Members

- Amartya Ayushi   -   2019B2A41467H
- B Shreyas Bhat   -   2019B1A80969G
- Vaibhav Rana   -   2019A1PS1113G
- Jathin Narayan   -   2019A7PS1001G
- Himanshu Jain   -   2019A3PS0423G

What is Frost Insights for K-12?

Frost is a suite of integrated platforms designed to provide the digital infrastructure for your educational content. From instructional design to data to assessments, Frost can handle it **all in one place**, simplifying your entire content ecosystem.

The Problem

Whether it be classrooms in Schools or an Ed-Tech platform, keeping the students engaged & connected is a common pain point which they all face.

Hence, **Student Engagement** is one of the pain points we are trying to solve. The other problem which we are trying to solve is **Identifying the Risk of a Student dropping out of School.**

Student Engagement

Student engagement plays a crucial role in education and benefits students, the college, and education partners.

It empowers students with the ability to **acquire and practice the necessary skills** to build a successful future. Helps in building better relationships with other students, staff, and faculty and helps the students understand governance within the institution's education system. As a result, it improves student personality and enhances their skills that are necessary for driving change.

The Bigger Picture

The goal of student engagement is said to be achieved when education partners also benefit. With the implementation of student engagement plans, there will be better bonding between education partners. This provides a better understanding of student's varying needs to boost the campus experience and promotes a more responsive education system.

Risk Labelling

Student dropout is a significant concern throughout America at both the high school and college levels due to numerous factors. These statistics demonstrate the current landscape of student dropout rates.

- 1.2 million secondary students leave high school each year without degrees — the equivalent of one dropout per 26 seconds.
- Male high school students made up approximately 55 percent of dropouts, according to a 2014 study.
- Considering only undergraduates who left school with college debt, nearly four million college students dropped out during fiscal year 2015 to 2016. The number who left with no debt is unknown

Hence it is important to identify the students at risk of dropping out!

Value Proposition

Frost Insights gives a panoramic view into each step of a student's learning journey by **leveraging rich contextual data insights**, allowing teachers administrators to make the right decisions at the right time.
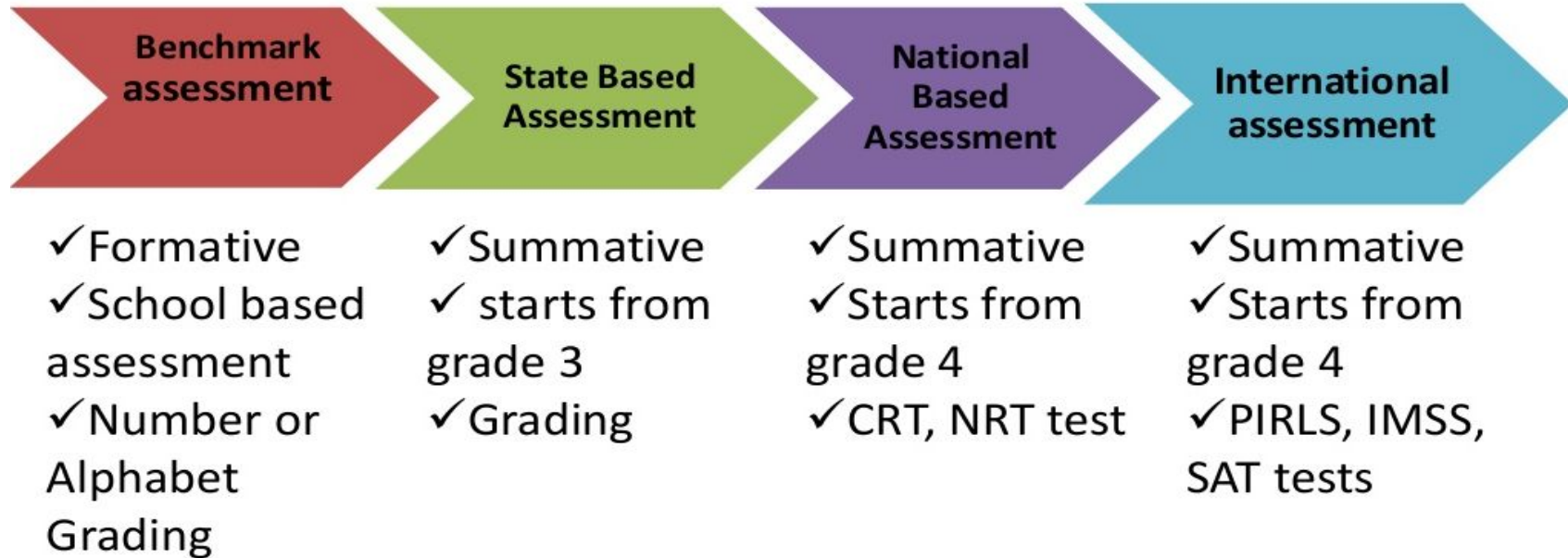
K-12 Education System In U.S

