**Chapter 2**

**LITERATURE REVIEW**

The Breast Cancer Wisconsin data set is a multivariate data set consisting of 570 records which is split into training and testing samples. Each sample contains 33 features/attributes such as radius, smoothness, compactness, diagnosis, ID, etc. which describes the various biological features of a patient. Similar to human learning behaviors, machine learning relies on past experiences to make judgements. The more experiences it accumulates the better judgements are made.

**2.1 Anaconda**

Anaconda is a free and open source distribution of the Python and R programming languages for data science and machine learning related applications that aims to simply package management and deployment. Package versions are managed by the package management system Conda. The Anaconda distribution is used by over 6 million users, and it includes more than 250 popular data science packages suitable for Windows, macOS and Linux.

Anaconda distribution comes with more than 1,000 data packages as well as the Conda package and virtual environment manager, called Anaconda Navigator, so it eliminates the need to learn to install each library independently.

The open source data packages can be individually installed from the Anaconda repository with the Conda install command or using the pip install command that is installed with Anaconda. Pip packages provide many of the features of Conda packages and in most cases, they can work together.

We can also make our own custom packages using the Conda build command, and we can share them with others by uploading them to Anaconda Cloud, PyPI or other repositories. The default installation of Anaconda2 includes Python 2.7 and Anaconda3 includes Python 3.6. However, we can create new environments that include any version of Python packaged with Conda.

* **Anaconda Navigator**

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows users to launch applications and manage Conda packages, environments and channels without using command-line commands. Navigator can search for

Department of CSE, RNSIT 2018-2019 5