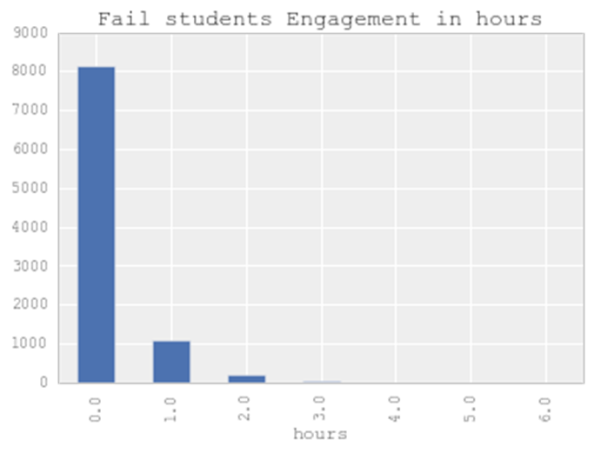
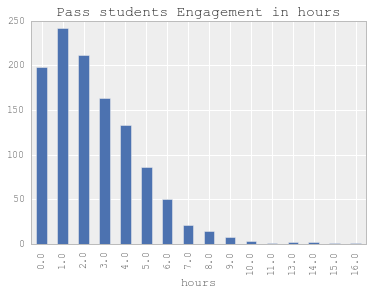
Case Study: MOOC Data Set

# **Experiment Objective**:

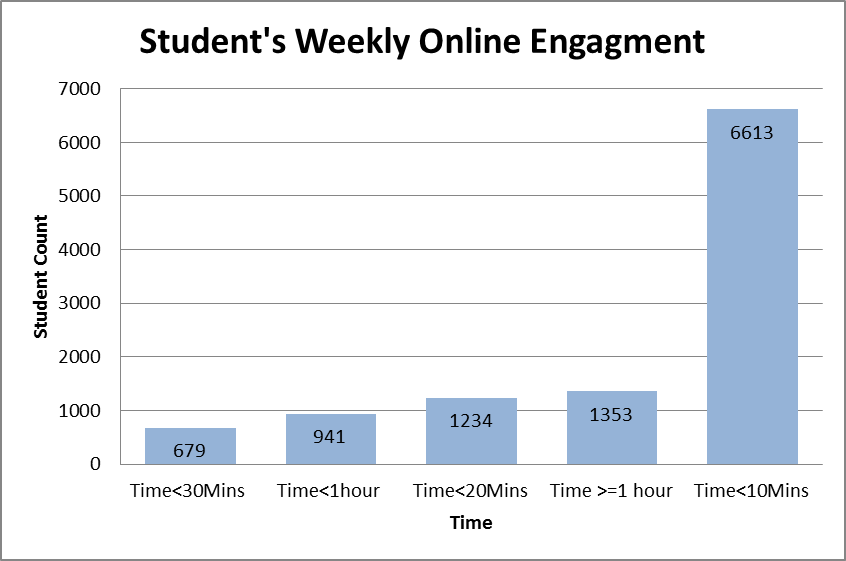
Classification of engagement level of student’s into two categories

1. Engaged
   1. Weekly time spent is more than 10 minutes
2. Not Engaged
   1. Weekly time spent online is less than 10 minutes

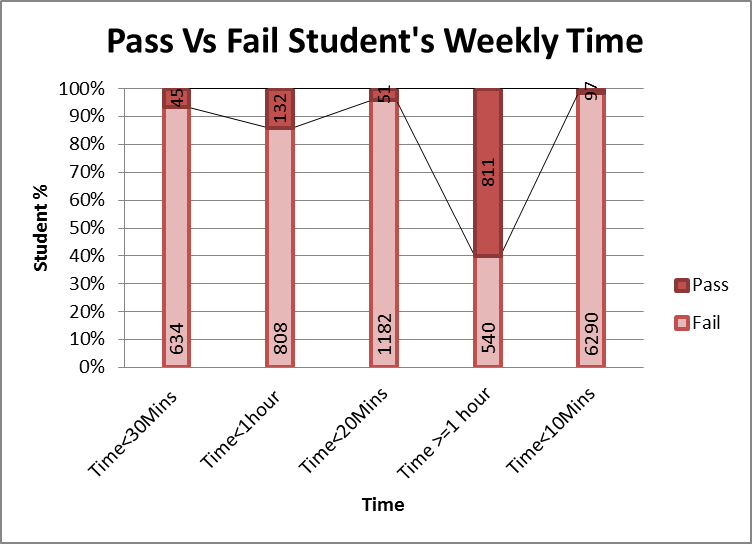
# Comparison of Pass and Fail students w.r.t Time spent online