In the United States of America, Education is compulsory for children from age 6-8 to 16-18 years. This education system is supported by the federal government as around 7.5% of the GDP goes to the education sector(budget 2016).

87% of students attend public schools in America and only about 10% goes to private schools, the rests are home schooled.
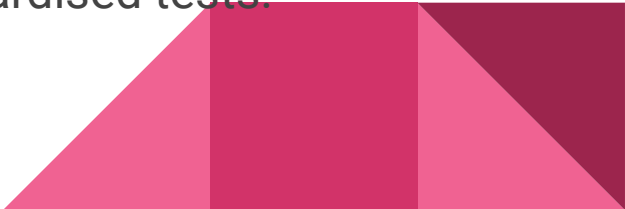
AGE			
18 17 16 15	GRADE 12 GRADE 11 GRADE 10 GRADE 9	HIGH SCHOOL 4 YEARS	
14 13 12	GRADE 8 GRADE 7 GRADE 6	JUNIOR HIGH SCHOOL 3 YEARS	
11 10 9 8 7	GRADE 5 GRADE 4 GRADE 3 GRADE 2 GRADE 1	ELEMENTARY SCHOOL 5 YEARS	
6 5		KINDERGARTEN 2 YEARS	
4 3		PRE-SCHOOL 2 YEARS	

Assessment Systems

Benchmark assessment	State Based Assessment	National Based Assessment	International assessment
✓Formative ✓School based assessment ✓Number or Alphabet Grading	✓Summative ✓ starts from grade 3 ✓Grading	✓Summative ✓Starts from grade 4 ✓CRT, NRT test	✓Summative ✓Starts from grade 4 ✓PIRLS, IMSS, SAT tests

Standardized State Tests

A state wise test is conducted every academic year in the USA to ensure that students attending public schools are getting the desired minimum level of education. This test non compulsory for students attending private schools and home schooled students. The average scores of all students belonging to a particular is calculated to check the overall performance of the schools. Most of the time the funds allocated by the state government for these schools is based on the overall performance of the schools in these standardised tests.

Grade Inflation

During recent years the average grades of Students in the State Standardized Tests and Internal schools are on rise but their performance in the SAT or other international tests is on decline , this indicates that there's been a inflation in grades of recent students which is a cause of worry for the schools and the govt.. This inflated grades also means that students are not able to perform on par at University level and thus their Risk of Dropping Out increases leaving them huge educational debt. Thus labelling students at risk have become even more important.
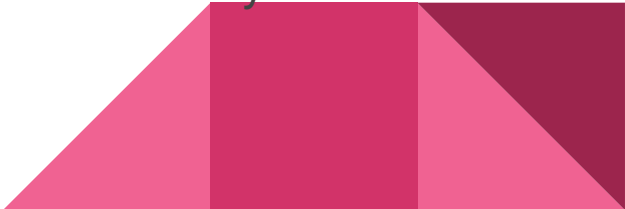
Student Engagement

A meta-construct that includes behavioural, emotional, and cognitive engagement

Behavioural engagement draws on the idea of participation and includes involvement in academic, social, or extracurricular activities

Emotional engagement focuses on the extent of positive (and negative) reactions to teachers, classmates, academics, or school.

