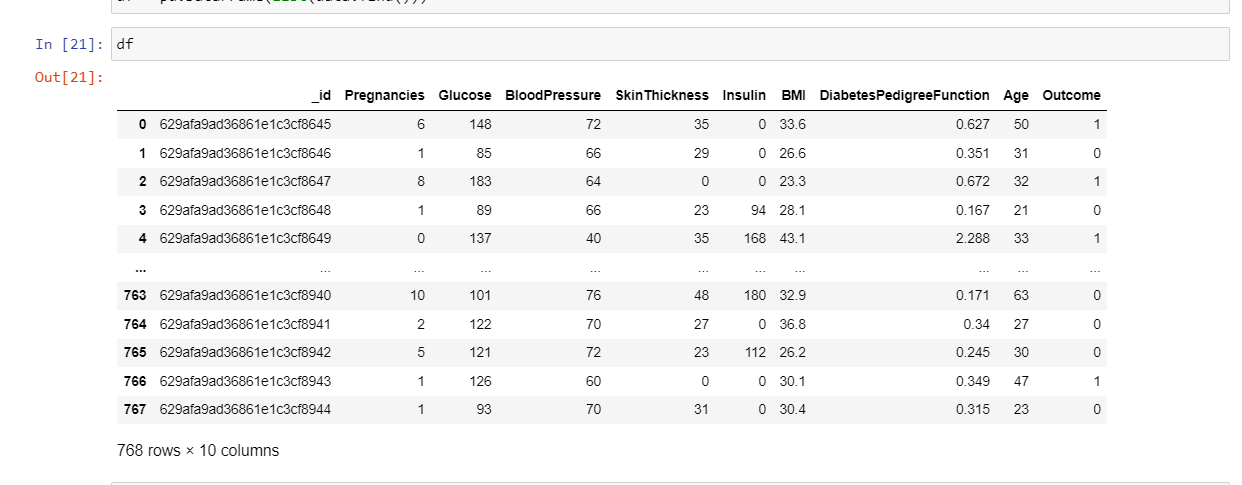
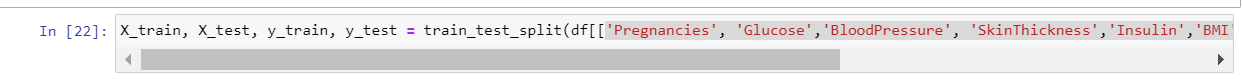
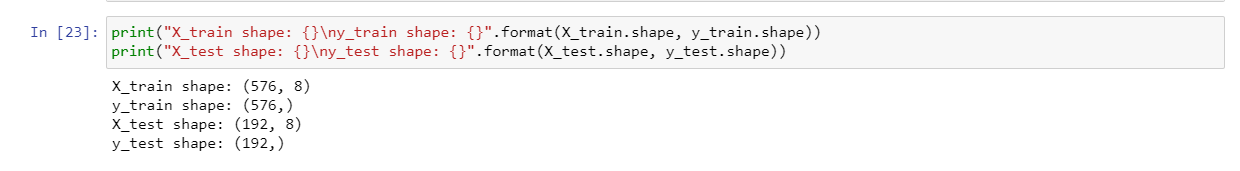
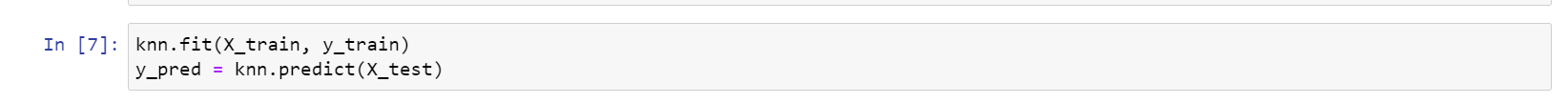
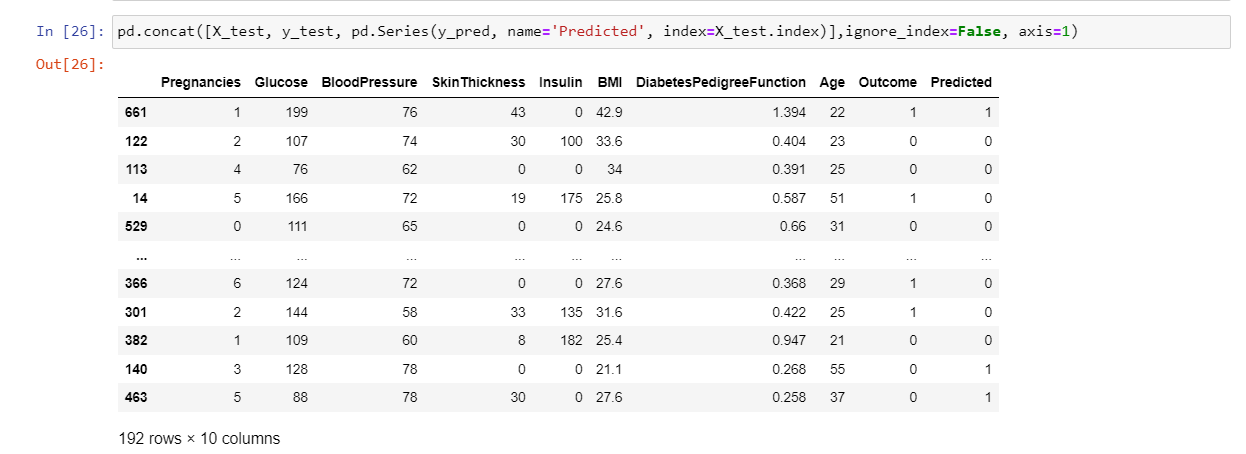
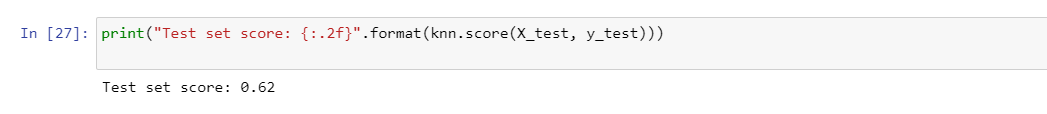
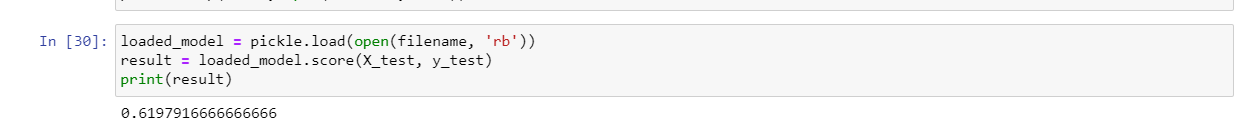
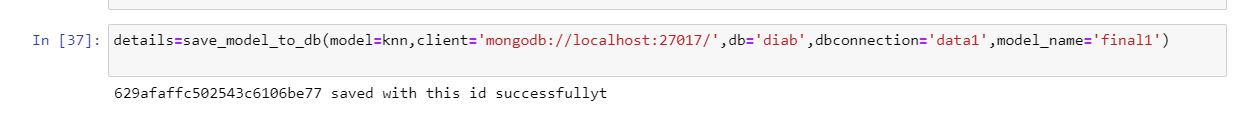
[**https://www.mongodb.com/try/download/community**](https://www.mongodb.com/try/download/community)