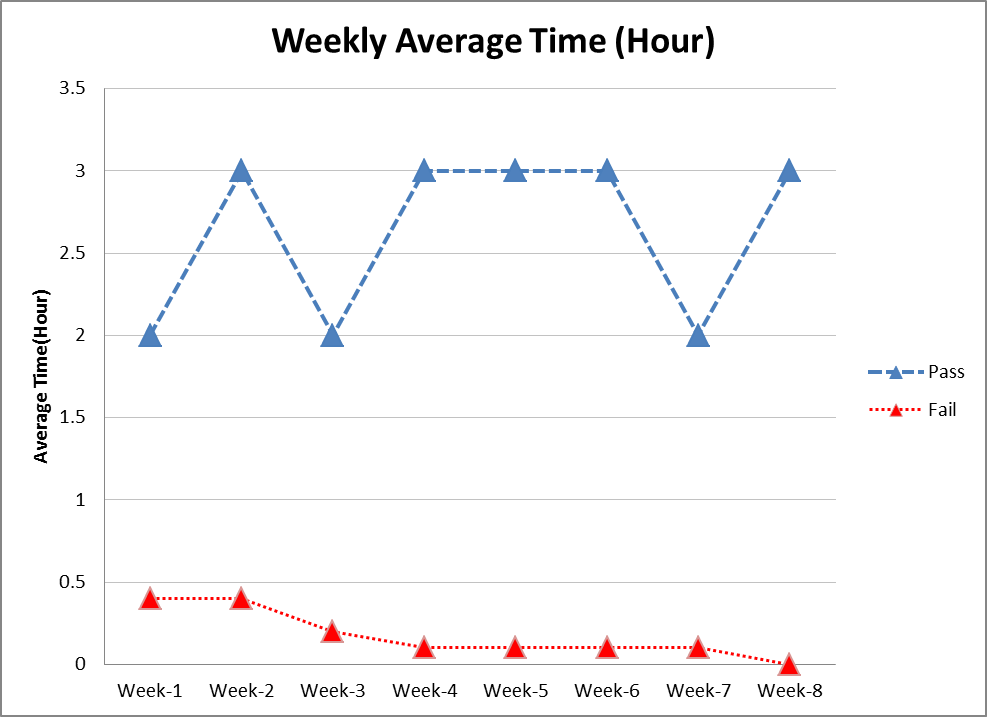
## Further Division of Time overall students comparison