Cognitive engagement is defined as student's level of investment in learning. It includes being thoughtful, strategic, and willing to exert the necessary effort for comprehension of complex ideas

Reviewing methods for measuring engagement

Student Self-report: critical to collect data on students' subjective perceptions, as opposed to just collecting objective data.

Experience Sampling: report on their thoughts, feelings, behaviors, and/or environment on multiple occasions over time, reduces problems with recall failure.

Teacher Ratings of Students:useful for studies with younger children who have more difficulty completing self-report instruments due to the reading demands and limited literacy skills

Interviews: can provide insight into the reasons for variability in levels of engagement in more open-ended and unstructured ways

Observations: composite of academic behaviours such as reading aloud, writing, answering questions, participating in classroom tasks, used by school psychologists to screen individual children into special need programs.

Parameters considered

- **Frequency_interaction** : number of clicks on the Virtual learning environment.

- **Performance_index** : Index providing performance of student.

- **studentAssessment['score']** : Scores on final assessment.

- **Student_info['attendance']** : Attendance.

- **Registration Dates** : Early registration/Late registration..

Engagement Algorithm

- Combining faculty perception, student participation, student self perception and incorporating Performance index calculated by another team

- Engagement Index = **(K1*frequency_interaction) + (K2*_Performance_Index_) + (K3*student_assessment) + (K4*(Date_registration - Date_unregistration)) + (K5*attendance)**

- Where K1 , K2 , K3, K4 and K5 are weighting constants.

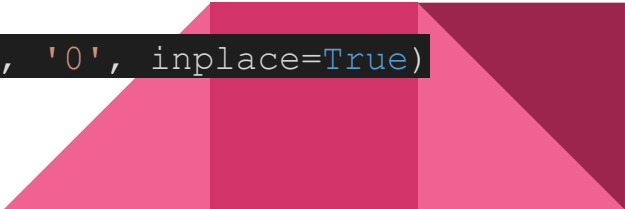- Normalized to keep score between 0 and 1

Handling Data

Missing Data

In real world data, there are some instances where a particular element is absent because of various reasons, such as, corrupt data, failure to load the information, or incomplete extraction.

Deleting Column not needed for analysis

```
studentVle.drop(['date'], axis =1)
```

Replacing missing data

```
studentRegistration['date_unregistration'].replace('NaN', '0', inplace=True)
```

Generating Parameters by Data Manipulation

In a real world dataset, more often than not, we find that we can't directly obtain all the necessary information and parameters.

	id_assessment	id_student	date_submitted
0	1752	11391	18
1	1752	28400	22
2	1752	31604	17
3	1752	32885	26
4	1752	38053	19
...	...	...	...
173907	37443	527538	227

	Unnamed: 0	id_assessment	Mean_date	id_student	date_submitted	late_submission
0	0	1752	19.356546	11391	18	N
1	1	1752	19.356546	28400	22	Y
2	2	1752	19.356546	31604	17	N
3	3	1752	19.356546	32885	26	Y
4	4	1752	19.356546	38053	19	N
...	...	...	...	...	...	...
173907	173907	37443	221.572674	527538	227	Y

By using some already existing functions in the pandas library of python we were able to perform some data manipulation and extract relevant parameters.

