

# STUDENT LEVEL CLASSIFIER



---

*Team name-  
Coding Warrior  
Name:- Gaurav  
Kumar daharia*

## Purpose of developing this project

- Education is an important concern and it is duty on an institution that they should take care of the performance of every student.
- This particular model it not only predicting the performance but also showing the region of interest of a particular student.
- This model also cover the point like no. of leaves a particular student takes.
- Through number of leaves we can predict the performance of student which is being discussed in this report.
- This particular project not only displays the performance of a student but also tells the part upon which an education authority have to work in order to improve the performance of student.

## Outcome of this project

- Performance can easy measured.
- Region of interest of student can be detected.
- Defaulter can easily we classified.
- Performance level can be increased by having idea of stage level of any student.

## Technique Used for building this Machine Learning model:

**Ensembling Technique:** In this technique we combine more than two machine learning algorithm in order to achieve good accuracy.

**Accuracy achieved:** 100 % on training data and 77 % on testing data.

**Algorithm used to train Model:** Support Vector Machine, Random Forest Classifier, Decision Tree Classifier and Extra Tree Classifier. 4

# Dataset Information

This is an educational data set which is collected from learning management system (LMS) called Kalboard 360. Kalboard 360 is a multi-agent LMS, which has been designed to facilitate learning through the use of leading-edge technology. Such system provides users with a synchronous access to educational resources from any device with Internet connection.

The data is collected using a learner activity tracker tool, which called experience API (xAPI). The xAPI is a component of the training and learning architecture (TLA) that enables to monitor learning progress and learner's actions like reading an article or watching a training video. The experience API helps the learning activity providers to determine the learner, activity and objects that describe a learning experience. The dataset consists of 480 student records and 16 features. The features are classified into three major categories:

1. Demographic features such as gender and nationality.
2. Academic background features such as educational stage, grade Level and section.
3. Behavioural features such as raised hand on class, opening resources, answering survey by parents, and school satisfaction.

The dataset consists of 305 males and 175 females. The students come from different origins such as 179 students are from Kuwait, 172 students are from Jordan, 28 students from Palestine, 22 students are from Iraq, 17 students from Lebanon, 12 students from Tunis, 11 students from Saudi Arabia, 9 students from Egypt, 7 students from Syria, 6 students from USA, Iran and Libya, 4 students from Morocco and one student from Venezuela.

The dataset is collected through two educational semesters: 245 student records are collected during the first semester and 235 student records are collected during the second semester.

The data set includes also the school attendance feature such as the students are classified into two categories based on their absence days: 191 students exceed 7 absence days and 289 students their absence days under 7.

This dataset includes also a new category of features; this feature is parent participation in the educational process. Parent participation feature have two sub features: Parent Answering Survey and Parent School Satisfaction. There are 270 of the parents answered survey and 210 are not, 292 of the parents are satisfied from the school and 188 are not.

## Attributes

1. **Gender** - student's gender (nominal: 'Male' or 'female')
2. **Nationality**- student's nationality (nominal :Kuwait ,Lebanon , Egypt, Saudi Arabia, USA, ,Jordan, Venezuela, Iran, Tunis, Morocco, Syria, Palestine, Iraq , Lybia )
3. **Place of birth**- student's Place of birth (nominal: Kuwait, Lebanon, Egypt, Saudi Arabia, USA, Jordan, Venezuela, Iran, Tunis, Morocco, Syria, Palestine, Iraq, Lybia)
4. **Educational Stages** - educational level student belongs (nominal: lower level, Middle School, High School).
5. **Grade Levels**- grade student belongs (nominal: G-01,G-02,G-03,G-04,G-05,G-06,G-07,G-08,G-09,G-10,G-11,G-12 )
6. **Section ID**- classroom student belongs (nominal: A, B, C)
7. **Topic**- course topic (nominal: English, Spanish, French, Arabic, IT, Math, Chemistry, Biology, Science, History, Quran, Geology)
8. **Semester**- school year semester (nominal: First, Second)
9. Parent responsible for student (nominal: mom, father)
10. **Raised hand** - how many times the student raises his/her hand on classroom (numeric:0- 100)
11. **Visited resources** - how many times the student visits a course content(numeric:0-100)
12. **Viewing announcements** - how many times the student checks the new announcements(numeric:0-100)
13. **Discussion groups**- how many times the student participate on discussion groups (numeric:0-100)
14. **Parent Answering Survey**- parent answered the surveys which are provided from school or not (nominal: Yes, No )
15. **Parent School Satisfaction**- the Degree of parent satisfaction from school(nominal: Yes, No)
16. **Student Absence Days**-the number of absence days for each student (nominal: above-7, under-7)

