



# QUESTION PAPER ANALYSER

## MACHINE LEARNING PROJECT



### Introduction:

Often during exam days, it is important for a student to know what topics she should prepare the best so as to score big in the examination. But sometimes it so happens that she might not be able to go through every past year question paper. This is the exact problem the project is targeted at. The project aims at analyzing a given question paper and produces an outcome, giving the weightage of different topics involved using concepts of machine learning. The idea of naïve bayes theorem is what propels this project in producing an accurate output.

### Resources:

Programming language	Pyhton
Libraries	<ul style="list-style-type: none"><li>• Numpy</li><li>• Pandas</li><li>• Sklearn</li></ul>
Ide	Jupyter notebook
Dataset	<a href="#">Math dataset</a>

### Working:

#### Preprocessing:

The above-mentioned dataset was used for working of this model. It contains JSON files which were then converted to csv format using an online platform [gigasheet](#). The dataset comprises of 1400 rows. Each row contains a problem related to the field of mathematics and the corresponding topic. The problems belong to seven different topics namely:

- Algebra

- ii. Prealgebra
- iii. Intermediate algebra
- iv. Geometry
- v. Counting and probability
- vi. Precalculus
- vii. Number theory

The topics were then assigned with numbers ranging from 1 to 7 as labels and the dataset was then split into a testing dataset and a training dataset in the ratio 1:3. Further, the counter vector was used to convert the string present in the problem column of each row into a question number array.

## Preparing the model:

The project is essentially a derivation of the naive bayes classifier model. It predicts the topic for each question input and then releases an output stating the percentage of each topic involved by identifying the number of problems belonging to each topic.

## Naive Bayes:

naive bayes classifiers are a collection of classification algorithms based on bayes ' theorem.it is not a single algorithm but a family of algorithms where all of them share a common principle, that is every pair of features being classified is independent of each other.

## Performance:

The model yields an astonishing accuracy of 82% given that it has been trained for a dataset containing only 1400 rows of information.

## Team Members:

1. **Rinku Ahirwar (20103124):**  
Laid out the basic idea for the project and carried out the preprocessing of the dataset.
2. **Saksham Karnatak (20103127):**  
Prepared the model while also debugged the program.
3. **Snehal Dhawan (20103142):**  
Created the dataset using JSON files and came up with the analyzing part for the project.