Students at risk

Students in this category are determined to be at risk of dropping out of their courses. It is crucial to have in place a dropout warning framework which would preemptively identify K-12 students who fall into this category.
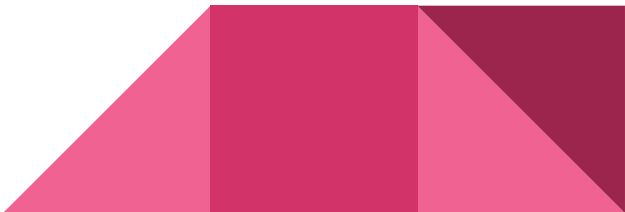
Parameters considered

After comparing the existing factors on a correlation heatmap, the following parameters were deemed to be the most efficient and linearly independent.

- **Date** - date of registration
- **Sum_click** - number of clicks on the VLE
- **Engagement_index** - index calculated from the previous section
- **Attendance -** number of days attended
- **Frequency_interaction** - number of times students interact with VLE

Challenges

It is important to study approaches to identify at-risk K-12 online students and build an effective yet practical warning system. However, this task is rather challenging due to the following characteristics:

- Multiple modalities - Difficult to determine uniform metrics due to subjective factors like quality of teachers, emotional attachment, etc.
- Length variability - Should be able to differentiate students on length of learning history and whether they are newly enrolled
- Data imbalance - K-12 online interactive courses have relatively low dropout rates in comparison to other forms of learning.

Predicting Student at Risk

- Multi-class classification into 'Low Risk', 'High Risk', 'No Risk' of dropping out
- Can be done using 'K-Means Clustering', 'Decision Tree', 'RandomForest', 'Naive Bayes Classifier' based on performance, attendance and engagement index
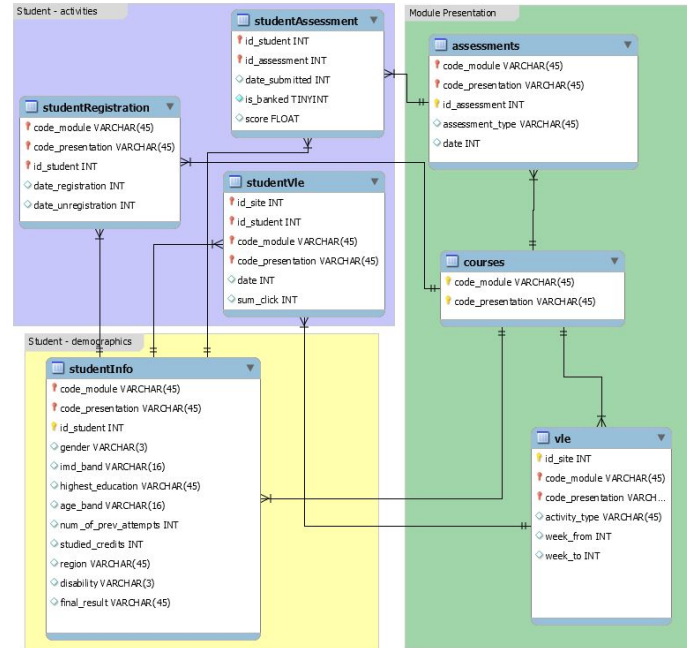- Defining thresholds in engagement index corresponding to the classification
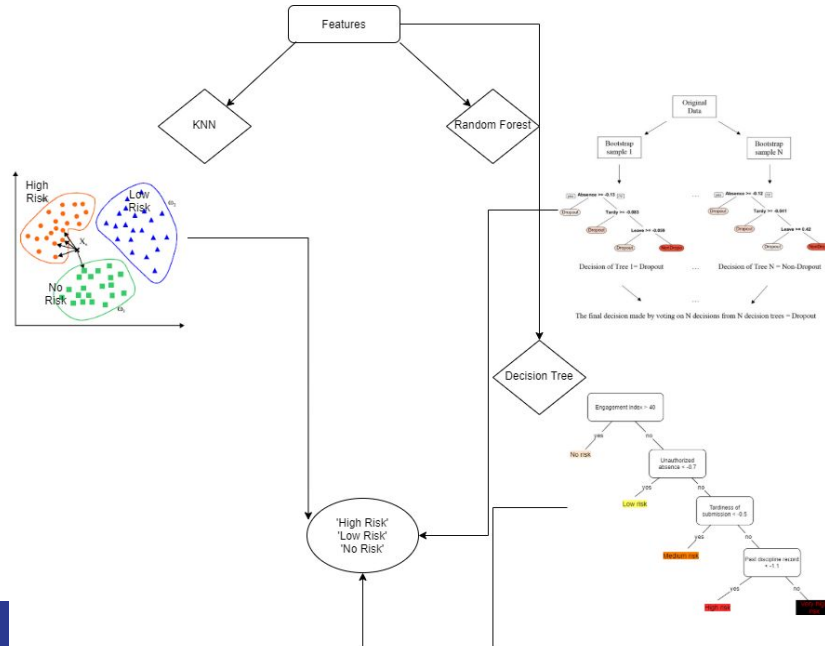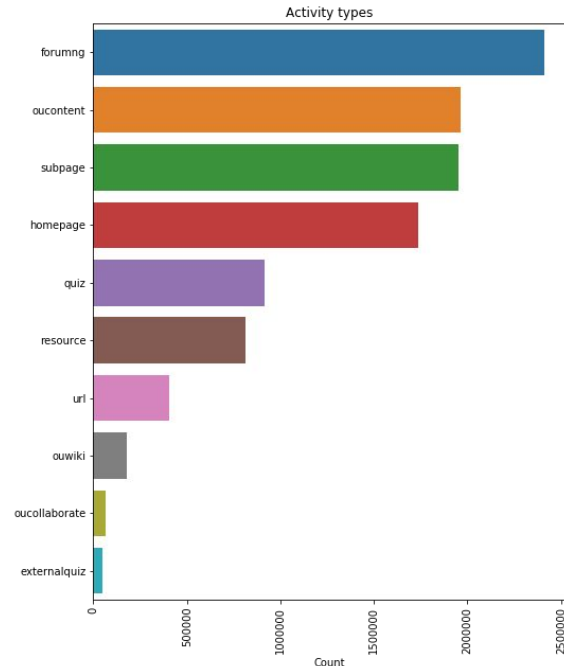
Results