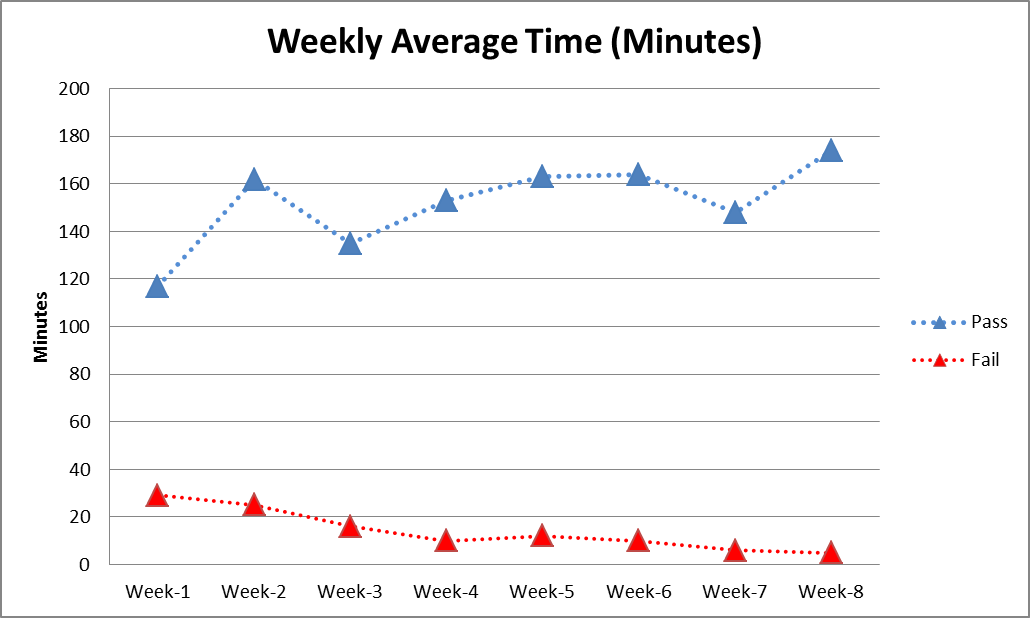
## Pass Vs Fail Students Time spent



## Weekly Average Time(hours) spent Pass VS Fail



## Weekly Average Time (Minutes) spent Pass VS Fail

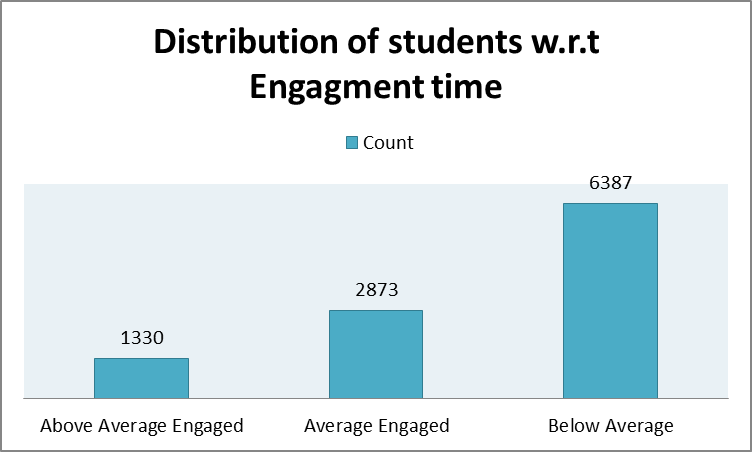


# Engagement Categories

* Below Average ( time spent is less than 10 minutes)
* Above Average ( time spent is between 10-60 minutes)
* Average Engaged ( time spent is more than 1 hour)

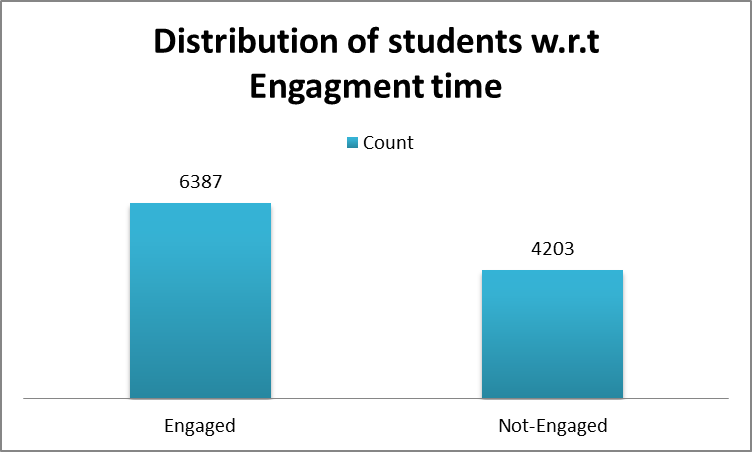
## Problem

Classification accuracy was poor due to imbalanced data



## Plan B

* 1. Engaged: Weekly time spent is more than 10 minutes
  2. Not Engaged: Weekly time spent online is less than 10 minutes