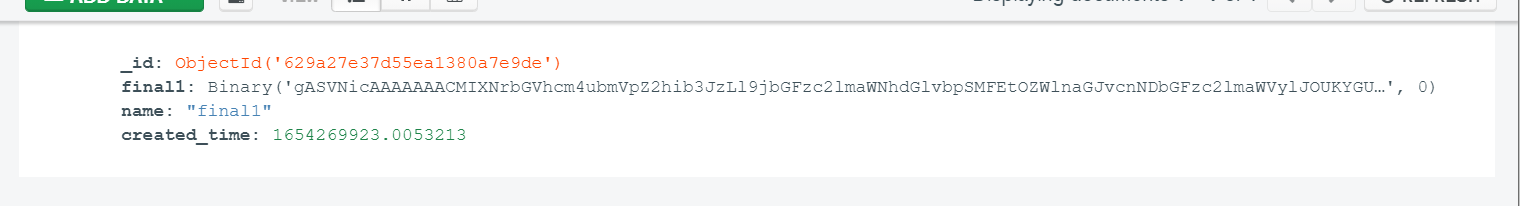
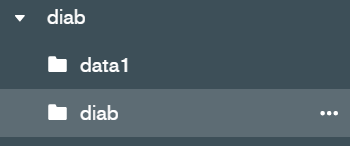
****

**PROCEDURE:**

1. Install pymongo module to connect the shell and mongodb and work with the database in the python shell.
2. Insert the data to mongodb using pymongo as follows.



1. Import the necessary libraries.
2. Connect the python shell with mongodb with the help of pymongo as follows.
3. The dataset has 768 rows and 9 columns.
4. Split the dataset into training and testing dataset.
5. Print the train and test shape.
6. Initialize the Estimator object.
7. Fit the model to the training set in order to predict classes.
8. Based on the training dataset, the model predicts the following for the test set:
9. Print the test score.
10. Save the model to disk.
11. Load the model from disk.
12. Save the model to db as mentioned below.



1. Get the saved model from the database.