# Project Proposal

## CS306: Data Analysis and Visualization

**Pratvi Shah**
201801407@daiict.ac.in

Instructor of the course: Prof. Pankaj Kumar
Teaching Assistant: Umang Patel

## 1 Objective:

Verify the hypothesis or reject the null hypothesis $H_0$ = Marks distribution of a large group of students is Gaussian/Normal distribution.

## 2 Rationale:

The marks distribution and the grading thereafter is done by fitting the marks of students over the normal distribution curve due to the underlying assumption that a given class has students of all calibre. This project will be a comment upon whether this assumption of the marks following the normal distribution is a valid argument or is it only applicable in specific cases.
This hypothesis is a topic of research due to the concerns of grade inflation which makes the data skewed yet the grading is done assuming normal distribution.

## 3 Approach:

Under this project, my aim would be to start from verifying the basic criteria that normal distribution satisfies (empirical rule) and then move on to perform specific tests to understand the level of error we would make if we reject the hypothesis (kstest, etc.). Furthermore, my attempt would be to perform the specific tests with respect to few other distributions or a combination of multiple gaussian distributions in an attempt to find a better approximation of the data(if applicable).

## 4 Timeline:

Following will be the flow of tasks that would be delivered within the timeline of this project:

- Clean the data and perform basic tests
- Research about other tests that would help analyze the data better
- Perform the tests and reject or verify the null hypothesis
- If null hypothesis is rejected, perform the tests for other possible distributions (log normal, logit normal ,etc.) that would give a better approximation.

## 5 Possible Issues:

The mode of examination may affect the distribution of marks so the analysis performed will be with respect to a particular set of data and may not be applicable to all marks distributions in general.

## 6 Datasets:

- Datasets obtained from Professor of DA-IICT with data of more than 150-200 students.

- Other dataset that might be of use: De La Hoz, Enrique (2020), "Data of Academic Performance evolution for Engineering Students", Mendeley Data, V1, http://dx.doi.org/10.17632/83tcx8psxv.1

# 7 References:


- https://www.journals.uchicago.edu/doi/10.1093/bjps/axs046
- https://stanford.edu/~cpiech/bio/papers/gradesAreNotNormal.pdf