# Experiment 1:

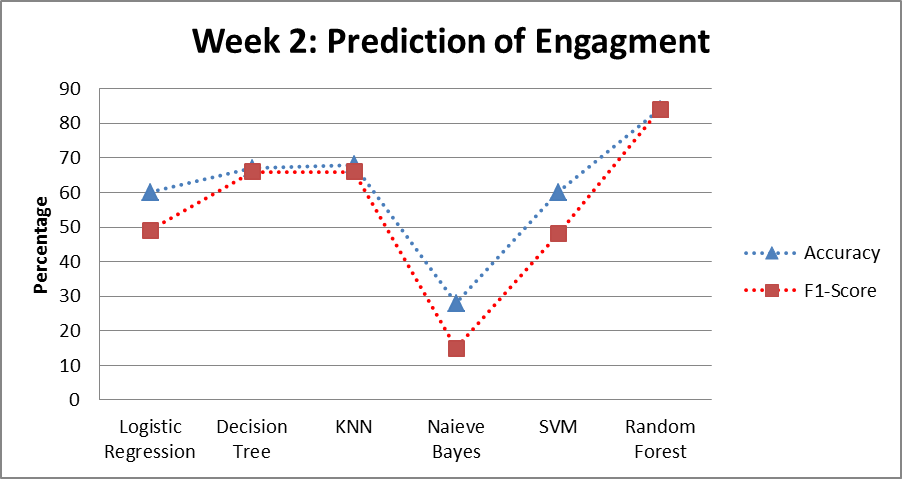
Comparison of following classifiers

1. Logistic Regression (LR)
2. Decision tree (DT)
3. Random forest (RF)
4. Support Vector Machine(SVM)
5. Naive Bayes(NB)
6. K- Nearest Neighbors (KNN)

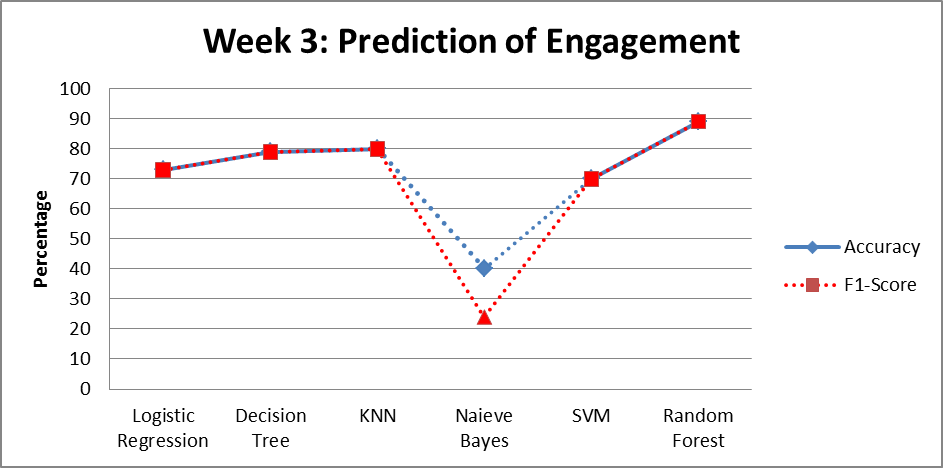
## Features used

|  |  |  |
| --- | --- | --- |
| S# | Feature Name | Definition |
| 1 | QuzAttmptAvg1 | The mean value of number of attempts that student’s made for quizzes during week 1 |
| 2 | QuzCount1 | Total number of quizzes students attempted during Week1 |
| 3 | QuzAvSco1 | Average score of Quizzes during week1 |
| 4 | QuzLagMean1 | Mean lag time between quizzes during week1 |
| 5 | LecLag1 | Mean lag time between lectures access during week1 |
| 6 | LecCount1 | Total number of lectures accessed during week 1 |
| 7 | VidActCount1 | Total number of activities performed during lectures of week 1 |
| 8 | eff1 | Total time spent online during week 1 |

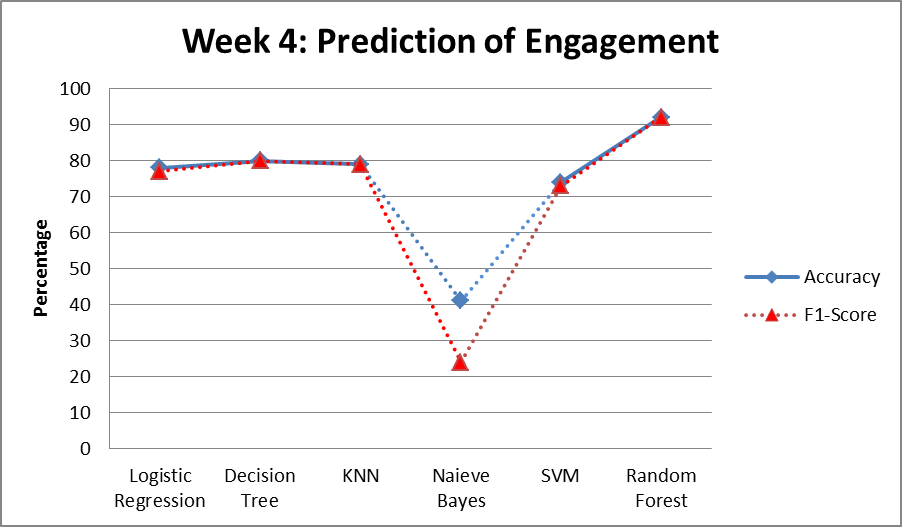
1. **Prediction of overall engagement based on Week 1 Data**
2. **Prediction of overall engagement based on Week 1 and Week 2 Data**



