Project

- Work individually.
- Submit a code and a long, detailed report along the code. The names of the files should be YOURNAME.py (or YOURNAME.ipynb) and YOURNAME.docx (e.g. OrsanOzener.py, OrsanOzener.docx)
- Any type of plagiarism will not be tolerated and will lead to disciplinary actions.
- **Due Date: 7th of February 2020, 13:00. Please submit your report through LMS. Only ONE submission per person.**

1 Introduction

In this assignment, you are supposed to predicted success of the students in different courses as well as an overall success based on the GPA value. There are all together 11 data files. The short explanation of the files are below. There might be missing data in the files as well but the imputation is most cases is a straight forward one (if the student is missing an exam or a homework assignment, then the corresponding value should be zero). The data is supposed to be clean and you are not allowed to drop any data unless instructed otherwise. If there are abnormalities in the data, please let me know.

- GPANew.csv: This file contains "StudentID", "StartYear", "TotalCredits", "GPA" fields. StudentID is the unique identifier of the students in all the given data files (Note that two files of the same course corresponding to different semester may contain the same StudentID as that particular student may be retaking the class (even with a passing grade).
- A2013New.csv: This file is for Course A, Fall Semester 2013. contains "StudentID","MT1", "HW1", "MT2", "HW2", "HW3", "MT3", "TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams and HW1, HW2 HW3 are the homework assignments. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.3(MT1 + MT2 + MT3) + \frac{0.1}{3}(HW1 + HW2 + HW3)$.

- A2014New.csv: This file is for Course A, Fall Semester 2014. contains "StudentID","MT1", "HW1", "MT2", "HW2", "HW3", "MT3", "TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams and HW1, HW2 HW3 are the homework assignments. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.3(MT1 + MT2 + MT3) + \frac{10}{3}(\frac{HW1}{70} + \frac{HW2}{60} + \frac{HW2}{60})$.

- A2015New.csv: This file is for Course A, Fall Semester 2015. contains "StudentID","MT1", "HW1", "MT2", "HW2", "HW3", "MT3", "TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams and HW1, HW2 HW3 are the homework assignments. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.3(MT1 + MT2 + MT3) + \frac{1}{7}(HW1 + HW2 + HW3)$.

- B2013New.csv: This file is for Course B, Fall Semester 2013. contains "StudentID","MT1", "MT2", "MT3", "PRJ", "Q1","TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams, PRJ is the project assignment and Q1 is the quiz. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.25(MT1 + MT2) + 0.3MT3 + 0.15PRJ + Q1$.

- B2014New.csv: This file is for Course B, Fall Semester 2014. contains "StudentID","MT1", "MT2", "MT3", "PRJ", "Q1","TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams, PRJ is the project assignment and Q1 is the quiz. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.25(MT1 + MT2) + 0.3MT3 + 0.15PRJ + 0.05Q1$.

- B2015New.csv: This file is for Course B, Fall Semester 2015. contains "StudentID","MT1", "MT2", "MT3", "PRJ", "Q1","TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams, PRJ is the project assignment and Q1 is the quiz. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.25(MT1 + MT2) + 0.3MT3 + 0.15PRJ + 0.05Q1$.

- C2013New.csv: This file is for Course C, Spring Semester 2013. contains "StudentID","MT1", "MT2", "MT3", "PRJ", "Q1","TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams, PRJ is the project assignment and Q1 is the quiz. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.25(MT1 + MT2) + 0.3MT3 + 0.15PRJ + Q1$.

- C2014New.csv: This file is for Course C, Spring Semester 2014. contains "StudentID","MT1", "MT2", "MT3", "PRJ", "Q1","TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams, PRJ is the project assignment and Q1 is the quiz. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.25MT1 + 0.3(MT2 + MT3) + 0.1PRJ + Q1$.

- C2015New.csv: This file is for Course C, Spring Semester 2015. contains "StudentID","MT1", "MT2", "MT3", "PRJ", "Q1","TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams, PRJ is the project assignment and Q1 is the quiz. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.25MT1 + 0.3(MT2 + MT3) + 0.1PRJ + Q1$.

- C2018New.csv: This file is for Course C, Spring Semester 2018. contains "StudentID","MT1", "MT2", "MT3", "Q1","TOTAL", "Grade" fields. MT1, MT2 and MT3 are the exams and Q1 is the quiz. Total is the final score of the student and Grade is the corresponding letter grade of the student. Note that Grade F means that the student fails the class. Keep in mind that the grading in some of these items may not be out of 100 as it may include bonuses or they might be alternative grading structure. The formula for the final score is as follows: $0.3(MT1 + MT2) + 0.375MT3 + 0.025Q1$.

You are suppose to perform the following tasks and in all of these tasks, the performance criteria will be the "Accuracy Score".

1. For all the courses data given, use only the data of the respective course and predict the pass/fail condition of the students given that the MT3 score is not available. Please use a 80/20 test train split with a random state parameter equal to 42. Report you accuracy score for each course and provide a detailed explanation of your solution method.

2. For Course C, given that 2013, 2014 and 2015 data is fully available and for Spring semester 2018, only MT1 score is available. You are supposed to predict the pass/fail condition of the students in 2018 given the available data. You may use any information regarding Course C (that is 2013, 2014 and 2015 data) as the training data and your test data will be the 2018 student data. Note that the grading scheme and the student body has changed over the years. Once again report your accuracy score and provide a detailed explanation of your solution method.

3. Assume that you have the full data available for Course A and B for years 2013, 2014 and 2015. You are supposed to predict the pass/fail condition of the Course C students for years 2013, 2014 and 2015 with no information regarding Course C. Note that the students are mostly the same students in these courses although there might a semester shift (so one particular student taking Course A in 2013 but Course C in 2015). You prediction will be based solely on the information of the student in previous courses. Once again report your accuracy score and provide a detailed explanation of your solution method.

4. Similar to the previous question, but this time MT1 score is available for all the respective semester for Course C. Therefore, your predictions will not be solely based on student previous performance but this partially this course's data as well. Once again report your accuracy score and provide a detailed explanation of your solution method.

5. Finally, given the grade performance of the student in all these courses, you are supposed to predict the GPA bracket of the students (only the students taking at least one course out of A,B and C in years 2013, 2014 and 2015). The brackets are defined as 2.00-2.50, 2.50-3.00, 3.00-3.50 and finally 3.50-4.00. Besides the accuracy score, please consider Mean Square Error as another criterion while brackets are ordered as 1, 2, 3, and 4 respectively and predicting a close bracket is preferable to otherwise. Once again report your accuracy score, MSE score and provide a detailed explanation of your solution method.

If you have any questions, please send an email to: orsan.ozener@ozyegin.edu.tr