**Chapter- 1**

**Introduction**

**1.4 Aims and Objectives of the thesis**

Our aim to ascertain a result that defines a precise result of those 500 students, whose data has been trained by KNN Classifier method. The data is genuinely a combination of their CGPA's and results in their SSC, HSC & 1-12th semesters in their university period & evaluate the performance of each student predicated on the data they provided or predicated on their educational information.

It is a data mining technique for dividing data into predefined groups and if groups have label or name that is called supervised learning methods [3]. These kinds of method used to specify all data that have located to one of existence class, so it called supervised. If mining can be executed in educational environment that is educational data mining (EDM).

Data Classification can be viewed as a two-step process which consists of learning phase and actual classification. In the learning (training) phase, a general model is constructed to walk through to study patterns found available in a sample dataset to understand the relationships that exists amongst the data. Information gained from inter-class and intra-class relationships can then be used in the second phase to label and classify entities in test data [23]. Accuracy of classification models can be evaluated on a certain test data as the amount of instances that are correctly classified by the model.

In order to predict the student’s performance at university based on high school grades, and to predict the courses that mostly effect the performance of students in the first two years of university. Several data mining techniques were used for Classification. [8] Such as KNN, Logistic regression, SVM, Naive Bayes, Apriori, Adaboost etc.

**1.5 Contribution**

We tried to implement and experiment with existing ideas in our thesis work. In our system we proposed a way to gain more accuracy than previous works which can be said as the most important proposal of our work. For this we proposed a technique that we use for implementation known as **KNN Classifier method.**

It should be noted that it is not yet formally proven the correctness or falsehood of our proposed model but as we came to gain certain good outputs and by our calculation we can verbalize that this proposal is adequate for the next level of precision. Instead, this thesis is inhibited to contributing, hopefully strong, evidence for or against its validity.

Here are our contributions of this thesis:

1. This thesis avails for generating new conception for higher precision of prognosticating the information and the proposed technique is much efficient than other technique that can be seen.
2. Our thesis will help to understand classification and recognition data in an easy way and we tried to implement the system in a simple manner so that anyone can use it for different purpose.
3. Supervised classification was used for getting different patterns on different datasets for better results and accuracy and for analyzing the system performance in different situations.
4. Besides, gaining accuracy and correct information also testing it for the best possible outcomes.

**1.6 Conclusions**

In this research paper, we represented the KNN Classifier method. k-Nearest Neighborhood (k-NN) classification is a method adoptable for classifying entities based on closest training examples in a feature space [9]. k-NN is a lazy learning classifiers that adopts instance-based learning hence having prediction done in two stages. Firstly, it undergoes minimal operations of analyzing the attribute values of individual instances in training dataset [23]. This paper provides some basic fundamental ideas about the KNN method for mining a data set.

For future works, how to quickly and accurately adapt to more new samples in online classification systems should be researched, and choosing a more efficient assessment method that can reasonably assign the training set and the testing set is necessary. We have majorly focused on more the accuracy & gaining correct information. The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.