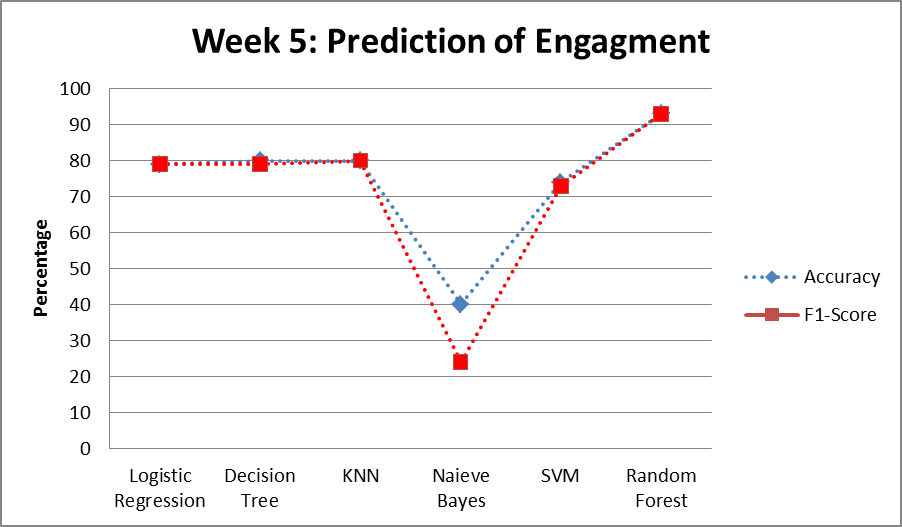
1. **Prediction of overall engagement based on Week 1 to Week 3 Data**