## The students are classified into three numerical intervals based on their total grade/mark:

**Low-Level:** interval includes values from 0 to 69.

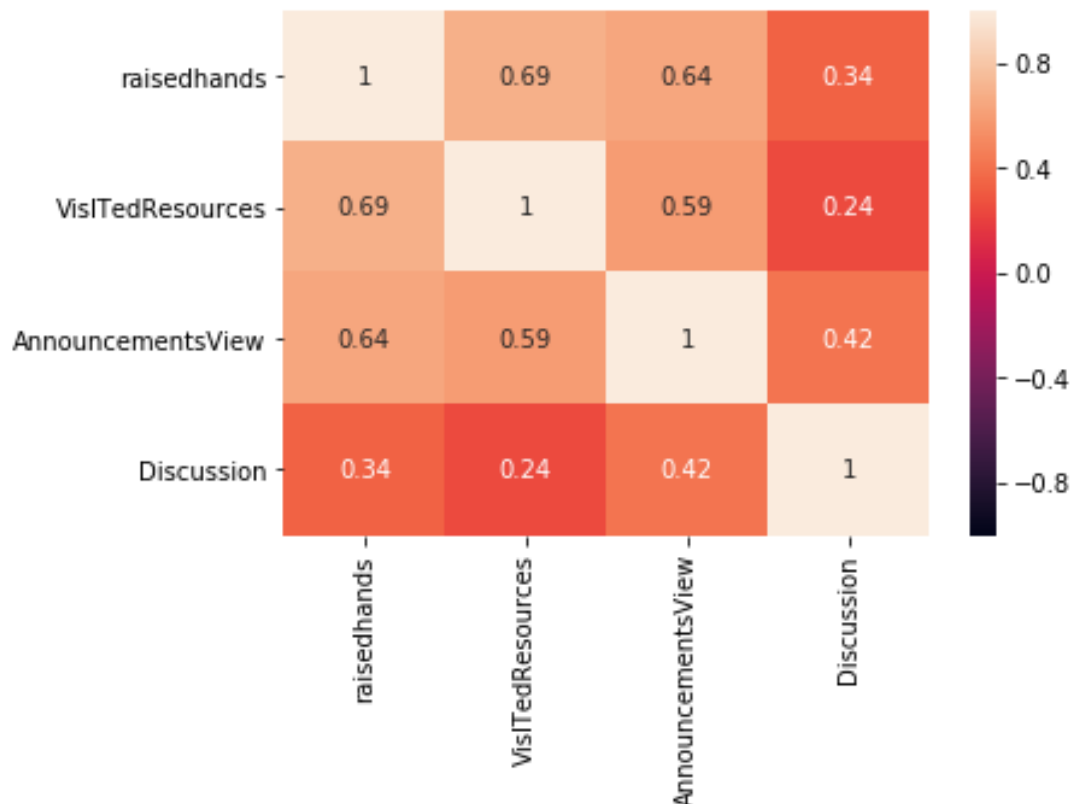
**Middle-Level:** interval includes values from 70 to 89.