- K- nearest neighbours test accuracy = 99.91 %
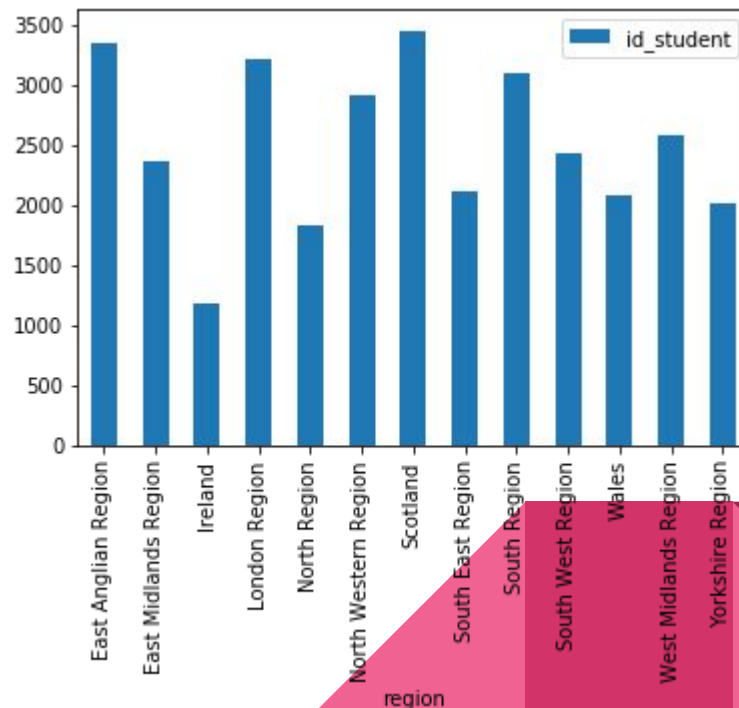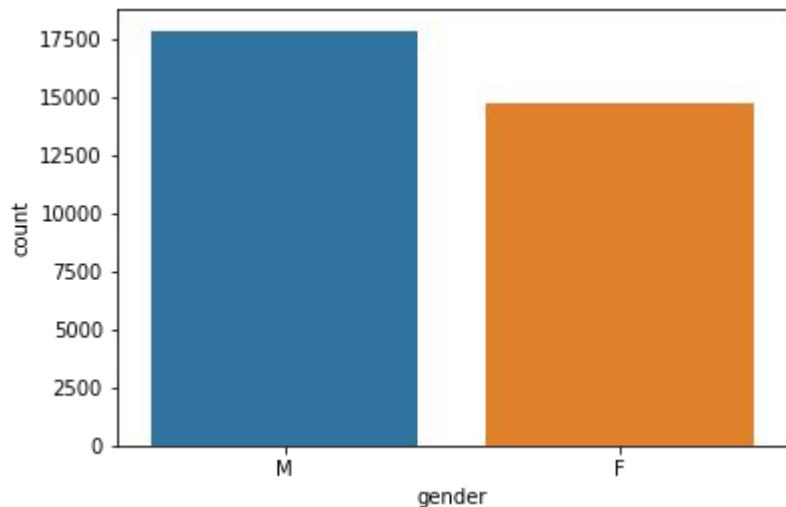- Random forest classifier test accuracy = 99.11%
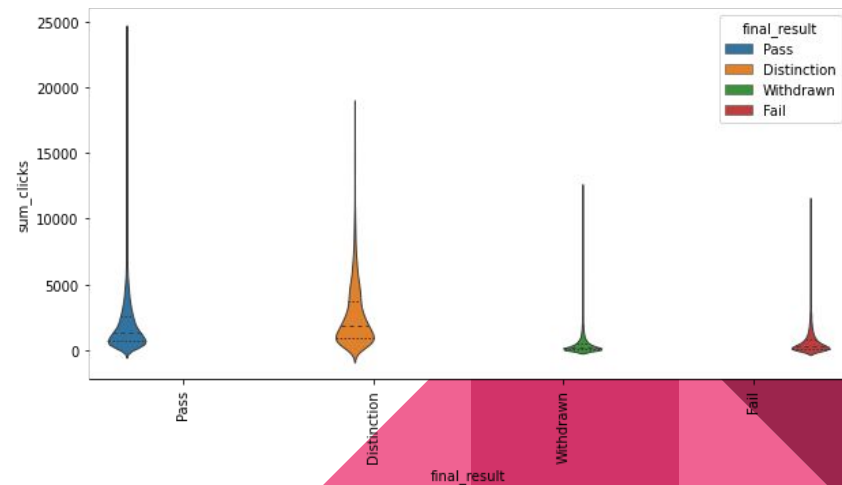
Datasets
OULAD: Open University Learning Analytics Dataset

- Anonymized data about courses, students and their interactions with Virtual Learning Environment(VLE) for 7 courses

- Tools used:
- Python language, Scikit Learn, Pandas, numpy, matplotlib, seaborn
- Trained on Google Colaboratory

Data Anonymization

Data anonymization is the process of protecting **private or sensitive information** by erasing or encrypting identifiers that connect an individual to stored data.

To ensure the safety of users, we have various compliances which a company needs to follow.

PII (Personally Identifiable Information)

Personally identifiable information (PII) is information that, when used alone or with other relevant data, can identify an individual. PII may contain direct identifiers (e.g., passport information) that can identify a person uniquely, or quasi-identifiers (e.g., race) that can be combined with other quasi-identifiers (e.g., date of birth) to successfully recognize an individual.

GDPR (General Data Protection Regulation)

GDPR's primary aim is to give individuals control over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU.

Thank You!