1. **Prediction of overall engagement based on Week 1 to Week 4 Data**

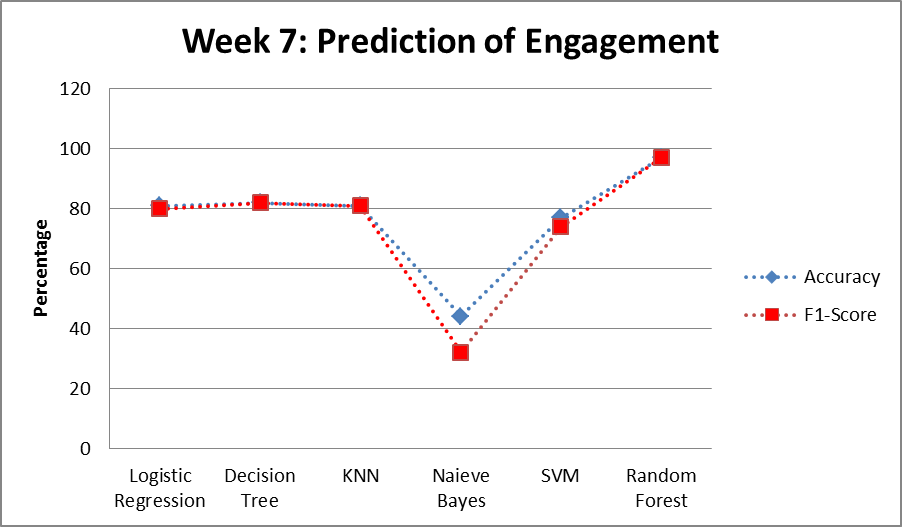


**Prediction of overall engagement based on Week 1 to Week 5 Data**

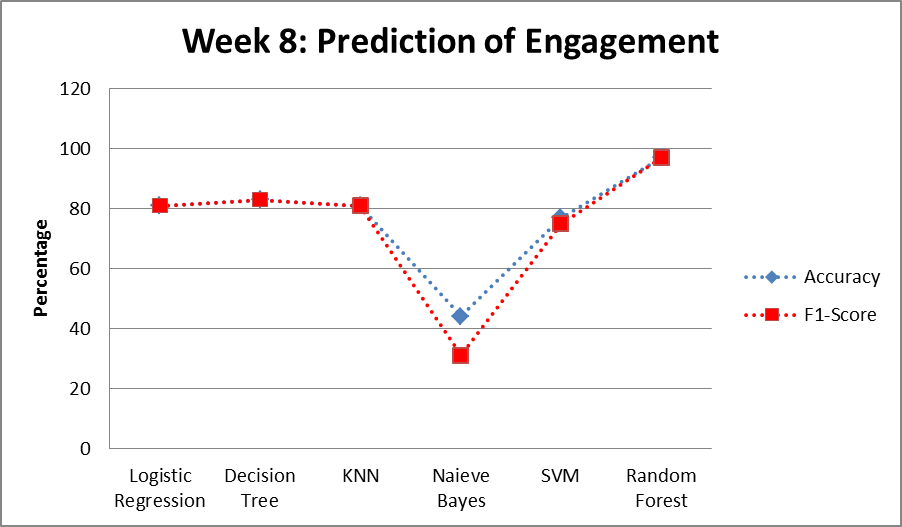


1. **Prediction of overall engagement based on Week 1 to Week 6 Data**

**Prediction of overall engagement based on Week 1 to Week 7 Data**



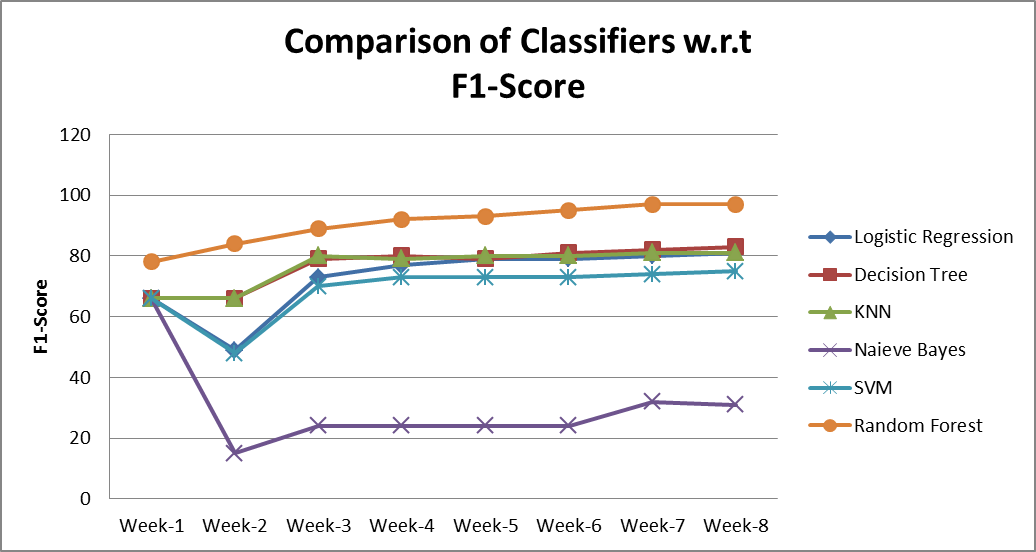
1. **Prediction of overall engagement based on Week 1 to Week 8 Data**