**High-Level:** interval includes values from 90-100.

## Data Visualization related to correlation of dataset

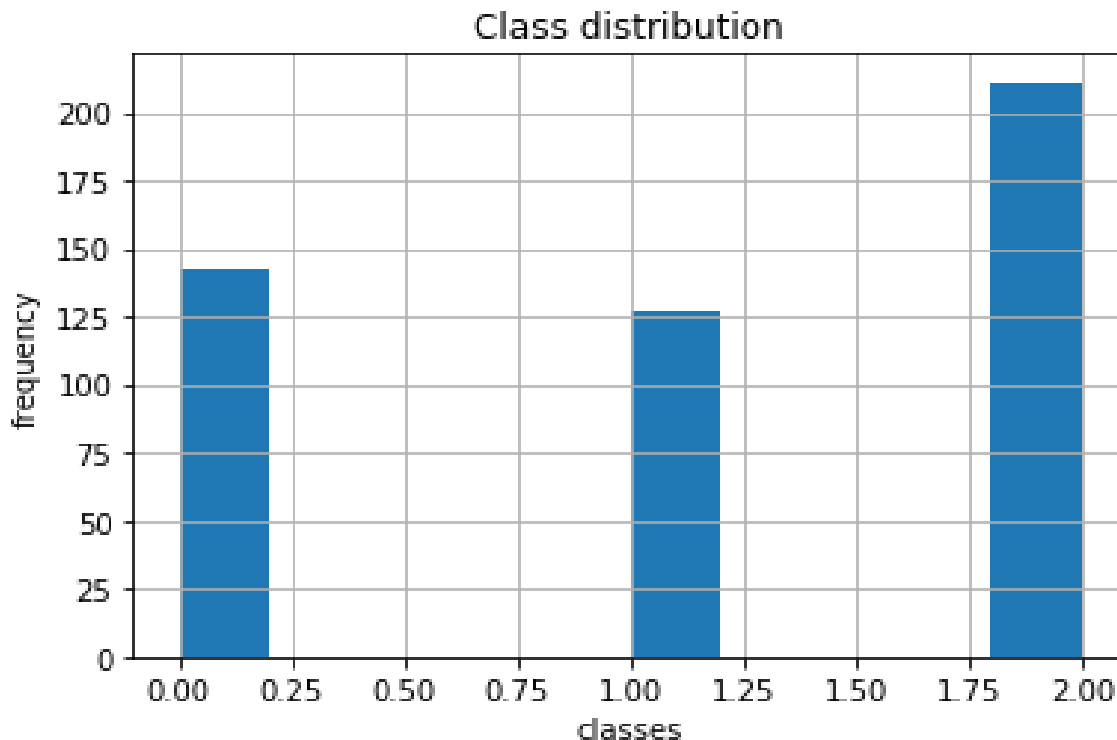
Calculating correlation of dataset is an important task for preprocessing of any data because tells us about how each a feature of dataset is related to each other and also tells us about which feature is correlated with other feature by how much percentage and does it relevant to put those features in training set to train machine learning model.

### CORRELATION MATRIX



This is correlation matrix of dataset which tells about correlation of data and the numeric inside is telling us about the percentage to which they are correlated with each other.

## Data Visualization related to target classes of dataset



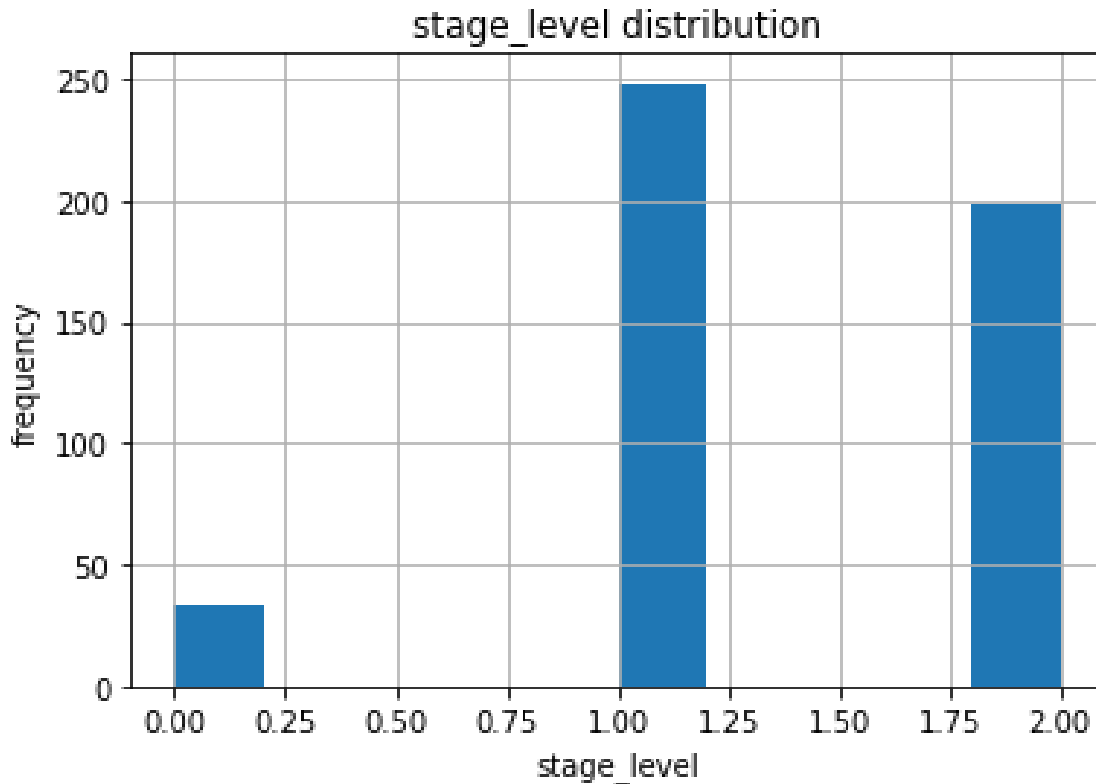
Above visualized data is related to our target classes on which I have trained my model to predict on which class a particular data point is lying. This visualization of data tells us about the frequency or occurrence of a particular class in dataset as we can see from above the following outcome:

Target Class	Frequency	Label Encoded Value
Low	148	0
Medium	126	1
High	205	2

There are some other category are also there from which we can find the region of interest of particular student and those factors that tells us about the interest of student are :

1. No. of days a student is absent in class.
2. There stage or can say there study level i.e. low level, middle level, high level.

## Data Visualization related to stage level of student

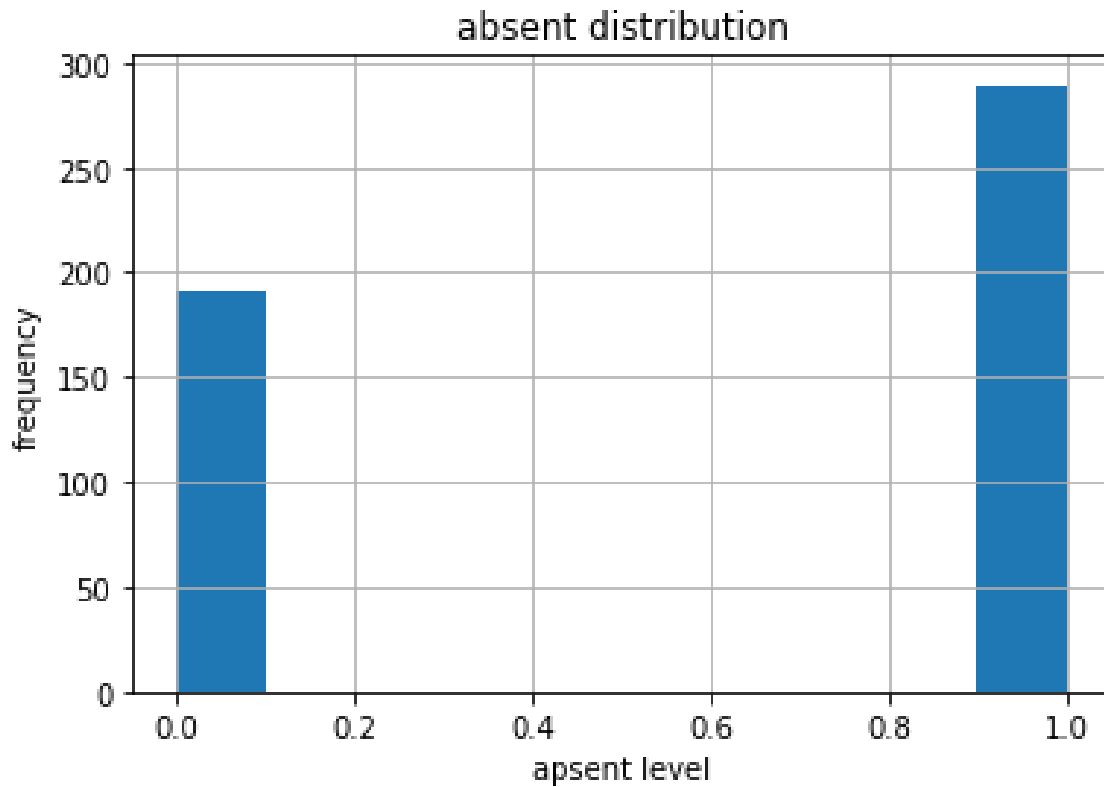


Above visualized data is reflecting the level of study or up to the mark they are performing in there academics. This is an important factor which tells how many number of student are lying in different categories of study level.

Study Level	Frequency	Encoded Value
Low level	47	0
Medium Level	249	1
High Level	200	2

From this distribution we can conclude that the majority of level of concentration is belonging to medium and high level so from here we can say that here at the low level there require more concentration and this is the only purpose of visualizing this data.

## Data Visualization related to No. of absent days of student



This is one of the important visualization of data from where we can predict the interest of a particular student towards a particular subject. This tells us about there are majority of student who are taking leave more than a week. The following are the values related to it:

Absent days	Frequency	Encoded Value
Under- 7 days	198	0
Above – 7 days	298	1