# Comparison of Classifiers based on Accuracy

# Comparison of Classifiers based on F1-Score

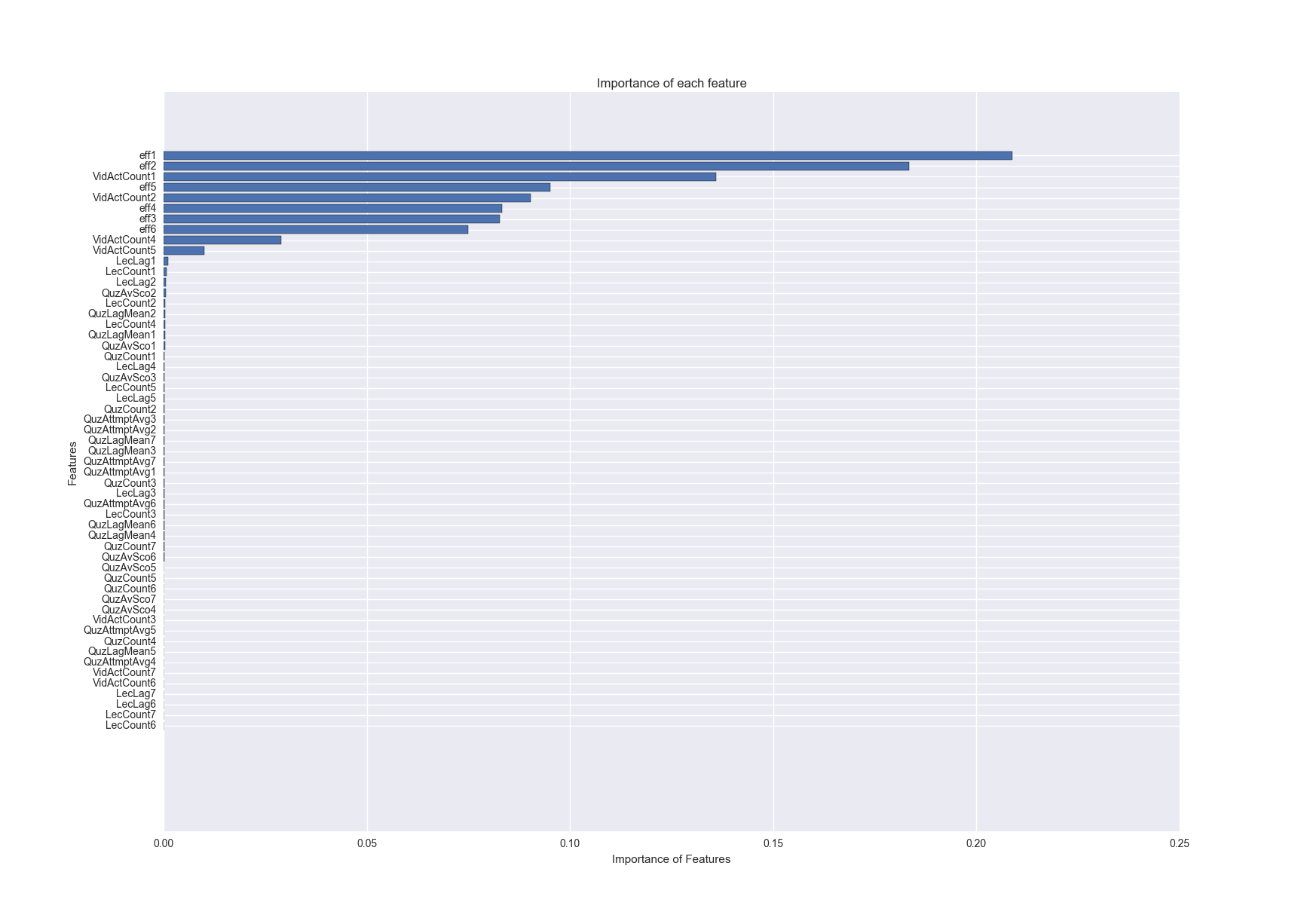
# 



# Random Forest Classification results

# Feature Importance

|  |  |  |
| --- | --- | --- |
| S# | Feature Name | Definition |
| 1 | QuzAttmptAvg1 | The mean value of number of attempts that student’s made for quizzes during week 1 |
| 2 | QuzCount1 | Total number of quizzes students attempted during Week1 |
| 3 | QuzAvSco1 | Average score of Quizzes during week1 |
| 4 | QuzLagMean1 | Mean lag time between quizzes during week1 |
| 5 | LecLag1 | Mean lag time between lectures access during week1 |
| 6 | LecCount1 | Total number of lectures accessed during week 1 |
| 7 | VidActCount1 | Total number of activities performed during lectures of week 1 |
| 8 | eff1 | Total time spent online during week 1 |



# Experiment two

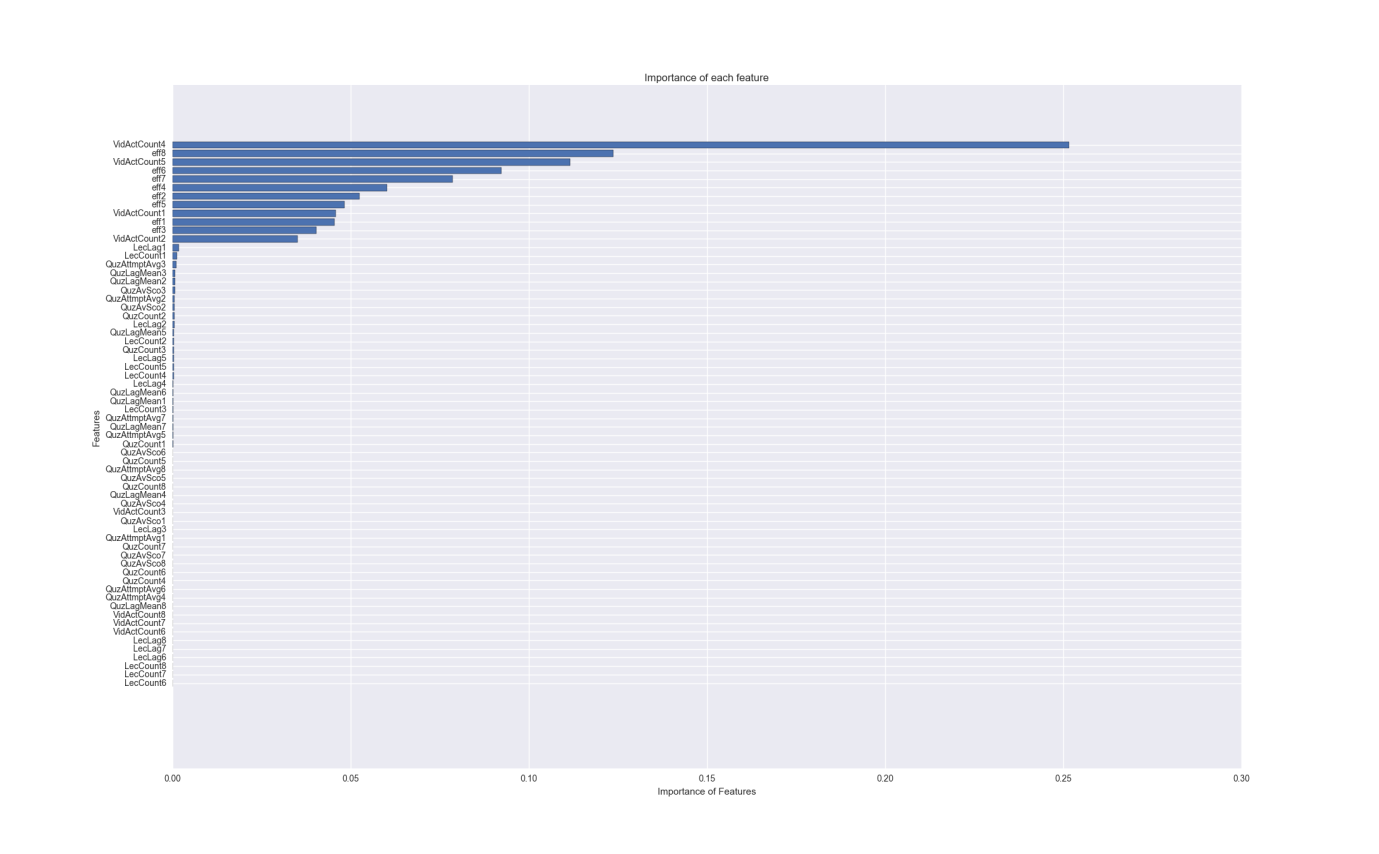
Prediction of Final outcome

Pass:

Fail :

|  |  |  |  |
| --- | --- | --- | --- |
| Data Set | Overall Accuracy | F1-Score Pass | F1-Score Fail |
| Week 1 | 88 | 20 | 93 |
| Week 2 | 88 | 40 | 94 |
| Week 3 | 90 | 44 | 95 |
| Week 4 | 95 | 72 | 97 |
| Week 5 | 95 | 76 | 97 |
| Week 6 | 95 | 76 | 97 |
| Week 7 | 95 | 74 | 97 |
| Week 8 | 95 | 76 | 97 |

# Importance of Features Classification of Final Score



**Study Objectives**

1. We investigate the Machine learning algorithms and focus on the prediction of student’s performance.
   1. Course outcome (Fail or Pass)
   2. Engagement status( In case of MOOC environment)
   3. Assignment and Quiz submission
   4. How early it is possible to detect student’s at risk? We will generate classifiers for one course offering and then apply them on the following years offering to verify reliability and robustness.
   5. To investigate the contribution of participation data in prediction.
2. Ensemble Methods of Multiple classifiers
3. To find the best features in process of prediction
4. We will investigate the process of learning through Process mining techniques. We will analyse the variation of the processes based on student’s grades.
5. We will investigate the opportunities the “real time conformance checking” offer for prediction of student’s at risk as early as possible.